TRANSPORTATION RESEARCH

RECORD

No. 1314

Rail

Advanced Train Control Systems 1991

Proceedings of a Symposium
June 17–19, 1991
Denver, Colorado

A peer-reviewed publication of the Transportation Research Board

TRANSPORTATION RESEARCH BOARD
NATIONAL RESEARCH COUNCIL
WASHINGTON, D.C. 1991

Transportation Research Record 1314
Price: $28.00

Subscriber Category
VII Rail

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data
National Research Council. Transportation Research Board.

Advanced train control systems, 1991 : proceedings of a symposium, June 17–19, 1991, Denver, Colorado.
p. cm. — (Transportation research record, ISSN 0361-1981 ; no. 1314)
"First International Symposium on Advanced Train Control . . . sponsored by the Association of American Railroads, the Railway Progress Institute, and the Transportation Research Board of the National Research Council."
ISBN 0-309-05151-7
1. Railroads—Automatic train control. I. International Symposium on Advanced Train Control (1st : 1991 : Denver, Colo.) II. Association of American Railroads. III. Railway Progress Institute. IV. National Research Council (U.S.). Transportation Research Board. V. Series.
TF638.A88 1992
625.1'0285—dc20 91-38478
CIP

Transportation Research Record 1314

Contents

Foreword

The First International Symposium on Advanced Train Control was held in Denver, Colorado, in June 1991. The principal purpose of the symposium was to bring together a broad cross section of railroad and supply company officials to demonstrate the unique application of advanced train control to all aspects of railroading, to share ideas and technology, and to enumerate the benefits of these systems. Advanced train control means the use of microprocessors and digital data communications to provide new command, control, communication, and information tools to improve railroad operations. The most well developed of these systems are the North American Advanced Train Control Systems (ATCS). The rapid spread of microprocessor use promises to promote the development of these systems worldwide. The use of the acronym ATCS implies a generic category of systems that use real-time communications and information or data processing to provide a tool to improve railroad productivity, safety, service quality, and market share.

The symposium was sponsored by the Association of American Railroads, the Railway Progress Institute, and the Transportation Research Board of the National Research Council. The program structure was designed to present the business case for advanced train control followed by technical sessions that described the design and implementation of these systems. Papers were solicited for the latter sessions and are published in this Record, along with papers from a session on advanced train control in Europe. There are no published papers from the sessions on the business case for ATCS. To provide a complete picture of the symposium to the readers of this Record, a summary of the initial four sessions is presented. The first three defined the need for ATCS, the business case for ATCS, and the financing of these systems. These sessions took a corporate view of ATCS and how these systems would benefit the railroads in an increasingly competitive freight transportation marketplace. The fourth session focused on the implementation and application of ATCS from the viewpoints of the railroads, the railroad suppliers, and the government.

The first of these sessions, Role of Technology in Meeting the Demands of the Marketplace, was designed to define the requirements for the freight transportation marketplace in North America. In an increasingly competitive market, the freight railroads have lost significant revenue marketshare to truckers. Over the past 10 years in a deregulated environment, railroad revenues have remained static in the range of $27 billion to $30 billion per year. Meanwhile, interstate trucking revenues have increased from $92 billion to $150 billion.

Two significant issues were brought out by all of the first-session panelists. First, the primary need of shippers is for consistent, reliable service, and the railroads lag behind their competitors, the trucks, in providing such service. Second, railroads can provide this service and have done so in specialized markets. The question left to the industry is whether ATCS is one of the tools—or perhaps the key element in a corporate strategy—for meeting the customers' service demands and for expanding the market for railroads. This strategy will not be directed only at advanced train control, but also will include other process control and information technologies such as automatic equipment identification and electronic data interchange.

The second session was on the Business Case for Advanced Train Control, featuring representatives from five North American railroads that are developing or have developed business cases for ATCS and a panelist from the Harvard Business School. The railroad presentations pointed out the different approaches to ATCS among the railroads, ranging from a corporate strategy of implementing several ATCS applications to open skepticism about the benefits of ATCS.

ATCS evolved from an early concept that focused on the use of advanced microprocessors and digital data communications to provide a better and less costly means for positive train separation and a broader range of commercial benefits by using the digital data communications link as a platform. The applications identified to date will affect nearly every railroad

department and will need a buy-in by the affected departments. This will require ATCS to be managed as a total railroad program.

Senior railroad managers must become familiar and comfortable with ATCS technology before they will buy in. This takes time and effort. The time for corporations faced with the choice of implementing technology of this type from first exposure to the decision to implement is from 3 to 6 years.

ATCS benefits are great, with high rates of return and with significant hard-dollar benefits such as fuel savings and improvement in locomotive reliability. However, there are also significant soft-dollar benefits such as improvements in quality of service. These soft-dollar benefits are difficult to justify, in part because they are difficult to measure. Another aspect of these systems is the potential to provide options for additional benefits from as yet unanticipated applications developed as railroad managers become familiar with the technology. ATCS should not and cannot be justified on one application alone or on the safety benefits alone, but on an overall strategy to build on the early commercial applications and then to pick up the options. With the technical and financial risk involved, most railroads will use this building block approach to ATCS.

ATCS are, first, communications systems and railroads will be trying to justify the installation of these communications systems using digital radio and from this base adding applications to take advantage of this communications link. Work-order reporting and locomotive health monitoring are examples of initial applications used to justify the installation of the data communications link

ATCS will require sophisticated integration with operations and information management systems. This is a very important part of any implementation strategy. ATCS also offer the potential to transform the industry to open up new ways to compete. In other words, ATCS may provide a competitive advantage to the railroads.

When computing the benefits of ATCS, most analyses assume flat revenues for the railroads over the past 10 years and project flat revenues in the foreseeable future. This is not the case. In fact, railroads have seen declining revenues in constant dollar terms, so ATCS benefits when computed should take this into consideration.

As Julie Hertenstein of the Harvard Business School, one of the panelists, stated,

> New technologies don't volunteer benefits; wrenching the benefits from these technologies requires commitment, hard work, and competence from individuals throughout the organization. Without these, time and money spent on technology are wasted.

The third session, Financing Advanced Train Control, covered many options for financing from inside and outside the company. Financing ATCS is an issue because of competing demands for capital and the difficulty railroads have in making their cost of capital. A return on investment (ROI) that exceeds the corporate hurdle rate is a must. The work-order reporting application on the Union Pacific is financed from internally generated funds with a ROI of 33 percent per year.

The panels in the fourth session were composed of railroad, railroad supply company, and government officials from Canada and the United States. The topic was Implementation and Application of the New Train Control Systems. This session brought out the diversity of implementation approaches and the variance of applications apparent in the railroad industry. The railroads tended to cast ATCS in terms of "when" and not "if." The suppliers who need to make investment decisions, wanted to know "when and what for" ATCS will be deployed. And government wanted to know if and when the safety applications would be implemented. This session confirmed that the railroads are approaching implementation of ATCS from a broad diversity of applications, using principally the same architecture. In addition to differences in application, there is a difference in pace and scope of application. This reinforces the building-block approach to ATCS. There was evidence from the diversity of ATCS approaches on the panel that the technology is not completely accepted by all the railroads and the railroad suppliers as providing the full range of benefits stated by the railroad participants in the business session. Skeptics are waiting to be convinced of the need for this technology.

Concern over the pace and scope of application of ATCS will continue as the railroads determine which applications are financially justifiable and best suit their needs. Safety is of

concern for all of the railroads and the government. There are several applications that have safety benefits that will fit into the strategy for installing ATCS. ATCS must be evaluated to include all of their applications for options, and be treated as a system, not as a series of unrelated investments. In this manner the interests of the railroads and the government should coincide.

The papers published in this Record provide differing approaches to advanced train control. The first section of papers, on advanced train control applications in Europe, will point this out. The published papers from the succeeding five sessions on information systems, train control, communications and computers, applications and implementation, and application and system design will reinforce the diverse approaches. In addition, these papers will demonstrate the continuing effort to marry advanced technology with railroad operations to provide more efficient and safer railroads in the decades to come.

Howard G. Moody
Association of American Railroads

Train Control on French Railroads

J. P. GUILLOUX

The French National Railways' (SNCF's) first high-speed line has been in service between Paris and Lyon for 10 years now. A second high-speed line came into revenue service in September 1989, bringing with it substantial improvements to passenger services to the west and southwest of France. At the same time, SNCF is engaged in construction work for the Northern Train à Grande Vitesse (TGV) line, which will link Paris, Brussels, and London via the Channel Tunnel in less than 2 years' time. A loop line around the eastern outskirts of Paris will link the Northern and Southeast high-speed lines. The Southeast high-speed line will be extended southward (initially as far as Valence). High-speed electric multiple units (emus) will be operated on the new lines at 300 km/hr (187 mph), whereas train speed on the Southeast high-speed line is currently limited to 270 km/hr (168 mph).

For the French National Railways (SNCF), the highest speed at which drivers can be sure of properly observing lineside signals, especially under difficult operating conditions (e.g., fog), is in the 200/220 km/hr bracket. For higher speeds, it is necessary to design a system that does not depend on correct observance of lineside signals because signaling information is transmitted directly to the driver in the cab. In addition, SNCF works on the principle that trains should be driver operated, with drivers responding to the information provided in the cab and adapting tractive and braking forces accordingly. It is, of course, common knowledge that, however reliable, human beings are more error prone than sophisticated automatic systems, whence the need to protect against driver failure.

A continuous speed-control system has therefore been installed in conjunction with the signaling system. This comes into action to generate braking should train speed differ too greatly from the curve set by the speed-control system.

SIGNALING ON THE SOUTHEAST AND ATLANTIC HIGH-SPEED LINES

The signaling system designed in relation to line throughput, vehicle braking characteristics, and specific operating conditions consists of the following:

- A continuous data transmission system (18 data items), and
- An intermittent data transmission system (14 data items), with track-to-train transmission.

The type and technical characteristics of the transmission systems selected were the result of recognition, early in the research phase, of the fact that it always takes time to design a signaling system because of its vital safety function and the many and serious environmental constraints (interference, wide-ranging temperatures, atmospheric agents, vibrations, etc).

In addition, SNCF considers it necessary to have a continuous control system for detecting broken rails, whence the choice of track-circuit–based signaling technology.

Continuous Data Transmission

Under the circumstances described, the continuous data transmission system selected consists of an alternating current (AC) track circuit. There are 18 modulation frequencies between 10 and 29 Hz, depending on the data item to be transmitted. To protect against crosstalk, four carrier frequencies are used.

With track circuits of this type, it is possible to use resonant blocking circuits instead of the insulated joints required with most track circuits, which, in turn, means that jointless, long, welded rails can be used with all their advantages in terms of track performance and vehicle stability.

Intermittent Data Transmission

The intermittent data transmission system consists mainly of a 10-m cable loop laid in the track. Frequencies range between 1,300 Hz and 3,700 Hz, depending on the data items to be transmitted.

Figures 1 and 2 show the braking and stopping sequence for block sections and for protection at mandatory stopping points.

Adaptations have been made to handle cases of access to track changeover points, turnouts, and points contiguous with lineside signaling when the high-speed line connects with conventional lines.

SIGNALING ON THE NORTHERN HIGH-SPEED LINE

Line throughput requirements for traffic expected on the Northern high-speed line, which will begin revenue service in 1993, are incompatible with the signaling used on the new Southeast and Atlantic high-speed lines. Whereas trains operated at 270 km/hr on the Southeast line have 5-min headway, it has been cut to 4 min on the Atlantic Train à Grande Vitesse (TGV) line for trains operated at 300 km/hr. The objective on the Northern high-speed line is a 3-min headway (in practice) for an operating speed of 300 km/hr.

Moreover, provision has been made for working the line at 320 km/hr (200 mph). To achieve these objectives, SNCF

Signaling Department, SNCF, 162, rue du Faubourg Saint-Martin, 75010 Paris, France.

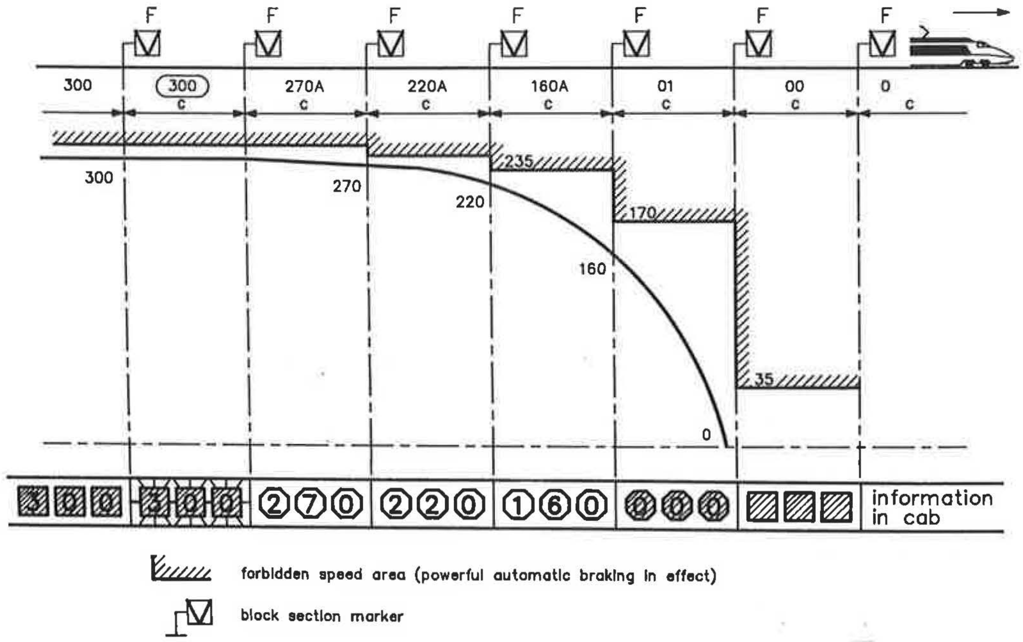


FIGURE 1 Stopping sequence for block sections.

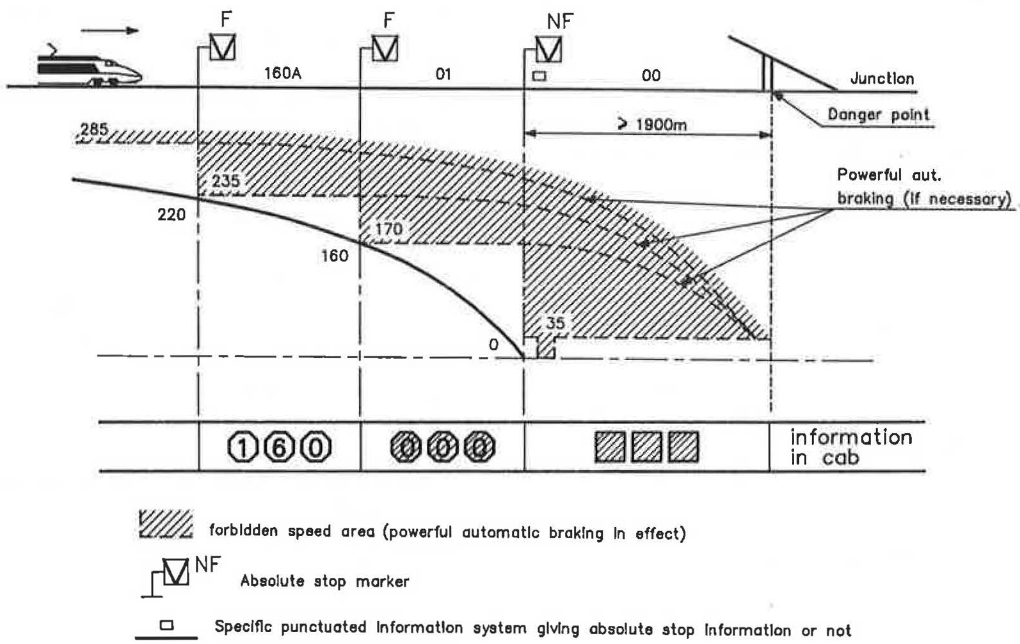


FIGURE 2 Principle of braking before an absolute stopping point.

has developed a new signaling system called TVM 430, derived from the system in operation on the Paris Southeast and Atlantic TGV high-speed lines (called TVM 300), but with better performance levels.

Functional Description of the TVM 430 System

Three measures have been adopted to meet throughput requirements. First, the block sections are shortened and have a length of 1500 m on the flat. Second, to limit system reaction time if necessary, the driver receives continuous advance warning information according to the following principle:

- If information is displayed continuously (not flashing on and off), the speed to be enforced in the next block section is not more restrictive,
- If the information displayed flashes on and off, the driver knows that the speed applicable in the next block section is more restrictive.

Third, the speed-control mechanism features a higher degree of precision than that possible with TVM 300. At any moment, the train data in the plan (distance, speed) are compared against a speed-control graph in the train-borne computer, this graph being based on the characteristic parameters of the train and information from TVM. Emergency braking is triggered when the curve on the graph is exceeded (Figure 3).

Technical Description of the TVM 430 System

Continuous transmission of 18 data items has proved to be insufficient; it is therefore necessary to enhance the system to transmit more information. With the new approach, the frequency modulated UM71 track circuit still acts as carrier, but from the very low frequency a signal is worked out from which 27-bit messages can be established. The structure of the message (Figure 4) gives the following information:

- Railroad address: This allows the train-borne system to determine the operating mode (North TGV, Channel Tunnel, etc.).
- Speed: This enables the train-borne computer to determine speed indication to be shown on the cab display; whether that indication should flash; maximum speed limit allowed in the block section; and automatic train protection (ATP) speed limit at the end of the block section.
- Distance: The distance between the start and end of a block section is called the "target distance." This distance is quantified and the data are transmitted to the moving vehicle for use in speed-control graph compilation.
- Coding: The message coding ensures data integrity.

The ground-based computer compiles messages for each track circuit several times per second. The information given is fail-safe.

Ground-Based TVM 430 Equipment Architecture

Two central units guarantee backup. One central unit is operational and receives input, processes information, and generates output. The other is on standby; it is initialized and is ready to take over if the central unit in use fails. A switching unit transfers from one central unit to the other if a failure occurs or in response to an external command.

Each unit is structured around two processors that process information in parallel. A single coded processor compares

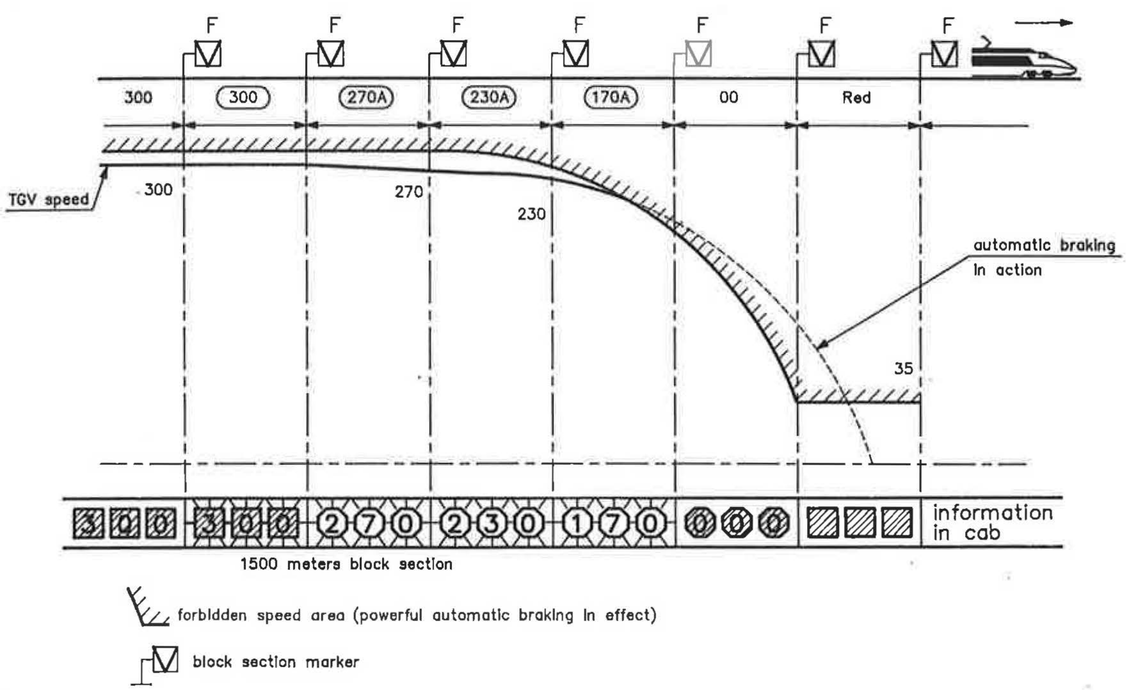


FIGURE 3 North TGV: automatic braking in case of excessive speed.

	TVM 430	North TGV			TVM 430
Field name	Operation	Speed	Distance	Gradient	Coding
Length (number of bits)	3	8	6	4	6
Safety information	YES	YES	NO	NO	

FIGURE 4 Message structure.

the results and cuts the power controlling the output if there is a disparity (Figure 5).

On-Board Equipment Architecture

The message received is decoded several times per second. The computer checks the display units and establishes the speed-control graph. This calculation is initialized each time a track circuit joint is passed; in particular, another target distance is then set.

Figure 6 shows the architecture of train-borne equipment.

Sensors are located over rails and have two windings (each linked to a digital receiver). The signal processing chain is made up of two identical units (digital receivers) each of which receives the signal sent by the sensors.

Although processor software is based on the same algorithm, it is different because the two data-capture chains are nonsynchronous. The central unit compares the results given by the two digital receivers and validates the message. A signal-processing chain is associated with each central unit. The two central units are structured around a single coded processor and operate simultaneously. Only one of them is selected by the switching unit and supplies the power needed to light the displays to which it is connected.

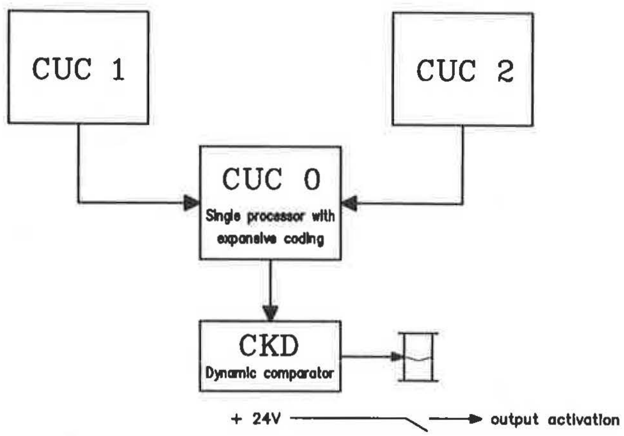


FIGURE 5 Safety data processing.

Safety

For both train-borne and ground-based equipment, data are constantly safeguarded in the capture stage by coded information, and, in the processing stage, either by coded information or duplication and by rereading at the output stage. The single coded processor system is based on use of a single processor and a single program for data processing. Safety is ensured by the very high level of data coding. The code dedicated to each variable has its own signature that relates to the previous process and the exact date of that process. For each cycle, an electronic fail-safe control unit compares all the signatures obtained from codes against a given word.

Intermittent Transmission

The carrier for the intermittent transmission system is a loop placed in the track carrying current modulated by phase jumps.

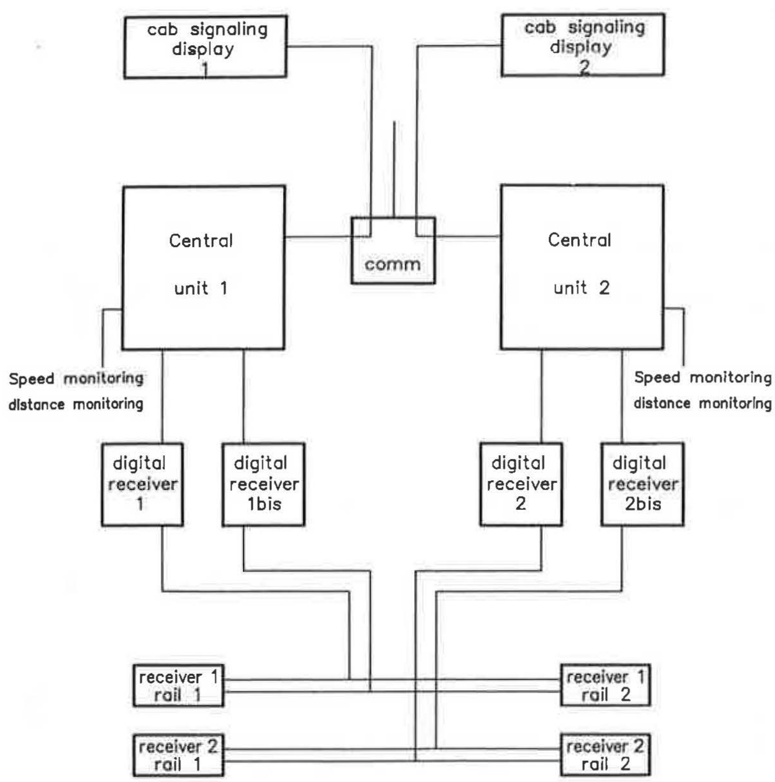


FIGURE 6 TVM 430 on-board equipment.

With this technique, a large number of telegrams can be sent to the train. Among the possible functions that can be performed are the following:

- Activate/deactivate the system,
- Ensure that no absolute stop signals are passed,
- Radio channel and system transmission,
- Cut power, and
- Lower pantograph.

VERY HIGH-SPEED LINE OPERATING SYSTEMS

High-speed lines connect with the existing network at specific points. In addition, to maintain suitable service standards even if traffic is, for example, stopped at a particular place or on a particular line, crossovers have been provided between the two tracks every 20 to 25 km (12.5 to 15.5 mi) and single-track working is to be possible on each of these sections. To be economically viable, signaling systems in the case of single-track operation must ensure a smooth flow of traffic with a minimum of constraints by comparison with normal conditions.

The maximum speed in the diverging direction on these changeover points is 160 or 170 km/hr. In certain places SNCF also uses points that can be negotiated at 220 or 230 km/hr in the diverging direction.

Control Center

Each very high-speed line is controlled from a control center that fulfills the following functions for the whole of the line:

- Points operation (remote control of signals and switches),
- Traffic control, and
- Remote control of the substations supplying power to the overhead lines.

Traffic control and power supply control are housed together in the same room. For example, the 280-km (174-mi) TGV Atlantic line is controlled from a single work station, although at peak periods or in cases of disruptions to traffic, dual manning is possible.

Power supply control facilities are single-manned for the whole line. Both work stations—traffic control and power supply control—have input keyboards and multicolor display screens. Only overall data (and train-describer data) are displayed on the visual control panel.

Traffic Control Systems

The traffic controller has a wall-mounted visual control panel containing the train describer, signal indications so the controller can be sure everything is working normally, and general monitoring data (warning systems, hot-box detectors, radio).

The traffic controller also has four multicolor screens plus keyboard for interfacing with the computerized control and monitoring systems. The screen's visual control panels display a magnified image of each station for safety purposes. Emergency control data are also displayed. The train describer system indicates remote train-announcing data, train schedule discrepancies as well as hot-box detector data or indications for the devices installed to stop trains automatically if road vehicles fall over bridge parapets. The computer-based, route-setting system also interfaces with the same screens and keyboard.

Route setting is automatically controlled from the control center using a computer-control module of similar design to that used in SNCF's computerized signal boxes.

CONCLUSION

In conclusion, this description of the systems to be used on very high-speed lines in France shows the factors and developments research engineers took as a basis in their design work to ensure better, even more cost-effective operations than on the Southeast and Atlantic lines. Today's sophisticated signaling techniques, especially those that are microprocessor-based, will be one of the many features underlying the commercial success of the Northern and other subsequent TGV lines.

UIC European Train Control System

W. R. SMITH AND B. J. STERNER

In an effort to make European rail networks more efficient and competitive with other forms of transport, the International Union of Railways (UIC) has begun developing a European Train Control System (ETCS) that must overcome the complexities produced by a host of different operating conditions, equipment, languages, and signaling cultures. Although the ETCS project is still in an early stage, it seems likely that it will use some of the higher levels of functionality of the Advanced Train Control Systems (ATCS) being developed in North America and rely on a wider range of transmission techniques than ATCS. A staged approach to implementing ETCS is being considered to avoid compatibility problems and the need for dual equipment on track.

The International Union of Railways (UIC), although it has members outside Europe, is primarily concerned with facilitating international rail traffic within Europe. It is the body that sets standards and makes recommendations for rolling stock and fixed equipment used in international traffic—passenger and freight. The UIC's standards and recommendations are also widely used for national traffic. The Office for Research and Experiments (ORE) is the research arm of the UIC, responsible for organizing research work for the joint benefit of the railways.

The UIC has undertaken the development of a European Train Control System (ETCS). Although the project is still in an early stage, some similarities to and differences from the North American Advanced Train Control Systems (ATCS) can be discerned.

EUROPEAN OPERATING PATTERNS

Operating patterns on the European railways differ very considerably from those in most of North America, and this dictates a different approach to the ETCS from that proposed for ATCS, where the emphasis seems to be very much on improving operation of lines with low levels of traffic and little or no signaling. In much of Western and Central Europe, the density and distribution of population (Figure 1), with major centers of population a few hundred miles apart, are favorable for intercity passenger transport by rail.

The size and density of development in and around major centers are also favorable conditions for rail to handle a significant proportion of local traffic, especially commuting. Government has often regarded subsidizing rail passenger transport as socially desirable. This means that in Europe there is a much greater emphasis on passenger traffic than in North America, and on some national railways this is much more important than the goods traffic.

W. R. Smith, O.R.E., Oudenoord 500, 3513 EX Utrecht, Netherlands. B. J. Sterner, Banverket, P.O. Box 1070, 17222 Sundyberg, Sweden.

There is a wide range of speeds and densities of traffic on European railways. In most countries there is an intercity network carrying mixed traffic, including high-speed intercity trains. In the last 15 years or so, some countries have built new, very high-speed lines for speeds up to 185 mph; some of these are dedicated to high-speed passenger traffic only, while others carry mixed traffic. There are also suburban networks carrying very dense passenger traffic, rural branch lines with lower traffic levels but mostly carrying both passenger and goods traffic, and some lines carrying goods only.

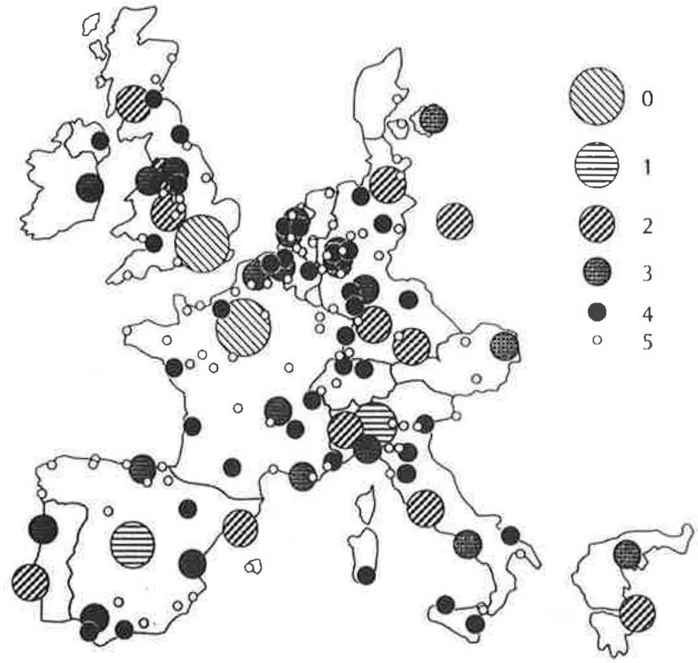
It follows that the European railways are "scheduled railways" and that most of the "management" applications of ATCS emphasized in North America are already looked after by Europe's national systems. The combination of these national systems into an overall European system for the management of international freight traffic is the aim of other UIC projects outside the scope of ETCS.

Most European railways were provided quite early on with comprehensive systems of lineside signaling, using semaphore signals, mechanical interlocking, and block telegraph. Over the last 30 or 40 years, the more heavily trafficked lines have generally been reequipped using relay technology, extensive track circuiting, and mainly with colored light signals. However, it was often uneconomic to apply this technology to lines with lower levels of traffic. Until recently there was no economic alternative to retaining the old equipment on these lines, so many have signaling that is obsolete—in some cases downright antiquated—and labor intensive to operate and maintain.

In most cases, warning, cab signaling, or speed supervision systems have been overlaid on the lineside signaling system. Only recently, on some very high-speed lines or very densely trafficked urban lines, have lineside signals been dispensed with. These warning or supervisory systems are of various ages and provide various levels of protection against driver error or inattentiveness. Some of the older ones are now considered to be obsolete or inadequate for present needs and require replacement.

INCOMPATIBLE NATIONAL SYSTEMS

Language problems, the political situation in Europe up to the middle of this century, and the different topographical and meteorological conditions in different parts of Europe led to very different solutions to signaling and train control problems in the past. Systems were, in general, developed to meet national needs, and compatibility with systems in adjacent countries is the exception rather than the rule. (In the context of compatibility it is interesting to note that no two adjacent countries in Western Europe have adopted the same



D'après RECLUS, 1989

Les villes européennes		
0	Population supérieure à 6,4 Mio d'habitants	
1	Population entre 3,2 et 6,4 Mio d'habitants	
2	Population entre 1,6 et 3,2 Mio d'habitants	
3	Population entre 0,8 et 1,6 Mio d'habitants	
4	Population entre 0,4 et 0,8 Mio d'habitants	
5	Population entre 0,2 et 0,4 Mio d'habitants	*Mai 1989*

FIGURE 1 Major population centers in the European Community.

electric traction system.) An attempt by ORE in the 1960s to produce an international track-train information transmission system was not really successful; the system produced was considered by most railways to be too heavily biased towards the requirements and practice of its country of origin.

This incompatibility has not been a serious problem until now, because the locomotives of most international trains are changed at or near the borders. However, these stops at borders have been a significant factor in limiting increases in the commercial speed of international passenger (and freight) services, a problem that has contributed to a decline in the competitive position of rail relative to road and air travel. Now not only do the railways wish to improve their competitive position, the European Community itself, faced with severe congestion on the roads and in the air, sees high-speed rail services as essential to its future development and has produced a plan for a network of high-speed railways (Figure 2).

This network will consist partly of new lines and partly of upgraded existing lines, some of which are already in service, (although only carrying national traffic). Others are under construction. On this network, stops at the borders will no longer be acceptable; border formalities are being greatly simplified. Those formalities that are not abolished will be carried out on the moving train.

The very high-speed passenger trains of the future will be made up of permanently coupled consists, including the motive power. Modern power electronics makes it feasible to construct multisystem tractive units. Limited multiple-equipping of locomotives for different train control systems is used on some routes and will be used on others in the future (e.g., Paris–Brussels–Köln–Amsterdam and London–Paris–Brussels), but this is limited by constraints of space, weight, and cost.

The above considerations have provided the stimulus to look at the possibility of a common range of train control equipment for the whole of Europe. A further incentive is that a common range of equipment would allow production on a larger scale and increased competition among suppliers, all of which should reduce the cost of equipment to the railways and should also simplify maintenance of international trains. Over the last few years a great deal of consideration has been given to the best way of tackling this question, and the evolution of ATCS has been of great interest. The result of this consideration has been a project to develop a system capable of meeting the full range of the European railways' requirements for internal as well as international traffic.

COMMUNICATION BARRIERS

Another aspect of incompatibility among countries' railways—and a difficulty that must be reckoned with in the

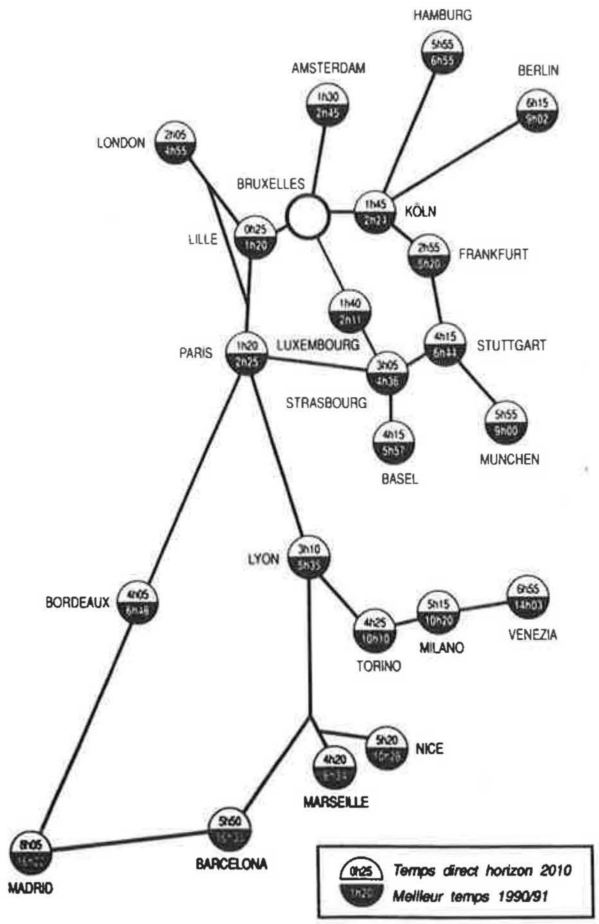


FIGURE 2 Future and present rail travel times from Brussels.

creation of a common ETCS—is the communication problem among the signal engineers or operators of the various countries. It is not just a language problem, although this problem is big enough, with more than 17 languages being spoken in the countries where the system may eventually be installed.

Just three of these languages (German, French, and English) are the official working languages of the UIC. Thus, for only a minority of the signal engineers is one of the three UIC working languages their native tongue. However, even in those cases, the communication problem is not solved, as the signaling cultures in the German-, French-, and English-speaking parts of Europe are so different that many concepts exist in just one of these languages and cannot be translated one-to-one into the signaling cultures of the other two.

To avoid some of these problems, the UIC has decided to use English as the only working language on the ETCS project, but this is not a complete solution. The now-active signal engineers were brought up in isolation in different subcultures, a fact that has led to great reluctance to adopt solutions or principles developed outside their own countries. In some cases, there is a reluctance, or even an inability, to understand the principles and solutions of other administrations. However, working closely together in a small group for a large proportion of the time should make it possible to overcome this problem, and this, in fact, is how the project will be organized.

THE ETCS PROJECT

It is self-evident that to cover the very wide range of the European railways' requirements adequately and do it at a cost commensurate with the different levels of traffic, a multilevel system, comparable to ATCS, is needed. Another point of similarity is that the UIC is also aiming at a modular system, specifying the modules in sufficient detail to allow interoperability between those from different manufacturers. The UIC is not aiming to develop equipment itself; this remains the manufacturers' task.

Because the project is in an early stage, the number and functionality of the levels have not yet been defined. This will be one of the first tasks after the full range of the railways' requirements has been established—work that is being done at present. It seems likely that the levels will be displaced upwards in comparison with ATCS; there is no interest in an equivalent of Level 10 and only limited interest in an equivalent of Level 20.

There is, however, interest in an equivalent of Level 30 that could be applied to secondary lines, especially where the signaling equipment is obsolete, but the greatest immediate interest is in the equivalent of Level 40—and possibly a level above this—for lines at the higher end of the range of speeds and traffic densities. The essential functions would be to enforce limits of movement authority and speed limits. The UIC will have to investigate to what extent such functions as train detection, train separation, and interlocking should continue to be carried out by conventional means or be incorporated into the new system; it is quite likely that both alternatives will have to be provided.

At the lower levels, corresponding to Levels 30 and 20 of ATCS, it seems likely to be economic for these functions to be incorporated into the new system. In particular, the low levels of traffic and light trains on many of the lines to which these levels would be applied make track circuits both uneconomic and unreliable. On many of these lines, train detection is still visual (i.e., done by signalmen). An economic technique for locating trains and proving them complete is needed that does not require significant additional equipment on-train, because a large proportion of rolling stock must be able to use such lines from time to time. This rules out the use of satellite navigation systems, which would not be applicable to the more heavily trafficked lines and so would require train equipment that would be underused.

The above considerations mean that the UIC will have to consider a wider range of transmission techniques than those proposed for ATCS. A number of techniques are used in the various systems installed on different railways, and it is likely to be a question of choosing which to use rather than of developing new technology. For discontinuous track-train communication, ETCS will probably use fixed-message transponders, as does ATCS, at the lower levels. But at the higher levels, ETCS will probably also use switchable transponders (i.e., the message can be varied according to information from other signaling equipment on the ground) or active beacons.

For radio transmission, the UIC has decided to concentrate its efforts on using the new train radio system it has decided to develop. This will be an all-digital voice/data system operating in the 900-MHz band, using as much as possible of the technology being developed for the public mobile radio networks. Provision will have to be made in the higher levels

of the system for supplementing radio transmission by continuous track-train transmission, possibly two-way, to provide for cases in which the intensity of messages would become too great for the available radio channels or there are problems of breaks in continuity of radio coverage.

A major question is how to make the transition from the present multiplicity of systems to the new standard one. A suggested approach is to provide a data bus to which are connected a central processing unit (probably itself of modular construction), interface modules to the pickups for existing and new systems, train radio, brakes, and the driver's display/control unit. Trains would be equipped only with those modules needed for the lines they run on. The relevant modules would be added when the standard system is installed on those lines and removed as old systems are withdrawn. This would enable networks to be equipped with the new system in a staged manner, without compatibility problems or need for dual equipment on track. The new system would often be introduced in conjunction with resignaling on heavily used lines, while on more lightly used lines it would replace the old signaling completely.

It should be emphasized, however, that this project is only at an early stage. These preliminary thoughts could change as ETCS develops.

Automatic Train Protection on British Rail: Present Plans and Future Possibilities

K. W. Burrage

British Rail has embarked on a 3-year effort to test and select an automatic train protection (ATP) system for national implementation beginning in 1992. Two systems are being pilot tested—one on a high-speed mainline, the other on a suburban railway. For economic reasons, both give intermittent ATP coverage with information transmitted at each signal, rather than continuously as with some metro or high-speed applications in continental Europe. Whichever system is chosen, it must be possible to fit the equipment to a wide variety of vehicles and lineside signaling with minimum disturbance to existing equipment. British Rail plans to finalize its national ATP specification late in 1991 and is committed to a 10-year installation program. Although the program will require considerable investment, British Rail expects the ATP system to yield a number of benefits besides reversing the rail system's rising signals-passed-at-danger (SPADs) statistics.

During the 1980s the number of signals being passed at danger (SPADs) has shown a significant increase and is a matter of serious concern to British Rail:

- Since 1982, SPADs have been on the rise.
- In 1988, there were 843 such incidents, 87 of them derailments or collisions.
- In 1989, there were 963 such incidents.

This concern has been brought into sharp focus by a number of passenger train accidents—such as those at Bellgrove in Scotland and Purley in South London—in which a number of people were killed.

The phenomenon of SPADs has been thoroughly studied from human and technical viewpoints and the conclusion reached that human performance factors are the cause of 85 percent of all SPADs and of the rising trend. This has led to the search for a technical solution to prevent driver error in approaching signals and obeying speed restrictions, and for a solution that can be implemented relatively quickly: an automatic train protection (ATP) system.

The existing driver supervision system on British Rail, the Automatic Warning System (AWS), is based on 1950s technology and only alerts the driver to the presence of a warning or stop signal. It cannot differentiate between the two and has no speed supervision function. It was quickly concluded that this technology could not be adapted to fulfill the needs of ATP on British Rail.

Signal and Telecommunications Engineering, British Railways Board, CP 10 Macmillan House, Paddington Station, London W2 1FT, United Kingdom.

THE PLAN

In autumn 1988, British Rail made a decision to embark on a 3-year program to produce an ATP system that could be available for implementation by early 1992. Given the 3-year timetable, it was impossible to pursue original technical development of a completely new system.

Therefore, British Rail set out to produce a performance specification for the perceived needs of its network and, at the same time, to review the experience of other railway administrations with ATP systems. Then, two pilot schemes were to be run to assess in practice the performance of the proposed specification on British Rail lines, leading finally to the procurement of a system for national application (Figure 1).

PILOT SCHEMES

British Rail decided to engage in large-scale trials (pilot schemes) of two proprietary ATP systems to gain experience before choosing a national system. The choice of an ATP system for a network has significant implications for the capacity and performance of that network as well as for its safety. It can also profoundly influence future signaling and train control strategy. Mistakes could be extremely damaging to the business and very costly to rectify.

Two pilot schemes were devised on two very different routes, one a high-speed mainline, the other a suburban route:

- Great Western Main Line ATP Project
  - —Diesel-operated, high-speed mainline trunk railway,
  - —London (Paddington) to Bristol via Bath and via Bristol Parkway,
  - —143 route-miles, two tracks to be fitted,
  - —125 mph maximum speed,
  - —350 signals to be fitted,
  - —100 Class 253 high-speed diesel trains to be equipped.
- Chiltern Lines ATP Project
  - —Diesel-operated suburban railway,
  - —London (Marylebone) to High Wycombe and Aylesbury,
  - —10 trains per hour into London in the peak period,
  - —67 route-miles, mainly two-track,
  - —75 mph maximum speed,
  - —188 signals to be fitted, and
  - —66 Class 165 Network Turbo diesel multiple units to be equipped.

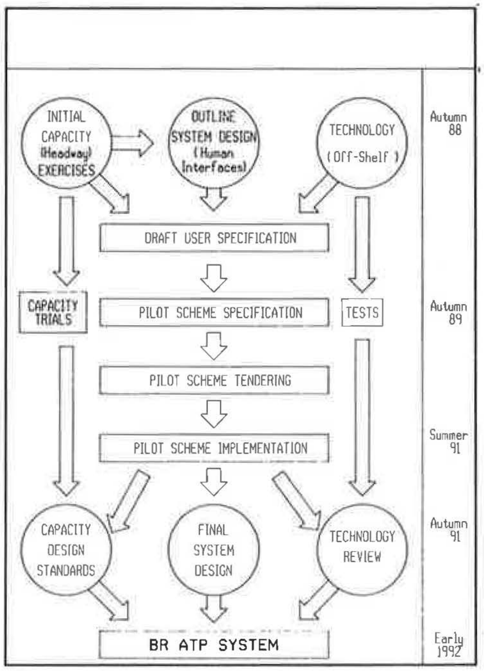


FIGURE 1 ATP development plan.

To select the systems for trial, a competitive tendering exercise was carried out. This used the performance specification as its basis, leaving industry to propose suitable technical solutions based on available technology. A detailed appraisal of the tenders was carried out against performance, technical, operational, and financial criteria.

The response of industry was most encouraging and the two successful companies selected for the pilot trials were ACEC of Charleroi, Belgium, which offered a system for the Great Western Main Line based on their TBL equipment in service with SNCB, and GEC-Alsthom Signals, UK, which offered the SELCAB system from SEL, Germany, for the Chiltern Lines. SELCAB is derived from SEL's continuous ATP products.

Both contracts were let early in 1990, giving the contractors a very demanding timetable to meet, for operation of parts of the pilots was needed by summer 1991. Apart from the needs for physical manufacture, some development of hardware and software had to be undertaken to meet the particular operating and technical environments of the two schemes. These challenges have been met, and installation commenced to plan in the first quarter of 1991. The rest of the year will see a major program of testing and evaluation.

THE SYSTEM

In outline and functionality, both systems in the pilot schemes are very similar. The vehicle equipment is shown in Figure 2 and the on-track equipment in Figure 3. The trackside equipment transmits to the train all the information required for the vehicle on-board computer to safely monitor the train's performance to the next transmission point. The information provided is either fixed or derived from the signaling system. Information provided to the train by the ATP at a beacon or loop is as follows:

- Aspect and type of signal,
- Distance to next beacon/loop,
- Distance to stopping point,
- Safe distance available beyond stopping point,
- Maximum line speed,
- Distance to speed restriction,
- Value of speed restriction,
- Length of speed restriction,
- Gradient,
- Differential speed restriction,
- Length of loop, and
- Beacon/loop identity (Great Western system only).

The systems employed give intermittent ATP coverage with information transmitted at each signal, rather than continuously as is the case for metro or some high-speed applications in Europe. The prime reason for this is economic—continuous coverage is much costlier than intermittent and could not be justified for other than relatively small portions of British Rail's network. The downside of an intermittent system can be a degraded performance of the network. This can be overcome by "infilling" part of the signal section to update a train in advance of the signal it is approaching. Both British Rail pilot scheme systems can achieve this with cable loops and, on one, extra beacons are an alternative. Extensive sim-

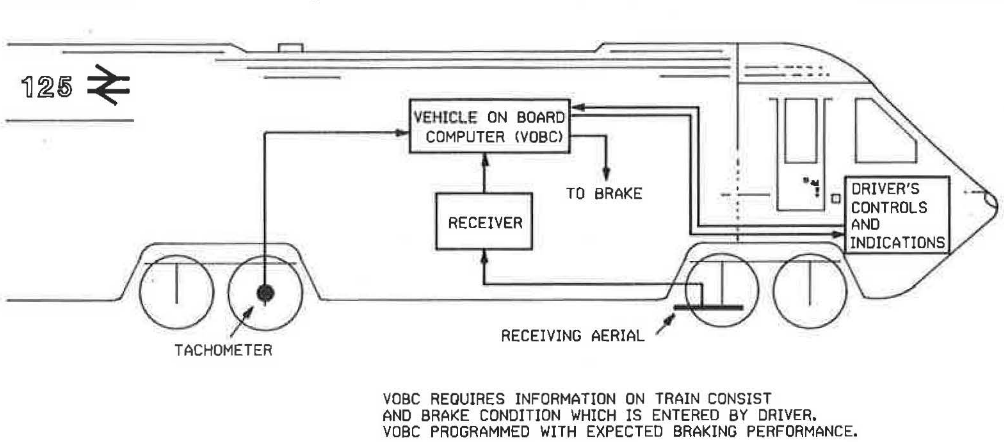


FIGURE 2 ATP train-borne equipment.

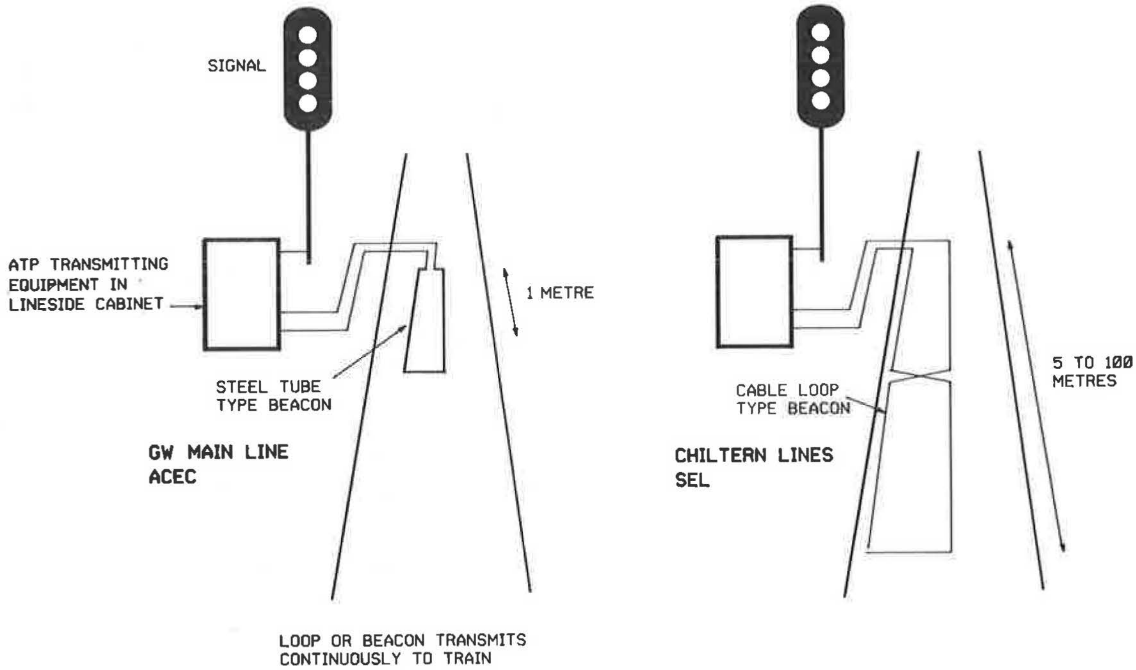


FIGURE 3 ATP trackside equipment (not to scale).

ulations have indicated that, by selecting the level of infill to suit prevailing traffic conditions, performance of the network can be maintained or even enhanced, if this is justifiable (Figure 4).

The loops or beacons being used on the pilot schemes are active (i.e., they transmit continuously to receivers on the trains and hence require the provision of lineside power). The systems only require transmission equipment at signals, so this presents no problems on the British Rail network because power is available at these locations.

One of the key requirements for both the pilot and the national ATP schemes is ease of fitting to the network. It must be possible to fit the equipment to a wide variety of types and ages of vehicles and lineside signaling with minimum disturbance to the existing equipment.

At the lineside, innovation has been used on the Great Western Main Line to allow the ATP electronics to "read" the signal aspects by means of a device that replaces a link in the lineside equipment cabinet. This will allow installation of ATP with minimum disturbance to existing wiring and limit the amount of the testing required of the relay-based signaling system in use there. On the Chiltern Lines, the ATP electronics connects directly to the outputs of the electronic module driving the lineside signal because this route was recently resignaled with British Rail solid-state interlocking equipment. After the choice of a system is made for national ap-

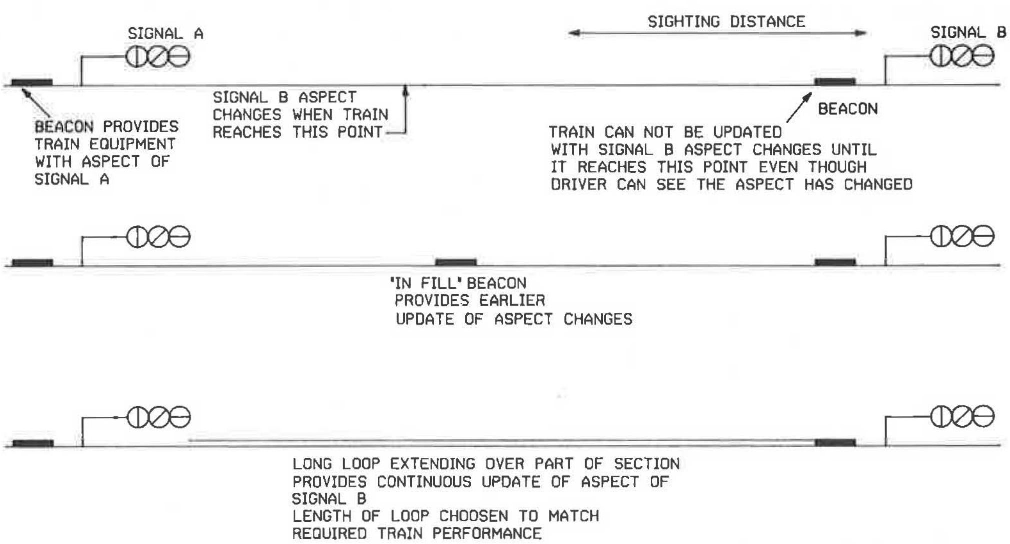


FIGURE 4 Provision of earlier updating of signal aspect changes "infill."

plication, development work will be undertaken to integrate the two sets of lineside electronics.

On board the vehicle, the receiving antenna is mounted on the leading bogie. The size of the antenna and its mountings is critical for economic fitment to vehicles whose bogies were not designed with this in mind. Space has to be found for the vehicle on-board computer, which can occupy typically 12U of rack space. This has been straightforward for the Intercity 125 trains on the Great Western Main Line, which have ample spare space in the parcels area. But on the Class 165 units on the Chiltern Lines, space is at a premium—as it will be on many of the types of multiple units to be fitted nationally. It is anticipated that the space requirements can be reduced for national application.

The cab display chosen has been based around the speedometer, and each vehicle will require a change of speedometer to accommodate ATP. This will, however, mean that the display can be fitted into all cabs and in a place readily observable by the driver. Figure 5 shows the display and associated driver's control.

The driver also has a key pad for entry of the train consist data and braking condition. Both this and the display are undergoing considerable ergonomic analysis to ensure that the optimal safe solution is found. It is quite possible that a redesign of the display could occur before national implementation.

NATIONAL IMPLEMENTATION

At the end of 1991, British Rail plans to be in a position to finalize the national specification for ATP and to engage in the selection of the system to meet that specification. The selection process will be vigorous and will cover many technical, safety, and performance issues as well as cost.

The problem of how much of the network and fleet is to be fitted will also have to be addressed. It is likely that all locomotives and multiple units will be fitted in the program—the nature of the British Rail network and use of vehicles means that all vehicles traverse busy sections during the course of their regular duties. To mix fitted and unfitted trains would create a significant gap in the protection afforded by the system. Similarly, gaps in the infrastructure fitted must not be frequent or widespread—the driver must be given a consistent level of security on the journey. Again, the nature of the network is such that only very minor routes are likely to remain unfitted in the longer term.

Having decided the level of fitment and chosen a system, the major challenge will then be to implement those decisions

- On the shortest practicable timetable,
- With maximum impact on safety, and
- With the optimum use of scarce resources.

British Rail is committed to fitment on the minimum practical timetable—the current judgment is that it will take 10 years. The reason for this is primarily one of resources, especially with regard to rolling stock. Each cab will take time to fit for ATP, and a float of vehicles from British Rail's wide variety of fleets will be required. It is British Rail's ability to generate this float and still carry the traffic on offer that is currently seen as a key factor pointing towards a 10-year program.

Current estimates of the cost of fitting ATP to the British Rail network are around £600 million (£1 = $1.67 U.S.); the final total will be clearer after the results of the pilot trials. This is a significant sum—especially when set alongside British Rail's current investment in signaling and train control of around £100 million per annum. This signaling investment is also vital to the continuing safe and efficient functioning of the network as it will replace aging 1950s and 1960s equipment on some key routes. ATP, therefore, requires additional fund-

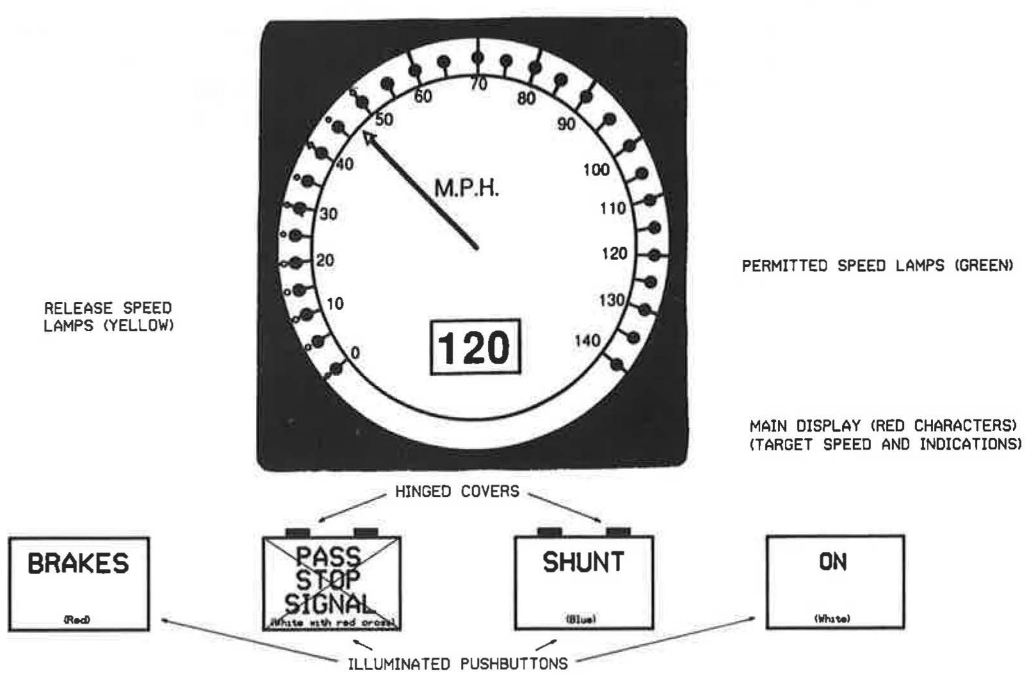


FIGURE 5 Driver's controls and indications.

ing and British Rail is in debate both internally and with government as to the preferred way of generating these funds.

FUTURE DEVELOPMENTS

The ATP system chosen for British Rail will have considerable potential for future enhancement, both in itself and by enabling other developments to be considered that were previously impractical or uneconomic. All the information required to control the train will be resident in the on-board computer and a communications link established between the signaling and the train.

To illustrate the strategic potential of this system, a 1987 study indicated that a move toward in-cab signaling generally on the British Rail network would not be economic, primarily because of the cost of fitting all the requisite rolling stock. With the advent of a national ATP system and the equipping of track and trains with most of the necessary equipment, the balance of the equation shifts. This is especially so if the cab display requirements for ATP and cab signaling can utilize the same hardware—this aspect is undergoing investigation by British Rail.

Other benefits to flow from a move to cab signaling on British Rail include the following:

- Reducing signaling installation costs by replacing a signal with a marker board,
- Reducing maintenance costs by eliminating lamps at the lineside,
- Flexibility over block lengths and layout design,
- Eliminating complex aspect sequences and controls, and
- Potentially reducing driver route-learning needs.

Another development enabled by ATP will be the advent of trains running at speeds over the current maximum permitted in the United Kingdom of 200 km/hr. The InterCity business has identified potential benefits of being able to run at 225 km/hr over some existing routes and is considering 250 km/hr. The United Kingdom's regulatory body has stipulated that ATP is a safety requirement for speeds over 200 km/hr.

The ATP system will also enable authority to travel at higher speed to be given to the driver in the cab, rather than by introducing extra lineside aspects. However, some modifications will still be required to the lineside signaling to provide the ATP ground equipment with the information that would otherwise have been displayed by a signal.

Over and above ATP's safety signaling role, it may be possible to include advisory speeds for train regulation purposes over the ATP track/train link. Benefits of energy saving and a quicker service recovery from disruption could result. The control center computers and the communications link to the lineside both require development for this to be realized, but the advent of ATP is, again, an enabling factor.

It will also be possible to add train/track communication to the ATP infrastructure. This opens up the following possibilities:

- "Moving block" in heavily trafficked areas such as the proposed Crossrail tunnel under London, and
- Train detection moving away from lineside systems such as track circuits and axle counters toward a train-based system.

The economics of all these future possibilities depend on British Rail's identifying a business benefit from adopting them. The lessons of the past and of recent studies into innovation on the United Kingdom's network have indicated that, because of British Rail's intensively signaled railway and integrated networked operations, this is best achieved by an evolutionary process.

This latter point is particularly important to bear in mind when considering moves to international standards, especially for the European networks. The economics of railway operation in Europe are such that infrastructure assets need to be exploited to the end of their natural working lives. The funds and the resources are not generally available to engage in premature renewal. The new high-speed lines provide a rare opportunity to start with a clean slate and it is unlikely that the crowded United Kingdom will see many of these. Exploitation of new techniques and possibilities will come from the use of intelligence on the train, enabling it to run over a variety of signaling systems. ATP provides that intelligence on the train for the first time on British Rail.

Value of High-Quality Service: How Should the ARES-Equipped Railroad Operate?

Michael E. Smith and Randolph R. Resor

The North American railroad industry is beginning to implement new, advanced train control technologies that will significantly change railroad operations. Collectively, these technologies are referred to as the Advanced Train Control Systems (ATCS). Burlington Northern Railroad's specific version of ATCS is called the Advanced Railroad Electronics System (ARES). In an extensive operations analysis, Burlington Northern found that it could use ARES to greatly reduce costs or improve service, or a little of both. To determine the optimal course of action, the railroad conducted a market study to determine the value of better service. The study results indicated that custo.ners were willing to pay much more for small improvements in service. Burlington Northern's market managers disputed this, believing that better service would not significantly increase prices or market share. Meanwhile, ARES operations analysis indicated that operating improvements could be targeted very precisely using the new technology. Travel time improvements could be allocated at will among various classes of trains. Given these results, Burlington Northern should concentrate its initial implementation of ARES functionality on reducing cycle times for bulk commodity trains. This will result in the need for fewer coal sets to move a given amount of coal. Then, the railroad will receive the certain payoff of reduced assets as opposed to the uncertain payoff of increased revenue. As implementation proceeds, Burlington Northern should use ARES capabilities to test the value of improving service. Then implementation strategies can be adjusted to improve the outcome.

The introduction of new technology into an existing operation often provides the opportunity to improve the processes making up the operation. Advanced Train Control Systems (ATCS) have the potential to improve railroad operations. Through investments in this new technology, railroads have the opportunity to lower costs and improve service simultaneously.

This happy state of affairs does not frequently present itself. Usually, providing an improved level of service requires that more resources (more cost) be used in the operation. An existing process can be used to translate input resources into desired outputs. More output, or higher-quality output, requires more input resources.

Now, suppose the process is changed. A new set of assets, better equipment, is used to produce the output. Now, less input is required to produce the same output. An example is a tailor using a sewing machine instead of hand stitching to make clothes. The tailor can now produce a garment in an hour when it used to take a day. This is a cost-reducing approach to using the new equipment. There is, however, a quality improvement approach. With a sewing machine, the tailor may be able to spend all day on the garment and produce one of higher quality.

Which approach is best? Should the machine be used to produce the output using fewer input resources (and reduce cost)? Or should it be used to produce the same quantity, but higher-quality, output (and improve quality)? These same questions are equally valid for ATCS.

The Burlington Northern Railroad is struggling with the decision as to whether it should install its version of ATCS, the Advanced Railroad Electronic System (ARES). In the process of making that decision, the railroad analyzed both the cost reduction and quality improvement potential of ARES. The way in which ARES is used and designed depends on the relative values of improving quality and reducing cost.

Two alternatives present themselves for the use of ARES on Burlington Northern—one that improves quality and reduces cost, and one that only reduces cost. The different methods of looking at the value of higher quality need to be weighed along with the impact on design and operations that the alternative ways of using ARES will impose.

DESCRIPTION OF ARES

In the process of developing ARES, Burlington Northern has spent considerable effort in determining how the system affects existing processes from the top down. This discussion presents a top-down flow of the control and process envisioned by ARES and then describes the individual components used in that process.

For a top view of ARES, consider Figure 1. This shows a hierarchical planning and command structure for a railroad operation. First, a planning and scheduling group establishes a schedule for operating the system. Depending on the level of sophistication, this could include train schedules, terminal schedules, equipment rotation plans, and maintenance schedules.

The first component of ARES is shown just under the dotted line in Figure 1. This top-level component is called the strategic traffic planner (STP). The function of the STP is twofold. First, it will translate the train schedule into time goals and priorities for trains handled by each dispatcher. Second, if schedules cannot be met, the STP will assist system

M. E. Smith, ARES Traffic Planning Systems, Burlington Northern Railroad, 9401 Indian Creek Parkway, Overland Park, Kans. 66201. R. R. Resor, Costing and Pricing, Zeta-Tech Associates, P.O. Box 8407, Cherry Hill, N.J. 08002.

FIGURE 1 Hierarchy of ARES command and control functions.

control personnel in selecting alternatives that do the least damage to the network.

The second component of ARES is shown as the next level down in the hierarchy. This component is called the tactical traffic planner (TTP). The TTP assists the dispatcher in finding the most efficient meet-and-pass plan for each train in the dispatcher's territory. Further description of how the STP and TTP are intended to work can be found in Smith and Resor (*1*).

Once the TTP has generated a meet-and-pass plan, it passes that information to a computer-aided dispatching tool that automatically generates authorities to vehicles in the field. Using global positioning system (GPS) location technology and digital data radio, field vehicles are continuously tracked to ensure that they are operating according to plan.

If a train is not operating according to plan, the engineer will be prompted to speed up or slow down, as appropriate, by the display on board. If the train is falling behind schedule, the engineer will be prompted to give a reason for the delay. If the on-board computer determines that the train will not arrive at the next significant event location (e.g., meet location or terminal) within the time required, a signal will be sent to the TTP requesting a new plan. This provides a closed-loop process that ensures the system responds to external disturbances in an optimal way.

The response of the system depends greatly on how "optimal" is defined. The TTP can provide meet-and-pass plans that minimize travel time, minimize deviation from schedule, minimize either of those with different weights for each train, minimize fuel, or any combination of the preceding. Which should be done?

When Burlington Northern evaluated ARES, it was assumed that the system would be used to minimize travel time, weighted by the value of each train. The results showed that this approach could reduce the cycle time of bulk trains and increase the reliability of carload freight trains. That left the task of determining how much each of these things was worth.

The worth of decreased cycle time for bulk trains was easy to understand. It simply amounted to the savings generated by requiring less equipment sets to move the same amount of commodity. The worth of increased reliability for carload freight was more difficult to determine. It depended on how much better service was worth to the customer and on how much of that value the railroad may be able to extract through increased marketshare or price.

To address this difficulty, Burlington Northern evaluated ARES with two different objective functions for the TTP. In the first objective function, the evaluation minimized train travel time weighted by the cost of train delay for each train type. This train-delay cost was based on both the value of the railroad equipment and the value of delay to the lading (*2*). This approach assumes implicitly that reliability to the customer is important and valuable. In the second objective function, the evaluation minimized travel time on bulk trains with no increase in the delays experienced by other trains. This approach implicitly assumes that additional reliability to the customer has no value to the railroad.

Which approach is right? How should ARES be used when it is put into service? That depends on how much reliability can be improved by reducing train delay and how much that is worth to the customer. Further analysis pointed toward the answers. Simulations showed how the operation would improve with the two different objective functions and the value of improved service to the customer was estimated.

THE IMPACT OF ARES ON SERVICE

ARES, as designed by Burlington Northern and Rockwell International, will incorporate a sophisticated optimization program that will provide dispatchers with a mathematically "best" dispatching plan. However, users must specify the inputs—such as the cost of train delay—that are used to determine the optimum. Different train-delay costs and even different delay functions (e.g., quadratic versus linear) can produce radically different dispatching plans.

The meet/pass planning algorithm within ARES is intended to devise an optimal meet/pass plan for each "planning line" (segment of railroad). This plan will minimize a weighted combination of fuel consumption and running time, taking into account the differing priorities of different trains. It is important to note that ARES will use weighted priorities rather than "cardinal" train priorities. That is, rather than being ranked by category, with trains in the highest category always receiving priority over lower-ranked trains, trains will be assigned varying delay costs. Thus, an intermodal train may sometimes be delayed for a lower-priority train if by doing so the total delay cost is minimized.

Computer-aided meet/pass planning is expected to yield large benefits for Burlington Northern when used in conjunction with the real-time location information provided by ARES. Areas of benefit include an increase in line capacity due to less time lost in making meets, improved reliability, and fuel savings through the avoidance of unnecessary stops and through "pacing" of trains so that they arrive at meet points exactly on schedule, rather than proceeding at maximum speed, arriving early, and having to sit and wait for an opposing train. However, these benefits cannot all be realized simultaneously; that is, a railroad cannot achieve increased capacity, improved reliability, and reduced fuel consumption together.

To evaluate the effect of this optimization program on Burlington Northern operations, a series of simulations was carried out using actual train movement data. Although the optimization model used (the SChedule ANalyzer or SCAN, developed for the University of Pennsylvania) is not identical to the model developed for use in ARES, it was thought that SCAN would provide a good approximation of the benefits to be expected from the use of dispatching optimization within ARES.

Data used in this study were originally gathered in spring 1988 from dispatchers' sheets and centralized traffic control (CTC) on switch (O/S) reports for 12 "lanes" covering a total of about 3,500 route-miles. They include a mix of CTC, automatic blocking system (ABS), and dark territory. On each lane, train movement data were gathered for a period of about 24 hours; on lanes with light volume, the period was longer, while on lanes with very heavy volume the period was as short as 16 hours. The number of trains on each lane varied from 7 (Madill to Irving) to 45 (Alliance to Edgemont). A total of 846 trains were included in the analysis.

To use the SCAN model, several data elements are required:

- Scheduled, or desired, running time for each train;
- Location and duration of all delays to trains;
- A delay cost for each train type;
- Route topology, including speed limits, grades, and siding locations;
- An estimated unconstrained (minimum feasible) running time for each train; and
- A cost per gallon of diesel fuel.

Train-delay costs were developed with the help of Burlington Northern accounting and marketing personnel and the Association of American Railroads. Burlington Northern calculates an hourly ownership cost for each piece of railroad-owned equipment; to this cost was added an estimated value of the lading carried. To estimate lading value, four broad categories of trains were created:

- Loaded bulk (coal, grain, ore),
- Empty bulk (empty return movements),
- Mixed freight (carload traffic), and
- Intermodal (trailers and containers).

For each category, an "average" train was defined. Table 1 shows the cost per hour of equipment, of lading, and the total cost used in the simulations.

Fuel cost was assumed to be 50 cents per gallon in all cases. (The analysis predated the recent upheaval in the Middle East.) Energy consumption for each train was calculated from gross weight, horsepower, and the route topology, and was then converted to fuel consumption by the use of appropriate factors.

Delays, along with train consist information, were obtained from dispatching records. Route characteristics were taken from track charts and timetables. Minimum feasible running times were estimated by use of Burlington Northern's train performance simulator.

The most difficult problem was the development of schedules for all trains. Many trains on Burlington Northern do not have schedules, and many that do will often run hours ahead of or behind scheduled times. For the purposes of this analysis, then, the desired schedule was assumed to be the same as the actual time each train operated. Actual running time was further assumed to be an upper bound; in determining the benefits of computer-aided dispatching, the first analysis tried to better the actual running times of all trains in the lane. Later, in a sensitivity analysis, running times of certain train types were held constant. These results will be discussed later.

Table 2 shows the results of the analysis by train type. Two kinds of mixed freights have been defined here; "priority" freights carry more time-sensitive commodities and have more locomotive horsepower assigned per ton than "secondary" freights.

These results clearly show that the greatest improvement in both mean running time and in variability of running times occurs for low-priority trains. This finding seems intuitively correct; a normal human response in situations where many different items must be considered simultaneously is to handle the most important ones first. Human dispatchers are simply letting the low-priority trains sit until there is time for them.

Table 3 summarizes the percentage changes in mean travel times by train type that are expected to occur with the installation of ARES.

When the results of this optimization were considered, a question arose about the different ways in which optimization might be used. The initial focus had been on minimizing total travel time for all trains. But suppose for a moment that the high-priority trains were already running as fast as market conditions required. To put it another way, let us say that there is no additional revenue or marketshare to be gained from shorter transit times for intermodal and mixed freight trains. Can running times of bulk commodity trains be reduced even further?

Most coal on Burlington Northern moves under contract in fixed annual volumes. The cycle time (round-trip time plus loading and unloading time) is known approximately for each coal movement. Thus, although these trains do not have schedules as such, equipment needs are determined by the number of train sets (cars plus locomotives) required to move the contractual coal volume. The longer the cycle time, the more equipment is required.

Burlington Northern moves a great deal of coal, and Burlington Northern and shippers have dedicated a large fleet of cars and locomotives to this service. Significant reductions in the running times of coal trains could reduce equipment requirements very substantially. Potential savings are very large.

TABLE 1 AVERAGE DELAY COST BY TRAIN TYPE

TRAIN TYPE	LADING DELAY COST ($/train hour)	EQUIPMENT DELAY COST ($/train hour)	TOTAL DELAY COST
Loaded Bulk	10.37	172.00	182.37
Empty Bulk	0.00	172.00	172.00
Mixed Freight	35.98	127.00	162.98
Intermodal	136.47	130.00	266.47

TABLE 2 COMPARISON, MEAN, AND STANDARD DEVIATION OF RUNNING TIME FOR ALL TRAINS BY CLASS (IN MINUTES PER TRAIN)

TRAIN TYPE	BASE CASE		OPTIMIZED CASE	
	MEAN	STANDARD DEVIATION	MEAN	STANDARD DEVIATION
Bulk	253.96	125.39	190.95	102.37
Intermodal	196.09	114.60	175.61	105.77
Priority Freight	203.71	149.96	163.31	120.57
Secondary Freight	245.90	132.29	179.13	94.08
All Trains	227.52	130.76	179.37	105.08

TABLE 3 SUMMARY OF CHANGES IN MEAN TRAVEL TIMES—BASE CASE

TRAIN TYPE	TRAVEL TIME CHANGE
Intermodal	(10.4%)
Priority Freight	(19.8%)
Secondary Freight	(27.1%)
Bulk	(24.8%)
All Trains	(21.1%)

SCAN allows the user to specify delay-cost functions for each class of trains. In the original analysis, a large cost penalty was assigned to lateness (running time exceeding the actual time each train took to traverse a lane), while a linear savings was assigned to earliness. Thus, savings could be achieved by running trains ahead of schedule.

A second and more limited analysis was carried out in which a large penalty was assigned to intermodal and mixed freight trains for both lateness and earliness. Thus, the SCAN model tended to try to run these trains as close to schedule as possible. For bulk trains, by contract, a linear benefit was assigned to earliness, just as in the original analysis. By doing this, it was hoped that all the benefits of optimization on bulk trains could be conferred.

The results of the analysis, shown below, were not surprising.

Train Type	*Travel Time Change (%)*
Bulk	32.9
All other	0.9

Although the total benefits were less in this constrained case (as might be expected), the net reduction in travel time for the bulk trains was nearly 33 percent—without any penalty to the mixed freights or intermodal trains.

A one-third reduction in the running time of every coal train on Burlington Northern will produce a very substantial benefit in terms of reduced equipment purchase requirements. Clearly, dispatching optimization within an ATCS can produce very substantial benefits. How the benefits are allocated among traffics will be determined by the perceptions of railroad managers regarding the value of service. Faster and/or more reliable schedules can be operated for high-priority trains, or for all trains, or alternatively the benefits can be taken entirely in the form of reduced cycle times (and therefore reduced equipment requirements) for the low-priority trains. There is a tendency to regard service quality improvements as a "soft" benefit, while savings in equipment requirements are a "hard" benefit. Historically, railroads have favored the hard benefits—and cost minimization—over the soft benefits of service quality improvements.

VALUE OF IMPROVED SERVICE

It became obvious from the operations analysis that the benefits of ARES, as well as the best way to use it, would depend greatly on the value of good service to the customer. Therefore, the marketing department on Burlington Northern was consulted in estimating this value. Based on the recommendation of the marketing department, the John Morton Company (JMC) was retained to perform the study.

JMC used a method called conjoint analysis to interview customers and map their preferences. This analysis method is not described here; there is sufficient description in Johnson (*3*) and in Johnson and Squeo (*4*). Briefly, conjoint analysis is a very sophisticated, computer-based interview technique that ensures that the respondent's preferences have been accurately mapped. The one weakness of the technique is that it maps what customers say they will do, not what they actually do (more on this later).

With the assistance of the marketing department, five commodities were selected for study: pet foods, aluminum, plastics, paper, and tires. The marketing department believed that these commodities were representative of the carload freight market.

The marketing department was then asked to assist in defining service variables that should be measured in the survey of the customers. Although there were nine service variables defined, only the three most important are presented here:

1. Reliability of cargo delivery—the percentage of time that a loaded car arrives at the customer's dock within the time window desired by the customer;
2. Reliability of empty equipment delivery—the percentage of time that a customer's request for an empty car is satisfied with an acceptable car within the time window desired by the customer.
3. Dock-to-dock transit time—the time required for the shipment to move from the shipper's dock to the consignee's dock.

In their survey of the customers, JMC intended to develop estimates of elasticity. That is, for each service dimension, how much marketshare gain could be expected from a 1 percent improvement in that dimension? Also, JMC estimated price elasticity. That is, for each 1 percent increase in price, how much marketshare would be lost? From these two numbers, service-price cross elasticities could be estimated. That is, for each 1 percent improvement in service, how much can price be increased without losing marketshare?

The results of the JMC study are shown in Table 4. These results were quite startling. If the customers have revealed what they really would do, there is tremendous potential for increased revenue from even small increases in service levels. For example, if the reliability of cargo delivery is improved by just 1 percent, these surveyed customers say they would be willing to absorb an average price increase of 4 percent. This would imply that ARES should be used primarily for improvements in reliability, not reductions in operating cost.

The marketing department at Burlington Northern was skeptical that these elasticities were truly representative of customer behavior. Sure, the customers may say they will do that, but will they really? And even if they were willing to pay extra for improved service, would they be sure enough that the service had actually improved? And, were we perceptive enough to capture all that a customer would be willing to pay when it came time to negotiate the price?

These nagging doubts led to a search for studies that had been done based on data representing actual customer choices as opposed to customers' stated preferences.

Kansas State University (KSU) had performed some analyses in the 1970s that used regression techniques to correlate rail and truck services to actual customer choices. After some discussion, we decided that insufficient data existed to support KSU's efforts to prepare similar models specific to Burlington Northern. However, KSU did provide their estimates of service elasticities from their previous studies (see Table 5).

But KSU did not provide price elasticities or cross elasticities. These price and cross elasticities were estimated based on an assumed range for price elasticity that will be discussed later. (Assumed quantities are shown in the table as italics.) The KSU numbers are based on a different definition of service. They defined service as car-miles per car-year. The presumption is that if the railroad is providing better service, cars will turn faster. KSU used this surrogate for service because there was no better measure from publicly available data.

Because KSU was unable to develop a Burlington Northern–specific analysis, and because their existing analysis used a questionable definition for service, Burlington Northern's market managers were interviewed and asked to forecast the market gains that could be achieved from improved service. The results of that survey are shown in Table 6.

TABLE 4 RESULTS OF JOHN MORTON COMPANY SURVEY

COMMODITY	SERVICE ELASTICITIES			PRICE ELASTICITY	CROSS ELASTICITIES		
	CARGO RELIAB.	EMPTY RELIAB.	TRANSIT TIME		CARGO RELIAB.	EMPTY RELIAB.	TRANSIT TIME
Paper	6.0	3.3	-1.1	-1.1	5.5	3.0	-1.0
Pet Food	6.9	2.8	-1.4	-1.5	4.6	1.9	-0.9
Aluminum	4.3	1.9	-1.3	-1.3	3.3	1.5	-1.0
Plastics	4.7	2.1	-0.9	-1.6	2.9	1.3	-0.6
Tires	6.2	2.3	-1.6	-0.9	6.9	2.6	-1.8
AVERAGE	5.3	2.5	-1.2	-1.3	4.1	1.9	-0.9

TABLE 5 RESULTS OF KANSAS STATE UNIVERSITY RESEARCH

COMMODITY	SERVICE ELASTICITY	PRICE ELASTICITY		CROSS ELASTICITY	
		LOW	HIGH	LOW	HIGH
Food Products	0.7	1.3	3.0	0.2	0.5
Tobacco Products	1.4	1.3	3.0	0.5	1.1
Textile Products	1.3	1.3	3.0	0.4	1.0
Lumber & Wood Products	0.6	1.3	3.0	0.2	0.5
Furniture	1.1	1.3	3.0	0.4	1.2
Paper Products	0.6	1.3	3.0	0.2	0.5
Chemicals	1.0	1.3	3.0	0.3	0.8
Stone, Glass & Clay	1.5	1.3	3.0	0.2	1.2
Primary Metal Products	1.8	1.3	3.0	0.6	1.4
Fabricated Metal Products	3.0	1.3	3.0	1.0	2.3
Non-Electrical Machinery	4.3	1.3	3.0	1.4	3.3
Electrical Machinery	1.7	1.3	3.0	0.6	1.3
TOFC/COFC	1.2	1.3	3.0	0.4	0.9

TABLE 6 RESULTS OF MARKET MANAGER SURVEY

BUSINESS UNIT	SERVICE ELASTICITY	PRICE ELASTICITY	CROSS ELASTICITY
Industrial Products	0.1	3.0	0.03
Forest Products	0.2	7.0	0.03
Food & Consumer	0.5	1.2	0.4
Automotive	0.04	infinite	0
Agricultural	0.1	3.0	0.03

TABLE 7 ELASTICITY ESTIMATES

SOURCE	SERVICE ELASTICITY		PRICE ELASTICITY		SERVICE-PRICE CROSS ELASTICITY	
	LOW	HIGH	LOW	HIGH	LOW	HIGH
John Morton	4.3	6.9	0.9	1.6	2.9	6.9
Kansas State	0.6	4.3	1.3	3.0	0.2*	3.3*
Market Mgrs.	0.01	0.5	1.2*	infinite*	0	0.4

The market managers were asked to provide service elasticities and cross elasticities. The price elasticities were derived by dividing the service elasticity by the cross elasticity. The definition of service used for this exercise was equivalent to the one used by JMC, that is, the percentage of shipments that arrive in the customer's desired time window. The high estimate on price elasticity is infinite. This implies that Burlington Northern could capture the entire market for that commodity by a very small cut in its price. We felt that this might be unrealistic.

The most common estimate of price elasticity from the market managers was 3.0. This value was then used as the high estimate for price elasticity to derive implied figures in the KSU work. The low value used for that purpose was 1.3, equal to the mean price elasticity reported by JMC.

The results of all three studies are reported in Table 7. There is not much guidance from this table. The range of numbers here is so broad that it is impossible to determine with any degree of reliability whether an investment in good service will pay off.

IMPLICATIONS FOR ARES IMPLEMENTATION

The results of these analyses have significant implications for the implementation of ARES. First, what value of service should be used to determine whether ARES should be pursued? Second, how should the railroad implement ARES? The second question has two parts: (a) What form should the objective functions take? and (b) What implementation strategy should be used for phasing in the functions of ARES over different parts of the railroad?

Because there was a very wide range of estimates on the value of service, we believe that ARES implementation should follow three principles:

1. The initial fielding of ARES capability should concentrate on its capabilities for reducing cost.
2. The ARES design should allow for flexible objectives. That is, the same system should be usable to meet a wide variety of business goals.
3. The implementation process should include a plan for testing the value of improving service.

In following these principles, we have recommended that ARES be installed first on routes where traffic is predominately unit coal trains. If ARES succeeds as projected, substantial reductions in coal train cycle times should be achieved. This will allow Burlington Northern to reduce the number of train sets in service while hauling the same amount of coal. Alternatively, the railroad will be able to haul more coal without putting additional train sets in service.

Once this has occurred, Burlington Northern will then be well assured that ARES can deliver its promised benefits. As the railroad continues to spread ARES capability across the system, it will use ARES' inherently flexible methods for establishing objectives to test the value of service. When ARES is installed in areas where carload freight trains predominate, the railroad can set objectives for the system that allow for more reliable service. Then Burlington Northern will be able to test that more reliable service in the marketplace and see if it pays off.

REFERENCES

1. M. E. Smith and R. R. Resor. Keeping Trains on Schedule: On-Line Planning Systems for the Advanced Railroad Electronics System (ARES). *Journal of the Transportation Research Forum,* Vol. 31, No. 1, 1990, pp. 17–24.
2. M. E. Smith and R. R. Resor. The Influence of Train Delay Costs on Optimal Train Dispatching Plans. Presented at 5th World Conference on Transport Research, Yokohama, Japan, 1989.
3. R. M. Johnson. Trade-Off Analysis of Consumer Values. *Journal of Marketing Research,* Vol. 11, May 1974, pp. 121–127.
4. R. M. Johnson and D. G. Squeo. A Proven Technique for Product Development. *Bank Marketing,* April 1982, pp. 20–24.

Management and Information Systems Components of Successful ATCS

PATRICK T. HARKER AND JEFFREY WARD

Much of the focus in the development of and debates surrounding Advanced Train Control Systems (ATCS) has centered on the technical aspects of the various hardware and software components that constitute such a system. However, numerous failures of advanced technologies in the service sector point to the need for careful consideration of the organizational and strategic needs of such a system before final design. One way of determining these needs is by looking at how ATCS can be used to support the overall strategy of the railroad. Once this relationship between ATCS and the railroad's strategy has been defined, a hierarchy of intelligent information systems components vital to linking strategy and implementation within the ATCS context can be formulated.

The development of Advanced Train Control Systems (ATCS) has primarily focused on the technical aspects of train and track hardware. Although such research is important, it often fails to "see the forest for the trees"; that is, the basic question of why ATCS are necessary and how they will affect the management of rail operations has received very little attention in the literature to date. The purpose of this paper is to analyze the management information system (MIS) needs of ATCS from the viewpoint of management strategy. Before this analysis, however, it is useful to first understand why this viewpoint is of vital importance.

As a component of the service economy, the railroad industry must carefully consider the implication of massive technological investments in terms of their impact on profit or loss. As described by Roach (*1*) and Hackett (*2*), the service sector has been a major consumer of computer technology in the past decade—it consumes at least 80 percent of all such investments and spends in excess of $3,000 per employee per year on computers. Unfortunately, the growth of productivity has not been even close to that which one would predict from such massive investments in technology. In fact, this lag in productivity has caused Hackett (*2*) to coin the term "service-sector sinkhole."

Why has this investment not been fruitful? As Hackett (*2*) and others (*3*,*4*) have noted, much of the problem lies in automation of poorly planned and executed production processes. The simple automation of a production system that is (a) outdated, (b) poorly organized, or (c) inappropriate for the type of service being produced will lead to little productivity growth, if any. The moral of this story is that management must first decide upon an operating strategy and related implementation plan before investing in technology. As Hackett (*2*, p. 403) states, "Companies that persist in developing their operating strategy solely from a technology perspective can look forward to tougher competition as their rivals realize the incremental benefits of an integrated, holistic approach" to technology and process change.

Failing to match strategy, organization structure, and information systems can be a serious pitfall. Only by forging congruent strategies, technologies, and structures can the effort succeed. To illustrate this relationship, two opposing scheduling strategies for freight railroads will be defined and the implications of these strategies on organization structure and information systems will be discussed.

TWO SCHEDULING STRATEGIES

The design implications of two polar cases of scheduling strategy are considered: a master scheduling strategy and a real-time scheduling strategy. The extreme cases are selected to illustrate the point that scheduling strategy defines the appropriate choice of organization structure and information systems for the scheduling function. In practice, scheduling strategies incorporate characteristics of each extreme, and therefore hybrid designs are warranted.

A *master scheduling strategy* is defined to be the periodic establishment of timetables that govern arrival and departure times based on a periodic review of demand levels and resource availability. Once established, the schedule would be in effect until the next scheduling period and subject to only minor revisions in the interim. This strategy is analogous to airline scheduling strategies. Airline departure and arrival times have long planning horizons and are based on forecasts of demand and resource availability. Although these schedules are subject to some revision, generally speaking, the flight will take place even if volume is low. Like the airlines, railroads pursuing a master scheduling strategy would publish timetables, which would serve as a marketing tool as well as a guide for operations supervisors.

Real-time scheduling operates over a considerably shorter planning interval. Under this strategy, timetables are continually revised as capacity on the network and demand in the marketplace change. Arrival and departure times would be set for an indefinite period. Real-time scheduling strategies are analogous to trucking schedules, which are determined daily according to demand and resource availability. Publication of timetables, because of the short planning horizon, would not be feasible.

A lack of flexibility characterizes master scheduling strategies because network interdependencies are only evaluated

Fishman-Davidson Center for the Study of the Service Sector, The Wharton School, University of Pennsylvania, Philadelphia, Pa. 19104-6366.

when the master schedule is developed. Deviations from the planned schedule would have dynamic consequences for the rest of the network; therefore, schedule revisions are discouraged. For example, operations supervisors would be discouraged from consolidating trains when volume is low.

Real-time schedules are highly flexible; the dynamic implications of changes to planned arrivals and departures are evaluated continually. For example, the impact of a proposed consolidation would be evaluated immediately and new timetables would be generated for dispatchers and yardmasters affected by the consolidation.

Master schedules must build large amounts of slack time into planned arrivals and departures (*5*). Because the schedule cannot be adjusted daily, enough slack time must be added to ensure that unforeseen events, such as unanticipated demand levels or equipment failures, do not affect system performance. In other words, because master scheduling strategies do not provide an efficient means of handling schedule exceptions, slack time must be added so that unforeseen events do not create such exceptions. Slack time in the schedule implies slack resources on the network in terms of equipment cycle times and excess infrastructure capacity (*5*).

A real-time scheduling strategy is characterized by small amounts of slack in arrival and departure times. The network schedule has tighter timetables, which means unforeseen events that would be absorbed by slack in a master scheduling strategy would constitute an exception under the real-time strategy. However, because real-time scheduling constantly updates timetables, these exceptions can be processed and new arrival and departure times produced for portions of the network affected by the schedule disruption (*5*). In short, real-time scheduling requires less slack resource, because the capacity of the network can be continually reallocated.

The need for coordination and communication is high under a real-time scheduling strategy. Exceptions to planned schedules must be quickly transmitted to a management group that will evaluate their impact and issue revised timetables to operations supervisors. Coordination and communication demands are low under a master scheduling strategy, because slack absorbs most exceptions and the timetable is revised infrequently.

Real-time scheduling strategies transfer the authority over local scheduling decisions to a group with network perspective; in contrast, master scheduling strategies are characterized by high autonomy for local supervisors. Because the master scheduling process is not designed to address variability in daily operations, dispatchers and yardmasters must be empowered to make immediate judgments. Granting this autonomy to supervisors surrenders control over some decision making (*6*). Control over local decisions can be exercised by establishing rules and procedures for operations (*5*). The master schedule, by prescribing arrival and departure times, functions as a system of rules and procedures. Rules and procedures cannot, however, anticipate all events.

To illustrate these differences, consider the respective response to a major schedule disruption. A master scheduling strategy would rely on local supervisors to recover from the disruption as much as possible and allow the impacts of the disruption to work through the system. The master schedule would again control operations after all the impacts of the disruption had been absorbed. The same disruption in a real-time scheduling environment would prompt immediate intervention by the network control group. Revised timetables would be established that more quickly restore the system to its performance objectives.

Under a master scheduling strategy, interaction with the marketing function would occur only at the master scheduling level. Marketing's input would be incorporated into the master schedule. But once the schedule has been established, revisions to accommodate new business or customer requests would be restricted. Real-time scheduling strategies, on the other hand, interact with the marketing function continually to evaluate the impact of scheduling decisions on customer relations or on proposed new business on the network.

In terms of reliability, master scheduling strategies yield more consistent performance (i.e., lower variance) than real-time strategies. Because slack time in the master schedule absorbs most disruptions, adherence to planned arrival and departure is high. Real-time scheduling, because of the continual reallocation of resources, produces inconsistent arrival and departure times. For example, under a master scheduling strategy, a train set to depart at 6 o'clock would do so each day. Under a real-time scheduling strategy, the same train departure might be delayed, advanced, or canceled on a given day.

By other measures, such as transit time or cost, real-time scheduling strategies may produce better results. If the objective is to minimize mean transit time, a real-time strategy produces better results, because of the reduction in slack time. For example, real-time scheduling allows trains to be released early, thereby providing more opportunities for cars to make tight time connections. If the objective is to minimize cost, then a real-time scheduling process produces better results, because of the reduction in slack resource. For example, real-time strategy provides opportunities to make rational train consolidations. Consolidations, of course, generate delays for some traffic. Thus, variance is higher under real-time scheduling. The differences between the two scheduling strategies are summarized in Figure 1.

ORGANIZATION DESIGN ISSUES

In the scheduling of freight railroads, three design elements (i.e., management functions) within the scheduling process are subject to organizational design decisions. The *master scheduling group*, located at the highest level of the organization, establishes timetables for the entire rail network. Im-

Master Scheduling ↔	Real Time Scheduling
- Low Variance	- Low Mean Transit Time
- Flexibility Low	- Flexibility High
- Slack Resource High	- Slack Resource Low
- Information System Cost Low	- Information System Cost High
- Scheduling Organization Costs Low	- Scheduling Organization Costs High
- Planning Horizon Long	- Planning Horizon Short
- Resource/Demand Information Uncertain	- Resource/Demand Information Certain

FIGURE 1 Strategy-structure continuum.

plementation of these schedules on a daily basis is the responsibility of *district dispatchers and yardmasters*. In between these two groups is a role for a group of managers who coordinate the interdependent actions of dispatchers and yardmasters, process exceptions to the schedule, and, in general, enforce the master schedule. Because of its position in the scheduling hierarchy, this function will be referred to as the *intermediate group*.

The emergence of three design elements in the scheduling process is attributable to the following dimensions along which these groups naturally differentiate themselves: their planning horizons, their goals and objectives, and the degree of uncertainty inherent in their task (*7*).

In terms of time, yardmasters and district dispatchers have short planning horizons (*6*). They schedule, over the length of their shift, switching in a yard or meets and passes in their district. The master scheduling function periodically establishes timetables that will govern the railroad's daily operations over a long time horizon. The intermediate group looks forward over several shifts or days and assesses the impact of unforeseen events on the planned schedule. It then revises the schedule to keep the system close to the performance objectives.

Temporal differences between the groups are also evident in the timeliness of feedback from decisions (*7*). Dispatchers and yardmasters receive prompt feedback on the results of their decisions—departure and arrival times were either met or missed. Intermediate group members can determine within a few days whether schedule adjustments have been effective. Master schedule developers evaluate the results of decisions by reviewing average system performance over the planning period.

The planning horizon of these groups also differs in terms of scope (*7*). Dispatchers and yardmasters have a comparatively narrow scope; they are concerned with the district or yard to which they are assigned. In contrast, the master scheduling group has a global scope. It must consider the network implications of scheduling decisions when setting timetables. The scope of the intermediate group synthesizes those of the others, because it must coordinate series of local decisions with network objectives.

Dispatchers and yardmasters have task-specific goals (*6*). For example, yardmasters must assemble a block of cars to make an impending departure and dispatchers must ensure on-time arrival to the next yard. The objectives of the master scheduling group are not easily identified with specific dispatches. Master schedule developers must incorporate the strategic objectives for performance and asset utilization into the timetables planned for the system. The goals of the intermediate group lie between task-specific and strategic, because the function must ensure that the daily traffic movement tasks are accomplished while attempting to maintain the integrity of the overall scheduling plan.

Railroad operations supervisors arguably face a highly uncertain subenvironment that is, for example, subject to variations in daily demand and equipment failures. The master schedulers, on the other hand, confront gradual changes in conditions over time and, for the most part, ignore daily variability (*7*). For example, the master scheduling group would consider a trend that indicated an increase in the average demand for service between an origin/destination pair, but ignore the daily variability in the demand level. The intermediate group must make schedule revisions that absorb the daily variability in local operations, while protecting the objectives of the master schedule.

In addition to the differences described above, there is a natural flow of information among these three groups that supports the differentiation argument. Master scheduling communicates planned timetables to the intermediate group and receives from the intermediate group information regarding the performance of the system. The intermediate group communicates schedule revisions to dispatchers and yardmasters and receives information from these line supervisors on schedule exceptions.

In summary, there is sufficient differentiation between the tasks in the scheduling function to classify each as a design element. Because of the pyramid-like flow of information and scope of responsibility (i.e., a small group of managers, the master schedulers, have global responsibility and communicate information to a large decentralized group of managers, the line supervisors), the scheduling function has a hierarchical structure. The crucial design decision is how to structure the design variables (i.e., the amount of authority and the communication links) for each design element in this hierarchy (*8*).

The amount of authority and autonomy assigned to each element is an important design variable because it will determine the degree of integration in the network. Assigning high autonomy to supervisors for scheduling decisions determines that the system will be decentralized and less integrated. On the other hand, establishing ultimate scheduling authority in the master scheduling group provides for better integration, because scheduling decisions will be made from a network perspective.

The degree of control that a master scheduling group is able to exercise over the daily scheduling decisions made by supervisors will be limited by the communication links established between the groups. For example, if the daily decisions of local supervisors are to be monitored by the master scheduling group, an expedient communication process must be established. Interaction between the scheduling function and other departments in the organization, likewise, depends on the type of communication process. In this paper, communication links between the sales and marketing function and the scheduling hierarchy will serve as a paradigm for this aspect of organization design.

DESIGN IMPLICATIONS OF SCHEDULING STRATEGY

It will be argued here that the appropriate choice of design variables depends on the scheduling strategy adopted by the railroad. The relationship between strategy and structure is based on the concept that organization design makes an economic difference (*9*). For example, if a firm adopts a strategy that is inconsistent with its structure, administrative problems will arise that decrease economic performance. Only after adjusting its design does the firm operate efficiently.

Galbraith (*5*) views organizations as information-processing systems. Therefore, the amount of information that must be processed for the firm to complete its tasks must be considered

in the organizational design process. The two scheduling strategies outlined above demand different information-processing capacities and, therefore, beget different organizational designs.

According the Galbraith, increasing the amount of uncertainty in the performance of the firm's tasks demands increased information-processing capacity in the organization. He defines uncertainty as the difference between the amount of information the firm has and the amount it requires to make decisions that accomplish the organization's goals. If a railroad pursues a real-time scheduling strategy, operations supervisors require increased information regarding the impacts of their decisions on the overall schedule. Likewise, they need to know the local impact of decisions made by other supervisors. Furthermore, tighter schedules decrease the number of disruptions that are absorbed by slack time and increase the number of exceptions that force schedule revisions. Real-time scheduling strategies, therefore, increase the amount of information needed to perform subtasks and demand that the organization increase its capacity to communicate information.

Galbraith (*5*) proposes two strategies to increase the information-processing capacity of the firm: investment in vertical information systems and creation of lateral relations. Lateral relations establish direct communication links between interdependent subtasks that cut across normal hierarchical lines of authority. Decision making is transferred from the normal hierarchical process to the lateral process. The speed of communication between the groups increases, because information no longer needs to be transmitted through the hierarchy.

Establishing such a communication link is crucial to a real-time scheduling strategy. The intermediate scheduling group can be designed to provide this important lateral communications link. In this design, the intermediate group communicates directly with line supervisors, circumventing the operations hierarchy. Information on schedule exceptions flows directly to the intermediate group where the global impact of the deviation can be analyzed. The intermediate group makes revisions to minimize the impact of the exceptions and communicates new timetables directly to the line supervisors affected by the change.

Without this design, railroads would have to coordinate interdependent supervisory decisions by processing the exception information through the operations hierarchy. Hierarchies have many disadvantages. Among them is a time-consuming decision-making process (*6*). For example, coordinating decisions between dispatchers would be the responsibility of the first common supervisor in the hierarchy. Each dispatcher would have to transmit information through the hierarchy to this supervisor, who in turn would make a scheduling decision and inform the dispatchers of the revised schedule. In a network as large and as interdependent as a railroad, this supervisor would be far up the hierarchy. Therefore, substantial time would be required to transmit information. A large number of schedule exceptions would quickly overload this communication and decision process.

A master scheduling strategy does not require the establishment of a distinct intermediate group with direct communication channels to supervisors, because it does not increase the amount of information required by the firm to perform its task. Uncertainty in the supervisor's task is not reduced under a master scheduling strategy; rather, the difference between the amount of information available versus what is needed is reduced. Slack time in the master schedule absorbs most exceptions. Therefore, dispatchers pass on only a limited amount of information. Because there is no increase in information flow, special designs to increase communication are not warranted.

The functions of the intermediate group under a master scheduling strategy are absorbed by the operations hierarchy. The hierarchy coordinates the actions of supervisors by enforcing the master schedule. It communicates the information required for the master scheduling process by periodically passing aggregate performance data to the master scheduling group. The speed of communication is not crucial to a master scheduling strategy.

Once the needed communication links have been established within the scheduling structure for each strategy, the degree of authority and autonomy must be assigned to the design elements. As previously stated, real-time scheduling strategies transfer authority over daily scheduling decisions to a group with a network perspective. It is also essential that the group with this authority have adequate communication links to supervisors. Therefore, scheduling authority is transferred from local supervisors to the intermediate group that possesses both a network perspective and adequate communication links. In addition, because the planning horizon has now become indefinite, the role of the master scheduling group loses importance. Scheduling authority is, thus, entrusted solely to the intermediate group under a real-time scheduling strategy.

Because the functions of the intermediate group are absorbed by the operations hierarchy under a master scheduling strategy, scheduling authority that requires a network perspective must be allocated to the master scheduling group. However, because this group relies on a time-consuming hierarchical process to communicate with line supervisors, some autonomy must be entrusted to the line supervisors. For example, the master scheduling group must delegate the handling of exceptions, because it lacks the necessary communication capacity.

The final design variable is to establish a communications link with the sales and marketing group. The objective is to design a communications link between the marketing function and the primary scheduling authority that circumvents unnecessary intermediaries. Under a real-time scheduling strategy, a direct communications link should be established between the intermediate group and the sales and marketing function. By selecting this link, the marketing department can easily evaluate the systemwide implications of new traffic and be informed of delays.

The primary link between scheduling and marketing under a master scheduling strategy is at the master scheduling level. Communication between these groups allows the firm to incorporate marketing objectives into its scheduling strategy. There is no communications link that would allow marketing to easily revise the schedule once established. However, informal communications at the line-supervisor level might allow for adjustments within the available slack time.

The alternative designs are presented in Figures 2 and 3. Because the master scheduling strategy requires no changes

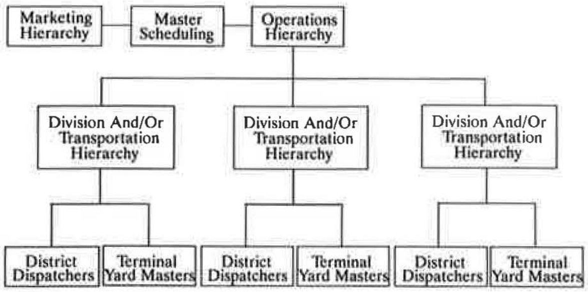


FIGURE 2 Organizational structure for a master scheduling strategy.

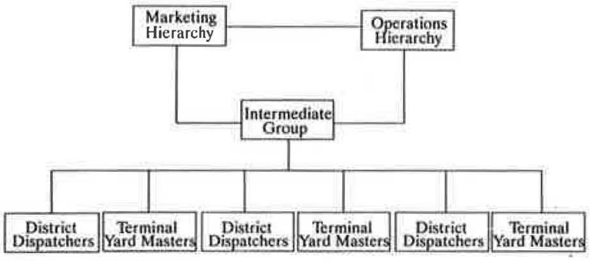


FIGURE 3 Organization structure for a real-time scheduling strategy.

in organization structure, it can be implemented inexpensively. The real-time scheduling strategy demands a structure with an independent intermediate group established outside the operations hierarchy. Costs associated with this design include personnel expense, organization friction arising from dual reporting for line supervisors (line supervisors report to both operations and the intermediate group), and increased communications cost.

The primary organization design impact from the choice of scheduling strategy is on the intermediate group. Under a master scheduling strategy, the role of the intermediate group is delegated to the operations/transportation hierarchy; whereas, under a real-time scheduling strategy, the intermediate group must be established outside the operations hierarchy and circumvent the normal channels of communication. Note the role of the master scheduling group is obviated under real-time scheduling, and the intermediate group has no distinct form under a master scheduling strategy. In summary, what has happened is that the integrating function (i.e., the group of managers with network perspective) is moved closer to the line supervisors as the planning horizon is shortened and slack resource is reduced.

IMPLICATIONS FOR MIS DEVELOPMENT

The selection of vertical information systems, Galbraith's (*5*) second approach to increasing information capacity, is likewise affected by the scheduling strategy decision. Accepting Galbraith's view of organizations as information-processing systems, it is impossible to separate decisions on design from those on information systems. Therefore, if the structure of the scheduling group is affected by the scheduling strategy decision, then the choice of information systems should be likewise affected.

Information systems increase the information capacity of the organization in two ways. First, a data-processing system increases the speed of data collection. It assembles data from the lowest levels of the hierarchy and transmits the data directly to the decision-making level. Second, expert systems can make decisions more quickly than managers.

Scheduling information systems can be categorized along two dimensions: scope of the data base and frequency of the decision (*5*). The scope of the data base from which scheduling decisions are made may be local or global. ("Local" has a relative meaning in this context because scheduling decisions made at the regional or divisional level are local relative to a networkwide data base.) The primary deficiency of expert systems that use local data bases in that network interdependencies are not considered. At the other extreme of this continuum is an expert system that uses a global data base and, thereby, ameliorates suboptimization problems inherent in decisions based on local data bases. The main trade-off is the higher cost of an expert system that utilizes global data bases versus the cost of suboptimal decisions arising from decisions based on local data bases.

In terms of scheduling frequency, information systems can assemble data and make decisions on a periodic or continuous basis. Periodic systems would be appropriate for performing a master scheduling task. Although these systems account for interdependence between subtasks when developing schedules, they do not overcome the inherent problem of long time-horizon planning. That is, master schedules begin to decay as unplanned events occur and, therefore, must contain adequate slack to eliminate the need to process exceptions (*5*).

Information-processing and decision-support systems are useful in the master scheduling process, because they allow the schedule planners to quickly evaluate alternative master plans and provide a means of incorporating networkwide data in the scheduling process. Expert systems using local data bases may also find applications in master scheduling environments. For example, computer-aided dispatching could be used to minimize cost within the time constraints established by the master schedule.

Continuous decision-making systems allow for a truncated planning horizon within which data are assembled and scheduling decisions made. These systems employ a global data base with sufficient decision-making capability to process exceptions quickly and make schedule adjustments. Continuous decision systems are most appropriately implemented with real-time scheduling strategies, because a real-time scheduling strategy requires the continual evaluation of the network implications of schedule deviations and adjustments.

In summary, the master scheduling strategy is best supported by a periodic decision system that utilizes a global data base. The real-time scheduling strategy should be supported with a continuous decision-making system that utilizes a global data base. Continuous decision-making systems are, of course, more expensive than periodic systems. Therefore, a real-time scheduling strategy requires a more expensive information system than a master scheduling strategy. The trade-offs among various choices of information systems are summarized in Figure 4.

The choice of information system can have organization structure implications beyond the design of the scheduling group. Choosing to schedule based on a local data base sys-

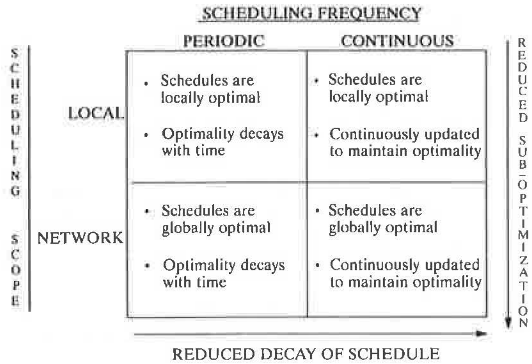


FIGURE 4 MIS implications of a scheduling strategy (*5*).

tem, for example, determines that the firm will be decentralized, with interdependence managed through hierarchical channels (*1*). Continuous decision making using global data bases allows the firm to truncate its hierarchy, because an automated data-collection and decision-making system allows for an increased span of control, thereby eliminating the need for large numbers of managers.

The power of decision-support systems to supplant portions of a hierarchy was demonstrated by Dawson and McLoughlin (*10*), who studied the impact of introducing TOPS (Total Operations-Processing System) on the role of supervisors at British Rail. The system was designed to provide accurate information about local operations to a central control facility. Some organization theorists had suggested that such a system would erode the traditional supervisory role by enabling higher-level managers to make local decisions. Others suggested that computerization would enhance the role of supervisors and create a further decentralized organization.

Dawson and McLoughlin (*10*) found that the system did increase the control of the operations by high levels of management. However, the role of the supervisor was enhanced rather than diminished. The new information flows and communication channels provided by the computer obviated the need for a division hierarchy. Information about operating conditions and performance at remote locations became available to headquarters management, and access to a real-time data base was provided to the local level. This meant that much of the decision-making responsibility at the division level could be delegated to the supervisors, whose positions could be redefined as area freight assistants. (British Rail was ultimately frustrated by labor unions in its attempt to eliminate the divisional hierarchy.)

NEED TO MATCH STRATEGY, STRUCTURE, AND DECISION ANALYSIS

The design of the scheduling group, choice of information system, and scheduling strategy must be matched if the strategy is to succeed. Without increasing the information-processing capacity of the organization, a real-time scheduling strategy is destined to fail, because the firm will be unable to process the needed information and make appropriate decisions. As a result, slack must be introduced into the schedule; and the firm, by default, will adopt a master scheduling strategy (*5*). Similarly, the investment in the necessary information systems and organization redesign to support a real-time scheduling strategy would generate needless expense for a firm pursuing a master scheduling strategy. Only by selecting the information system and organization structure that matches its scheduling strategy does the firm realize the economic benefits of its scheduling choice.

Contingency theory of organization structure suggests that there is no one best design for all organizations; rather there is one best design for each organization (*11*). Intermediate group designs that circumvent the operations hierarchy are not necessarily better. The are best for real-time scheduling strategies, but completely inappropriate for master scheduling strategies. Similarly, continuous-scheduling decision systems using global data bases are not necessarily better than systems that make periodic decisions. They are, however, best for real-time scheduling strategies.

According to the argument presented here, the structure of the scheduling function follows the scheduling strategy decision. But if an organization refuses to adjust its structure, then its scheduling options are limited. The firm is, therefore, also constrained by its structure (*12*). The divisional hierarchy structure found in the railroad industry has many rational features well founded in organization design theory. In addition, railroads have a long history of employing the operations hierarchy to implement schedules. Firms may be unwilling to disrupt these established reporting procedures by introducing the intermediate scheduling group. Therefore, it may be difficult to implement the necessary changes to adopt the real-time scheduling strategy.

Assuming real-time scheduling strategies are not precluded by inertia, how will firms decide which scheduling strategy to adopt? The principal cost and service trade-offs to consider are as follows:

1. The cost of the organization structure dictated by a real-time scheduling strategy, including personnel, communications, and organization friction costs,
2. The higher cost of the decision-support systems needed to support real-time scheduling versus master scheduling,
3. The costs of the slack resource created by a master scheduling strategy, and
4. The higher reliability of master scheduling strategies versus the lower mean transit times of real-time scheduling.

Real-time scheduling has higher variable costs associated with organization structure and decision-support systems; whereas, the master scheduling approach has greater fixed costs in terms of excess capacity or slack resources. Another factor to consider is how much of the cost of slack resource is avoidable. For example, it may be difficult to decrease the fixed cost of infrastructure even if a scheduling strategy that reduces the need for the excess capacity is adopted. Because adoption of a real-time scheduling strategy will surely increase short-run variable costs, the uncertainty of eliminating the cost of slack resource may make firms reluctant to change from a master to a real-time scheduling strategy.

SCHEDULING STRATEGY FOLLOWS MARKETING STRATEGY

Both revenue and costs must be considered when making strategic decisions. Choosing a scheduling strategy based on

a comparison of the cost to introduce real-time scheduling versus the cost of maintaining slack resource ignores the revenue component of the profit equation. Each of the proposed scheduling strategies emphasizes different service attributes. Therefore, the key strategic decision may be the choice of marketing strategy, because different service attributes will appeal to different markets. That is, once a marketing strategy has been selected, the scheduling strategy, design of the intermediate group, and choice of information system follow naturally.

Chandler (*9*) proposed that organization structure follows the firm's growth strategy. He identified four key growth strategies: volume expansion, geographic expansion, vertical integration, and product diversification. Volume expansion entails increasing the marketshare of a single product in one market. The increased volume overloads the existing structure and requires a more extensive functional hierarchy. Creating multiple field offices and duplicating a portion of the hierarchy is warranted when the sales strategy incorporates entry into new geographic markets with a single product. New functions or departments are added to the hierarchy when the firm pursues a forward or backward integration strategy. The firm becomes decentralized when it expands into new product markets. Ultimately, it adopts a holding company form when diversification leads into unrelated product markets. Strategic planning may incorporate portions of all the above strategies that lead to complex, hybrid firm structures.

The design implications of scheduling strategy were analyzed above in terms of two polar cases. The impact of marketing strategy on the choice of scheduling will now be illustrated in the same way. Again, this argument is intended to demonstrate a relationship between marketing and scheduling strategy. It suggests a reason, other than cost, for a firm to choose one of the two scheduling strategies. It does not preclude wholly different or hybrid marketing strategies; rather, it suggests that there should be some congruence between the firm's scheduling strategy and its marketing strategy. Two marketing strategies will be considered: focus and expansion. It will be argued that the focus strategy affords no role for real-time network scheduling, but that the expansion strategy cannot succeed without a real-time scheduling strategy.

A *focus growth strategy* is when the firm concentrates on traditional rail markets where rail has a distinct economic advantage. Demand for rail service is well established in these markets and operational efficiencies are the key source of competitive advantage. Although the market may be elastic in terms of reliability, it is comparatively inelastic in terms of transit time. In short, dependable rail service satisfies the freight transportation demand in these markets.

Selecting a real-time scheduling strategy would be inconsistent with this marketing strategy. In terms of the scheduling strategy following the marketing strategy process, if a railroad adopts a focus strategy and real-time scheduling strategy, the market would not compensate it for the resulting improvements in transit time. The firm would be, therefore, inefficient because it could maintain the same revenue base without the variable expense created by real-time scheduling.

However, the market for freight transportation service is much larger than that of the traditional rail shippers. A railroad could, alternatively, adopt an *expansion growth strategy* in which the firm aggressively pursues intermodal traffic and, generally, seeks to capture traffic currently moved by truck. A real-time scheduling function would be required to deal with this market's demand for reduced transit time, increased integration between operations and sales, and an expanded customer service function.

Improved transit time would require elimination of slack time from the schedule. In addition, the network would have to recover quickly from disruptions, thereby minimizing delays arising from unusual events. Without adopting a real-time scheduling strategy, the firm cannot make the necessary reductions in transit time. A focus strategy is adopted by default because only the focus strategy market base will be sated by the performance level.

The expansion strategy also demands increased integration of the operations and marketing functions. The master scheduling strategy provides for formal interfunctional relations only at the highest level. The functional hierarchy would be quickly overloaded with the volume of decisions regarding the feasibility of proposed services for new shippers in the expanded market. Because a time-consuming decision process is likely to reduce the railroad competitiveness in these markets, structural adjustments to increase the organization's ability to process decisions are warranted. The intermediate scheduling group would be positioned to evaluate the feasibility of proposed services quickly by virtue of their global perspective and access to line supervisors.

Entry into new markets is likely to increase the amount of price and service negotiations between the marketing department and customers. To negotiate price effectively, marketing executives need accurate cost information. Because of different shipment characteristics, some business fits well with existing traffic patterns, while some generates high amounts of variable cost. The dynamic implications of additional traffic must be evaluated to price service correctly. The intermediate real-time scheduling function would be positioned to provide the relevant cost information quickly. Processing these pricing decisions through the existing hierarchical structure would be too time consuming.

The expansion strategy also increases the importance of the customer service function. Customers in more service-elastic markets will require more information about shipments in transit and increased responsiveness from the operations department. For example, customers using just-in-time inventory systems need accurate shipment-tracing information as well as reliable delivery. In addition, customers may require increased flexibility such as in-route reconsignment. Customer service personnel need accurate information about location and about the possibilities of changing current plans. Therefore, the interface between customer service and operations must avoid bureaucratic delays typical of hierarchical designs. The intermediate scheduling function would possess the network information to address these customer service demands.

In summary, a real-time scheduling strategy and the design adjustments and information systems needed to support it are essential for a marketing strategy that relies on expansion or entry into markets presently controlled by motor carriers. The less-expensive master scheduling strategy is preferred for marketing strategies that focus on traditional rail markets.

INDUSTRY SURVEY

To see if the arguments presented above were realistic, a group of five large North American railroads were surveyed

regarding their scheduling practices. The railroads were queried as to whether they employed master scheduling or real-time scheduling and were asked if real-time scheduling groups were positioned outside the normal hierarchical channels. They were asked about the communication links and information systems in use or under development to support the scheduling function. The reasons they chose their scheduling strategy were also discussed. Finally, the connection between scheduling strategy and marketing strategy was investigated.

Figure 5 presents a table that summarizes the responses of the railroads to questions regarding their scheduling practices. Except for Railroad C, all the railroads in the survey had master schedules in use or under development. Except for railroad B, all the railroads that included master schedules as part of their scheduling strategy intended to override the master schedule regularly. Railroad E described the master schedule as a rough guideline that was rarely adhered to in practice. Both Railroads C and E identified the inaccuracy of forecasts of demand and resource availability as the primary limitation of a master schedule.

Railroads that either chose to override the master schedule or had no master schedule were able to identify centralized, real-time scheduling groups in their organization. The scheduling groups were generally located at the companies' operations headquarters and were composed of representatives from the divisions plus system train control managers. Interestingly, all the railroads that employed real-time decision making established a reporting structure outside the operations hierarchy. Each of these railroads identified the intermediate group described in this paper as having the same characteristics, in terms of organizational structure, as their real-time scheduling group. It should be noted that real-time scheduling for these railroads consisted of the development of daily or twice-daily scheduling plans.

Several of the railroads expressed a need for improved data collection and expert systems to support the real-time scheduling function. Presently, the real-time scheduling process makes use of telephone conference calls to speed communication. Approximate demand levels are developed by the division representatives. Managerial experience was cited as the primary scheduling decision tool.

None of the railroads had optimization systems for use in either the master or real-time scheduling process. However, all expressed a strong interest in such systems. The railroads in the survey appeared to have technology, either in place or in development, that would support scheduling decisions. Railroad C, for example, reported that a system to be used to make daily scheduling decisions was under development. Railroad B has several systems under development that could support real-time scheduling. Most of the railroads had car-monitoring and trip-planning capabilities, and most had aggregate data on performance available. Finally, some of the railroads had computer-aided dispatching systems in place.

Both marketing and cost control were cited as reasons behind the choice of scheduling strategy. Railroad B, for example, believed that a master scheduling strategy would improve reliability, which would result in increased customer satisfaction. Several of the railroads surveyed identified special scheduling processes for time-sensitive traffic, such as intermodal and automotive traffic. For example, master schedules with very low slack are used for many intermodal trains. The railroads augment these low-slack master schedules with real-time support in the event of schedule disruptions. Several railroads identified flexible scheduling policies to protect automotive traffic. In addition, some firms assign the responsibility of protecting service on important traffic to a member of their real-time scheduling group.

Cost reduction from the consolidation of trains was cited as a primary reason for real-time scheduling. Railroad C expressed the opinion that the railroad was compensated to "move freight, not trains." Therefore, they tried to let demand dictate the service level. In general, most of the railroads believed that because demand fluctuated widely from day to day, the evaluation of demand and available resources (i.e., real-time scheduling) was necessary to control variable cost.

Thus, the alternative scheduling strategies described in this paper appear to accurately describe the scheduling strategies in practice. The strategies are not, however, mutually exclusive. Many firms employ hybrid scheduling processes that have master and real-time scheduling characteristics.

A PROTOTYPE MIS STRUCTURE

It has been argued that strategy determines the appropriate organization design and information system. However, this is not intended to diminish the role of technological innovation, which can make new strategies and structures feasible. The argument is intended to imply that, for a firm to capitalize on advances in technology, it must adjust its strategy and structure accordingly.

Research currently under way could provide new opportunities for railroad scheduling and marketing strategies, and for innovative organization structures. Harker (*13*) has presented a structure for the development of intelligent (i.e., model-based) information systems that enable the railroad to adapt to its chosen organizational and marketing strategies:

Schedule Policy Evaluation
⇅
Tactical Scheduling of Trains
⇅
Real-Time Scheduling
–trains
–locomotives
–crews
–cars
⇅
Computer-Aided Dispatching
⇅
Optimal Train Control

Once an overall strategy has been decided on how schedules (e.g., local or networkwide) will be generated, one must implement this policy on a weekly or monthly basis. This tactical scheduling of trains differs from the above strategic question in that all trains at the tactical level will have some type of

	A	B	C	D	E
Master Scheduling?	Yes	Yes	No	Yes	Yes
Override?	Yes	No	N/A	Yes	Yes
Real Time Decision Making?	Yes	No	Yes	Yes	Yes
Intermediate Group with Unique Reporting Structure	Yes	N/A	Yes	Yes	Yes

Source: Interviews with 5 Class I North American Railroads

FIGURE 5 Survey results.

schedule. Thus, for those trains that must be scheduled (passenger, intermodal, etc.), the tactical scheduling procedure will create a set of feasible schedules; that is, a set of schedules that are logically consistent in the sense that an operating plan exists that can achieve the times stated in the schedules with high probability given the delays encountered by each train as a result of random occurrences (wind, breakdowns, etc.) and interference with other trains. For trains that run on a tonnage basis, scheduled slots would exist. That is, trains would not be permitted to depart at random but must depart within a stated time window if they are to be operated on a given day. Thus, a tactical scheduling system must also have the capability to create such slots and check that they are feasible when considered alone and when combined with the other scheduled traffic.

Given the tactical schedules, the purpose of the real-time models is to develop operating plans that will achieve the stated schedules as best as possible given that events have occurred (breakdowns, crew shortages, etc.) that disrupt the plan of operations on which the tactical schedules are based. For trains, one wishes to develop a plan of arrival and departure times at each major yard or, more generally, at each point where the planning of the train operations changes (i.e., a boundary of the dispatchers' territories). For crews, locomotives, and cars, one attempts to plan their movements to guarantee that sufficient resources are available at each yard to achieve the tactical schedule plan. These models are the "heart and mind" of the intermediate group described previously. That is, these real-time models serve as the vital link between the overall strategic mission of the railroad and its implementation on a day-to-day basis.

After defining the arrival and departure times of the trains at the boundaries of the dispatchers' territories (i.e., a planning line), the computer-aided dispatching system attempts to schedule the meets and passes along a rail line with planned arrival and departure times at intermediate points (sidings, beginnings and ends of double track, etc.) to ensure compliance with the times from the train scheduling model.

The dispatching system provides each train with a specific goal in terms of the time and velocity at which it should reach each point on its path. The engineer and the on-board computer system must then calculate a velocity profile (a combination of throttle and dynamic/air brake settings) that will achieve this goal in a safe and fuel-efficient manner. The train must solve a pacing problem that is quite more complex due to the nature of train forces and handling techniques.

The above discussion has described the flow of information down the model hierarchy. Of course, the reverse flow is also very important. The train must constantly inform the dispatching model of its location and performance, the dispatching system must inform the network control model of the status of planning lines, and the performance of the network control system (the interline planner) must be monitored to assess the long-term viability of various schedule policies. It is precisely this flow of information that must be coupled with the organizational design of the railroad.

CONCLUSION

The research program underway at the University of Pennsylvania described by Harker (*13*) is attempting to build model-based information systems technology to deal with all of the issues outlined in the previous section. In fact, the argument presented in this paper implies that the tactical and real-time scheduling systems and the computer-assisted dispatching methodology are the most important pieces of ATCS in that they link together the strategic goals of the railroad and resulting organizational structure. That is, the type of MIS structure described above is the vital lubricant in Galbraith's (*5*,*14*) flow of information. Without such systems, it seems clear that ATCS will not achieve their promise.

Thus, much more attention must be given to the development of the MIS component of ATCS. However, there will never emerge one standard for the entire industry as many envision because of the need to adapt such systems to the particular strategy chosen by each railroad. The framework presented by Harker (*13*) and outlined above is the closest to a standard that will be achieved in this area. It is the goal of the research program described by Harker (*9*) to develop MIS technology that can be adapted to the individual strategies each railroad elects to implement. This is a challenge, but one that must be met if the "service-sector sinkhole" is to be avoided.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation Presidential Young Investigator Award and by a grant from the Burlington Northern Railroad. Many individuals at various railroads, who will remain anonymous, have contributed to this research through surveys and personal interviews; their help is gratefully acknowledged. Special thanks are extended to J. Pennings of the Wharton School and R. Gunkel, V. Guy, and T. Polland of the Wharton M.B.A. class.

REFERENCES

1. S. S. Roach. Technology and the Services Sector: America's Hidden Competitive Challenge. In *Technology in Services: Policies for Growth, Trade, and Employment* (B. R. Guile and J. B. Quinn, eds.), National Academy Press, Washington, D.C., 1988, pp. 118–138.
2. G. P. Hackett. Investment in Technology—The Service Sector Sinkhole? *Sloan Management Review*, Winter 1990, pp. 97–103.
3. J. L. Heskett. *Managing in the Service Economy*. Harvard Business School Press, Boston, Mass., 1986.

4. J. B. Quinn, J. J. Baruch, and P. C. Paquette. Technology in Services. *Scientific American*, Vol. 257, No. 6, December 1987, pp. 50–58.
5. J. R. Galbraith. *Organization Design*. Addison-Wesley Publishing Co., Reading, Mass., 1977.
6. D. Robey. *Designing Organizations*, 3rd ed. Richard D. Irwin, Inc., Homewood, Ill., 1991.
7. P. R. Lawrence and J. W. Lorsch. Differentiation and Integration in Complex Organizations. *Administrative Science Quarterly*, June 1967, pp. 1–47.
8. R. Duncan. What Is the Right Organization Structure? *Organizational Dynamics*, Winter 1979, pp. 59–79.
9. A. D. Chandler. *Strategy and Structure*. MIT Press, Cambridge, Mass., 1962.
10. P. Dawson and I. McLoughlin. Computer Technology and the Redefinition of Supervision: A Study of the Effects of Computerization on Railway Freight Supervisors. *Journal of Management Studies*, Vol. 23, No. 1, 1986, pp. 116–131.
11. J. L. Gibson, J. M. Ivanovich, and J. H. Donnelly, Jr. *Organizations: Behavior, Structure, and Process*. Business Publications, Inc., Plano, Tex., 1988.
12. D. D. Wyckoff. *Railroad Management*, D.C. Heath and Co., Lexington, Mass., 1977.
13. P. T. Harker. The Use of ATCS in Scheduling and Operating Railroads: Models, Algorithms, and Applications. In *Transportation Research Record 1263*, TRB, National Research Council, Washington, D.C., 1990, pp. 101–110.
14. J. R. Galbraith and D. A. Nathason. *Strategy Implementation: The Role of Structure and Process*. West Publishing Co., St. Paul, Minn., 1978.

Decision Support System for Train Dispatching: An Optimization-Based Methodology

DEJAN JOVANOVIĆ AND PATRICK T. HARKER

The authors argue that the primary purpose of dispatching tools is to allow trains to arrive on time rather than to minimize train delays in general, and present a new methodological framework for the role of computer-aided train-dispatching (CAD) systems. One of the biggest obstacles to the successful implementation of an optimal CAD system is the combinatorial nature of the optimal train-dispatching problem and the need for optimization algorithms that could provide good solutions in real-time environments. Lower-bound–based algorithms for the minimum tardiness-cost train-dispatching problem can be used to handle larger traffic volumes and cover longer planning horizons and larger dispatching territories than is possible with the current state of the art.

Interest in optimal train-dispatching systems has been revived in recent years by the development of Advanced Train Control Systems (ATCS) technology (*1–3*). Besides improved safety, the potential for lower fuel consumption and improved railroad operations are often cited as the major benefits of ATCS. However, there seems to be a lack of a conceptual framework defining the goals and flow of information between the optimal dispatching system and other components of ATCS that would lead to improved railroad operations.

TRAIN-DISPATCHING PROCESS

Train dispatching is a demanding and complex task. The survey paper by Petersen et al. (*4*) gives a brief description of the issues involved. Dispatchers monitor and control the movements of trains over railway lines and resolve potential conflicts between trains. The primary conflicts arising on single-track lines with passing sidings and partially double-track lines (such lines represent more than 90 percent of all railway lines in the United States) are meets of trains going in opposite directions. Two trains traveling in opposite directions cannot occupy the same single-track segment at the same time, or a collision would occur. Meets are resolved by switching one of the opposing trains onto a side track, where it waits until the other train passes. On fully double-track or multiple-track lines, there are no conflicts between trains going in opposite directions, but conflicts may arise between a fast train and a slow train that is traveling in the same direction in front of the faster train. Such conflicts can be resolved by overtaking, that is, by switching the faster train onto a parallel track on which it can pass the slower train, provided that this parallel track is not occupied by a train going in the opposite direction. Overtaking can also be done on a single-track line by switching the slow train onto a passing siding clear of the main track, where it stops and waits to be passed by the faster train. The number of ways in which conflicting train movements can be resolved is an exponential function of the number of trains and track segments.

Other duties of a train dispatcher include safe coordination of movements of roadway maintenance gangs, signal maintenance crews, and industrial switch engines, as well as a host of clerical duties concerned with maintenance of various operating statistics. Quoting from Sauder and Westerman (*5*), "Safety is the paramount consideration in all of the dispatcher's tasks."

Besides safety concerns, train dispatching is of paramount importance in the operation of a railroad network for another reason: dispatching decisions, through meet/pass delays, greatly influence train transit times and on-time performance. According to one study (*4*), 45 percent of the variance of train arrival times is due to the variance in over-the-line transit times. Unfortunately, dispatchers do not have at their disposal the information that shows systemwide effects of their decisions; their main incentives (besides safety) are to avoid delaying a "hot" (high-priority) train. As reported by Sauder and Westerman (*5*), a common response of dispatchers was to clear low-priority trains into a siding far in advance of incoming hot trains, thus minimizing the chance of delaying such a train while causing unnecessary delays for low-priority trains. During periods of very dense traffic, this strategy can often backfire. Delaying a cluster of low-priority trains can soon create an area of congestion in which all trains are delayed regardless of their priority.

Due to a heavy workload and insufficient information concerning future traffic, dispatchers are forced to cope with incoming traffic as it arrives and have little ability to make plans; that is, the function of a train dispatcher, is, at present, *reactive* rather than *proactive*. There is, however, a methodology designed to put dispatchers in a more active role in which they would work toward the common operating objectives of a railroad system, without increasing their workload.

D. Jovanović, Network Planning, Burlington Northern Railroad, 777 Main Street, Fort Worth, Tex. 76102-5304. P. T. Harker, Fishman-Davidson Center for the Study of the Service Sector, The Wharton School, University of Pennsylvania, Philadelphia, Pa. 19104-6366.

STATE OF THE ART IN OPTIMAL TRAIN DISPATCHING

In their article, Petersen et al. (*4*) gave a survey of computer-aided dispatching (CAD). The authors note that most of the information-gathering and record-keeping activities (on which dispatchers once spent 75 to 80 percent of their time) and other routine tasks can be computerized, and that commercial systems to handle these tasks are available. Petersen et al. (*4*) state that, as new CAD technology becomes available to simplify the dispatcher's tasks (such as automated signal clearing and calculations of expected train arrival times), the next important step is the development of optimal train-dispatching systems. The biggest obstacle towards development of such a system (*4*) is the combinatorial nature of the problem; for example, even for moderate traffic intensities and an 8-hr time horizon, it could take 12 days for a supercomputer to evaluate all possible meet/pass plans. Similar problems are reported in the only published account of a computer-aided optimal train-dispatching system implemented in the United States beyond the testing stage (*5*). Algorithms that significantly improve on the existing enumeration-based optimal dispatching algorithms were recently developed by Jovanović (*6*); the new algorithms allow for real-world dispatching problems to be solved optimally or near-optimally in less than a minute, thus removing the obstacle described above for most practical purposes.

However, an issue that has received little attention in the literature or from the vendors of CAD equipment concerns the goal of or the potential benefits from an optimal CAD system. As was lucidly noted in the paper by Duncan et al. (*7*), which describes the development and testing of an optimal CAD and train control system on an ore-hauling Australian railroad, "The difficulty in achieving the 'best' meet/pass plan is defining what is meant by 'best.' " Minimizing the sum of weighted train delays is the objective that is encountered most often; e.g., this objective was proposed in the pioneering work by Szpigel (*8*) and in the survey paper by Petersen et al. (*4*). The objective of the Norfolk Southern system is to minimize the sum of train delays and the priority-weighted sum of train lateness. A similar objective is proposed by Rockwell International (*9*) for the meet/pass planner under development within the Advanced Railroad Electronics System (ARES) project: in addition to minimizing the weighted sum of train delay and lateness, the ARES meet/pass planner attempts to minimize fuel consumption as well. The objective of the "driver-assist" optimal CAD system under development in Australia (*7*) is to minimize some combination of fuel consumption and train delay; however, the authors emphasize that railroad management must supply the objective or the "strategy" the system attempts to implement. Finally, the Union Switch and Signal CAD system, recently installed at CSX (*10*) and Union Pacific, does not have any explicit objective function in its meet/pass planner. Issues of how to calculate the weights or costs associated with train delay and lateness, whether they are always linear and constant in time and by train type, and what reference times should be used for the calculation of lateness are seldom discussed in the published literature.

COST MINIMIZED BY AN OPTIMAL COMPUTER-AIDED TRAIN-DISPATCHING SYSTEM

Fuel Consumption

A decrease in fuel consumption has often been mentioned as one of the major benefits of ATCS. Besides a CAD system that determines target times for trains as they traverse the line, it is necessary to have on-board train controllers that monitor train movements and advise the engineer on the most fuel-efficient manner for reaching the next target point within the allotted time. (We assume that the problem of fuel-efficient control of an individual train can be solved; see the report by Milroy et al. (*11*) for a promising approach.) Note that some fuel savings can be achieved through the use of on-board controllers, even if the minimization of fuel consumption is not incorporated into the CAD system objective function, as long as the planned meet/pass times are passed to the train controllers. A proprietary 1989 study based on research by the authors estimates these fuel savings at about 2.5 percent. To achieve greater fuel savings, it is necessary to trade off fuel versus train transit time and to incorporate fuel consumption minimization in the objective function of the CAD system. In the same proprietary study, it was established that the value of time lost by the rolling stock and lading of a train would be higher than the value of the fuel saved by decreasing the speed of the train in almost every instance. In another part of the study, a nonlinear optimization algorithm described by Kraay et al. (*12*) was applied to historical dispatching data sets with the objective of minimizing fuel consumption subject to the constraint that no train should arrive at its destination later than it actually did; under these conditions, the average fuel saving was about 7 percent. The main reason for the relatively low potential fuel savings lies in the shape of the fuel consumption curve and the ratio of train weights: those trains that incur the largest amount of dispatching delay are mainly the heavy and slow bulk-commodity trains whose fuel consumption is relatively insensitive to a decrease in speed, while most of the trains whose fuel consumption curve is sensitive to speed reduction are the fast high-priority trains that usually incur fewer dispatching delays.

An indirect benefit associated with decreased fuel consumption is decreased rolling stock and right-of-way wear and tear resulting from lower train speeds and the lower braking and tractive forces.

Another issue associated with the decrease in fuel consumption is that of the "robustness" of meet/pass plans. A train that is slowed down by the on-board controller to save fuel is more likely to be late for the planned meet than were it running at full throttle. This lateness may cause delay to the other meeting train, which was planned to go through the meet without any delay; the unanticipated delay can have a domino effect on the other meets and, thus, make the entire plan invalid. This trade-off between fuel and the reliability of a meet/pass plan, which directly influences the reliability of the trains' on-time performance, points to the need for the reliability of on-time arrival at the planned meet-point to be a primary goal of the on-board train controller; this issue has received very little attention in the literature.

It can be concluded that although at least 2.5 percent of total fuel consumption can be saved through the use of on-board train controllers coordinated by the CAD system, the reduction of fuel consumption should not be the primary goal of an optimal CAD system because of the high value of train transit time that must be traded for fuel. Once an optimal meet/pass plan is chosen, it may be possible to modify slightly the planned target times to achieve additional fuel savings as long as this does not significantly decrease the probability that the chosen meet/pass plan can be achieved. This approach can bring the total fuel saving up to 7 percent, depending on the desired trade-off between fuel cost and increased running times.

Cost of Train Delays

It should be noted that on a fully signalized railroad territory, there is no great need for new technology (such as ATCS), other than software, to achieve reductions in train tardiness and delay. Although the installation of on-board controllers and train-positioning systems envisioned by ATCS is necessary to realize fuel savings, adequate information and control capabilities may be provided initially by the existing signal system.

The ability of CAD systems to decrease significantly the amount of over-the-line delay that trains incur waiting for meets or overtakes or for a preceding train has been proven in practice and by numerical experiments. Since the optimal CAD system described by Sauder and Westerman (*5*) was installed by the Norfolk Southern Railroad, the delay per train has decreased by an average of 12.4 percent during the first year of operation and by 25.4 percent during the second year of operation when compared with the year preceding installation. The proprietary study of benefits from optimal CAD systems estimates (through the analysis of historical dispatching data) that train over-the-line transit times can be decreased by an average of 17 percent without involving any overtaking. The time savings that could be realized by individual trains vary greatly, depending on the amount of delay incurred and the importance (weight) assigned to a particular train in the CAD objective function; for example, the delay of several high-priority intermodal and mixed-freight trains was decreased by 1 to 3 hr out of a 9-hr transit time, with little additional delay to the other trains.

The goal of minimizing train delay and tardiness, suggested by most developers of optimal CAD systems, implies that certain costs are incurred by delaying a train and that certain benefits are gained if a train arrives early. We will attempt to analyze the potential sources of the implied costs and benefits.

Shipment Transit Time

The time when a train arrives at its next terminal can influence the total shipment transit time in two ways. First, if this terminal is the final destination for the shipment (e.g., in the case of passenger, intermodal, and bulk-commodity unit trains), then shipment tardiness is a direct function of the train arrival time. Second, if the car containing the shipment is scheduled to be transferred to another train, the late arrival of the inbound train relative to the departure time of the outbound train can increase the total shipment transit time by hours or days, depending on when the next appropriate outbound train will depart.

The actual cost of increased shipment transit time varies greatly with the type of shipment and can be very hard to quantify; the shipping contract may or may not require the railroad to pay penalties for late deliveries and the amount and form (linear, progressive) of this payment varies. Another, less tangible cost associated with late or unreliable shipment deliveries is customer dissatisfaction and loss of business to other carriers or, inversely, the loss of potential revenue that the customer would be willing to pay if the quality of service was higher. These costs depend on the customer's sensitivity to disruptions to his distribution and/or production process (which can be quite high if he is using just-in-time inventory control), as well as on the time-value of the commodity being shipped.

In most cases, there are no benefits from delivering the shipment early; on the contrary, the customer may not be ready to receive it and early delivery can cause unnecessary congestion in either the customer's receiving and inventory system or in the railroad terminal. The latter case can be observed when trains arrive before the planned time at a classification yard, occupying capacity planned for other trains and causing longer processing times and late departure of those trains.

The above discussion shows that the cost of train delay with regard to shipment transit time can only be calculated relative to some *scheduled* time: either the scheduled shipment delivery time (the train's published scheduled arrival time in the case of passenger and intermodal trains) or the latest arrival time necessary for the shipment to make its connection. The cost of shipment transit time for a train composed of several blocks of cars, each block destined for a different outbound train, will most likely be a nondecreasing staircase-shaped function of train arrival time, with each step corresponding to the connection time of a group of shipments (i.e., a block of cars) destined for the same outbound train.

Rolling Stock Value

It is obvious that significant reductions in rolling-stock capital cost could be achieved if the turn-around times of trains were decreased, because a smaller number of cars and locomotives could produce the same output, or the same equipment could produce a higher output. These reductions, however, can only be realized if the *planned* or *scheduled* train transit times are decreased to take advantage of the faster train movements made possible by optimal CAD systems. For those trains that are not scheduled (e.g., unit trains), the target arrival and departure times should be set in real time by a systemwide operating plan and passed down to dispatchers as objectives, alongside target arrival/departure times for scheduled trains. Only some local trains may be left to the complete discretion

of the dispatchers with regard to their arrival and departure times. Thus, the goal of increased fleet utilization cannot be directly incorporated within the CAD objective function at the real-time level, because it belongs within the systemwide train scheduling process. Equipment-related costs that could be incorporated within the CAD objective function are related to the exceptions from the schedule. (For example, the inability to form or adequately power a new train and/or the lack of empty cars to be delivered to the customer may be caused by the late arrival of trains whose rolling stock was scheduled to be used for this purpose.) These costs are related to the planned train arrival time, which can be changed in the operational plan to be earlier than the published scheduled time if the train is bringing equipment (cars, locomotives) urgently needed for another train. On the other hand, *unplanned* early arrival does not bring any benefit in terms of equipment utilization and may cause yard congestion.

Crew Costs

The wages paid to train crews for a particular leg of a train trip depend on the type of agreement between the labor union and the railroad; this discussion deals only with those crew-related costs that are a function of the train's arrival time. Although the current practice in North America is to pay crews primarily be the mileage of the trip, it is not inconceivable that they could be paid for the actual hours worked or for the scheduled train travel time plus overtime pay if the train is late. Thus, crew-related costs are likely to increase with late train arrivals, either through direct pay or through worsened labor relations.

Another important crew-related cost is associated with the federal rule that prohibits crews from operating a train after 12 continuous hours spent on duty; when excessive train delays over the scheduled travel time cause this limit to be reached, the train must be stopped on the line and a fresh crew brought in. The cost of bringing in the relief crew and the resulting delays to the train and the blockage of the line should be incorporated within a CAD system.

OBJECTIVE OF AN OPTIMAL CAD SYSTEM

In the previous section it was argued that, at the operational level, the costs related to train transit time could be defined only with reference to some scheduled or planned target arrival times. These target times are not necessarily the published schedule times, in part because not all freight trains have published scheduled arrival times, and in part because various disturbances in the schedule may require a new operating plan that differs from the tactical (published) schedules. For example, if a high-priority intermodal train were delayed during previous legs of its itinerary to the point where it could not reach its final destination on time, then its planned target arrival time must be shifted forward, and the plan for other affected trains (e.g., those waiting for the locomotives from the late train) must be adjusted accordingly.

In the case of unscheduled trains (e.g., unit coal trains), once a decision is made to run such a train, the desired running time and the desired arrival time at its destination should be made a part of the plan, and shipment-, equipment-, and crew-related costs of exceeding the planned time could be known. For every train planned to enter the dispatcher's territory the operational cost as a function of the train's arrival time could be known.

The objective of an optimal CAD system is to ensure the implementation of the systemwide operating plan over a given dispatching territory or, if necessary, to ensure that the fewest cost deviations from this plan are achieved. Thus, rather than being governed by standard operating procedures that state, for example, that a low-priority train should always be put on a siding when meeting a high-priority train, dispatchers will be guided by the CAD-generated meet/pass plans that minimize systemwide costs even if that means delaying a high-priority train running early in order to get a medium-priority mixed-freight train to arrive on time for a connection. The importance of high-priority trains is maintained through the high costs associated with the *late* arrival of such trains rather than by fixed operating rules. It should be noted that track maintenance work can also be assigned a cost function and be scheduled like a train, admittedly a very slow one. Rather than putting the track to be maintained out of service at the convenience of the maintenance-of-way gangs, dispatchers could evaluate possible "windows" of the required length and choose those that disturb the traffic flow the least.

A systemwide operating plan, then, consists of target arrival times for all trains in the system, planned track maintenance, planned car-block transfers at the yards, planned distribution of power and empty cars, crew rosters, etc., which should be updated periodically using feedback on the train status as generated by the CAD and yard information systems. Such operating plans are used by railroads at present, except that instead of setting the target train arrival times for the dispatchers, the operating plan is based on the expected time of arrival. The proposed methodology aims to change the position of railroad management, dispatchers, and other field officers; rather than being reactive, they can be proactive by setting clear operating objectives in the form of a systemwide operating plan and by minimizing the systemwide effect of disturbances when these objectives cannot be realized.

Note that this methodology does not require all trains to be scheduled at the tactical level. The only requirement of this methodology concerning the operating plan is that it should be realistic (i.e., feasible). The more "robust" the tactical schedules are, the easier it is to maintain a feasible operating plan and vice versa; if the tactical schedules are infeasible, then an operating plan based on those schedules would be impossible to maintain. Note that the feasibility of an operating plan should not be evaluated solely on the basis of over-the-line operations, but also by considering yard operations and capacity, systemwide flow of locomotives and empty cars, etc.

One of the main benefits of the use of optimal CAD systems within this framework will be the increased reliability of service through (a) better on-time performance for high-priority trains, (b) fewer missed connections and less deviation in the transit time for mixed-freight shipments, (c) faster turn-around and more predictable service of bulk-commodity unit trains, and (d) more evenly distributed workload and more reliable

operations at yards. An example from the proprietary study of benefits of optimal CAD systems can be used to illustrate the potential of well-designed CAD systems to improve on-time performance. During 16 hr of operations on a 300-mi-long line, 7 high- and medium-priority trains out of 25 total trains accumulated a total of 724 min of tardiness, ranging from 3 hr to 40 min per train; if these trains were dispatched with the support of an optimal CAD system, the total amount of tardiness for all 25 trains could have been only 14 min.

DESIRED FEATURES OF AN OPTIMAL CAD SYSTEM

Flexibility and Speed of the Optimal Meet/Pass Planning Algorithm

The optimal dispatching algorithm should be able to handle various types of cost function associated with train lateness; some obvious examples include piecewise linear, piecewise quadratic, and step functions, none of which are differentiable. The lower-bound–based algorithms described by Jovanović (*6*) can handle any nondecreasing function of train arrival times. (The constraint on the monotonicity of the cost functions is not overly restrictive: trains simply can be constrained not to arrive before a certain time, subject to congestion levels in the terminals.) In addition, the meet/pass planning algorithm should be able to dispatch trains between any two points on the line, even if there are no terminals at these points, in order to be able to handle maintenance-of-way gangs as special trains.

Dispatchers should also be allowed to exercise their judgment and expertise by manually constraining certain train meets to particular points or by preventing some meets from occurring at certain points. Finally, the algorithm should be fast enough to provide reasonably good solutions in enough time to allow the dispatchers to test the sensitivity of the plan to some random events, and to generate a revised meet/pass plan quickly if some of the input parameters change (e.g., if one of the locomotives in a certain train has broken down and this train can no longer achieve planned speed). At the same time, to allow systemwide operational planning, the algorithm should be able to handle long time horizons (at least 12 hr) and large dispatching territories. For example, if trains take on the average 9 to 10 hr to transit a particular line, a meet/pass planning horizon of 8 hr is obviously inadequate, because it cannot guarantee on-time arrival of all trains that are now in the system, let alone be used to plan for trains that have not yet entered the territory.

The heuristic algorithm described by Jovanović (*6*) is shown to satisfy all of the above requirements.

Display of Information

Time-distance diagrams ("string-line" diagrams, in railroad jargon) have traditionally been used by railroads to depict the progress of trains over the line. Thus, it is natural, as in most proposed CAD systems, to use time-distance diagrams to communicate suggested meet/pass plans to the dispatchers. One such diagram depicting a meet/pass plan over a predominantly single-track Whitefish to Spokane line is shown in Figure 1; the current plan time is 4:00 a.m. and the diagram shows planned train movements until midnight—the end of planning horizon. A track schematic is given along the left side of the diagram with sidings and double-track sections represented by rectangles. (There are only two short double-track sections in this example, one from Sandpoint to Algoma and the other one from Irvin to Spokane.) All train meets occur either at a siding or over a double-track section.

The information content presented to the dispatcher in the meet/pass diagram is very important. To illustrate this point, compare Figures 1 and 2. At first glance, the two plans are almost identical; yet the total amount of train lateness for the first plan is 230 min versus 40 min for the second plan. The cost of lateness (e.g., missed connections) is more than three times higher for the plan in Figure 1 than for the plan in Figure 2 (the latter plan was optimized using algorithms described by Jovanović (*6*)). All this is not obvious from the diagram: a relatively small change in meet locations for some train pairs, such as those marked with circles in Figure 1, eliminated or reduced late arrivals for six trains. The difference becomes more apparent if one observes that late trains in both diagrams are drawn using dashed lines from the point where they were made late; the number of late trains has decreased from seven in Figure 1 to only two in Figure 2. It would be even easier to differentiate the meet/pass plans if the trains were drawn with increasingly intense shades of red corresponding to the increasing cost of the trains' lateness.

The train-tardiness cost information embedded in the graphical interface of an optimal CAD system clearly presents to the dispatchers the effects of their actions and how well they are realizing the objectives set for them. In this way, the dispatchers can work toward achieving the common systemwide operating plan rather than attempting to move trains from their territory as fast as they can just so the trains can become somebody else's responsibility.

Assigning Track Time to Maintenance-of-Way Work Gangs

Maintenance-of-way (MOW) gangs could be treated as special trains within the context of an optimal CAD system because they occupy track capacity in space and time and cause congestion in a similar manner as trains. The main scheduling difference between trains and MOW gangs is that the activities of the latter are much less time sensitive; i.e., the value of the output of MOW gangs often remains constant when the completion time of the activity is shifted a few hours later or earlier; the same is not true for trains. The only direct costs associated with track maintenance that are time sensitive are those connected to labor (including nighttime and overtime) and equipment utilization. These costs are often not high enough to justify the hours of lateness caused to high-priority trains, but they may be high enough to give higher priority to the MOW gang over a low-priority coal-hauling train. Train dispatchers should have the final say regarding the exact time windows assigned to the MOW gangs and should be able to assess the effects that the assignments will have on train performance.

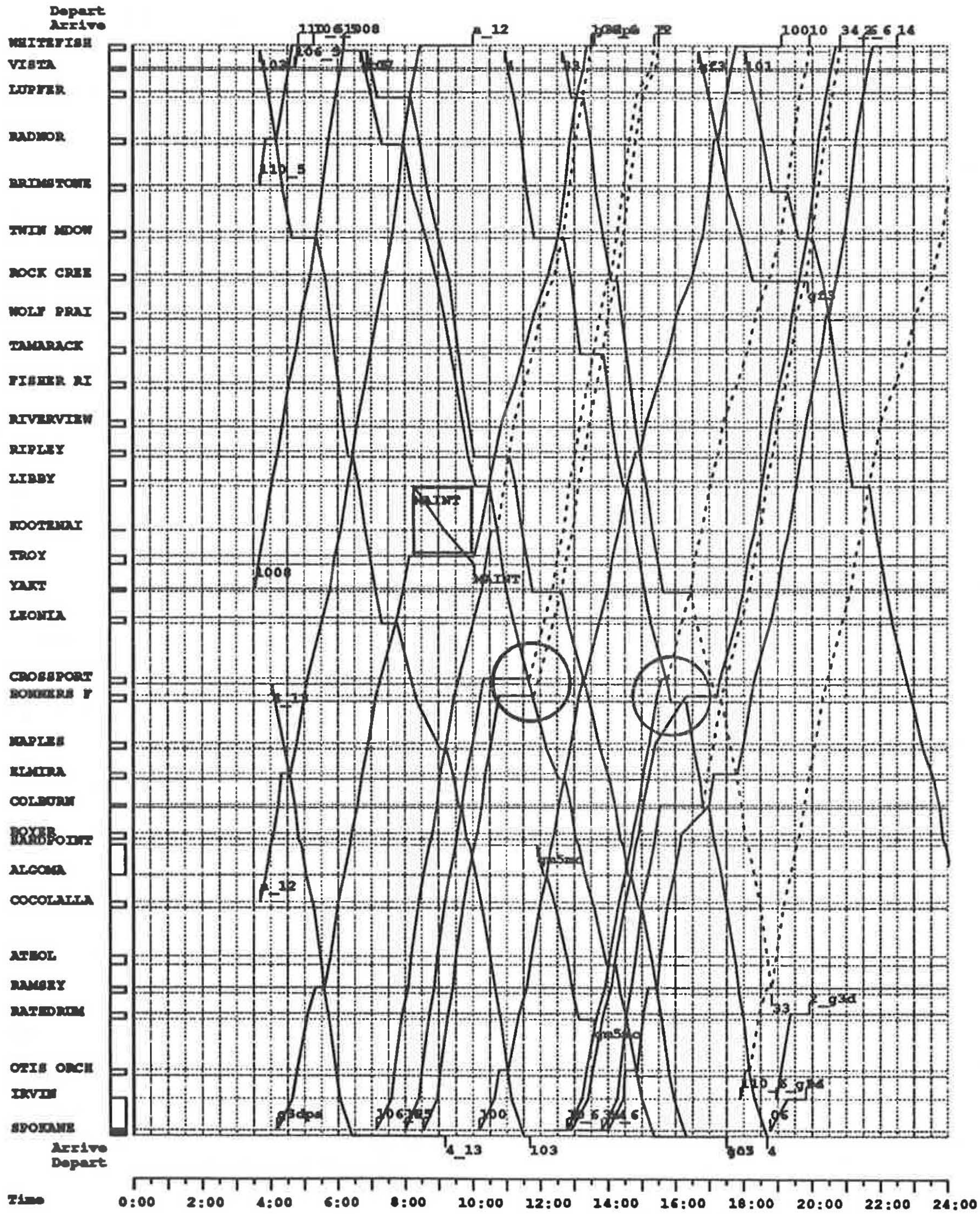


FIGURE 1 Unoptimized meet-pass plan.

The importance of the efficient allocation of time for track maintenance and its optimization by a CAD system is illustrated in Figures 2 and 3. In Figure 2, the outage of the track segment between Libby and Troy, represented by a rectangle marked "MAINT" that occupies this segment between 8:10 and 10:00 a.m., produces a significant delay and 30 min of lateness to Train "g3dps" (this is the train that departs from Spokane toward Whitefish starting shortly after 4:00 a.m.). If the maintenance block were treated as a special train that could be shifted in time rather than as a given constraint and the interval allocated to maintenance were shifted just 15 min ahead to 8:25 a.m. (10:15 as shown in Figure 3), Train "g3dps" could traverse the Libby–Troy segment before it was closed for maintenance and arrive on time at Whitefish.

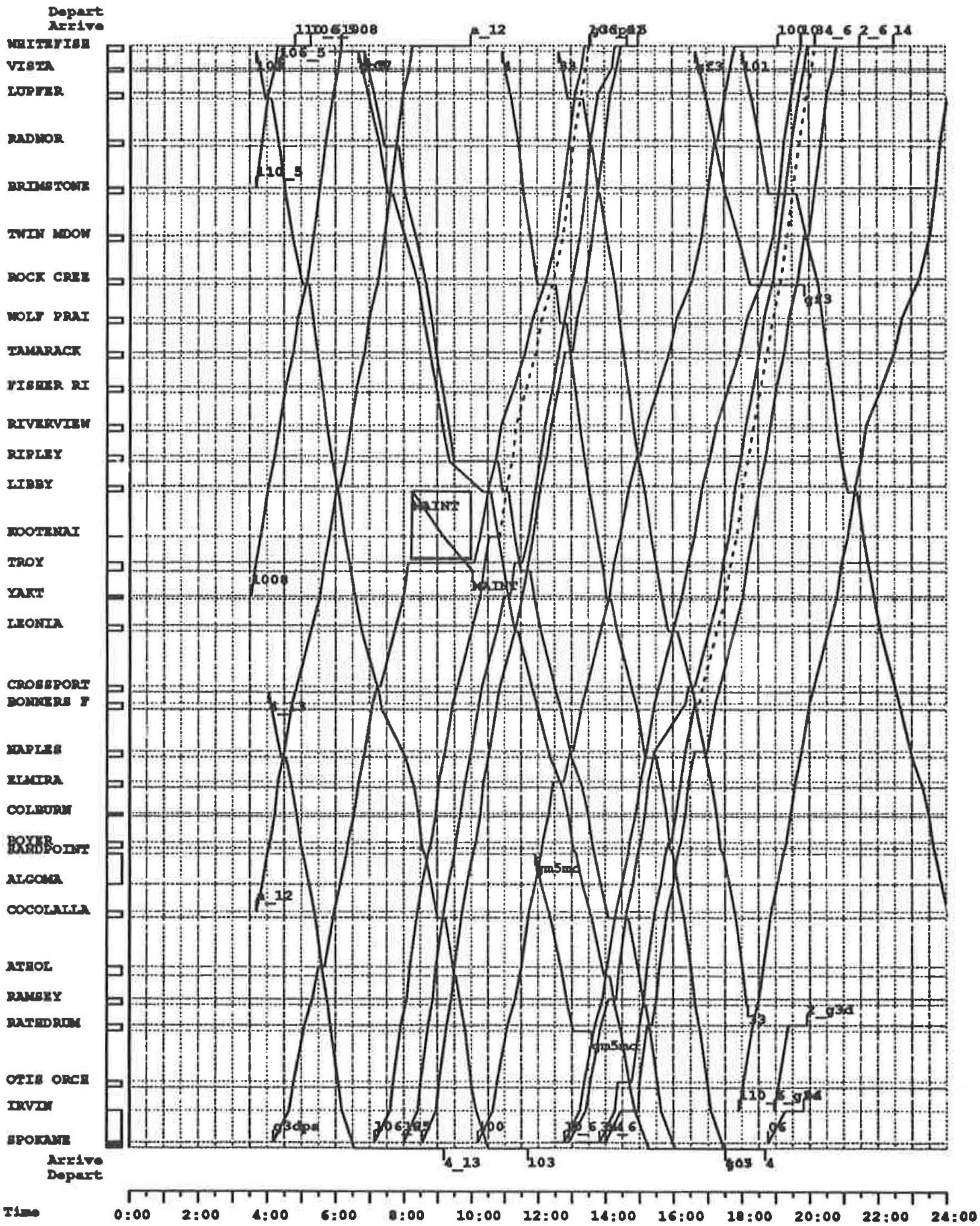


FIGURE 2 Optimized meet-pass plan with track maintenance treated as a constraint.

Control over Trains Entering the Line

In addition to having the control over time and track segments assigned to MOW gangs, dispatchers should control when and in what order new trains are permitted to enter their territories. The present situation in which dispatchers very often do not know when and which trains are going to enter their territory at a junction or from another line is analogous to air-traffic controllers having new aircraft appear unexpectedly in the middle of the air space that they control.

A similar argument can be applied to the order in which trains entering the line leave the yard; letting a slow train with a loose schedule depart in front of a fast high-priority train with a tight schedule either will delay the faster train and probably make it arrive late or require the faster train to overtake the slower train in a time-consuming and track

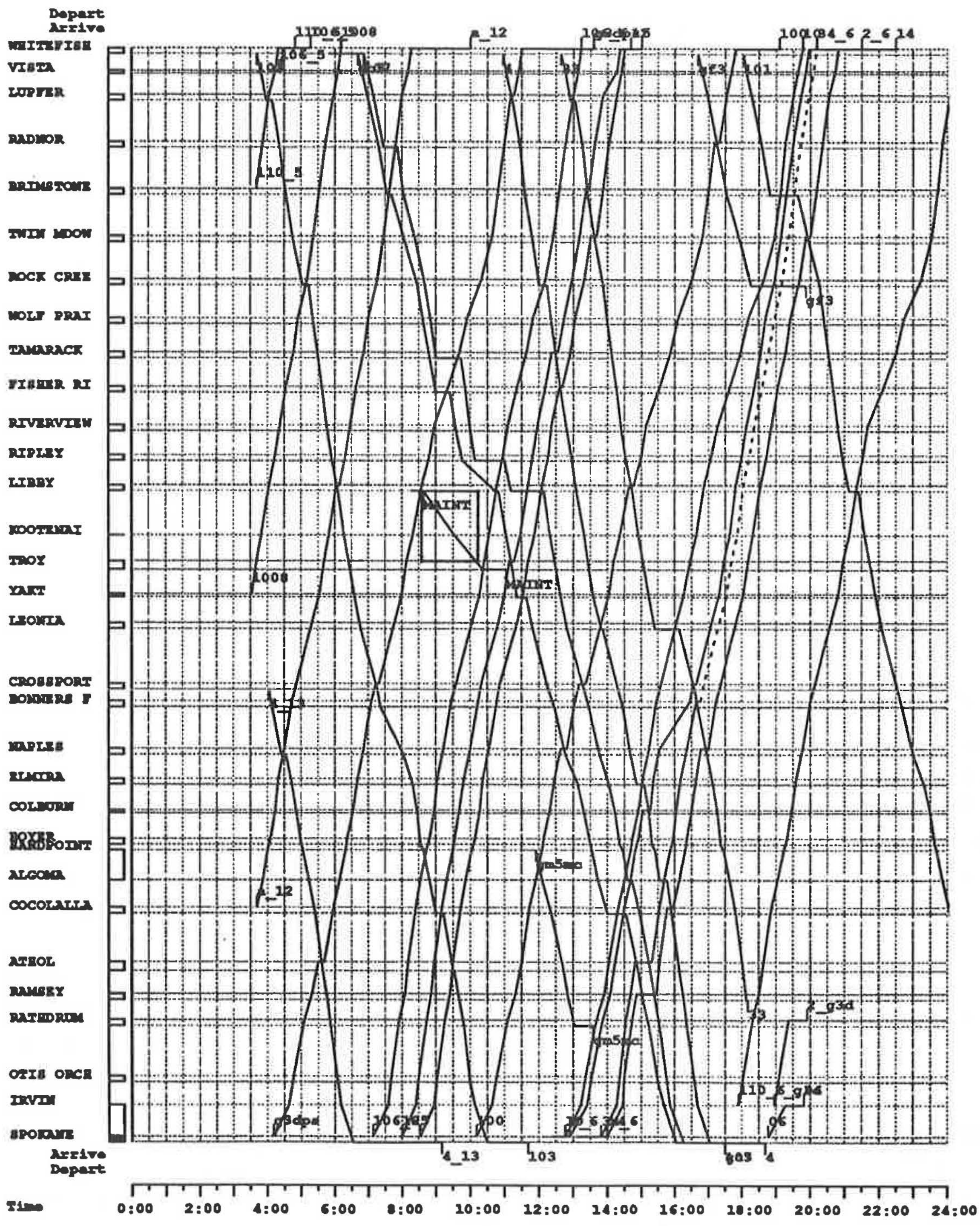


FIGURE 3 Optimized meet-pass plan with track maintenance scheduled as a train.

capacity-intensive process. However, delaying a train that is ready to depart from a yard so that some other trains can depart from the yard first can cause congestion in the yard. Hence a trade-off between line delays and yard delays may be required.

The importance of the ordering of trains departing from the yard is shown in Figures 3 and 4. The only late train in the meet/pass plan in Figure 3 (dashed line) is delayed and made late by a slower train in front of it. Although the late train departs from Spokane almost 1 hr after the preceding slow train (1:00 p.m. versus 1:45 p.m.), this time gap is quickly eliminated by a difference in speed between the two trains. Normal railroad practice would be to allow the late train to overtake the slow train in front; however, overtakes can often be eliminated by planning the order in which trains depart from terminals. In Figure 4, the slow train in front was ordered to depart from Spokane at 2:55 p.m., behind the train that is late in Figure 3, and both trains arrive on time at Whitefish.

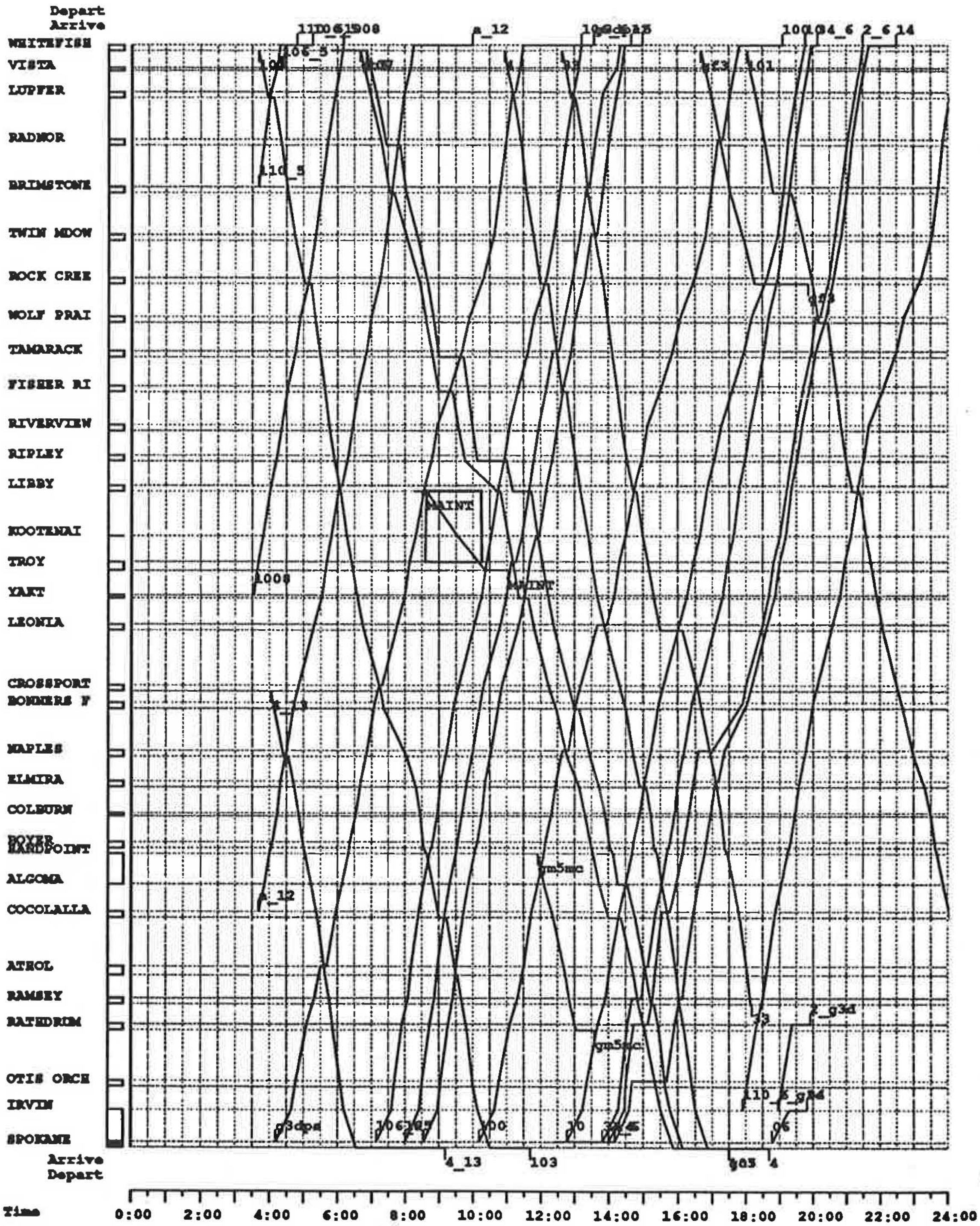


FIGURE 4 Eliminating late arrivals by reversing order of departure.

If this change in departure time is planned far enough in advance and there is enough yard capacity, almost 1 hr of the slow train's on-duty crew time can be saved along with gaining more time to assemble, power, and inspect the train.

One possible approach toward incorporating the line/yard trade-offs would be to include a simple yard model in the CAD algorithm. This model would account for the limited yard storage capacity by imposing additional constraints and treat the line, branches, and yards as a continuous dispatching territory covered by a single optimal CAD system. The algorithms described by Jovanović (*6*) can be extended to accommodate this approach. Another, probably less optimal approach would be to optimize yard and branch line plans separately and then have some higher-level model (or a decision-maker) modify these plans to ensure their compatability.

EXTENSIONS

The CAD methodology described in the previous sections assumes that no random events occur during the planning horizon. Of course, this is not realistic; equipment failures and adverse weather conditions influence the performance of trains and track availability, which in turn can render the current meet/pass plan infeasible. Another source of randomness in line-haul operations is the late departure of trains from their originating yards or the addition of new trains. Thus, the further one extends the planning horizon of a CAD system, the greater the probability is that the plan will have to be modified due to unforeseen events. There are several ways to approach this problem.

The approach used in the Norfolk Southern optimal CAD system (*5*) is to discount future costs associated with dispatching delays; i.e., each train's cost is multiplied by a discount factor $(T - a)/T$, where T is the planning horizon of the meet/pass planner and a is the interval between the current time and the expected time of the train's arrival in the territory ($a = 0$ for trains already on the line). One problem with this approach is that it tends to underestimate costs; i.e., although it is true that the actual meet/pass pattern to be implemented in the future and the associated costs are uncertain, it is much more likely that the actual costs and delays will be *higher* than the costs predicted by the optimal meet/pass plan because of potential accidents and the deterioration of the dynamic performance of some trains. Nondiscounted cost estimates associated with a minimum-cost meet/pass plan will already tend to be below the expected cost; the use of discounted costs will amplify this error.

Another approach might be to focus on the nature and causes of the stochastic events that affect line-haul operations. A belief has started to emerge in the railroad industry that a large number of these events can be controlled and significantly reduced or eliminated by better planning, preventive maintenance, and better work discipline. The examples of high-precision, punctual operations of Japanese, French, and Swiss railroads show that random events can be controlled and accounted for. The highly reliable operations of these railroads can be attributed in part to fully double- or multiple-track lines with fewer train conflicts; in fact, a well-designed CAD system can have an effect similar to that of adding an additional track in terms of the reduction of train conflicts. However, some practical way to handle the problem of stochastic events is needed before the long-term efforts aimed at better planning, maintenance, etc., produce the desired results. The reliability of a meet/pass plan can be increased by allowing sufficient slack in the plan. Thus, instead of decreasing the *average* train transit times, the capabilities of newly installed optimal CAD systems can be used to decrease the *variance* of train transit time. With sufficient reserves built into the minimal point-to-point train running times used in the meet/pass algorithm, the probability that trains will be able to achieve these times increases. As the random events become more controllable, the slack in the meet/pass plan can be reduced and, consequently, the scheduled transit times can be decreased to yield faster service and better rolling stock utilization.

Another important issue related to the implementation of optimal CAD systems is the problem of coordination of dispatching and yard activities throughout the railroad network through the generation of targets and objectives for the dispatchers and yardmasters. A systemwide, real-time information system containing positional data and status of all carblocks and trains is necessary to achieve the full benefit from optimal CAD systems. A need and a research opportunity arise in providing additional modeling and optimization tools to support the generation of systemwide operating plans. Only through efficient systemwide operational planning of railroad operations can the potential of the proposed ATCS technology be fully realized.

ACKNOWLEDGMENTS

This research was supported by a grant to the University of Pennsylvania by the Burlington Northern Railroad and by the National Science Foundation under their Presidential Young Investigator Program.

REFERENCES

1. *High Technology Rides the Rails*. Association of American Railroads, Washington, D.C., 1985.
2. *Revolution in Research Will Benefit Railroads and Shippers*. Association of American Railroads, Washington, D.C., 1986.
3. G. Welty. ATCS: On Time, on Target. *Railway Age*, June 1987.
4. E. R. Petersen, A. J. Taylor, and C. D. Martland. An Introduction to Computer-Assisted Train Dispatching. *Journal of Advanced Transportation*, Vol. 20, 1986, pp. 63–72.
5. R. L. Sauder and W. M. Westerman. Computer-Aided Train Dispatching: Decision Support Through Optimization. *Interfaces*, Vol. 13, 1983, pp. 24–37.
6. D. Jovanović. *Improving Railroad On-Time Performance: Models, Algorithms and Applications*. Ph.D. dissertation. Department of Systems, University of Pennsylvania, Philadelphia, 1989.
7. I. B. Duncan, K. M. Winch, and G. A. Bundell. Driver-Assist—Microprocessor Technology to Aid in the Scheduling of Trains. In *Computers in Railway Management: Proceedings of COMPRAIL 87 Conference in Frankfurt, West Germany* (T. K. S. Murthy, L. S. Lawrence and R. E. Rivier, eds.), Springer-Verlag, Berlin, Germany, 1987.
8. B. Szpigel. Optimal Train Scheduling on a Single Track Railway. In *Operational Research '72* (M. Ross, ed.) North-Holland, Amsterdam, Netherlands, 1973.
9. R. D. Burns and D. B. Turner. Safety and Productivity Improvement of Railroad Operations by Advanced Train Control Systems. In *Proceedings of a 1989 Joint IEEE/ASME Railroad Conference*, Philadelphia, Pa., April 1989.
10. CSX: 'Welcome to the 21st Century!' *Railway Age*, June 1989.
11. I. P. Milroy, B. R. Benjamin, R. L. Payne, J. Kautsky, P. J. Pudney, A. J. Skinner, and S. P. Rice. *Freight Locomotive Fuel Conservation Project—Final Report*. South Australian Institute of Technology, The Levels, Australia, 1987.
12. D. Kraay, P. T. Harker, and B. Chen. Optimal Pacing Of Trains in Freight Railroads: Model Formulation and Solution. *Operations Research*, 1990 (forthcoming).

Integration of ATCS with MIS

M. FRANK WILSON

Advanced Train Control Systems (ATCS) bring economic benefit to railroads through the automation of existing procedures and the provision of new features. Many of these benefits are realized through integration of ATCS with the railroads' management information systems (MIS). The integration of ATCS with MIS poses technical challenges to both the ATCS community and the MIS community. The ATCS and MIS must adopt compatible data communication protocols. ATCS must be able to adapt to changes in the MIS community—among them, a move toward distributed and diverse systems. Being able to adapt to these diverse systems may require adapting ATCS protocol standards. The interface protocols must also provide for appropriate implementation of security procedures. These procedures must be provided to prevent breeches of security in either the MIS or the ATCS. This is necessary because of the sensitive business nature of MIS data and because of the safety-critical nature of some ATCS functions. Another technical challenge is the formulation of an appropriate strategy for distribution of information throughout combined ATCS-MIS. This strategy must account for the fact that ATCS have limited capacity for data transmission and that MIS applications share that capacity with other ATCS applications. Overloading ATCS data networks can seriously degrade the operational benefits that are the prime benefits of ATCS.

Advanced Train Control Systems (ATCS) bring economic benefits to railroads through the automation of existing procedures and the provision of new functions that are oriented toward the safe, efficient control of trains.

Management information systems (MIS) provide for the control and distribution of vital business information. MIS continually face the challenge of change and growth to meet the changing needs of the railroads. MIS also must adapt to change in the technology of commercially available computer systems.

Many ATCS benefits are best realized through the integration of ATCS with the railroad's MIS. The integration of ATCS with MIS must combine ATCS requirements for timely, efficient, and safe control, and MIS requirement for security and change. Therefore, successful integration imposes requirements upon both ATCS and MIS.

PROTOCOLS

The first requirement for integration of ATCS and MIS is for the systems to communicate with each other. The two communities must implement common communication protocols. Modern MIS are implemented on an increasingly diverse set of platforms. These range from traditional, large, proprietary mainframe systems to personal computers or networked open architecture computers. These options represent significant challenges in constructing the communications protocol strategy.

One strategy is to modify MIS to communicate in ATCS protocols. This is a very effective strategy and can yield very efficient systems. It can, however, be expensive if ATCS protocols must be developed for many types of MIS computers.

Another potential strategy is to implement MIS protocols in the ATCS. This also can be a very effective solution. However, it can be expensive if a large number of MIS protocols must be implemented.

Another strategy to realize common communications protocols is to deploy protocol adapters (Figure 1). With this scheme, ATCS and MIS need not implement identical protocols. A protocol adapter's task is to convert the protocol of one system into the protocol of the other system. The protocol adapter function is analogous to that of a human interpreter.

A hardware protocol converter entails the expense of an additional machine. However, separate hardware also provides solutions for the translation of data formats and the adaptation of differing time schedules between ATCS and MIS. Many ATCS features are event driven and interactive, while some MIS processes may be batch oriented. Hardware can effectively arbitrate between these different modes of operation. When these functions of data translation and time schedule buffering must be implemented, a protocol translator is the most logical choice.

NETWORKING

Current trends in MIS are dominated by the distribution of computer resources. MIS are no longer entirely centralized. This distribution of MIS imposes its own set of challenges upon system integration.

Directory services are required by ATCS to identify the appropriate MIS entity in a distributed MIS environment. Implementation of directory services allows the MIS distribution strategy to change without major changes to the ATCS because changes can be isolated to the directory service.

The implementation of directory service in the integration strategy imposes the requirement that ATCS be capable of routing based on the entities that are identified in the directory (Figure 2). Implementation of directory services in ATCS requires further work.

Locomotive health monitoring is an example of how directory services may be used to integrate ATCS with MIS for motive power management and mechanical department planning. Certain locomotive health messages may be routed to either of these MIS as desired by a particular railroad. Further, it may be desirable to route messages to the mechanical planning system for a particular repair facility. These routing

Railroad Electronics, Rockwell International Corporation, 400 Collins Road NE, Cedar Rapids, Iowa 52498.

FIGURE 1 Protocols.

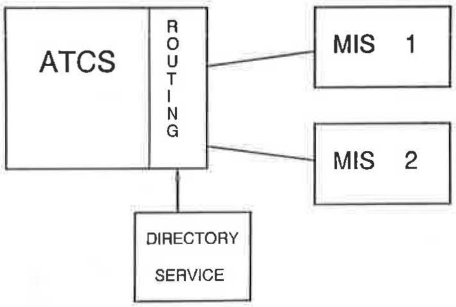


FIGURE 2 Routing with directory service.

decisions should be implemented independent of the health-monitoring logic.

SECURITY

Both ATCS and MIS have their own security requirements. An integrated ATCS/MIS solution must accommodate both systems' security requirements.

One major factor affecting ATCS security is availability of the system for operational use. ATCS will become indispensable for the operation of railroads. Consequently, ATCS architectures must ensure that noncritical administrative data do not overload the data network. This is accomplished by a combination of data traffic prioritization and load-limiting functions.

Another threat to ATCS security is the emulation or corruption of operational instructions such as movement authorities. ATCS protocols are designed to prevent inadvertent corruption of safety-critical messages by the data network. The integrated systems must prevent counterfeit safety-critical messages from entering operational ATCS by way of an MIS interface. Care must also be taken to prevent the corruption of data bases used by ATCS safety-critical processes.

The major security needs of MIS relate to the security of data; both data stored by MIS and made available to ATCS, and data provided to MIS by ATCS. Sensitive business information must not be disclosed to unauthorized parties via inadvertent access to MIS data or by interception within the ATCS network (such as at the mobile radio link). Appropriate measures to safeguard sensitive information should be taken. In some cases this may include encryption of data.

MIS rely upon information conforming to certain business rules or data constraints. Some examples of data constraints are restricting cars set out to be cars already part of the train consist or restricting crew members' hours of service to those labor categories for which they are qualified.

Data constraints must be properly enforced upon data sent from ATCS to MIS. One of the challenges of ATCS/MIS integration is to determine which data constraints should be enforced in ATCS and which should be enforced in MIS. Once this has been determined, methods for recovering damaged or improper data must be put in place. Proper systems integration can ensure inexpensive and efficient recovery methods.

In general, the threats to the security requirements mentioned above are from three sources: software errors, user errors, and malicious actions. Proper systems integration can ensure each security requirement is protected from each of these security threats.

INFORMATION DISTRIBUTION STRATEGY

ATCS networks and most MIS networks operate at significantly different speeds. Many MIS networks and networked applications are designed for use over local area networks operating at speeds of over a million bits per second. ATCS mobile data networks operate at speeds of a few thousand bits per second. As a result ATCS networks determine both the total amount of data that can be sent and the response times that can be expected.

One method that may be used to optimize the use of ATCS mobile data links is compression of data. This is especially appropriate for terminal emulation sessions in which the information is generally encoded as viewable text. Viewable text is a relatively inefficient scheme for encoding information and is easily compressed by available algorithms.

Another method to optimize use of ATCS mobile data links is the local storage of screen formats. With this method, a screen format need not be sent across the data network. This saves the bandwidth used for transmitting the screen definition. Many MIS terminal-oriented applications transmit screen formats as a normal operation and will require some modification to omit transmission of formats that are stored.

Another important factor is the triggering of data transfer transactions. Triggering conditions are generally event oriented. Examples of events that may serve as triggers include a locomotive equipment failure, a successfully completed set-out of cars, or acquisition of a locomotive by the data network (locomotive coming into coverage). Alternatively, a trigger can be a set time of day assumed to be a nonbusy time for the mobile network. There may be some data that should be sent daily and could be delayed until a nonbusy time. It is also conceivable that ATCS network services could be created that would identify low-traffic periods as they occur and notify MIS applications. In this manner, a nonbusy time would not be hard-coded into an application, and triggering could be adaptive to current conditions.

APPLICATIONS

There are many applications of ATCS/MIS integration that have already been tested and some that soon will be. Many more will be identified in the future as real ATCS are deployed.

MIS is often the most appropriate source for train brief (schedule and route) and consist information. Two of the key challenges for this application are the encoding of the information in a format acceptable to both ATCS and MIS, and

the relative timing of when the information is available and when it is needed. Planned information may be known far in advance of the departure of a train. However, accurate "as-departed" information may not be known until just prior to or just after departure. ATCS applications must account for the availability time of information by not requiring data that cannot be made available and by providing means to accept incremental updates as more accurate data become available.

Long-term archival storage of train sheet information is another good candidate for ATCS/MIS integration. Many MIS shops already have very secure data archival systems in place. Utilizing these same systems avoids the cost of duplication.

Estimated time of arrival information based on actual real-time conditions is data that can be provided to MIS by ATCS. This information is very valuable for terminal-area planning and customer service.

Work order applications are a natural application of ATCS/MIS integration. The benefits to the railroad are improved car tracking and improved responsiveness to customers. Customer car releases entered into MIS by customer service agents can automatically trigger the transmission of work order instructions to work order terminals. Also, work order completions entered by train crew members can trigger MIS billing systems.

Locomotive health applications provide a means to better utilize locomotives. Locomotive health information is useful in various MIS, as discussed earlier. Mechanical department systems require detailed failure and diagnostic information, whereas motive power management systems require status summary information. Depending upon the operational scenario selected by a railroad, the relative time urgency of this information may differ. The status summary information may be needed urgently to initiate traffic planning. The detailed failure information may not be needed as urgently, and therefore the triggering condition for transmission may be different.

Engineering and maintenance functions include payroll, material control, and productivity measurement. This information is available via ATCS track forces terminals.

Operational analysis is a task that can utilize several types of data generated by ATCS. These can range from the time to traverse regions of territory to the use of fuel. ATCS provide a mechanism to collect much of this valuable information automatically. Prior to ATCS, operations analysis tasks frequently required much manual collection of data.

Another exciting application for the integration of ATCS with MIS is the integration of electronic data interchange (EDI) applications. These applications could provide dramatic improvements in customer service as well as cost savings due to automation.

EDI applications can work in both directions. ATCS actions can trigger EDI transactions. For example, completion of a set-out action in ATCS can trigger an EDI "freight bill" transaction. EDI transactions can also trigger ATCS actions. For example, an EDI "shipping instructions" transaction can trigger ATCS pick up actions.

CONCLUSION

The integration of ATCS with MIS will create many identified ATCS benefits as well as many benefits not yet envisioned. Integration of ATCS with MIS will grow as the utilization of real-time operating information from ATCS permeates the railroad resource management structure. Proper systems integration resolves the issues of common protocols, network issues, security, and information distribution strategies, and yields dramatic benefits to the railroad.

Realizing Benefits from ATCS Using a Motive Power Information and Management Support System

MARK HORNUNG, HOWARD ROSEN, JOHN SZYMKOWIAK, AND DAN DION

The motive power management function at a railroad can be significantly improved thanks to more timely and accurate information. With earlier and more reliable knowledge of train and locomotive performance and demands for power, motive power managers can improve their forward planning, which leads to improved locomotive utilization and better on-time train performance. Advanced Train Control Systems (ATCS) can be an important source of information for motive power management. With their train location, locomotive health, and work order reporting systems, ATCS have the potential of increasing accuracy to near 100 percent and reducing to a matter of seconds the time lag between an event and when that event becomes known to motive power managers. For this more timely and accurate information to be exploited, it must be organized and presented to the motive power managers in an efficient manner. In addition, there must be established a mechanism for timely communication to field forces of the motive power managers' plans. To achieve this, a computerized motive power control system was designed and implemented at Canadian National Railways. It consists of graphic displays of current train and locomotive location and status, alerts that highlight critical new information, functions for motive power planning, and facilities for communicating plans to field forces. With the motive power system in place, and gathering its information from ATCS, managers know about and can respond immediately to changes in train and locomotive demand and performance. Although ATCS are not a prerequisite for achieving benefits from an improved management control system, an effective management control system for trains and locomotives *is* a prerequisite for achieving full benefits from ATCS.

Railroad motive power is an expensive asset that requires efficient management. Canadian National Railway (CN) operates a fleet of 2,000 diesel locomotives on a 50,000-km railroad that spans Canada from Halifax on the Atlantic coast to Vancouver on the Pacific coast.

Like many railroads, CN manages its motive power from one central control center. CN's motive power control center is responsible for monitoring and distributing locomotives, vans (cabooses), and end-of-train devices to meet the needs of more than 700 trains each day. It must operate these trains with as few locomotives as possible, but not delay any trains because of a lack of power. It must balance the flow of locomotives to take into account future demand and must meet locomotive maintenance requirements.

As of the mid-1980s, the center was principally a manual operation essentially unchanged since the conversion to diesel locomotives approximately 30 years before. Motive power assets were tracked using a large magnetic board covering one wall of the control center. The board contained a track schematic of the CN system. Each locomotive and van was represented by a moveable magnet. Reports of train and unit movements were received in the center via COMTEL (teletype) and telephone. The motive power distributors moved the appropriate magnets to correspond to a movement report. Colored tags were stuck on the magnets to indicate reports of abnormal condition, such as failures or inspections due.

Motive power control and CN's technological development department began looking at ways to modernize the operation center starting in 1985 and continued to do so off and on into 1987. Several studies were done during that time that highlighted the need for more timely and accurate reporting of events in the field. Coincident with CN's determination of this need, the railroad industry and its suppliers were investigating means for reliable automated tracking of trains as part of the effort to develop Advanced Train Control Systems (ATCS), and commercial demands were being made for automatic equipment identification systems to provide customers with more timely tracking of their shipments. As a consequence, CN committed to improvements by installing a railroad equipment identification system (REIS), on-board transponders, and wayside detectors at key locations to automate reporting (*1*, *2*). REIS would selectively replace reporting by field clerks in feeding CN's main computer system, TRACS. Field clerk reporting is typically hours behind reality, sometimes inaccurate, and therefore not relied upon by the motive power controllers. Obviously, the manual procedures in motive power control would have to be automated to take advantage of the improvement in timeliness and accuracy of reporting that ATCS technology would provide.

USER NEEDS

In the fall of 1987, a full-time user representative from motive power control joined with members of CN's technological development department to identify requirements for a new motive power control system. Questionnaires, studies, and in-depth discussions were used to produce an initial set of requirements. In April 1988, an outside consultant was retained to further refine the requirements and produce a system design.

CN's motive power is managed from a central control center, manned 24 hours daily, every day of the year. Normally

M. Hornung and H. Rosen, ALK Associates Inc., 1000 Herrontown Road, Princeton, N.J. 08540. J. Szymkowiak and D. Dion, Canadian National Railway, Technological Development, Pl. Bonaventure, 935 de La Gauchetiere St. West, Montreal, PQ H3B 2M9, Canada.

four distributors are on duty at any one time, each of whom distributes power to a particular geographic region. On the weekday shift, a coordinator is on duty who exercises a supervisory role and ensures that the supply of power is properly balanced between regions. The system operations control officer handles these duties as necessary when the coordinator is not on duty.

The principal task of the motive power distributors is to plan upcoming locomotive assignments. For each train departing each major yard, the distributor must decide which locomotives to assign. The distributor works with yard personnel to determine this assignment. At minor yards, the yard personnel usually "turn around" or "send on" power to the next major yard, but the distributor may also need to specify plans at minor yards in certain circumstances. The distributor's work product is a list of locomotives, by serial number, to be assigned to each outbound train, from each major yard.

The distributor's work is complex because of the large number of variables that must be considered. The supply of locomotives is limited. Not all locomotives are suitable for every train. Certain trains require one or more locomotives with one or more special features. Trains should not be overpowered, lest locomotives and fuel be wasted. Trains should not be underpowered, lest they be unable to meet their scheduled transit times. Requirements for power are often unbalanced by direction, by day of the week, or by season, so locomotives must be "repositioned" to be where they are needed when they are needed. Locomotives must be cycled to a repair shop as close as possible to when they are due for maintenance. Locomotives fail en route and must be "rescued." There are also requirements for local and yard power that the distributor must meet.

A need was identified for both a broad-based view of power movements for purposes of achieving balance in power flows and a close-up view of individual trains and units for the distributors to use in determining which particular units should power which trains. It was important to the broad-based view that the trains and units appear in their proper geographical relationship. The existing magnet board was providing both views for the users in that they could step back and get the overall view or they could walk up close to the board to see detail in a particular area. The controllers and coordinator wanted to have both these views in any automated system.

Another need was for the system to remind power controllers of work to be done. Placing paper stickers on unit and train magnets, placing magnets at odd angles (such as upside down or sideways), placing blank train racks on the magnet board, and piles of paper notes and COMTELs were all being used by the controllers to note work to be done. The controllers hoped that the automated system would organize work to be done in a helpful way and provide automatic reminders of things that needed attention. Certainly, the system could not rely on the controllers hunting and searching through a computer system for things that needed attention.

The users also pointed out a shortcoming in the visual symbology of the current board. Color coding was being wasted on static attributes such as builder class, number of axles, and so forth. After a short time working in the center, all of the controllers had these static attributes memorized based on ranges of serial numbers. The colorful symbology of the magnets was unnecessarily obscuring important information on temporary conditions. This information was being noted on small stickers stuck to the magnets and by placing magnets sideways or upside down. The new system should focus on highlighting temporary conditions and not be made busy with unnecessary static information. It should not be what Tufte (*3*) calls a "graphic duck."

It was desirable that a new system automatically detect problems with field-reported data or controller instructions. For example, if a controller tried to power a train with too few locomotives for its tonnage and train profile, the system should alert the controller and require confirmation before issuing the power consist instructions to the field. Similarly, if an impossible or contradictory locomotive reporting was received from the field, the controller should be automatically alerted to solicit a correction on the reporting.

Finally, the system should be expandable to take advantage of advances in expert system, optimization, and other technologies. CN was interested in the expert system and optimization logic of ALK Associates' Locomotive Distribution System (*4*). CN and its consultant agreed that it was essential to automate both inbound reporting and outbound instructions as a prerequisite to considering expert system or optimization technology.

SYSTEMS AT OTHER RAILROADS

Simultaneous with surveying user needs, other railroads were canvassed to see what solutions they had undertaken to automate motive power control.

The largest group of railroads had nonautomated systems similar to CN's then-current system. Canadian Pacific Railroad (CP), CN's principal competitor, has its control center a few blocks from CN's in Montreal, Canada, and the two centers appeared to operate nearly identically. CP's center did seem to rely more on the telephone than on COMTELs relative to CN's. The CP power controllers wore headsets so they could stand at the magnetic board and move magnets while listening to reportings on the telephone. Like CN, the CP controllers made limited use of mainframe computer terminals to perform enquiries.

Several railroads (Union Pacific, Conrail, and CSX) had recently replaced their magnet boards with nongraphic terminals linked to their mainframe systems. Motive power managers at all three railroads were relatively unhappy with this. The mainframe systems forced the controllers to do a lot of hunting through the system for locomotives, there was no queuing of reminders of work to do, no way to see power in its geographic perspective, and no way to get a broad view of how power was doing. In addition, the amount of typing involved was extensive and introduced numerous errors. (CSX has recently begun to address these shortcomings (*5*).)

Two of these railroads (Union Pacific and CSX) were in the process of installing large "cyclorama" dispatch centers, containing enormous (hundreds of projectors) computer-driven wall displays for dispatching the entire railroad system from one large room. To improve communications between controllers and dispatchers, the motive power controllers were to be relocated to these dispatch centers. Management at both railroads thought the controllers would obtain some value from the large wall displays. However, the displays were de-

signed for the dispatching function and are of little use to motive power control. Only trains, not locomotives, are shown on the wall displays. In any event, the displays are too dim and the controllers' desks are too far away for the displays to be readable.

The most extensive attempt found to automate motive power control was Burlington Northern's CAPMAC system. This system features an array of 42 CRT screens that contain a schematic of Burlington Northern's system and the trains on it. Like the dispatch centers, the schematic is in the form of straight lines and does not show the rail lines connecting in the proper geographic context. However, it is oriented to motive power control in that individual locomotives are shown, and the controllers can easily read the displays from their desks. A very complicated visual symbology is used to convey both static and transient locomotive attributes. Like the mainframe systems, however, there are no automatic reminders of work to do, no checking of work, and the system requires a lot of typing, making it prone to human error. Another disadvantage was that CAPMAC runs on a type of computer that has been discontinued and it would be expensive to port CAPMAC to more modern equipment.

Two railroads (Burlington Northern and Union Pacific) were beginning to install expert system and optimization technologies for distributing locomotives (*4*). Both railroads were reporting efficiencies from using these systems (*6*; Hornung, unpublished data), although acceptance was being hampered by problems in the timeliness and quality of data reporting. Like CN, as part of a broader effort to implement ATCS, both railroads are moving to automated reporting technology similar to REIS, which, by improving data reporting, will make such systems more useful (*7*).

DESIRED SYSTEM FEATURES

It was clear from the user interviews, and from the evaluation of other railroads' systems, that the magnet board had a number of attractive attributes that should be incorporated into an automated system. The system must be designed to meet the needs of motive power controllers, not dispatchers. It must support views for both broad-based power flow and detailed power consisting. It must show power in its proper geographic perspective. It should be as easy to use as moving magnets on a wall. It should not require lengthy typing of locomotive numbers or train identifiers.

To improve upon the manual system, the automated system should also contain a number of new features. It must update itself automatically without human intervention at any point in the process. It must automatically remind the controllers of work to be done, and it must check the integrity of both inbound reportings and outbound controller orders. It should be expandable to include expert system and optimization logic. Finally, it should use a widely accepted computer architecture so as to prevent unnecessary dependence on any one vendor and should use current technology to minimize the chance of early obsolescence.

HARDWARE SELECTION

To meet the desired system attributes and avoid the shortcomings of other railways' systems, a graphic user interface was deemed essential. This eliminated nongraphic and noninteractive graphics systems from consideration. The intensive graphics, size of data files, and desire for expansion ruled out low-end systems such as personal computers. It appeared that what are commonly referred to as "graphic work stations" would be the most appropriate platform. These are single-user microcomputers, usually connected in networks, that are more powerful than personal computers. Further, to support a large "system" view, high-resolution, wall-mounted projection screens were selected.

USER INTERFACE

System and Detailed Views

The completed motive power system (MPS) has one system view and a number of detailed views. An artist's rendering of approximately one-twelfth of the system view is shown in Figure 1. The system view is a schematic of the mainline trackage of the CN railroad system, shown as solid lines connecting major yards. Each train is shown, at its last reported location, as a sequence of small rectangles representing the train itself and the units on it. The complete system-level view is projected on a wall of the control center. The user can change which portion of the view is shown on a work station by clicking on the relevant area on the scrolling strip (bottom of Figure 1), which is a condensed version of the entire system display.

There are also four detailed views in MPS: yard-, outpost-, branchline-, and link-level displays. These can be invoked from the system-level display by clicking on the appropriate yard, link, etc. An artist's rendering of a yard-level view is shown in Figure 2. The name of the yard (Gordon in Figure 2) and various statistics are shown at the top of the view. There are areas in the view for outbound trains in two directions (top left and top right in Figure 2) and for inbound trains from two directions (bottom left and bottom right in Figure 2). The center of the screen shows units that are not on trains but that are at the yard. These are shown according to what facility they are at in the yard; for example, at a shop (in a repair facility) or on the ladder track (serviced and ready to go).

Unit and Train Symbology

On all MPS views, trains and units are represented by icons using consistent symbology. Unit icons are rectangles with a pointed end, which show the direction the cab on the unit is facing (the icon for cab-less units is pointed at both ends). The unit number is shown inside the icon. The color of the unit icon corresponds to its current condition, and the color of the icon tip indicates ownership and/or lease status. A unit icon may also have a white border indicating that the unit is preassigned to a train. Such a unit is displayed at its preassigned location with a solid border and at its actual location with a dashed border.

The MPS train icon is a rectangle containing the train's identifier (train I.D.), together with a small square that contains a one-letter status code. The color of the train icon indicates the train's on-time or lateness status. The unit icons for each unit on the train, in order, follow the train icon.

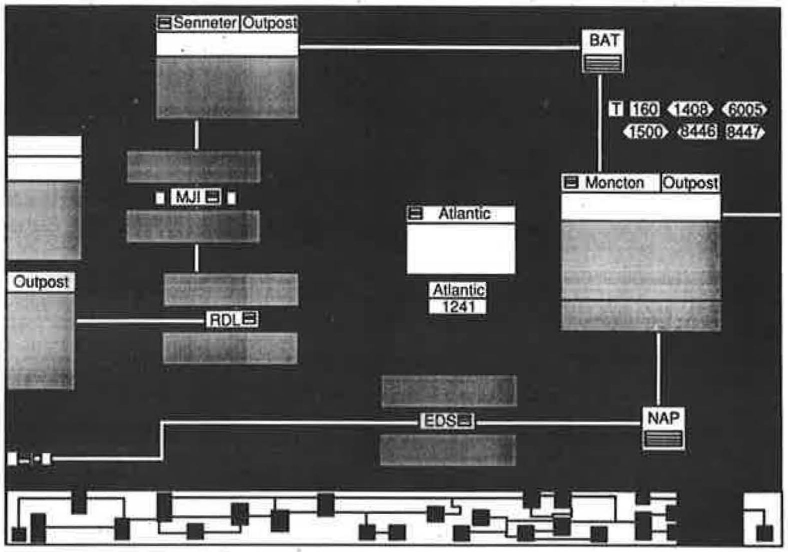


FIGURE 1 One section of the MPS system-level display.

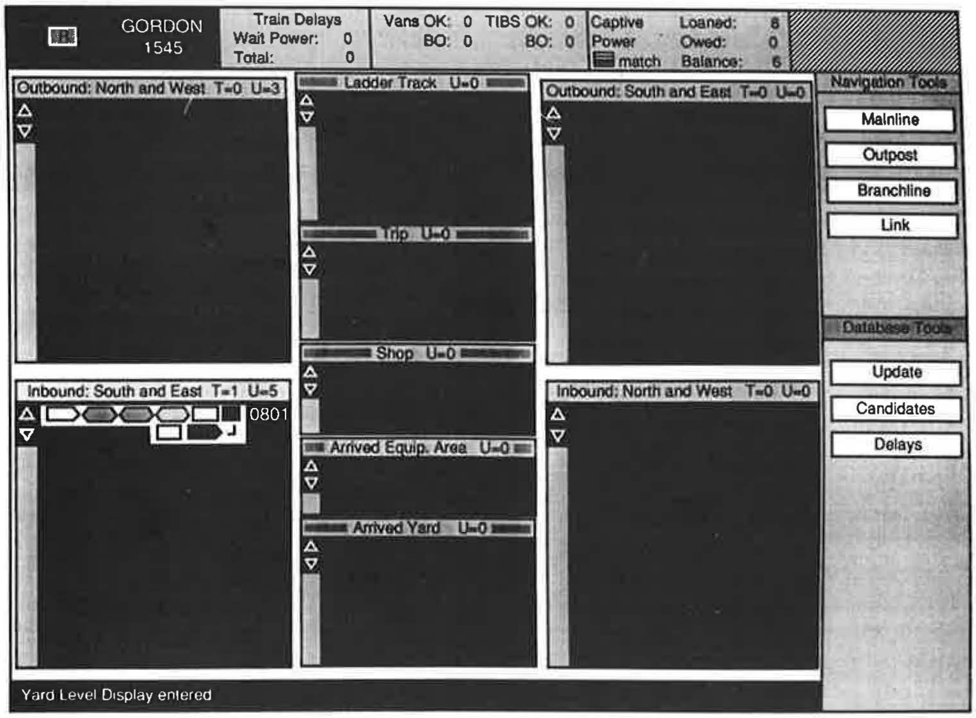


FIGURE 2 MPS yard-level display.

After the last unit icon is a rectangle listing the number of vans (cabooses) and/or TIBS (brake-sense units) on the train.

The symbology chosen is simple, yet communicates all the transient information necessary for the motive power controllers to make their decisions. To make the transient information more prominent, the displays are not overly burdened with static information (such as locomotive characteristics). A new controller can show the static information by making an object-sensitive query.

Alerts and Highlights

Alerts, in the form of blinking icons that look like flags, are used to draw the controller's attention to work to be done. For example, when a unit's status changes from noncritical to critical, the controller is alerted so action can be taken. Similarly, when a new lineup of outbound trains is issued by a yard, the controller is alerted so that a power supply plan will be prepared. A blinking border is used to highlight trains or units at the direction of the motive power controller.

Object-Sensitive Menus and User Dialogue

The MPS user interface is object-oriented. The user chooses the object of interest using a desktop "mouse" that controls the movement of an on-screen pointer. The user points at the object of interest and presses a button on the mouse to "pop up" a menu of functions pertinent to that object.

Many functions in MPS can be completed just by pointing at the object and selecting the function from the menu if the MPS needs no additional information from the user to complete the function. Other functions need more information than just the identity of an object and the function to be performed. For example, to move a unit or train, one needs to know two things: what to move and where to move it to. In MPS, a user can move a unit or train by pointing at it, then—while keeping a button on the mouse pressed—dragging the unit or train across the screen and releasing the mouse button at the location it is to move to. In MPS, objects are moved on the screen to indicate position in the same way that magnets were moved on a board in the system MPS is replacing.

MPS uses dialogue boxes for those functions that require a more in-depth dialogue with the user. The dialogue box is drawn in a "window" on top of the display from which it was invoked. For example, suppose the user pointed at train A216 and chose the "change train confirmation status" function. A dialogue box for this function would appear (Figure 3) drawn on top of the display with current information for train A216 displayed. Within the dialogue box, one or more items of information are displayed that the user can type over, as well as buttons to click on to perform certain functions.

PROTOTYPING EXPERIENCE

The design of MPS was derived not only from conceptual thinking but also from prototyping experiments. Screen layouts were initially sketched on paper, then refined using a computerized drawing program, and finally mocked up on the UNIX work station. The work station was attached to a projector to show the result projected onto the wall display. Often, a result that looked good on the work station screen did not look good on the projector and vice versa.

Prototyping was used to determine the size of the wall display, icon size, lettering fonts, colors, suitability of blinking, and so on. The concept that was subjected to the experiment usually worked in principle but was refined by prototyping. For example, the number of screens for the system-level view was increased from 10 to 12 to accommodate a larger, more readable lettering font. Alert flags were made to blink to make them more visually prominent. Additional timing, fuel, crew change, and power control points were added that were important to the controllers but had never been on the magnet board and were not identified by conceptual thinking.

A color editor provided by the manufacturer of the UNIX work stations proved to be of enormous benefit in selecting colors for each type of object. The prototype system was organized so that each type of object was assigned its own color index. Using the color editor, the designer pointed to an object, then used the mouse to slide on-screen bars corresponding to each of the three primary colors (red, blue, green) until the best color was found. The intensities of red, blue, and green for that object type were then noted for inclusion in the permanent data base.

Prototyping was also used to refine the interaction of the dialogue boxes. For example, MPS dialogue boxes provide the ability to "undo" the last action at two levels. At the broadest level, the user can cancel everything he has done since invoking the dialogue box by pressing the "abort" button in the dialogue box (such as in the center right of Figure 3). At a narrower level, the user can cancel entries for a particular object (such as a unit or train) by not pressing the "save" button (also see Figure 3) before moving on to the next object. Implicitly, then, while in a dialogue box there are two states: saved and not saved. The initial prototype did not visually indicate the state of the object being displayed, whether it was saved or not saved. Users who were interrupted by the telephone while using the dialogue boxes forgot what state they were in. It became obvious that a visual indication was necessary, and this was added (see the upper right corner of the dialogue box in Figure 3).

IMPLEMENTATION EXPERIENCE

MPS was installed and tested at CN during the winter of 1990–1991. One component of the testing was to compare the timeliness and accuracy of MPS information with that of the manual system, which continued to operate throughout the testing period.

Train and unit information was expected to become more timely due to the elimination of two existing time lags. First, there was a delay in the field between the actual time of an event and the time a clerk reported the event. This delay, which averaged several hours, would be eliminated at locations where REIS interrogators were installed. Second, there was a delay in the motive power center between the time information was received via COMTEL or telephone and the time the magnet board was updated to reflect this information. This delay, which also averaged several hours, would be eliminated entirely with MPS, regardless of how extensively REIS interrogators were installed in the field.

During the testing time period, only a handful of REIS interrogators were operational, so a significant impact from this improvement has yet to be measured. However, the improvement from eliminating the delay in posting information was immediate and dramatic. Motive power controllers who assisted in the testing were impressed with the improved timeliness of MPS, even though only one of the two sources of delay had been eliminated. The testing also revealed instances of trains and units that were missing from or incorrectly displayed on the magnet board, but were shown correctly in MPS.

A further test was conducted in which MPS was used manually, without benefit of its automatic information link to the field, to "shadow" the activities of a power controller. This test was conducted for two reasons: to make sure that all the functionality a power controller needed had been included in MPS, and to make sure that a power controller could keep up operation of the railroad in the event of a failure in MPS data links to the field. The test was successful in both cases. With the benefit of the field data links eliminating the work of updating the magnet board and the system of alerts to draw the controller's attention to work needing to be done, the controller's productivity is certain to increase over that of the manual system. The increased productivity can be used to reduce the number of simultaneous power controllers re-

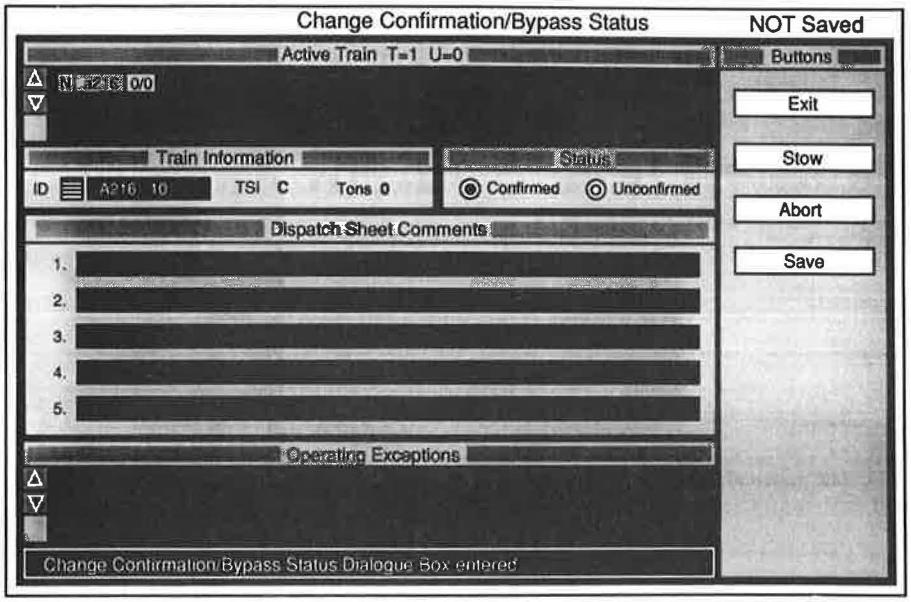


FIGURE 3 MPS dialogue box.

quired, to increase the amount of time and attention paid to motive power planning, or a combination of both.

CONCLUSIONS

The MPS testing period is just concluding as this is being written, so it is as yet impossible to measure quantitative improvements in locomotive utilization or one-time performance. However, based on what has been learned in design, development, and testing, the following conclusions can be drawn:

- Information on train and unit location and status is available several hours earlier in MPS than on the magnet board, even without benefit of automatic reporting technology.
- Automatic reporting technology results in information being made available a further several hours earlier, to the point where the motive power controllers know what happens in the field within a matter of seconds. This has been measured by CN at less than 1 min for those sites that were measured.
- The elimination of manual update tasks and use of MPS alerts to items needing attention will likely increase the productivity of the power controllers.

The new CN control center, including the MPS software, was dedicated on March 14, 1991, and is now operational.

ACKNOWLEDGMENTS

The authors especially acknowledge the contributions of Mike Beebe, Rick Greenidge, and Steve Rock of CN Rail, and Michael Bodden, Ingrid Brandle, Per Kreipke, Steven Prescott, Grace Ramus, David Seneko, and William Tamashunas of ALK Associates, toward the design and implementation of the motive power information and management support system.

REFERENCES

1. Electronic Railroad Rolling On. *Progressive Railroading,* June 1990, pp. 39–42.
2. L. Miller. ATCS Advances in Canada. *Railway Age,* March 1989, pp. 41–43.
3. E. R. Tufte. *The Visual Display of Quantitative Information.* Graphics Press, Cheshire, Conn., 1983.
4. K. C. Chih et al. Implementation of a Real-Time Locomotive Distribution System. In *Computer Applications in Railway Planning and Management,* Computational Mechanics Publications, Boston, Mass., 1990, pp. 39–51.
5. J. Sweeney. CSX on Mac Track. *Macintosh News,* Feb. 26, 1990, p. 2.
6. M. Walsh. UP Sets '90s Priorities. *Railway Age,* Feb. 1990, pp. 27–30.
7. G. Borzo. AEI's Newest Hurdle: Putting Data to Use. *Modern Railroads*, March 1990, pp. 29–32.

Advanced Train Control Systems Control Flow Development and Validation

ROBERT G. AYERS

The Railway Association of Canada and the Association of American Railroads initiated the Advanced Train Control Systems (ATCS) project in 1984. The railroads developed an operating requirements document and contracted a team of engineering firms (ARINC Research Corporation, Transportation and Distribution Associates, and Lapp-Hancock, Ltd.) to act as the system engineering team. The project then published a technology assessment, a number of draft specifications that defined a physical architecture, and a preliminary high-level assignment of functional requirements to physical components. It soon became clear that a method was required to document how the various components of ATCS should cooperate in carrying out railroad operations. In 1987, the first version of this documentation, known as Control Flows, was produced using the MacDraw software package and consisted of a set of high-level figures depicting the application logic. Since that time, railroad industry reviews have rapidly increased the detailed information contained in each flow. The project quickly outgrew the capabilities of the MacDraw tool as well as the flow chart format (Easy-Flow) and computer-aided software engineering tool (STATEMATE) that were subsequently adopted. The project abandoned STATEMATE in favor of ADA Syntax Program Design Language, which is readily convertible to software. Logic specifications written in ADA have been published and are designed to significantly reduce ambiguity, enhance maintainability, and provide a solid basis for future development by both the ATCS project and the supplier community.

The Advanced Train Control Systems (ATCS) project was initiated by the Railway Association of Canada (RAC) and the Association of American Railroads (AAR) in 1984. The purpose of the project was to develop a series of comprehensive and advanced operating systems for the control of train movement that are considered to be essential for improving safety, productivity, and energy efficiency of railroads. This paper describes the process by which the system logic was developed starting from operating concepts and proceeding to detailed software specifications.

OPERATING REQUIREMENTS

The project's first major endeavor was to develop a document entitled, "Advanced Train Control Systems Operating Requirements." This document enumerated the economic, operating, and safety objectives for ATCS and a set of operating requirements.

ARINC Research Corporation, MS 5-371, 2551 Riva Road, Annapolis, Md. 21401.

The operating requirements section of the document was structured as a hierarchical set of lists. At the lowest level of the hierarchy were short narrative descriptions of individual requirements. The top level of the hierarchy consisted of six items:

1. Presence detection, train identification, and location,
2. Track and route integrity,
3. Ancillary systems interface,
4. Switch control,
5. Train control, and
6. Management of train operations.

The operating requirements document did not detail techniques to be used to accomplish the requirements, the components of ATCS, or the interfaces among components.

The ATCS project, having established its requirements, engaged a team led by ARINC Research Corporation to act as the systems engineer on the project.

TECHNOLOGY ASSESSMENT

The first step in the process of developing a system architecture was to perform a technology assessment. This study looked at a number of technologies that might benefit ATCS, but concentrated on technologies supporting three major areas: vehicle locations systems, data communications systems, and display systems.

Having determined the likely technologies to be used in ATCS, the project set out to develop a system architecture and form, fit, function (F^3) specifications for the ATCS components. To facilitate this process, committees were established that included representatives from the railroads, system engineers, manufacturers, and system integrators. Participants from all groups provided input to the specifications; however, when a committee failed to reach a consensus on any issue, the railroad representatives voted to determine how the specifications would be developed.

The committee process led to rapid advances in developing a system architecture, hardware specifications, and data communications specifications for ATCS. It became clear by 1987 that a method was required to document how the various components of ATCS should cooperate in carrying out railroad operations. This led to the development of the first generation of Control Flows, which later became known as the

"Macintosh Flows" because they were developed on a Macintosh computer using the MacDraw package.

MACINTOSH FLOWS

The Macintosh Flows described each of a number of railroad operations in a two-page format (Figures 1 and 2). One page showed the flow of control (at a very high level) and the other showed data flows among components. These flows were reviewed by railroad operations and signal representatives and updated to reflect their inputs.

The Macintosh Flows were quite useful in validating the initial assignment of functionality to components and the high-level view of how ATCS should work with the railroad community. They had serious shortcomings as design documentation, however, because their high-level view abstracted out important decision-making processes within components, such as how the central dispatch computer determined if an authority request was safe to issue.

EASY-FLOW CONTROL FLOWS

The need to document additional details about the operation of ATCS led to the next generation of Control Flows. These were developed in a IBM-PC environment using a tool called Easy-Flow (see Figure 3). These control flows used a more

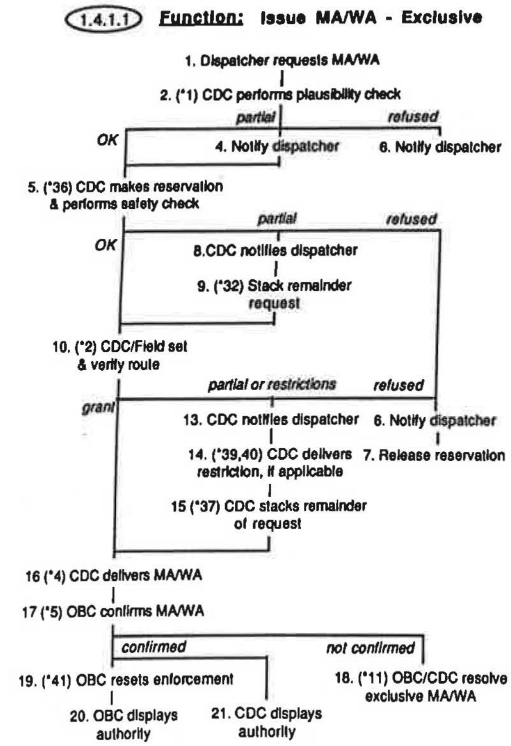


FIGURE 1 High-level flow of control in a Macintosh Flow.

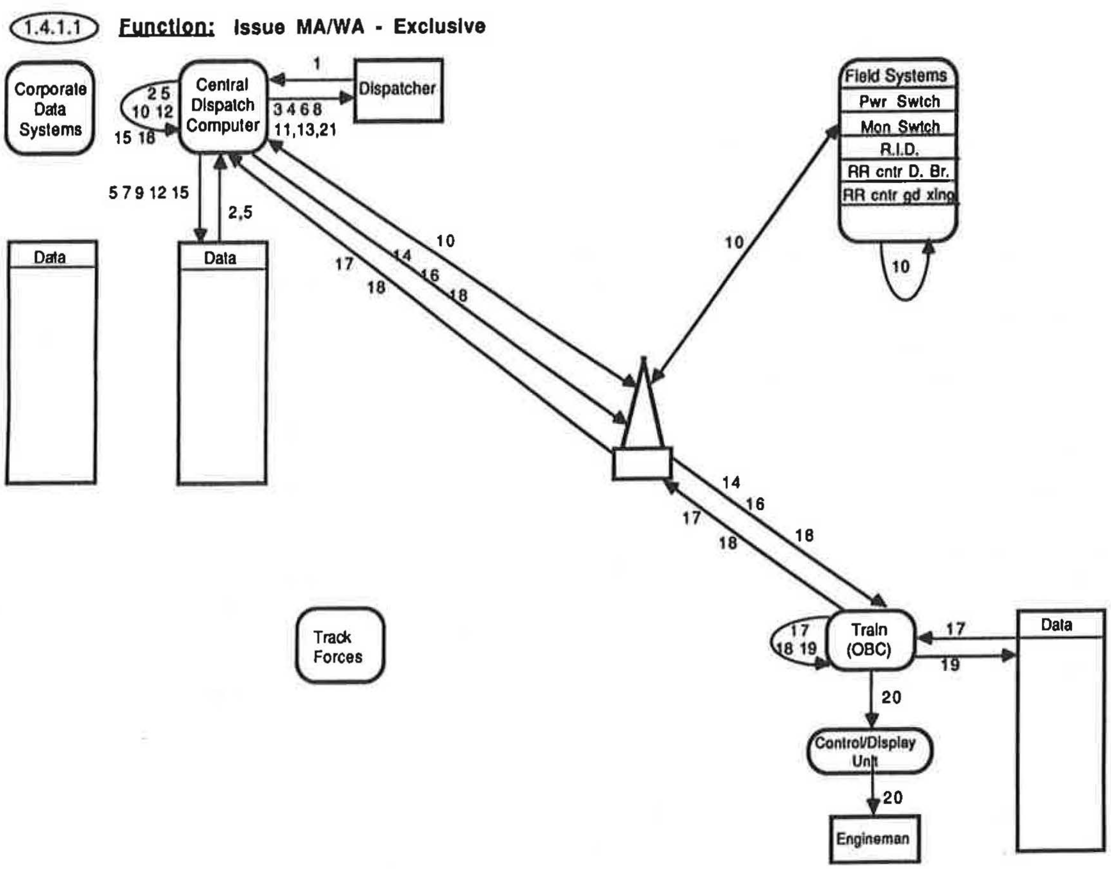


FIGURE 2 Data flow between components in a Macintosh Flow.

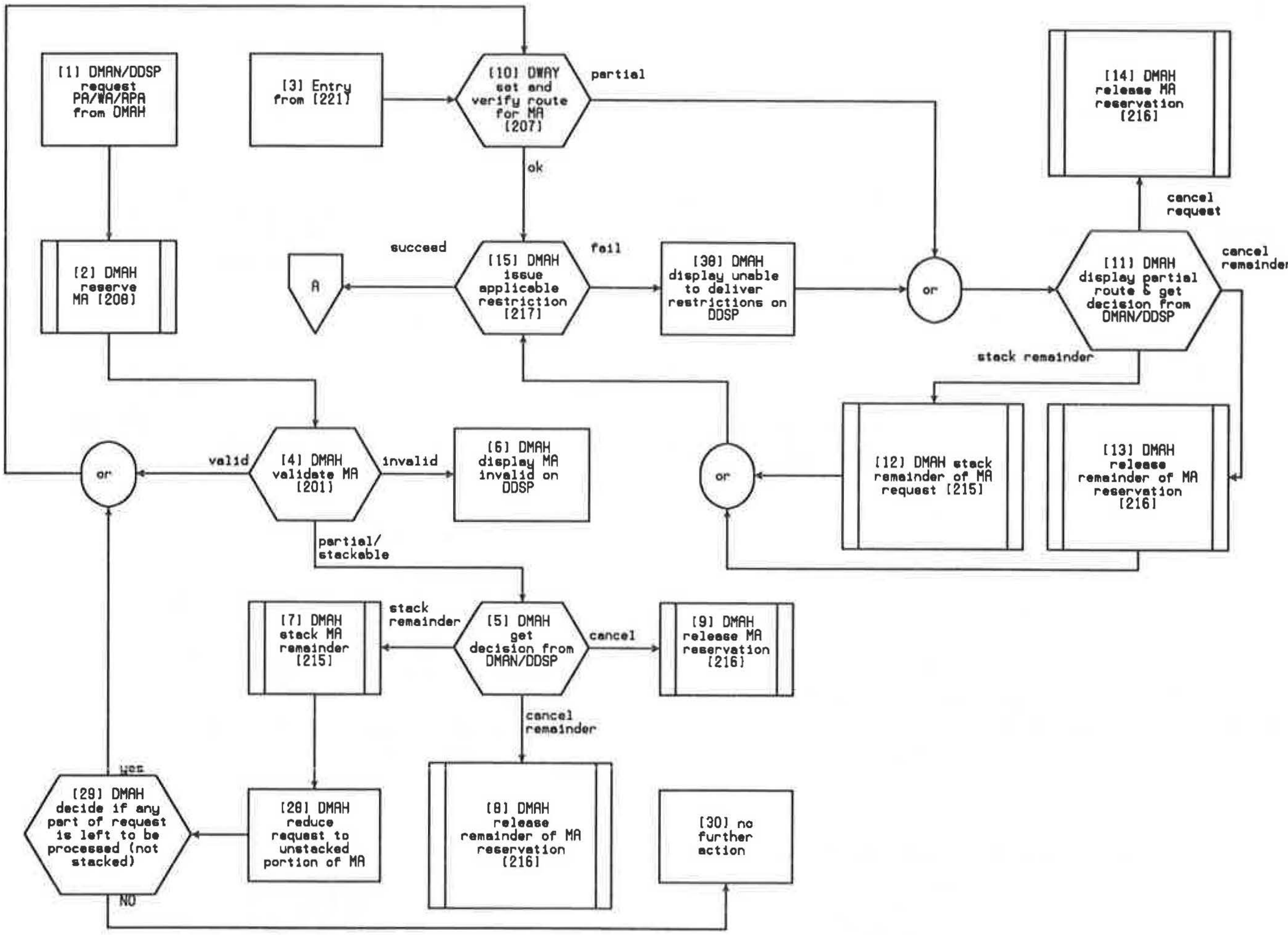


FIGURE 3 Easy Flow Control Flows. (*continued on next page*)

traditional flow chart approach to documentation and were published in Specification 100. The Easy-Flow control flows are usually referred to as simply "the control flows."

Along with additional details, the new flows included additional structure. The control flows had three levels of hierarchy:

1. Master Flows, which were a set of tables showing what events triggered what macro flows to execute;
2. Macro Flows, which were the direct descendant of the Macintosh flows, in that each macro flow represented how ATCS components cooperated to carry out a railroad function; and
3. Micro Flows, which provided additional detail on how a step in a macro flow was to be performed.

The conversion to Easy-Flow also involved adopting a set of conventions for describing ATCS activities. First, they identified the "addressable entities" or software modules in each component. Each step in these control flows was constructed in a subject-verb-object format. The subjects and objects were either addressable entities or railroad employees. The verbs were phrases describing the actions taken.

The external messages exchanged between ATCS components were identified at the places where these messages were sent in the flows.

The Easy-Flow control flows were also stored in dBase III, which allowed various searches and sorts to be conducted and reports to be generated.

The Easy-Flow control flows were used by the project to develop a manual simulation of ATCS. This simulation allowed railroad, industry, and government representatives to get a much more detailed understanding of how ATCS work. The simulation was conducted by having people act out the parts of the various ATCS components in an operating scenario. Message flows were simulated by exchanging paper forms.

The Easy-Flow control flows were also used as the basis for developing some pilot projects by North American railroads and have formed the basis for the AUSTRAC version of the ATCS being developed on the Australian National Railroad.

Although the Easy-Flow control flows were helpful in defining ATCS functions to a greater degree of detail, they still had some shortcomings. First, each flow contained a verb that had only an intuitive meaning. Second, the assignment of

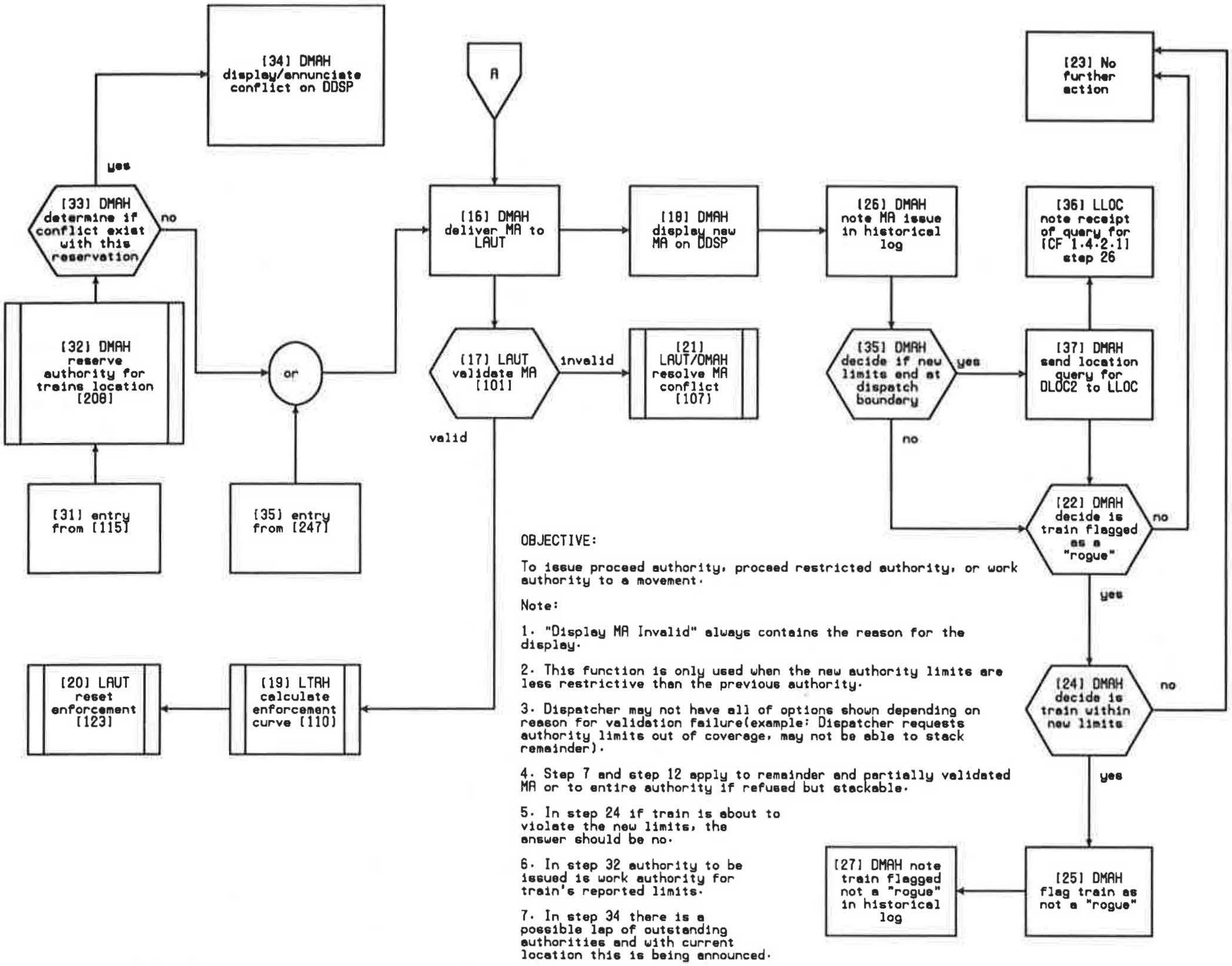


FIGURE 3 ***(continued)***

subjects to steps was only intuitive. Third, receipt of a single message by a component might occur in more than one flow, leading to ambiguity as to which logic to initiate. Fourth, there were inconsistencies in the definition of how entities within components cooperated. Finally, the intuitive and nonrigorous nature of the steps meant that the flows were not and could not easily be made machinable. Machinability is a key characteristic necessary for validation of the logic.

CONTROL FLOW VALIDATOR

These problems with the Easy-Flow control flows and the need to validate the logic led to the establishment of the control flow validation (CFV) project (within the ATCS project). The CFV project was an attempt to rehost the control flow logic into an off-the-shelf computer-aided software engineering (CASE) tool. The tool that was chosen was STATEMATE by Logix.

The CASE tool is also being used to prototype and test the communications protocols used by ATCS applications to access the communications network. This effort should lead to a completed prototype of the stack by the end of 1991, and a testing program is planned for 1992.

The first step in the rehosting process was to define the software and hardware architecture of ATCS to the CASE tool. The hardware architecture ("module definition") went smoothly; however, the attempt to define the software processes ("activities") soon led to the discovery of additional inconsistencies in the entity and function definitions in the Easy-Flow control flows.

The CFV team then went through a process of restructuring the function and entity definitions to produce an internally consistent software architecture for ATCS. The architecture that resulted (see Figure 4) was generic in the sense that it could be adapted to describe not only ATCS, but any distributed command and control system. This architecture can be conceptualized as a grid where a row of items constitutes all of the software in a single hardware *component*. A column of items constitutes the software items in various hardware components that cooperate to form a *function*. Each item is

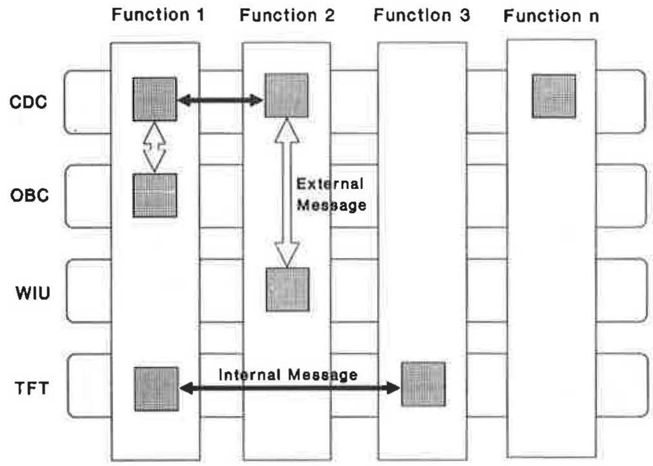


FIGURE 4 Components, functions, and basic units define the software architecture for ATCS.

the software module that performs a function within a component, and is termed a *basic unit*.

In this architecture, an addressable entity became a group of related (adjacent) items in a row of the grid that shared a data store. External messages (between hardware components) are constrained to flow within a column. Communication between entities in a hardware component can be accomplished in two ways:

1. By the exchange of internal messages (which are constrained to flow in rows only), and
2. By the use of utilities, which may use information from more than one entity's data store to determine a result (e.g., is Device A in front of Train B).

Although the effort to restructure the software architecture and to prototype the protocol stack using the CASE tool was quite successful, the attempt to rehost the actual control flow logic was much less so. The CASE tool provided minimal support for expressing multiple instances of either hardware entities or data items. Expressing the situation of two trains sending location reports was a significant problem for the tool. The CASE tool also required logic to be defined in a specific graphical form. The amount of detailed logic that was required to define ATCS adequately in this form required that either a very large number of figures be developed or that all of the accompanying text be moved into separate documents. Finally, even when the logic was defined in the required graphical form, it was virtually indecipherable as a software specification.

These problems with the CASE tool led the CFV team to look for an alternative method for specifying the system logic. The method had to make producing the specifications straightforward. There had to be a readily available method to publish the specifications. And the specifications ultimately had to be machinable. In the end, the CFV team decided to use ADA Syntax Program Design Language (PDL). This decision was reviewed and approved by the ATCS system engineering task force. The specifications being developed in this form are variously called the system logic specifications, or the control flow specifications.

CONTROL FLOW SPECIFICATIONS

The new control flow specifications are based on the ATCS software architecture developed with the CASE tool. A separate specification was developed for each component. Each specification contains a standards and conventions section, a description of the components functions, and a lower-level specification for each entity in the component.

The entity specifications describe the entity's purpose, list the basic units in the entity, and contain a lower-level specification for each basic unit in the entity. The basic unit specifications describe the basic unit's purpose and contain a list of all of the transactions in the basic unit, followed by a lower-level specification for each transaction in the basic unit.

The transaction specifications contain the real substance of the control flow specifications. Each transaction defines the logic to be executed upon occurrence of a unique event or *trigger*. The types of triggers in a basic unit are as follows:

1. Initialization of the basic unit,
2. Termination of the basic unit,
3. Receipt of an internal or external message,
4. Expiration of a timer (previously started by the basic unit), and
5. Other special-purpose triggers used by the stack and session manager basic units to interact with the communications protocols.

Each transaction specification (see Figure 5) contains a header defining the transaction's purpose, a definition of each condition affecting the flow of logic control in the transaction, and the logic of the transaction. The logic sequence defined for a transaction is assumed to be noninterruptible.

The transaction logic defines what steps (primitive events) are performed under what conditions and in what order. Low-level calculations and interactions with data structures are not shown in the transaction but are carried out by the logic in the primitive events. The primitive event logic is not developed and distributed with the specifications but must be developed by the component manufacturer. Primitive event names begin with one of a limited set of verbs (e.g., SEND MSG, GET MSG, or START [TIMERS]).

Each verb has a description in the specification of what type of activities it may perform. Each transaction specifies what conditions (used by the transaction) that a primitive event must set. The part of the primitive event name following the verb describes what is to be done (e.g., the SEND MSG verb is followed by a message number to form a primitive event [SEND MSG 6 2 1]).

Publication

The system logic was grouped into three major areas (control, monitor, and flexibility). The control flow specifications are being developed by area. The monitor functions were completed and published in March 1991. The control functions

```
CD_9_M_30_78          3-Apr-1991 07:43:56     VAX Ada V2.2-38                    Page   1
01                    28-Mar-1991 08:47:38    [USERS.TMANNING.TRANS]CD_9_M_30_78.ADA;1 (1)

with CD_9_M_30_78_PRIMITIVES;
-------------------------------------------------------------------------
-- Transaction: CD_9_M_30_78
-- Basic Unit : CD_ISSUE_AUTHORITY
-- Purpose    : This transaction processes message 30.78,
--              REPLY_VERIFY_CURRENT_TRAIN_AUTHORITY_AND_LOC_MSG, which is used
--              by the dispatcher to indicate that the current status of all
--              trains is correct as displayed.
--
-- Created    : 28 March 1991
-- Modified   :
-- Modified   :
-------------------------------------------------------------------------
use CD_9_M_30_78_PRIMITIVES;
procedure CD_9_M_30_78 is
--
-- abbreviations:
--
-- conditions:
-- VERSION_2 indicates the revision level of the message is level 2.
--   DEFAULT: TRUE
--   SET BY: GET_MSG_30_78
--
begin                                    -- CD_9_M_30_78 algorithm
  GET_MSG_30_78;
  if (VERSION_2) then                    -- process message version 2
    SEND_MSG_56_3;                       -- note results in historical log
    SEND_MSG_50_114;                     -- request dispatcher to verify gang
                                         -- authorities and locations
  else                                   -- invalid message version
    SEND_MSG_42_5;                       -- version error report to MON_HEALTH
  end if;
end CD_9_M_30_78;
pragma page;
```

FIGURE 5 Transaction specification.

are scheduled to be completed in December 1991. The flexibility functions are scheduled to be completed in 1992.

Due to the size of the control flow and message specifications, they will only be published and distributed on magnetic media and only upon receipt of payment from an interested party.

Concurrent with the development of the new control flow specifications, version control procedures were put into place. Each future publication of the message, data, and logic specifications will have a version number. The project will make efforts to ensure that versions are backward-compatible to the extent feasible. This will not include compatibility with draft specifications. A configuration management plan for ATCS is under development and will be released later in 1991.

A tool set to be used with the software and message specifications is also under development. This tool set will allow the project to maintain the specifications in a controlled manner and also provide facilities for vendors and railroads to extend the specifications in a consistent manner when adding proprietary features. This tool set will operate in a VAX/VMS environment using VAX RDB and VAX ADA. The tool set is scheduled to be available by the end of 1991.

Future Uses

The fact that the software specifications and message specifications are machinable means they can be used for a variety of possible future projects:

1. Automated consistency checking of the specifications—some of this will be included in the tool set.
2. Development of an ATCS system model—this would use the control flow specifications as the engine of a model that could be developed in a modular fashion. A completed model would allow ATCS-based railroad operations and scenarios to be executed and tested on a computer.
3. Development of prototype components—the addition of primitive event, transaction scheduling, and protocol stack logic to the control flows would allow a prototype of a component to be developed directly from the specifications.

4. Development of production components—the addition of vitality checking logic to a prototype component, along with proprietary features, and optimization of selected transactions for improved execution speed would provide a vendor with a production component. This technique would likely reduce both first cost (to develop) and collateral costs (to maintain and upgrade) the production software, when compared with conventional software development techniques.

CONCLUSION

The development of the control flows over a period of years has progressed from a relatively high-level conceptual overview of ATCS to a detailed software specification. This specification is designed to reduce significantly ambiguity, enhance maintainability, and provide a solid basis for future development by both the ATCS project and the supplier community.

Verification and Validation for Advanced Train Control Systems

GIDEON BEN-YAACOV

Verification and validation (V&V) techniques are used in military, space, nuclear power, and other industries to improve the quality of digital systems. V&V will ensure that quality advanced train control systems (ATCS) are implemented and that the ATCS will, indeed, provide the anticipated level of vitality. Railroads or system suppliers can use V&V programs when developing V&V procedures, plans, tasks, and activities for use during the implementation of ATCS.

Verification and validation (V&V) programs are vital for successful advanced train control systems (ATCS) design and performance. V&V processes can be applied to the overall ATCS, including the software, hardware, and user terminals.

The V&V program is defined in an implementation plan document that provides the recommended procedures and tasks to be followed when the V&V process is implemented. This document is also known as the system verification and validation plan document. The recommended V&V procedures apply to both vital and nonvital items of ATCS hardware and software. They are performed in parallel to ATCS development.

V&V should be performed by a group independent of the ATCS developer and supplier. The group should be designated at the beginning of the ATCS project and should be responsible for the following:

- Preparing a detailed V&V plan,
- Performing V&V activities as defined in the plan,
- Participating in development of a test plan, test procedures, and acceptance criteria,
- Witnessing system testings, and
- Evaluating test results.

If the ATCS are developed by outside suppliers and/or system integrators, the railroad companies should assign engineering staff as active participants on the V&V team.

The V&V activities can provide a comprehensive evaluation of each phase of the ATCS development project and help ensure the following:

- Deficiencies are detected and corrected as soon as possible in the ATCS life cycle.
- Probability of project risk, cost increase, and schedule delay is reduced or eliminated.
- ATCS quality and reliability are enhanced.
- Proper vitality features are provided with the ATCS.
- Management's knowledge of the ATCS developmental process is improved.
- Proposed changes/improvements and resulting consequences are assessed and monitored.

DEFINITIONS

The following are definitions of terms used in this paper.

- Required inputs—Items necessary to perform V&V tasks for each life-cycle phase of the system.
- Required outputs—Items resulting from performing V&V tasks for each life-cycle phase of the system.
- System integrated testing—Process of testing an integrated hardware and software system to verify that the system meets its specified requirements.
- Test plan—Documentation specifying the scope, approach, resources, and schedule of intended testing activities.
- Test procedures—Documentation detailing test procedures, inputs, predicted results, and execution conditions for each test item.
- Validation—Process of evaluating a system at the end of the development process to ensure compliance with functional and performance requirements. It is based on evaluating an integrated testing of hardware and software. The validation process provides assurance that the capabilities of the system, as described in the requirement specification document, are implemented in the hardware and software.
- Verification—Process of determining whether the product of a given phase of a system development cycle fulfills the requirements established for that phase. Verification is based on system development documentation. Verification ensures that an accurate translation of information from one phase of development to the next phase has been performed (for example, the review of design specifications ensures that the design meets functional requirements).

V&V IMPLEMENTATION PLAN

V&V tasks for a specific ATCS project are defined in an ATCS system verification and validation plan (SVVP). The tasks are intended to verify that the product of each ATCS development phase complies with the phase requirements for correctness, completeness, consistency, and accuracy and to validate that the completed ATCS complies with established requirements.

Automated Monitoring and Control International, Inc., 11819 Miami Street, Omaha, Nebr. 68164.

Figure 1 provides a summarized review of ATCS V&V activities as defined in a SVVP. The V&V input, tasks, and output for each ATCS life-cycle phase are identified. The ATCS life-cycle phases used in the SVVP are as follows:

- Concept,
- Requirement,
- Design,
- Implementation,
- Testing,
- Installation and checking, and
- Operation and maintenance.

Table 1 contains a description of the V&V tasks applicable to each project implementation phase as well as the inputs and outputs required to accomplish these tasks. The scope and contents of each section of the SVVP are described next.

1. *Purpose:* The purpose and scope of the proposed V&V plan and the project for which the plan is prepared are described.
2. *Reference documents:* Documents that support plan implementation are identified.
3. *Plan overview:* The organization, schedule, resources, responsibilities, and plan management methodologies necessary to perform ATCS V&V tasks are described.
4. *V&V tasks:* A detailed plan of V&V tasks throughout the ATCS life cycle is provided and includes the following:

(a) Specific V&V task(s) recommended for each phase,
(b) Methods and criteria used in performing each task,
(c) Inputs and outputs required for each task,
(d) Schedule for the tasks,
(e) Resources for performing the tasks,
(f) Roles and responsibilities of individuals responsible for performing the tasks, and
(g) Management of the tasks.

5. *V&V reporting:* The content and format of all V&V reports are described. The following types of reports are generated to support V&V activities:

(a) Deficiency reports are generated for each deficiency or abnormality detected by V&V. Each report contains a description of deficiency, the impact on ATCS operation and performance, the cause, the criticality, and recommendations.
(b) V&V phase summary reports contain summaries of the results of the specific V&V tasks performed for each life-cycle phase. Each report contains a description of the V&V task, a summary of task results, a summary of deficiencies and resolutions, an assessment of ATCS quality, and recommendations.
(c) V&V final report is issued at the end of the installation and checkout phase. It includes a summary of all V&V tasks, a summary of deficiencies and resolutions, an assessment of ATCS quality, and recommendations.

6. *V&V administrative procedures:* The following V&V administrative procedures are described: deficiency reporting and resolution, configuration management procedures, and standards, practices, and policies.

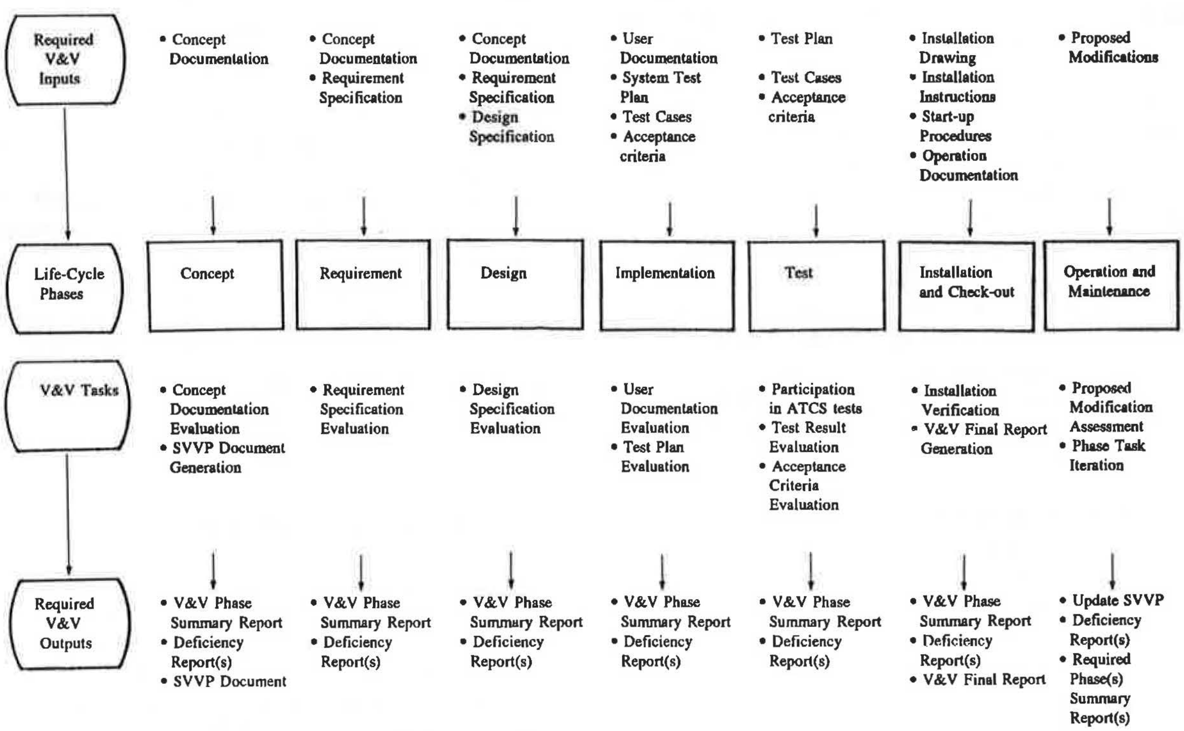


FIGURE 1 V&V plan overview.

TABLE 1 V&V TASKS, INPUTS, AND OUTPUTS FOR ATCS LIFE-CYCLE PHASES

V&V TASKS	REQUIRED INPUTS	REQUIRED OUTPUTS
	V&V PLAN	
ATCS Verification and Validation Plan (SVVP) Generation During Concept Phase, the SVVP is generated for all ATCS' life-cycle phases. The SVVP should be considered to be a "Living" document, and changes are made as needed.	• Project implementation Schedule • Concept documentation	• SVVP Document
Concept Documentation Evaluation Concept documentation is evaluated to determine if proposed concepts satisfy user needs and project objectives. Major constraints of interfacing subsystems and limitation of proposed approach are identified.	• Concept documentation (for example, Statement of Need, Advance Planning Report, Project Initiation Memo, Feasibility Study, ATCS Definition Document)	• Deficiency report(s) • V&V Phase Summary Report
	REQUIREMENT PHASE V&V	
Requirement Evaluation ATCS functional requirement specification is evaluated for correctness consistency, completeness, accuracy, readability, and testability. The ATCS requirement specification is to be assessed as to how well it satisfies railroad's objectives and user needs as defined in the Concept documentation.	• Concept Documentation • ATCS functional requirement documentation (for example, ATCS requirement specification, request for proposals, proposals, contract documentation, etc.)	• Deficiency report(s) • V&V Phase Summary Report
	DESIGN PHASE V&V	
Design Evaluation ATCS design specifications are evaluated for correctness, consistency, completeness, accuracy, and testability. The ATCS design specifications are assessed as to how well they satisfy ATCS objectives as outlined in the ATCS requirement specification.	• ATCS Requirement Specification document • ATCS Design Specification document	• Deficiency report(s) • V&V Phase Summary Report
	IMPLEMENTATION PHASE V&V	
ATCS Test Plan Evaluation ATCS test plan is evaluated to determine whether it provides a suitable plan for ATCS testings.	• ATCS Test Plan Document	• Deficiency report(s)
ATCS Test Procedure Evaluation ATCS test procedures are evaluated to determine whether the procedures are adequate for subassembly testing, integration testing, and acceptance testing.	• ATCS Test Plan Document • ATCS Test Procedure Documents	• Deficiency report(s) • V&V Phase Summary Report
	TEST PHASE V&V	
Testing ATCS testing is witnessed by a member(s) of the V&V team who would assist in recording all testing results as required by the Test Plan.	• Test Plan Document • Test Procedure Document	• Deficiency report(s)
Test Analysis ATCS test results are analyzed to determine whether the ATCS satisfies acceptance criteria.	• Test Plan Document • Test Procedure Document • Test Results	• Deficiency report(s) • V&V Phase Summary Report
	INSTALLATION AND CHECKOUT PHASE V&V	
Installation Audit ATCS Installation is audited to determine that all equipment, software, and cables are correctly installed and operational. instructions.	• Installation documentation (for example, installation drawings, installation procedures, installation tests, diagnostics, start-up procedures, operational procedures)	• Deficiency report(s) • V&V Phase Summary Report
V&V Final Report Generation All V&V activities and results are summarized in this report.	• All V&V Phase Summary Reports	• V&V Final Report
	OPERATION AND MAINTENANCE PHASE V&V	
V&V Plan Revision The ATCS SVVP document will be revised, as necessary, to be	• SVVP document	• Updated SVVP
Proposed Modification Assessment All approved modifications, enhancements, or additions will be assessed to determine the effect each modification would have on V&V tasks. The extent to which V&V tasks would be iterated will be determined. These V&V tasks are aimed to insure that each planned modification is implemented correctly.	• Approved modification(s)	• Scope of V&V tasks required to support the modification(s)
Phase Task Iteration For each approved modification, the specific V&V tasks required to support the modification will be performed.	• Approved modification(s)	• Deficiency report(s) • Required phased outputs of iterated tasks

V&V MANDATORY ACTIVITIES

V&V activities are an integral part of the quality assurance process used to finalize the functional specification requirements, design, building, testing, installation, and acceptance of the ATCS. V&V activities are, therefore, integrated into the ATCS project activities.

Some of the V&V activities are extremely beneficial and, therefore, are considered mandatory, whereas others are viewed as desirable and, therefore, considered optional. Mandatory tasks are performed during the requirement (verification of ATCS requirements), design (verification of ATCS design), testing (ATCS validation), installation and checkout (verification of ATCS installation), and operation and maintenance (V&V of ATCS modifications and upgrades) phases.

These mandatory activities are as follows:

1. *Verification of the ATCS requirement:* During this verification process the ATCS requirement specification document is reviewed. The main objective is to determine whether the ATCS requirements are complete, understandable, consistent, feasible, and testable. This review should be performed before the detailed design and development of hardware and software.
2. *Verification of the ATCS design:* This verification process consists of a review of the ATCS design specification document. The review of the hardware and software designs focuses on determining whether the designs are a correct implementation of ATCS requirements.
3. *ATCS validation:* The ATCS validation consists of testing to validate that the complete ATCS is a correct implementation of the ATCS requirement specification.
4. *Field verification:* The field verification consists of testing field installations to ensure that the ATCS has been properly installed.
5. *Operation and maintenance V&V:* This activity supports all future upgrading and modifications of the ATCS.

These five mandatory V&V activities for an ATCS development project are further described in the following paragraphs.

Verification of ATCS Requirements

The system requirements are the foundation on which the ATCS is designed, built, and accepted. System requirements are verified through a review of the ATCS requirement specification document for correctness, completeness, consistency, understandability, feasibility, testability, and traceability. The verification of the ATCS requirement specification is perhaps the most important V&V activity. Its principal goal is to determine independently that the ATCS requirements result in a practical solution to the railroad's objectives. The ATCS requirement verification addresses the following:

- Completeness and correctness in specifying the functional capability and performance requirements of the ATCS;
- Completeness and correctness in specifying the internal interfaces within the ATCS and the external interfaces to the railroad's corporate computers;
- Unambiguous, correct, and consistent description of the functional characteristics and performance for each major subsystem and/or software package;
- Reasonable, achievable, and suitable ATCS test requirements;
- Definitions of subsystems' physical characteristics, reliability and maintainability objectives, operating environments, and design and development standards;
- User interface requirements; and
- Definition of installation, operation, training, documentation, and maintenance requirements.

The results of the ATCS requirement specification document review are recorded in a V&V report. Typical contents of this requirement verification report are as follows:

- Background:
  - Identification of the ATCS requirement specification document,
  - Identification of other referenced documents, and
  - Review participants;
- Results of review:
  - Functional requirements,
  - Definition and interfaces,
  - Design requirements,
  - Operational environments,
  - Test requirements and acceptance criteria,
  - Supporting services requirements (i.e., training, documentation, etc.),
  - Performance requirements,
  - Installation requirements,
  - Operational capabilities,
  - Maintenance, and
  - Other requirements;
- Attributes:
  - Correctness,
  - Completeness,
  - Consistence,
  - Understandability,
  - Feasibility,
  - Testability, and
  - Traceability; and
- Summary of deficiencies to be resolved.

Verification of ATCS Design

Verification of ATCS design is performed through the review of the ATCS design specification. The ATCS design specifications are reviewed to ensure that the ATCS functional requirements, implemented by the hardware and software designs, are complied with and that there are no ambiguities or deficiencies. The design review examines the design of the hardware and software in terms of their ability to satisfy performance and functional requirements:

- Architecture (both hardware and software);
- Major equipment interfaces;
- Operating procedures (initialization, start-up, error detection, restart, etc.);
- Testability (use of test equipment such as simulations);
- Timing analysis (display response time, communication response time, etc.);

- Availability (reliability prediction report indicators);
- Information flow (data communication between major hardware components); and
- Human factors (analysis performed to evaluate layout and contents of user's interfaces).

The ATCS design specification documents are usually more detailed than the system requirement specification document. They detail how the requirements are met. One of the key objectives of design verification is to ensure that the ATCS design is consistent with the ATCS requirements. The design review, therefore, will provide an independent assessment of the ability of the design to meet functional and performance requirements. Such capabilities as response time, availability, man-machine interface, data quality, operating environments, and testability must be analyzed as part of the design evaluation. The V&V team will review the information contained in the design specification documentation for correctness, feasibility, and consistency. Some of the performance requirements may be difficult to ascertain, but it is more efficient to identify these issues during the design phase than later during the validation testing.

Another result of the design review is the evaluation of test requirements and criteria. This information is useful for the evaluation of the ATCS test procedures aimed at validating ATCS performance and capabilities. Furthermore, the evaluation of test requirements and criteria ensures that the ATCS is designed to be testable and that a test plan can be efficiently developed.

The results of the design verification activity are documented. Any deficiencies are identified. Information provided in the design verification report is as follows:

- Background:
  - –Identification of the ATCS design specification documents,
  - –Identification of other referenced documentation, and
  - –Review participants;
- Results of review:
  - –ATCS hardware and software architecture,
  - –Hardware subsystem interfaces (including isolation, signal type and rates, protocols, etc.),
  - –Hardware subsystem design,
  - –Software subsystem design (including operating systems, applications, utility programs, priorities, error control/recoveries, algorithms, etc.),
  - –ATCS hardware and software integrated testing, and
  - –Human factor engineering;
- Attributes:
  - –Design completeness,
  - –Design consistence,
  - –Hardware/software testability,
  - –Design traceability, and
  - –User acceptability; and
- Summary of deficiencies to be resolved.

ATCS Validation

ATCS validation is accomplished through extensive ATCS testing. This testing demonstrates that the integrated ATCS meets the functional and performance requirements described in the ATCS functional specification document. Four activities are associated with ATCS validation testing: (a) Test plan review, (b) ATCS test procedures review, (c) participation in ATCS test execution, and (d) analysis of test results. The test results, analysis, and nonconformances to acceptability criteria are documented in a validation test report. Typical contents of such a report are as follows:

- Background:
  - –Purpose of tests,
  - –Summary of test plan, and
  - –Reference documents;
- Analysis of test results:
  - –Test environment and configuration,
  - –Test results collected,
  - –Acceptance criteria,
  - –Test results analysis,
  - –Actions to be taken, and
  - –Conclusions;
- Summary:
  - –Capabilities demonstrated and
  - –ATCS deficiencies; and
- Recommendations:
  - –ATCS refinements and
  - –Further testing.

Field Verification

Field verification ensures that the ATCS has been properly installed. The installation procedures, drawings, and start-up procedures are reviewed by the V&V team. Field verification entails stationary tests of all system cables, mechanical constructions, and software installations to ensure that all ATCS components are properly installed.

Field verification activities are defined in a field verification plan, which is included in the SVVP. The plan identifies the methods used to verify proper installation and the environments of the various subsystems. Analysis of field verification results is documented in the field verification analysis report. Typical contents of such a report are as follows:

- Background:
  - –Purpose of field verification,
  - –Summary of field verification activities, and
  - –Reference documents;
- Analysis of results:
  - –Completeness of installation drawings and start-up procedures,
  - –Verification of proper installation of system cables,
  - –Verification of proper installation of hardware devices,
  - –Verification of proper mechanical construction, and
  - –Verification of proper installation of system software (including utilities, diagnostics, operating systems, firmware, etc.);
- Summary:
  - –Installation deficiencies and
  - –Actions to be taken; and
- Recommendations:
  - –Correction of deficiencies and
  - –Further verification of installation work that will be redone.

Operation and Maintenance Phase V&V

As experience is gained in the operation and maintenance of the ATCS, improvements and modifications will be made. Such modifications may entail a new requirement or capability, or a design modification that improves performance, usability, or reliability. Once the scope of a modification is defined and approved, the V&V implementation phases for the modification are similar to those used during ATCS development.

The V&V for ATCS modifications will depend on the type and extent of each modification. For example, for minor modifications such as a format change on the on-board display, it may consist merely of noting the result of the change. Modifications that significantly affect the performance or the functional capability of the ATCS will require more detailed V&V.

CONCLUSION

Providing state-of-the-art technology for train control and monitoring involves greater risk than using the traditional field-proven signaling system. A major reason for this is that ATCS subsystems are based on newly developed products and software and are built according to an individual railroad's specific requirements, vendors' specific design specifications, and the North American railroads' generic ATCS specifications. The V&V process described in this paper reduces the risk associated with newly developed ATCS.

V&V is a process of review, analysis, and testing employed throughout the ATCS development life cycle. It provides a way to ensure successful ATCS implementation, thereby providing an increased level of confidence in ATCS design, development, and implementation. Through a series of checkpoints and reviews, V&V, as a methodology, helps ensure that quality ATCS systems are implemented.

Verification, through documentation review, is performed at each phase of the ATCS development life cycle. It determines that each phase product is correct, complete, and consistent with itself and with predecessor products. Validation, through testing and analysis, determines the correctness of the end product (i.e., the entire ATCS) with respect to ATCS requirement and design specification.

The recommended V&V activities provide a comprehensive set of tasks aimed at enhancing the quality of the entire range of ATCS development phases. Proper implementation of the recommended V&V process will result in the following:

- Safe train operation under ATCS,
- Accurate cost estimating and schedule planning,
- Understanding of the ATCS requirements,
- Adequate testing of the ATCS,
- Proper documentation of the ATCS,
- Satisfactory ATCS performance,
- Use of sound human factors engineering practices,
- User participation in ATCS development,
- Management control during ATCS development, and
- Achievement of anticipated benefits.

The level of V&V effort appropriate for an ATCS implementation project described in this paper is considered adequate to provide confidence that the ATCS can perform its functions in a satisfactory manner. However, the effort level is not extensive as to create unnecessary delays in the development and implementation of the ATCS. The main goal of the recommended level of V&V is to provide results in areas where most benefits can be obtained. It is believed that the recommended level of effort is sufficient to meet the intent of the recommended V&V process stated in ANSI/IEEE Standard 1012 for developing V&V plans for both critical and noncritical software.

Prototype and Test Environment for ATCS Data Communications

Dorothy A. Colburn

Advanced Train Control Systems (ATCS) are distributed command, control, and communications systems that also provide an infrastructure usable for train control and numerous ancillary applications. A crucial stage in ATCS development is the verification of the logic that allows applications to access this infrastructure. The communications protocols used to connect applications in different physical components are described in ATCS Specification 200, which represents the protocols as finite state machines using CCITT Recommendation Z.101 notation. These state machine representations are converted into a statechart representation in a computer-aided software engineering (CASE) tool. This tool simulates the protocol operations and generates an ADA prototype of the protocols. The communications between applications within a physical ATCS component are not standardized. For prototyping purposes, however, these communications have been modeled as a software bus. The software bus allows applications to communicate with the prototype of the protocol stack, providing an integrated environment in which the interactions of applications, the protocols, and the bus may all be examined and monitored. The prototype code generated by the CASE tool is supplemented by ADA code developed to simulate events that would occur within the interactions of the communications protocol. The interdependency of the software bus, the CASE-generated ADA code, and the programmer-developed ADA code provides an environment for meaningful prototyping and testing of the access logic, as well as the communications protocols. Instances of this prototype may be connected through a network emulator for use in further testing and prototyping the ATCS communications architecture.

Advanced Train Control Systems (ATCS) are a joint U.S.-Canadian endeavor focused on increasing railroad efficiency and service performance. ATCS computers will automatically monitor and control train traffic, allowing railroads to improve service while reducing costs through more precise management of resources (i.e., people, capital equipment, and fuel). During the last decade, the ATCS program has evolved from the determination of requirements through the identification of required components and the writing of performance specifications, to the development of procedures for testing components and complete systems.

The ATCS data communications system structure is based on the International Standards Organization's Open Systems Interconnect (OSI) Reference Model (*1*). This seven-layer protocol model was selected to define the points of interconnection between systems. Figure 1 shows the relationship between the OSI model and the ATCS communications architecture for communications between a dispatch center and a locomotive computer.

ARINC Research Corporation, 2551 Riva Road, Annapolis, Md. 21401.

The ATCS seven-layer architecture can be divided into two subsystems: the lower layer protocols (LLP) and the upper layer protocols (ULP). The LLP provide for the switching and routing of information, media access control, synchronization, framing, error detection and recovery, and the actual movement of the bits across a transmission medium. The ULP provide data formatting and identification, interaction between applications, and end-to-end data integrity and quality of service.

TESTING THE ATCS DATA COMMUNICATIONS SYSTEM

The communications components of ATCS provide functionality within the lower three layers of the data communications system protocol. These are the mobile communications package, base communications package, front-end processor, and cluster controller. All seven layers of the protocol are implemented within the end-user computer system.

To test the functionalities of the lower layer communications components, a communications interoperability tester (CIT) was developed. This test system, written in the ADA programming language, allows for a stimulus-response, script-based–type testing of the four ATCS communications components in a simulation environment.

According to ATCS Specification 210, "A central objective of ATCS is to ensure safe train movements and track occupancies which avoid physical conflict, and other operational hazards of similar magnitude." To ensure that the implemented ATCS will meet this objective, development of a simulation model of an ATCS is imperative. As a subset of simulating an entire ATCS, a working prototype of the communications protocols is being developed.

SELECTION OF A PROTOTYPE MODEL

Some of the advantages of building a prototype are discovering design problems early, spotting system interface problems from the start, and using the prototype to experiment with how proposed enhancements and modifications will affect the system (*2*).

Fundamental to the success of a prototype is the ability to develop a good model of the system. As Hoover and Perry have pointed out, "Models are not true or false, but rather they are useful and appropriate for the analysis at hand" (*3*). The type of model required for the simulation of the ULP must have the following characteristics: descriptive, discrete,

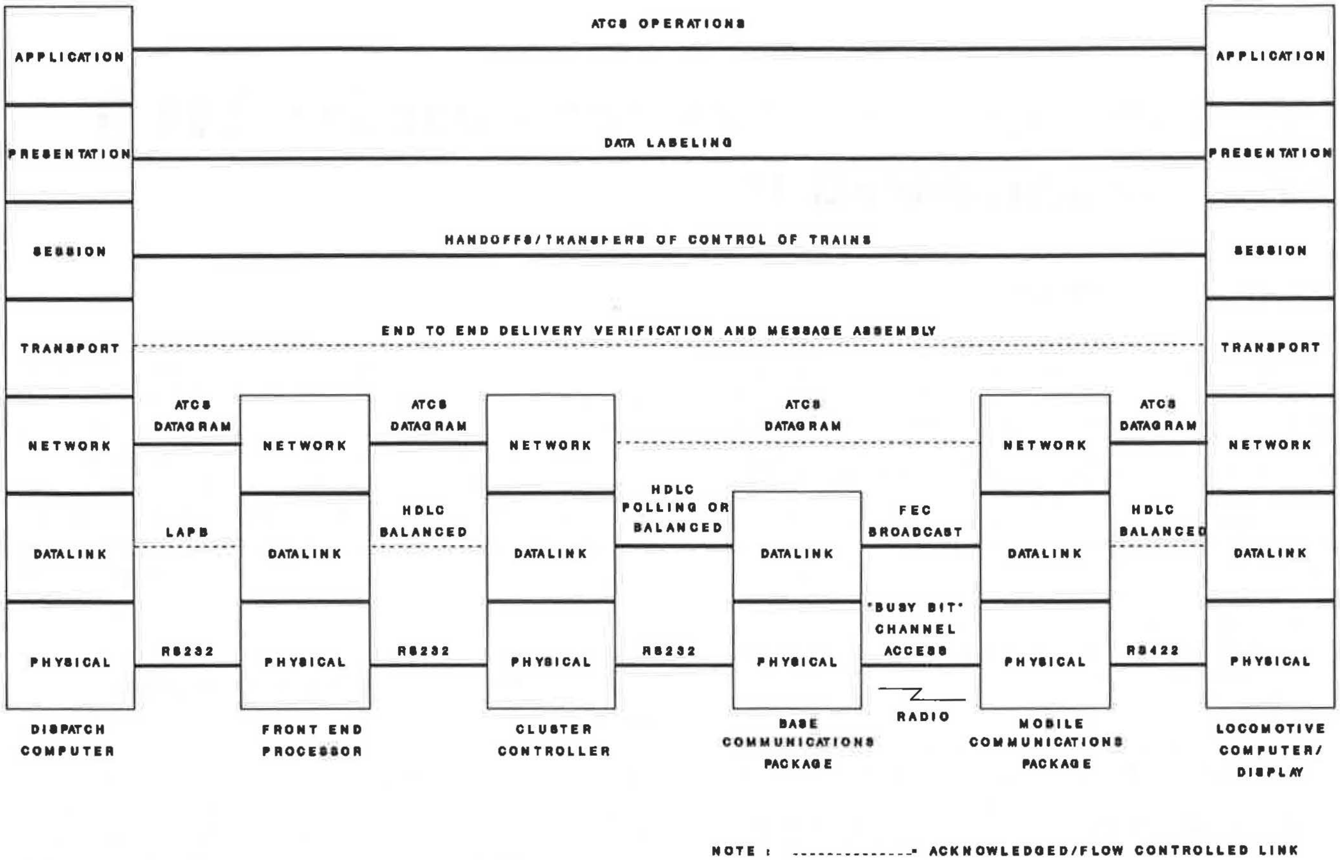


FIGURE 1 ATCS data network by OSI layer.

probabilistic, dynamic, and open-loop. The purpose of this prototype is to describe the system behavior rather than to optimize the system; therefore, a descriptive model of the system is required. Data packets, and messages, arrive at discrete points in time; hence, the system model must be discrete. The arrival times of these packets and messages cannot be predicted, requiring that the model be probabilistic. All communications networks are dynamic, with the number of messages/packets in queue at any given time being variable. Finally, a communications network is an open loop; the system output is not fed back into the network to modify subsequent outputs, but is consumed by the environment with which it interacts.

With the required model classified as a discrete event model, and the objective to simulate the model, the next step is to select a technique for developing and simulating the system model. The major tasks in model construction have been defined as (a) develop a computer program flow chart for the model; (b) select a programming language; (c) provide for generation of random numbers and collection of performance measure values in the program; and (d) write and debug the program code (*3*).

Two approaches for developing model charts of the system were identified: the physical flow and the state transition flow. The physical flow approach identifies and diagrams the physical entities of the system. For the communications network, this would be the receive and transmit queues, processing queues, and any internal queues between layers of the protocol (see Figure 2). The state transition approach uses state transition diagrams (*4*) to represent a system. This is an event-driven approach that uses state variables to describe the system (see Figure 3). Both of these approaches are helpful in understanding and diagramming the intricacies of a system.

The ATCS communications protocols are described in ATCS Specification 200 as finite state machines (FSMs) (see Figure 4) using CCITT Recommendation Z.101 notation (*5*).

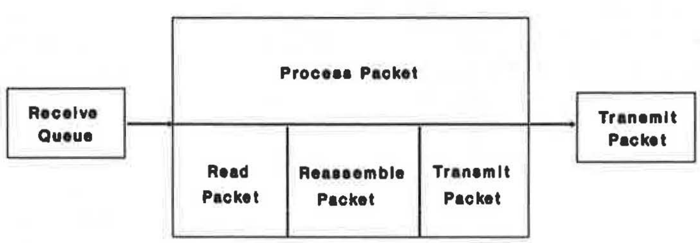


FIGURE 2 Physical flow chart example.

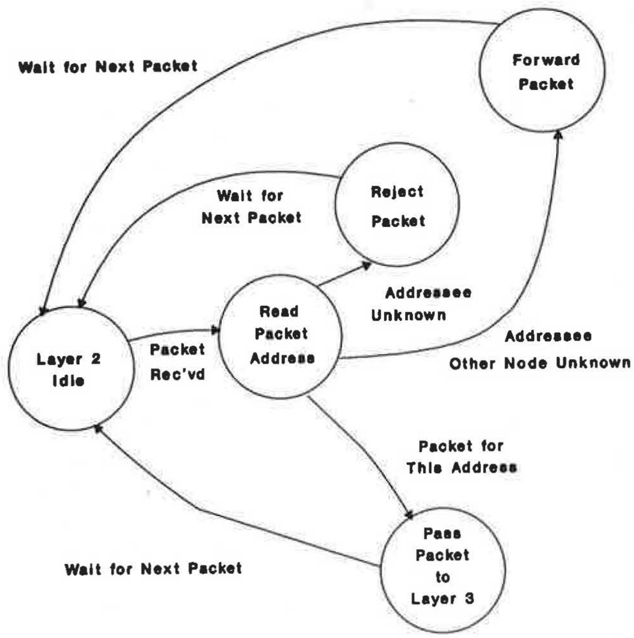


FIGURE 3 State transition diagram example.

Statecharts, extensions to FSMs, have been developed and implemented in STATEMATE, a set of tools intended for the specification, analysis, and design of reactive systems. STATEMATE was selected for the development of the prototype model for the simulation of the ULP.

STATEMATE allows for the development of the model from three separate, but related, points of view: structural, functional, and behavioral (*6*). Figure 5 illustrates how these views are linked together to create a logical model of the system. The structural view of the model is decomposed into modules. These modules may describe the actual hardware components of the system, or they may define the software components of a system: the subroutines, packages, or tasks. This view of the system is modeled using module-charts (see Figure 6).

The conceptual view of the model is decomposed into its functions and controls. The functional view of the system is described using activity charts. The system's functions are described as activities (see Figure 7). An activity typically accepts inputs and produces outputs. Activities that reside at the lowest level of the system may also be described as code in a high-level programming language. Activity charts may also contain data stores and control activities. Data stores portray a temporary or permanent storage area for data. Con-

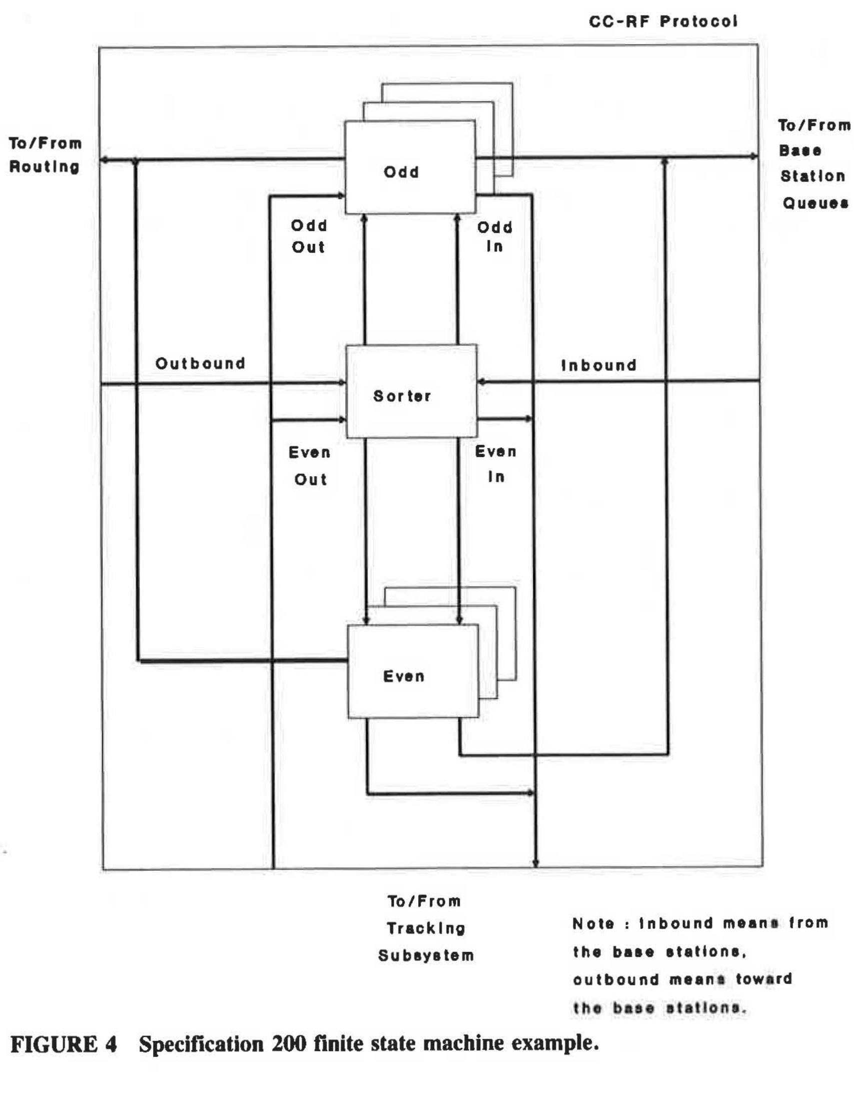


FIGURE 4 Specification 200 finite state machine example.

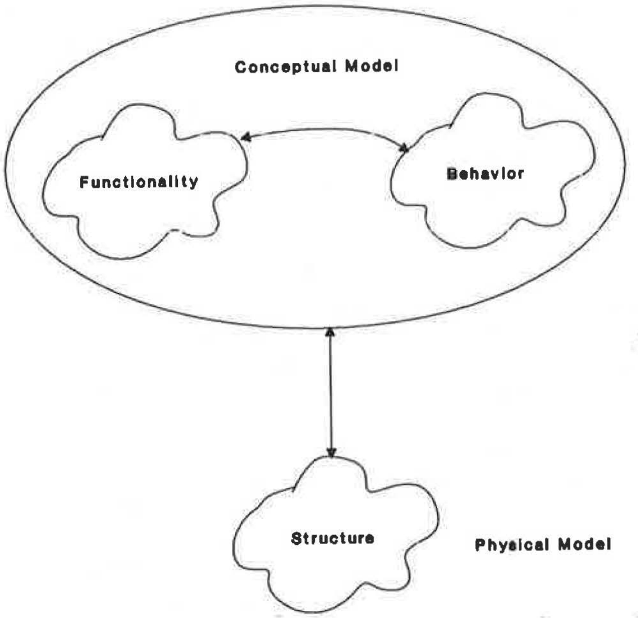


FIGURE 5 Structure of a STATEMATE model.

trol activities, the behavioral view of the system, are described as statecharts (see Figure 8).

The STATEMATE simulation tool can be used to simulate the dynamic behavior of the system at any stage during and after its development. The simulation tool may also be used to debug and verify that the system works as expected. It can also be used to demonstrate the behavior of the system or to run it through various scenarios and observe the system's interactions.

PROGRAMMING THE MODEL

After a model has been developed and analyzed, STATEMATE can then translate the developed system into code in one of two available high-level programming languages, ADA or C. This prototype code represents the executable conceptual model of the system. This code may be compiled and linked as is, or it may be enhanced to call other code developed by programmers.

For implementing the ULP, ADA was selected as the programming language for code development. One of the main reasons for the selection of ADA is that the CIT, which was built to test the lower three layers of the protocol, was developed in ADA. Should the ULP model and any of the CIT software need to interface, the process would be expedited because they were both developed in the same language.

Although there are arguments that the overhead associated with the use of ADA in OSI-style communications systems is significant (*7*), the intent is not to develop an efficient model of the system as much as it is to develop an *accurate* model of the system. Therefore, any overhead introduced by the language would be moot.

There are also arguments that code generated by CASE tools is cumbersome, poorly structured, and inefficient. The STATEMATE developers admit that the code may be inefficient (*7*). However, the advantages of having the CASE tool generate the code outweigh the disadvantages. Having the system generate the code that can then be compiled and linked to form the executable program ensures that the final program accurately and precisely correlates to the diagrammed model. The only potential areas of deviation from the communications specification are (a) incorrectly entered diagrams or (b) incorrect implementation of the manually produced primitive event code.

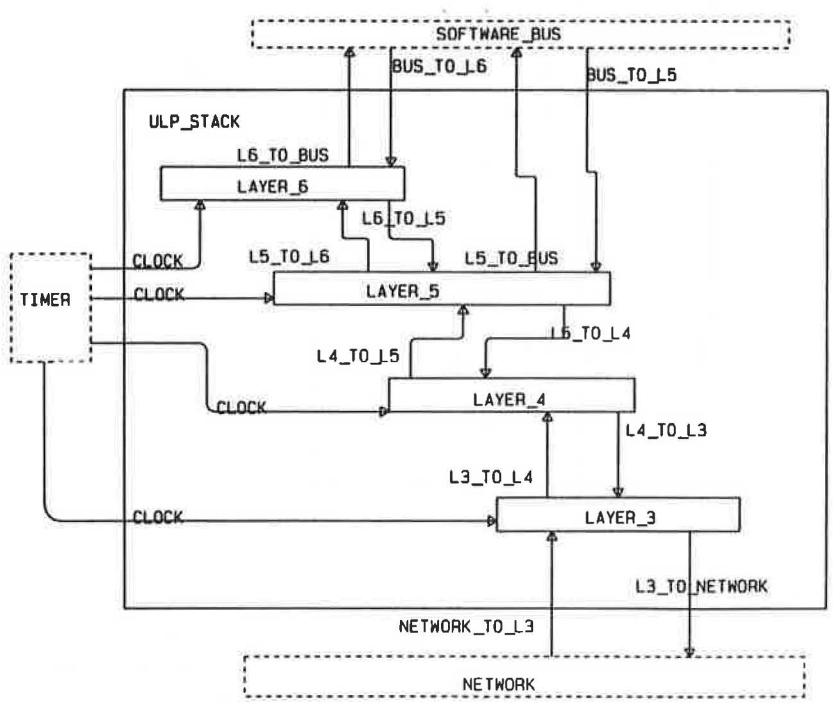


FIGURE 6 ULP module chart example.

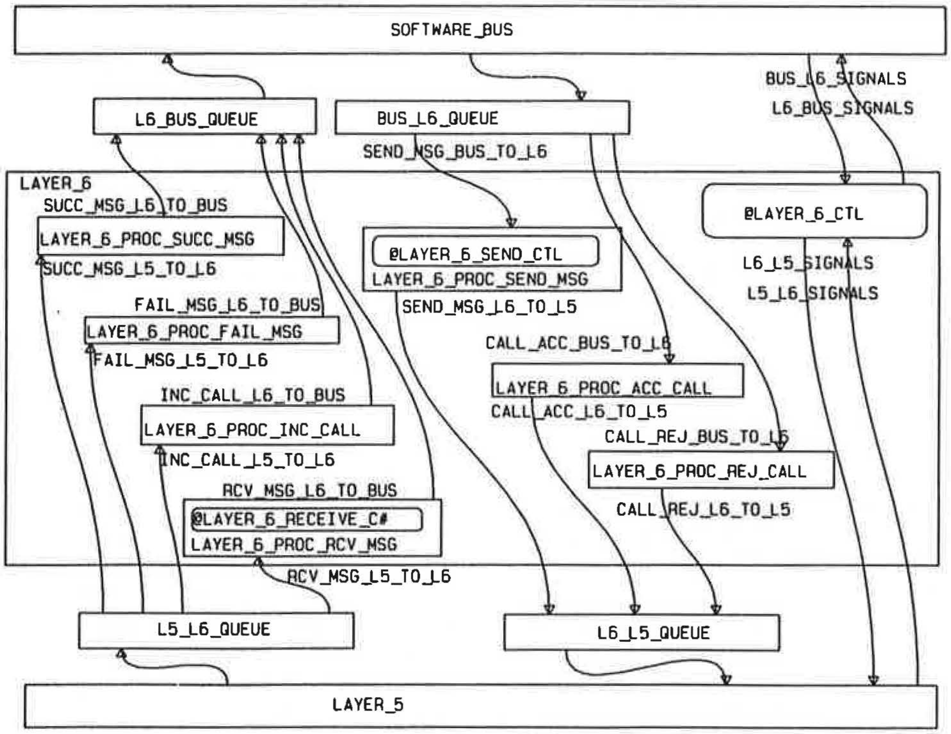


FIGURE 7 ULP activity chart example.

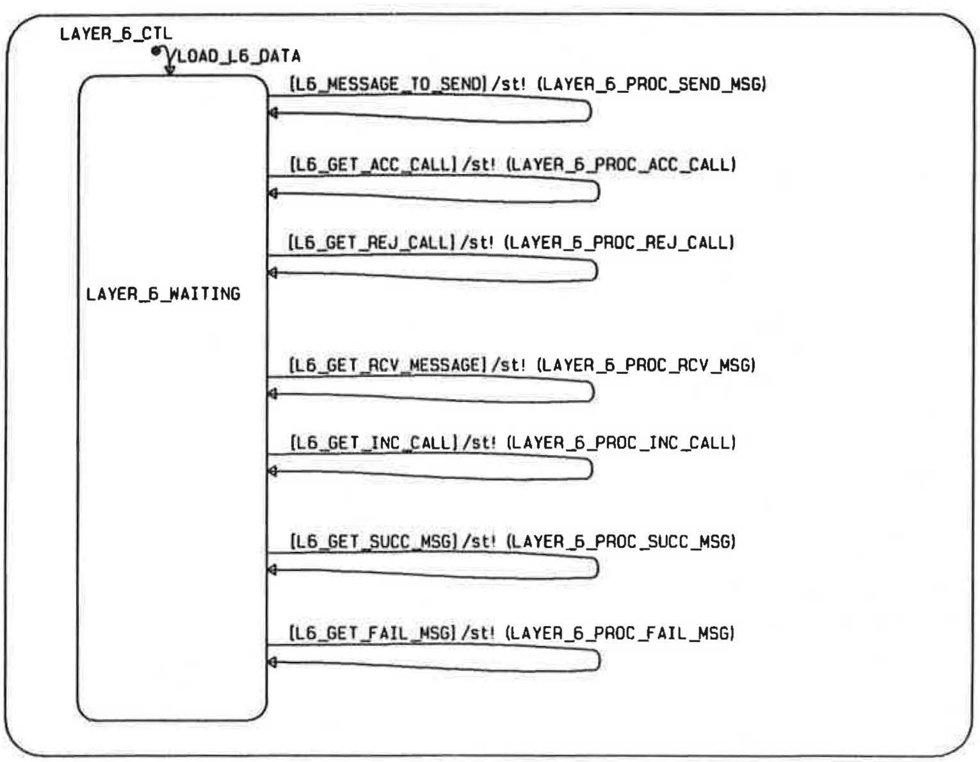


FIGURE 8 ULP statechart example.

Using the code-generation tool also reduces the number of hours required to develop the simulation model. For example, the STATEMATE-generated code handles the ADA tasking development and synchronization, a function that can take a programmer weeks to refine.

Because the purpose of this project is to develop a fully functional model of the system, the prototype code produced by STATEMATE is edited to call modules written by the programmer—referred to as primitive events. These primitive events are called to perform such things as storing a message header in a queue or retrieving a message header from a queue.

MODEL IMPLEMENTATION

At the current writing, Layers 5 and 6 of the ULP have been diagrammed in STATEMATE, verified using the testing tools of STATEMATE, and the ADA code generated for these two layers. The code has been edited to call the programmer-developed primitive events. After being compiled and linked together, the code was debugged by stepping through the logic of the model.

Many uses of the final product of this model have been suggested. One plan calls for further development of a limited version of the entire ATCS. This would require a simulation model that could simulate ATCS behavior in the command and control of trains. To pass messages from the ULP model to the planned models of ATCS components, a "software bus" has been developed. This software will pass simulated ATCS communications between the ULP model and the component models during simulation of the system.

Plans have also been made to use the model to test a ULP stack developed by other organizations. In this manner, the peer-to-peer relationships of the ULP under test could be verified.

MODEL VERIFICATION AND VALIDATION

The ULP model is being developed using modularity and stepwise refinement. The first segment of the model developed was Layer 6 of the protocol. This module was then interfaced with the software bus and a stub of Layer 5, debugged, and tested. Layer 5 was then developed, debugged, and tested in the same manner. Layers 3 and 4 will follow this same development method.

At completion of the entire model, the model will be validated. Verification and validation of a model is at best an inexact science, yet the credibility of the model must be established. Many formalized methods for model validation have been developed, such as the Delphi Method, the Turing Test, and the Structured Walk-Through. Although all of these methods have something to offer, and the exercise of any of them will, no doubt, turn up errors in the model's logic, each has its own drawbacks.

Plans are to test the ULP model by first stepping through it with the debugging tools available in the VAX implementation of ADA and the debugging tool in STATEMATE. This process will validate the layer-to-layer services of the model. The next test will be to perform communications between two executing copies of the model to validate peer-to-peer functionalities of the model. A third validation of the model will be its execution with the software bus at the upper end of the ULP stack and an ATCS network emulator at the lower end of the ULP stack. This process will validate the end-to-end functionality of the protocol model between the user application and the network emulator.

At each of these validation stages, the model should be tested not only to ensure that it functions properly under normal conditions but also to observe its behavior under extremes.

CONCLUSION

Prototyping is based on building a model of the system under development and then simulating the system's behavior using the model. The use of STATEMATE to visually formalize and then generate code to develop the ULP simulation model takes advantage of the benefits of prototyping. The use of ADA as the programming language exploits the tasking capabilities of the language and provides for future interfaces to the CIT. This approach maximizes the capabilities of the available technologies and the computer system. The result should be the development of a model that accurately represents the real system, its functionalities, and its behavior.

ACKNOWLEDGMENT

The author would like to thank Robert Ayers, Jack Bailey, and Denny Lengyel of ARINC Research Corporation, without whom this project never would have begun. Robert Ayers deserves special thanks for his technical support, assurances, and constructive criticism during the development of the model thus far and for his design and development of the software bus. Thanks also to Jennifer Miller for keeping the development computer system and software up and running during this project.

REFERENCES

1. The International Telegraph and Telephone Consultative Committee. *Data Communication Networks Open Systems Interconnection (OSI) System Description Techniques, Recommendations X.200–X.250*, Vol. VIII, 1985.
2. K. E. Lantz. *The Prototyping Methodology*. Prentice Hall, Englewood Cliffs, N.J.
3. S. V. Hoover and R. F. Perry. *Simulation: A Problem-Solving Approach*. Addison-Wesley, New York, 1989.
4. A. M. Davis. A Comparison of Techniques for the Specification of External System Behavior. *Communications of the ACM*, Vol. 31, No. 9, Sept. 1988, pp. 1,098–1,115.
5. The International Telegraph and Telephone Consultative Committee. *Functional Specification and Description Language (SDL), Recommendations Z.100–Z.104*. Vol VI, Fascicle VI.10, 1984.
6. D. Harel, H. Lachover. A. Naamad, A. Pneuli, M. Politi, R. Sherman, A. Shtull-Trauring, and M. Trakhtenbrot. STATEMATE: A Working Environment for the Development of Complex Reactive Systems. *IEEE Transactions on Software Engineering*, Vol. 16, No. 4, April 1990, pp. 403–413.
7. N. R. Howes and A. C. Weaver. Measurements of ADA Overhead in OSI-Style Communications Systems. *IEEE Transactions of Software Engineering*, Vol. 15, No. 12, Dec. 1989, pp. 1,507–1,517.

Advances in Flat Panel Display Technology and Applicability to ATCS On-Board Terminals

Chuck J. Karbowski, Gideon Ben-Yaacov, and David Blass

The introduction of Advanced Train Control Systems (ATCS) to the railroad industry changed the way operational information is communicated to locomotive crews. Voice communication can now be supplemented and may eventually be replaced by data communication via intelligent display terminals. Flat panel display screens are well suited for locomotive display terminals; the screens are compact and have the potential of being able to sustain reliable operation in harsh environments such as those found on board locomotives. In choosing flat panel display screens for locomotive applications, a number of factors must be considered: how the various flat screen display technologies operate, what their technical features are, and how they respond to harsh environmental conditions. Although no perfect display screen exists for locomotive ATCS applications, a comparison of strengths and weaknesses of the various technologies currently available points toward the electroluminescent and the thin film transistor displays as being the most appropriate.

Cathode ray tubes (CRTs) and flat panel display screens are the two types of display screens commonly used with display terminals to provide the operator-interface medium.

CRTs, which have been used in process industries for more than 25 years, have a wide range of commercial, industrial, and military applications. On the commercial side, CRT displays dominate the personal computer market. On the industrial side, CRTs dominate the industrial work station market. On the military side, CRTs provide operator interfaces. Presently CRTs are the only display type that can provide color at reasonable cost. But CRTs are bulky, susceptible to vibration, shock, and electromagnetic interference, and they require a moderate amount of power. If CRTs are to be used in ruggedized environments, they have to be packaged in such a way that they can withstand the harsh environments. Because CRT screens are bulky and fragile, they are not suitable for locomotive Advanced Train Control Systems (ATCS) applications.

Flat panel displays, on the other hand, are light weight, ruggedized, reliable, and compact; therefore, they can be mounted in tight spaces (i.e., locomotives). Flat panel monochrome displays are becoming popular in industrial and military applications because of technological advances in their production. Color flat panel displays are also becoming available at a reasonable cost.

A flat panel display screen that can meet all the environmental and visual requirements for a locomotive ATCS application does not exist. Each of the available display technologies has some limitations with its environmental and visual specifications. Nevertheless, flat panel display screens are considered the most suitable display screens for locomotive ATCS applications because they are ruggedized and compact.

Depending on the environmental and visual specification requirements considered by a railroad to be the most important, a compatible display technology can be identified. Such a display technology will have superior capabilities with those features considered most important, but also may have limitations in areas considered less important.

A review of the various types of flat panel display technologies, with detailed feature comparisons, is given below. The environmental characteristics of each display technology are then compared to the applicable specification requirements (ATCS Specification 110, Environmental Requirements). Key features of each flat panel display technology are outlined and each display technology ranked against the others. Finally, the display technology that the authors consider the most appropriate to support the visual and environmental requirements of ATCS locomotive applications is presented.

TECHNOLOGY REVIEW

Liquid Crystal Display

The operational theory of liquid crystal displays (LCDs) can be described as a shutter mechanism that blocks the transmission of light. The blockage of light produces a pixel whose contrast is determined by the difference between the blocked area of the pixel and the illumination (either reflected or transmitted) from a light source. The light blockage is achieved by a liquid material that is able to bend or twist light. This liquid crystal is sandwiched between two sheets of glass. Polarizers, which are oriented along each axis where the light is blocked, are placed on each layer of glass. When the two polarizers are placed on top of each other, no light can pass through the glass. When the liquid crystal is placed between the two polarizers, the twisting or bending of the light allows it to orient with the rear polarizer, so that a portion of the light can be reflected or transmitted back out. When a relatively low current is passed through the liquid crystal, the material twists again, taking the transmitted light out of orientation with the rear polarizer. This effectively blocks that area, producing the image of a black or dark pixel on a light background.

C. J. Karbowski and G. Ben-Yaacov, Automated Monitoring and Control International, Inc., 11819 Miami Street, Omaha, Nebr. 68164. D. Blass, Emerald Computers, Inc., 16100 S.W. 72nd, Portland, Oreg. 97224.

The grid of the display is scanned row by row. As each row is scanned, those pixels to be oriented are addressed and charged. The scanning panel then moves to the next row, and so on. The pixels can only "twist" when energized. This takes place during the scanning. Once the signal is removed, the crystalline material "relaxes" back to its original state.

The display must maximize the contrast of the pixel, so ample time must be given to allow the display to twist completely. Lengthening the time each row is given to twist the shutter will increase contrast of the pixel, but decrease the video refresh rate of the display. Either the display must be slowed well below the standard 60 Hz frame rate or a decreased level of contrast must be tolerated. All LCDs have had their driving schemes modified to optimize this conflicting issue.

Types of LCDs

There are four types of LCDs. Each successive display described is technologically advanced over the others. The major differences between the LCDs depend on the nature of the liquid crystalline material and the application of polarizers.

1. *Standard twist* (twisted nematic, TN). This is the original LCD technology. The distinguishing characteristic of the standard twist is that the twisting material orients at either 0° or 90°, and none other. TN is employed today in either low-information-content displays (such as digital wrist watches) or fixed legend displays, but rarely in any display that has a significant volume of information to convey. It will not be discussed further in this paper.
2. *Supertwist* (supertwisted nematic, STN). The supertwist advancement produces a liquid crystal capable of rotating 180° to 240°. This tighter rotation makes a darker pixel and increases contrast. In addition to the highly twisting nature of STN, the crystalline material also produces a birefringence effect. Simply put, birefringence allows the orientation of polarizers to transmit various frequencies of light that can produce colors. This effect permits different potential colors for both the pixel and the background, increasing contrast and flexibility.
3. *Double supertwist* (DST). This advancement lays two supertwist displays on top of each other. One layer is active, meaning that the drive mechanisms for the pixels are present. Through the birefringence effect, the second passive layer is used to cancel out interfering colors, producing a display that appears black on white. Because of the high cost of the double layering of the liquid crystal material and glass, a second method of DST, called double film supertwist, was developed. Rather than completely duplicating a layer of liquid crystal and glass, the compensating role of the second layer is performed by using a layer of specialized polarizers. The double film LCD has the same basic characteristics as the double layer, but costs substantially less.
4. *Active matrix* (thin film transistor, TFT). Active matrix displays have the same basic display characteristics as double supertwist displays. The key improvement is the elimination of (a) the row/column scan and (b) the opposing timing issues of time-to-scan versus the full rotation to the pixel. The essential principle of TFT is putting a drive transistor behind each of the pixels on the display. This allows the pixel to be turned on, and left on, until the data signal turns it off. This produces a display with high contrast, full viewing angle, and video resolution comparable to CRTs.

Backlights

Viewability in variable lighting conditions has always been a problem with LCDs. To solve this, backlights were added so the display could be viewed under night or low-light conditions. Three types of backlights are used with LCDs.

1. *Electroluminescent* (EL). An EL backlight is essentially a large area of phosphor material sandwiched between two sheets of mylar. When direct current is sent though the material, the phosphor glows. The backlight is very thin, but suffers from a short lifespan (2,000 to 5,000 hr) because of erosion of the phosphor. Additionally, an EL backlight does not emit sufficient light to be used at night.
2. *Cold cathode*. This method of backlighting consists of a fluorescent tube positioned behind the display. Fluorescent lighting yields one of the most efficient light sources. The cold cathode produces a higher light output for its power draw than nearly any other light mechanism.
3. *Hot cathode*. This technique is similar to cold cathode in that a fluorescent light source is used. It is much brighter than the cold cathode backlight, and draws more power. The hot cathode backlight is required for active matrix LCDs, which utilize the transmissive type of polarizer. The thin film transistors cut the light transmission by up to 40 percent, so the light source must be very bright.

Polarizers

All LCDs have at least two polarizers to produce the pixel image. Because LCDs work on the principle of light either passing through or being blocked by the display, the polarizers play an important role in maximizing the readability of the display.

1. *Transmissive*. This rear polarizer allows light to pass from the back of the display, through the display, and out the front. It is designed to operate best with overhead projectors and active matrix-types of displays.
2. *Reflective*. This polarizer reflects the ambient light that passes through the display and back out the front of the display. It reflects the maximum amount of light back through the display, making it the best polarizer for displays that are not backlit.
3. *Transflective*. The transflective polarizer not only reflects light like a reflective polarizer, but will also allow the passage of light from the back of the display. It offers the effectiveness of both functions of the polarizer (i.e., it will reflect less light than the reflective type and transmit less light than the transmissive polarizer).

Electroluminescent Display

EL displays work on the principle of sending a discharge of energy through a phosphor material, which produces light. EL is structured much like a sandwich of material deposited

on glass. The glass is placed at the front of the display and is viewed through, followed by the first conductor. Because the first conductor (the column conductor) is in front of the active material, it must be transparent. This material, indium tin oxide (ITO), is transparent. It is also relatively resistive and is used on the column conductor (the shortest axis). Behind the ITO row conductor is a dielectric barrier, followed by the active material, zinc sulfide doped with manganese (ZnM). This material, when energized, produces the amber color of EL. Behind the phosphor layer is another dielectric layer, followed by the row conductor. Because the row conductor is behind the layering, aluminum is used, which, although it is not transparent, has low resistance.

Through a row/column scan method similar to the other display types, each pixel can be addressed. One-third of the total voltage required to energize (light) the pixel is sent down the column, and the other two-thirds is sent across the row conductor. At the intersection (the pixel), the combined voltage creates a sufficient electrical charge to burst through the dielectric layer. This capacitive discharge causes an "electron avalanche" through the manganese phosphor, creating light. Because the pixel is essentially a capacitor and the column electrode is a resistor, this RC (resistor/capacitor) network creates a timing constant. In other words, there is a predetermined time for the signals to travel the display. This time constant prevents EL displays from being very tall (beyond 5 to 6 in.) without the implementation of costly and complex production techniques.

Types of EL Displays

There are two versions of EL displays, alternating current-driven thin film EL and direct current-driven, thick film EL. Both work on the same basic principle of operation.

1. *Alternating current thin film EL* (ACTFEL). Thin film makes EL displays truly solid state, not mere "material between glass." The films are deposited on the glass in a vacuum chamber using the same method used for making semiconductors.
2. *Direct current thick film EL* (DCEL). Instead of using the thin film process of depositing the phosphor material on the glass, a thick film paste is deposited on the glass with a silk-screening method. Also, a DC drive scheme is used, which has a lower cost than AC.

Like LCD (and plasma displays), EL displays are scanned row by row. Because the thin film display is quick to respond to the signal, pixel scanning occurs at regular video rates. In essence, an EL display has all the visual characteristics of a typical monochrome CRT.

PLASMA DISPLAYS

Plasma displays are flat neon tubes. The principle is similar to that used in standard neon signs. Neon gas is held within a vessel. A high-voltage current is passed from a cathode through the gas to an anode. The resulting excitation of gas molecules produces red light. Like EL displays, the row-by-row control of plasma displays is sufficiently fast to maintain standard video refresh rates.

Types of Plasma Displays

There are three types of plasma displays: DC, AC refresh, and AC memory. These three display types are differentiated by the electronic driving scheme of the display. By theory of operation, however, all three display types are the same.

1. *DC plasma.* DC plasma uses a DC drive scheme. The display state is always lit, even in a pixel "off" state. The lighting of a pixel is accomplished by turning it on even higher than the background. The nature of the background generates a steady glow over the entire surface of the display, resulting in a lower contrast level.
2. *AC refresh.* By means of an AC drive, individual pixels are scanned and refreshed in a fashion similar to an EL display. With this method, the background is turned "off" when a pixel is not lit. The scan refreshing method (turning a pixel on quickly and then leaving it to scan the rest of the panel), however, produces less light than the AC memory type.
3. *AC memory.* This type of plasma uses the same AC drive scheme as AC refresh, but includes a memory function. Simply put, true memory drive occurs when a pixel is turned on and remains on until a control signal turns it off. Smaller plasma displays use true memory, but in the large-area displays this method is modified somewhat. AC memory addresses every other row on every other frame. This leaves the pixels on twice as long as any other type of display. The longer the display is on, the brighter the display appears.

VISUAL CHARACTERISTICS

Feature Comparison

This section will itemize the main features and criteria for choosing the optimal display for locomotive ATCS application. To select the best display for the ATCS application, particular attention must be paid to the visual characteristics of the display, environmental specifications, and the latest state-of-the-art characteristics of various display types.

Visual characteristics are very important. The visual characteristics of the display are what the engineman and other railroad personnel will see. The combined total of all of the visual features will cause each railroad to select the display they "like." This subjective decision often has no engineering or mechanical merit. The railroad will simply select the display they personally prefer. Despite the biased nature of the decision, the criteria themselves are based on the very objective characteristics of the display. Viewing angle, color, contrast, and adjustments for low and high ambient light conditions are the main points of comparison.

Sunlight Readability

The definition of "sunlight readability" used by the U.S. Air Force is, "a readable display which exhibits a 3:1 contrast

ratio, filtered, in a 10,000-candlepower ambient light." The various types of displays are ranked below according to their sunlight readability.

STN—Excellent. The STN version of LCD is rated excellent in sunlight. The STN technology is available with either the reflective or transflective polarizer. The transflective type of polarizer is the optimum polarizer for both sunlight and backlit applications (see the polarizer section above). A reflective polarizer will maximize the readability in full sunlight, but it will not support a backlight, making the display useless in low to no light. Although the transflective display does lower the reflective qualities of the polarizer, it is still excellent for the sunlight readability requirements.

DST—Excellent. (Same as STN.) Reflective polarizers would render the backlight useless, so a transflective polarizer is the only reasonable choice.

TFT—Fair. This display depends entirely on a bright backlight and does not use any reflective/transflective polarizers. Only with the use of external sun shades would this display be usable in bright light.

EL—Poor to fair. This display produces the same light output as a monochrome CRT. Generally, an EL display is washed out by ambient light in the 5,000–candlepower range, which is far below minimum "sunlight readable" specs. Generally, it is used outdoors only with the benefit of shades or cowling. Also, sunlight readability can be enhanced with a circular polarizing filter.

Plasma—Poor to fair. Plasma displays generally have the same specifications as EL displays or CRTs. The display produces its own light, but not enough to hold up in sunlight. AC plasma is brighter, and is rated fair, but would still be found to be unacceptable for locomotive applications.

Low Light Conditions

STN—Fair to very good. With the use of a backlight, STN is fair to very good, depending on the type of backlight chosen. An EL backlight, because of its low light output, makes for an adequate lighting mechanism. However, it is not preferred. The cold cathode backlight does produce enough light to make a good display in low lighting.

DST—Good to very good. With a cold cathode backlight, the display has good to very good low light characteristics.

TFT—Excellent. The TFT is generally provided with the hot cathode backlight scheme. It therefore produces an excellent light source for low light conditions. The backlight can be adjusted downward in case the backlight is too severe. Given its bright light output and its adjustability, TFT produces the best display for lower light conditions.

EL—Very good to excellent. EL is very good to excellent in low light conditions. It produces a very bright display and can be easily seen in low light. The main drawback of EL is the lack of downward adjustability. There is no standard light adjustment offered by any of the EL manufacturers. The display brightness can be lowered by altering the high-voltage converter on the display, but tampering with display components is potentially a violation of the display's warranty and is not recommended (although not discouraged) by EL manufacturers. (The main reason for the lack of manufacturer support of brightness control is that EL emits uneven lumination across the panel when panel voltage is lowered.)

Plasma—Very good to excellent. As with EL, this light-emitting display technology does a very good job in low to no light conditions. Unlike EL, plasma can be lowered in brightness, and it will track in a uniform manner. The major drawback to the plasma display is the neon color red, which becomes, over viewing time, increasingly hard for the eye to tolerate. Low ambient light exacerbates the problem.

Color

STN—Good. The color of the STN is generally offered in blue on white. These colors, developed in the birefringence, are a good color combination. Blue is pleasing to the eye, but it does not offer the same contrast as the DST LCD. Typically, because of the lower refresh rate and the backlight, the blue color can appear to be washed out and not as saturated as would be desired.

DST—Very good. Often thought to be the best monochrome LCD color combination, DST is neither as dark as black nor as saturated as desired. The black-on-white display, even though slightly pale or washed out, does offer the highest contrast ratio of all of the monochrome LCDs. It does rate better overall in color over the STN, but is best categorized as very good.

TFT—Excellent. As a full-color display, the TFT's color is perfect. The display produces a range of colors similar to the saturation levels of a color CRT.

EL—Excellent. From the color standpoint, EL is excellent. The amber color emitted from the display is, physiologically, the second easiest color for the eye to see.

Plasma—Fair. The neon red color of plasma is rated fair. The eye tolerates the color red least of all, and the red color of a plasma display is a contributing factor to eye fatigue. The problem increases in low lighting conditions.

Contrast

STN—Good. The contrast between the pixel and the background of an STN display is no better than good, and that is only in the best conditions. When set to the maximum orientation, the display contrast is good, but can change when the display is viewed off axis. Also, the contrast setting reorients the display in a vertical adjustment (i.e, up and down). If the display is being used at different viewing heights, the contrast will be different for each viewing height. Also, the contrast is affected by variations in the ambient light or by variations in temperature.

DST—Good. The contrast of the DST is generally good, but, as in the case of the STN display, contrast is variable and is only good in a correctly adjusted position. Viewing angle and adjustments can vary the contrast to an unacceptable position. Correctly adjusted, a black-on-white display does offer good contrast. Also, the contrast is affected by variations in ambient light or by variations in temperature.

TFT—Excellent. Again, the display excels in all of the visual characteristics, including contrast.

EL—Excellent. The amber-on-black display of EL is excellent. But the excellent rating of the EL is only achieved with the use of a contrast enhancement filter. (A filter typically is a user option and is not supplied by the manufacturer.) A circular polarizer or neutral density filter satisfies the requirement. Without the filter, the contrast of the EL is fair to poor. However, the filter alters the light output affecting sunlight readability.

Plasma—Fair to excellent. The contrast of plasma varies, depending on the type of plasma technology chosen. The DC plasma display offers the poorest contrast ratio. The background of the display is constantly lit, and the lit pixel is simply driven harder. Both of the AC-driven displays have the background turned off, and it is considered black. The AC memory model has a greater contrast ratio because the pixels are brighter against the black background. The plasma panel's contrast ratio ranks DC fair; AC refresh, very good; and AC memory, excellent.

DST—Good. DST improves the color and contrast of the LCD over that of the STN, but suffers from a much narrower viewing angle. If properly adjusted for the current conditions, it appears better than STN, with a sharp, clear, black pixel. It will, however, fall out of adjustment more easily, and the effects of the misadjusted display are more pronounced. This display reacts to temperature the way STN displays do.

TFT—Excellent. TFT is superior in color and crispness of the display, as well as in viewing angle and contrast. This display reacts to temperature the way STN displays do.

EL—Excellent. As a monochrome monitor, EL offers an excellent display. The pixels are uniformly shaped, focused, and have an excellent contrast ratio and viewing angle.

Plasma—Good. Plasma receives a lower score because the pixels are fuzzy and generally out of focus. This is caused by the ionization of the gas that is not contained, producing more of a flaring dot without uniform edges. The color is a serious weakness for long-term viewing potential.

Viewing Angle

STN—Fair. The viewing angle of the STN is considered fair. Rated as 30° from perpendicular, this is adequate for most single-user applications.

DST—Poor. The viewing angle is poor. Rated at 15° from perpendicular, it has the lowest viewing angle of any of the display technologies.

TFT—Good. This LCD is better than the other LCD types, but its viewing angle is only 45° from perpendicular.

EL—Excellent. The viewing angle of the EL is greater than 80° from perpendicular. It is equal to or greater than that of TFT.

Plasma—Good to excellent. As with contrast, the viewing angle of plasma varies with the three types of plasma displays. DC plasma has vertical ribs that keep the gas discharge in a confined area. When viewed from the side, these ribs will eventually block the view of the pixel. This occurs at about 45° to 60° from perpendicular. Both the AC-drive plasmas share the same viewing angle as EL and TFT. A unique phenomenon, ionization of the neon gas, can cause the brightness and color of the plasma display to change, especially in the AC memory panel. For this reason, AC memory plasma panels fall slightly behind the other two technologies.

Image Quality

STN—Fair to good. The STN display suffers from a variety of display deficiencies. Individually, they are not severe, but when averaged together, the overall rating is only acceptable. The image quality is good only when the display is properly adjusted to the ambient light conditions and viewing angle of the operator. The pixels are well formed and focused. Off-angle viewing will wash out the pixel color and reduce contrast. There is a perceptible screen flicker from both the 60 Hz fluorescent backlight and from the scanning of the display. Also, temperature has an effect on the image quality. At colder temperatures, display response slows. At higher temperatures, the pixels orient to all on.

Gray Scale Ability

STN—Good. STN has the capability to be dithered. This is a gray scale technique that uses the refresh rate to change the intensity of the pixel. Instead of turning on a pixel the full number of frames per second, it can be controlled to turn the pixel on less often, producing a dimmer, lighter pixel. STN can produce eight levels of gray scale effectively.

DST—Very good. Using the same dithering technique as STN, DST has a more dynamic range because the pixels are darker to begin with. The display can generate 16 levels of gray scale with dithering. It should be noted that it is not necessarily easy to distinguish contiguous gray scale levels from each other.

TFT—Excellent. Even in the current limited models, TFT displays can produce 16 colors. Color-generation range will increase to 256 colors within the next 12 months. Because gray scaling is a monochrome technique intended to substitute monochrome shading for color, the TFT, being a color display, is clearly the superior display for color representation.

EL—Good. The EL display can produce gray scale through either dithering or with pattern gray scale. Because the display is fast, the dithering process can clearly be witnessed, and can be distracting. Pattern gray scale (also known as hatching) is fine for large-area graphics, but is not acceptable for text, because it turns off pixels in the character cell. True gray scale EL is relatively new. The technique of voltage modulation (16 levels of drive voltage for 16 levels of brightness) does not produce an acceptable range of definition. The lower levels are too dim, and there is very poor definition between consecutive levels. Even the newer voltage modulation/pulse width modulation (voltage plus variable signal duration) EL displays are better, but EL, in general, cannot produce 16 recognizable levels of light between threshold (the lowest level of pixel turn-on) and full pixel brightness.

Plasma—Fair to good. Generally, plasma follows the same results as EL. It can be patterned or dithered with the same limitations. Only one model of DC plasma is offered with true 16-level gray scale. Because of an increase in the dynamic range of the display, plasma is marginally better than EL, but

it, too, has a difficult time producing discernable, contiguous levels.

Size

STN/DST—Good. Both display types are available in the proper visual display size (10.5 in. diagonally). The models do have a relatively wide package in relation to the width of the active area. This extra width is wasted space. The depth of the displays is acceptable at approximately 1.25 in. The EL backlight models are incredibly thin, which makes a very thin and attractive display head, but the EL backlight method is not recommended because of its low light output, its lack of adjustability, and the rapid aging of the EL film.

TFT—Very good. The size of the TFT package is also 10.5 in. diagonally. The overall package has less wasted space than the STN/DST. TFT has the same depth concerns as STN/DST because of the backlight requirement. It is not offered with EL backlighting because of insufficient light output from the EL film. The hot cathode backlight provides the same packaging as the cold cathode type.

EL—Excellent. In the past, EL manufacturers have suffered from an inability to produce cost-effective displays larger than 9 in. diagonally. New gray scale models are now available with a display size of 10.4 in. diagonally. With the fact that the package is only 1.4 in. deep (including the power converter), the depth of the EL display is excellent.

Plasma—Very good. Plasma displays are available in large display sizes (10.5, 14, 17, and even 26 in. diagonally) and can produce high-resolution displays (up to 1,024 x 800). In spite of that, plasma suffers from two problems. The "dead" margin around the display is large in proportion to the active area. Plasma also requires a large power converter that must be placed somewhere in the terminal or on the back of the display. This adds considerable depth to the package.

ENVIRONMENTAL SPECIFICATION

The various types of displays have different strengths in terms of the environmental factors of locomotive ATCS application: temperature, humidity, shock and vibration, electromagnetic interference, power consumption and, of course, cost.

Temperature

STN/DST—Fair to good. LCDs have the narrowest temperature range, typically 0°C to 50°C and often 0°C to 40°C. The key to the display is the "liquid" crystal. The crystalline material loses its primary properties at temperature extremes. At less than 0°C, the material becomes thick and viscous before it freezes completely. At low temperatures, the display will draw an increasing amount of power to rotate the thickening crystal. Once it freezes, the display will not operate until it thaws. Also, the display becomes increasingly slow at low temperatures. High temperatures affect LCD differently. The liquid crystal will orient the pixels to an "all on" orientation, rendering the screen black. This condition will reverse, but only after the display cools down. Additionally, the display polarizers can be jeopardized if exposed to excessive temperatures and humidity. New, extended-temperature liquid crystal material and panels are making an entrance into the market.

TFT—Fair. TFT has the same problems with temperature as standard LCD models. There is a published warning regarding the operation of the panel at high temperatures (50°C). The warning states that operation at 50°C be less than 12 hr in duration for a 24-hr period. At this writing, there is insufficient data to predict extended temperature models.

EL—Very good to excellent. EL, as a display technology, is relatively unaffected by temperature. EL displays can actually operate from −55°C to +125°C, with only a slight brightness output change (±30 percent). The only true limitation to an EL application is the commercially available and specified integrated circuits (ICs) that make up the components of the panel itself. The component manufacturers rate them only at 0°C to 50°C. Though they have a relatively narrow range, the ICs have been tested and actually function without any problems from −20°C to 70°C. Below −30°C, most of the ICs fail to operate properly.

Plasma—Good. Plasma has a broader temperature range than LCD, but a narrower range than EL. Neon gas used in the display will not ionize or turn on below −5°C. The gas, too, can be affected by long-term exposure to very cold temperatures.

Humidity

STN/DST—Fair. LCD panels are rated at 85 percent humidity at 40°C. Because of the railroad industry environmental specifications, this is considered only fair. In many geographic regions, an 85 percent relative outdoor humidity will be exceeded during several months of the year. The polarizers of the display are generally most vulnerable. In excessive heat and humidity, the polarizers can peel off completely.

TFT—Fair. The preliminary TFT specification clearly spells out the warning that the absolute humidity must be lower than 85 percent relative humidity (RH) at 40°C.

EL—Excellent. EL is rated at 95 percent RH at 40°C. This is the highest humidity rating of any of the technologies.

Plasma—Fair. As with the LCD, plasma display manufacturers rate their panel at 85 percent RH. This seems low, because there are no apparent technical limitations of the technology to warrant such a specification.

Shock and Vibration

STN/DST—Fair. LCD models have low shock and vibration specs. Typically, they range from 3 *g* to about 20 *g* of shock. Vibration durability, too, is less than the other technologies'. The physical package of the LCD, made up primarily of plastic, is responsible.

TFT—Poor. Again, this type of LCD fares poorly in the shock and vibration specs. The shock specification is 3 *g* and vibration is significantly less than even the STN/DST models. Here, not only does the plastic package lend itself to the weak specification, but actually the entire TFT technology appears to be relatively delicate.

EL—Excellent. EL has a shock and vibration specification that is up to five times that of plasma or LCD. It is rated at

100 *g* of shock and a 3 *g* peak-to-peak vibration spec. Only the EL display has the integrity to withstand the shock and vibration specification established by the railroads.

Plasma—Good. Although more durable than LCDs, plasma displays' shock and vibration specs are much lower than those of EL. Plasma panels tend to average 10 *g* to 30 *g* shock and 0.5 *g* to 1.5 *g* vibration in specification.

Electromagnetic Interference

STN/DST—Excellent. The STN and the DST displays are rated as excellent for their low levels of reactivated electromagnetic interference (EMI). This can be attributed to the use of a nonemissive technology that utilizes low power and low scan rates.

TFT—Very good. The TFT display is rated very good for its low level of radiated EMI. It is not quite as good as the STN and the DST displays because its clock rates are much higher.

EL—Fair to good. The EL display is rated as fair to good because of its high clock rate and high-voltage operation.

Plasma—Fair. The plasma display is rated as only fair as it radiates even more than the EL because it, too, has a high clock rate and high-voltage operation.

Power

STN/DST—Excellent. The primary "claim to fame" for LCD is the extremely low power consumption of the display. LCD runs less than 5 W total, even with a cold cathode backlight on. The display itself draws even less power. This is pertinent for determining the packaging needed to meet the temperature specifications.

TFT—Good. The display portion of the TFT has the same very low power draw as the other LCD types, but the most voracious power draw within TFT is the hot cathode backlight. This type of backlight draws much more power than the cold cathode type, and the draw is actually slightly higher than that of EL displays. A TFT display will consume 18 W of power with the backlight on.

EL—Good. EL draws about 15 W in a typical mode. This represents a lower overall draw than that of plasma displays, but approximately three times more than that of a STN/DST LCD.

Plasma—Poor. Plasma can draw nearly 40 W of power. The display draws power linearly, so the more pixels lit, the higher the draw. In a typical screen, plasma will draw about 30 W. The worstcase power draw tops out at 50 W.

Cost

STN—Excellent. The STN is the lowest-priced flat panel display on the market today.

DST—Very good. The DST is only slightly higher in cost than the STN display.

TFT—Poor. The TFT is the highest-priced flat panel display on the market today. This can be attributed to its low production volume and the fact that it is not a mature technology.

EL—Good. EL displays cost about one-tenth of the price of TFT displays and about twice the price of the STN/DST displays.

Plasma—Fair to good. Plasma is generally priced equal to or marginally higher than the EL displays after factoring in the cost for the power converter.

CONCLUSIONS

A display that meets all the environmental and visual requirements for a locomotive ATCS application does not exist. Each of the display technologies has some limitations, yet in spite of these limitations, flat panel displays are considered the most suitable display screens because they are ruggedized and compact. A summary of all the visual characteristics, environmental specifications, and cost factors for flat panel displays is provided in bar chart format in Figures 1, 2, and 3.

The STN and DST LCDs are marginally qualified for use in locomotives. These displays, which can be read in sunlight, are rated only as poor to good for almost all of the other criteria. The temperature range is too narrow, the maximum humidity is too low, and the shock and vibration tolerance is

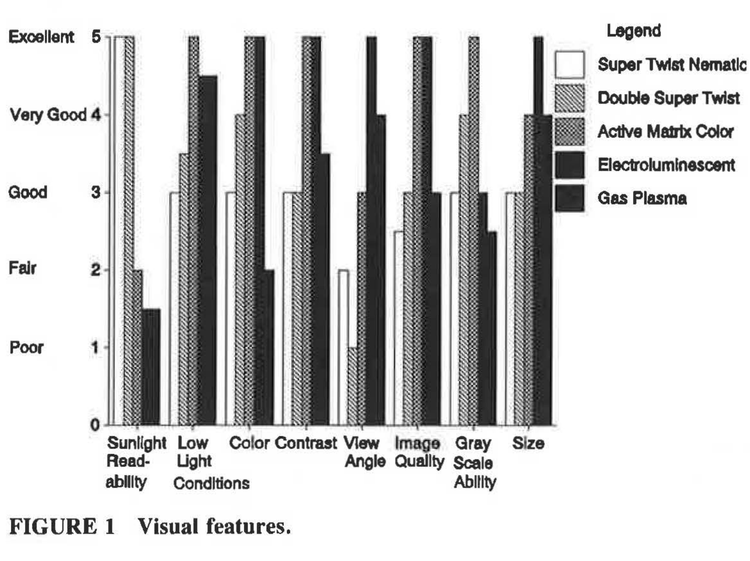


FIGURE 1 Visual features.

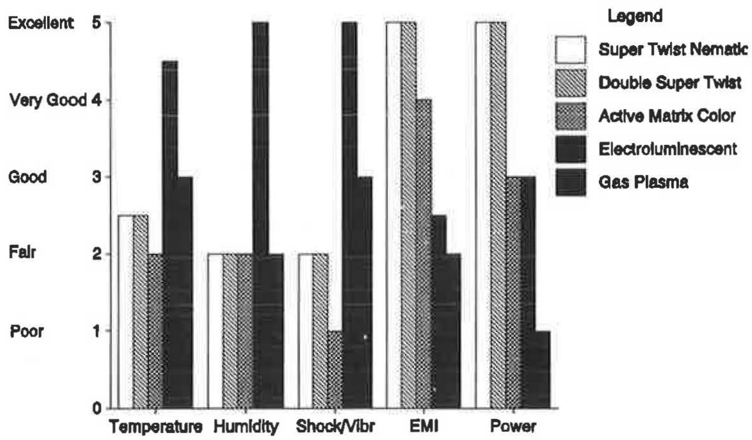


FIGURE 2 Environmental specifications.

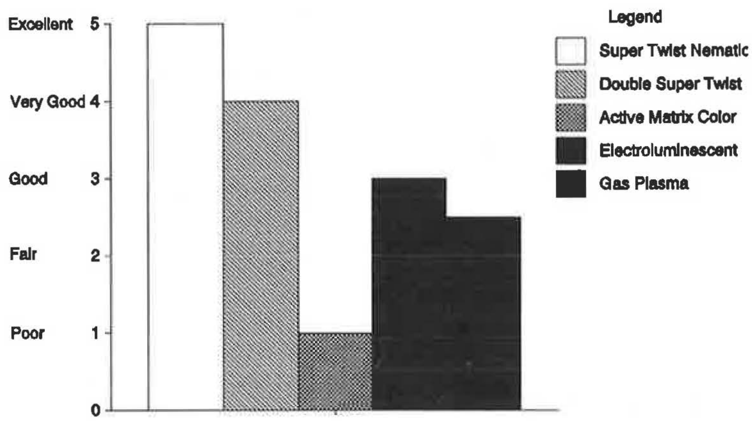


FIGURE 3 Cost.

below railroad industry specification requirements. The display does, however, meet sunlight readability requirements. If sunlight readability is given a higher weight than temperature, humidity, and shock and vibration, then these displays could be considered if properly packaged to reduce the effects from the environmental conditions found in a locomotive application. Also, these displays are only available as monochrome.

The TFT LCD dominates all visual specifications except sunlight readability and viewing angle. It is currently the only flat panel display that provides multicolor, and it produces a range of colors similar to the saturation levels of a color CRT. Simply put, it is beautiful. However, once one looks past the visual features and reviews the environmental specifications, it falls short of the mark. Much can be done in the packaging to support this display for shock and vibration, but the temperature and humidity problems are difficult to take care of. However, this display has been recently introduced into the flat panel display market, and as improvements are made in the ruggedization of this display it will become much easier to integrate into the harsh environment found in the railroad industry.

The EL technology is the most rugged of all the flat panel displays. Its shortcomings for EMI are manageable by using good packaging techniques and by selecting the right clock frequency. Other than EMI, it dominates all the remaining environmental specifications. Besides its high environmental ratings, it has received good to excellent ratings on its visual features. It does very well on the visual characteristics except that it only received a poor to fair rating for sunlight readability. The largest drawback to the EL technology is that it is not currently available in multicolor. Development is under way, but there is still difficulty in increasing the efficiency of the blue phosphor. Once this is accomplished, the EL technology will be able to provide a multicolor solution.

Finally, the plasma display's downfalls are its red color, sunlight readability, and the fact that it is completely overshadowed by the EL technology. There is simply no reason to use this display because the EL display is more rugged, provides better visual characteristics, and is priced lower.

In summary, the two top candidates for an ATCS-compatible flat panel display are the EL and the TFT. Both the displays have very good to excellent visual characteristics. The EL is the most rugged of the two, but it is not available in multicolor. On the other hand, the TFT is multicolor, but it is not nearly as rugged. As far as cost is concerned, the TFT is very expensive in comparison with the EL. All things considered, the choice comes down to the weight assigned to the multicolor capabilities, the environmental parameters, and cost differentials between these two flat panel displays.

Development of a Wayside Detector Open Communication Standard

HAROLD HARRISON

There is an emerging trend in the railroad industry to consolidate more wayside detectors at fewer installations. By combining several detectors and auxiliary devices, such as vehicle identification equipment (cameras and tag readers), considerable efficiencies are achievable, which, in turn, benefit the growth toward automatic train control systems (ATCS) integration. Given that the various devices are not generally available from a sole supplier, a standardized means of communicating among all devices is obviously needed. In building a framework for the development of such a standard, the primary goal is to separate information into relative groups by the nature of their respective time criticality, the quantity of information passed, and the relative capacity of each device involved to handle its task.

Most microprocessor-based detectors have at least some limited communication capacity via a serial port (usually RS-232–D, 8 bits, 1 stop, no parity). Because ASCII code is the most common form of information sent over serial communication lines, it makes sense to use ASCII as the basis for any message (whether command or data). ASCII characters also provide unique bit patterns that can be reserved for specific functions. Development and debugging are simple as well, using text editors and terminal emulators on personal computers. There are, however, fundamental limitations to this approach: the finite time necessary to transmit a message, and having sufficient "smarts" in any one particular device to communicate in this mode and to do it in a real-time manner.

There are several different criteria regarding timing and time delays or uncertainties that have some significance in defining "real time" as it applies to the wayside environment. For instance, for many activities, 1 msec is about the limit of "instantaneous response." This is equivalent to about 1 in. (25 mm) of uncertainty of position for train speeds of 60 mph (100 km/hr). One msec is probably the lower limit of timing control of separate devices and is coincidently the time it takes to transmit one ASCII character at 9600 baud. This arbitrary definition could be used as a starting point for several classes of activity:

1. Submillisecond timing—Presently this is beyond the scope of the proposed specification, except for dedicated control lines that could be set up between devices. (Salient has accommodated activity between its detector subsystems by using high-speed multidrop serial links and distributed reference timers that are synchronized to about 100μ sec.)
2. Millisecond timing—This is the practical lower limit of this basic approach using single ASCII character messages. Because addressing is not possible under these constraints, this mode would be restricted to one device on each end of serial link.
3. Near real time (multimillisecond)—Sufficient time is available to transfer more complete commands and small data messages. This timing uncertainty should be sufficiently tight to never misidentify specific axles. This mode is still too time restrictive to allow multidrop communication mode or receiver acknowledge.
4. Negative time—This class allows for events that cannot be handled in real time. For example, one detector may be located down the track so that its "real time" is much later than that of the device to which it talks. Obviously, this requires that the equipment receiving the message have buffering that accommodates the longest negative time specified.
5. Future time—Provision for preloading messages that apply at some point in the (near) future specified to 1 msec. This would allow two real-time devices to handle this communication at lower priority than that needed for immediate response. It further implies clock synchronization between the two devices, unless time is relative to transmission.
6. Posttrain activity—This is the worst-case condition of negative time that applies to most current detectors that may have only a posttrain mode of communicating over a serial link. In addition, many devices may pass final judgment on their respective alarms after reviewing their complete records. This indicates an important need for a cleanup phase of operation that may undo prior activity, as well as accommodate longer messages that would bog down the communication process during the time-critical period.

MESSAGE FORMATS

Borrowing from microcomputer instruction techniques, different classes of messages can be defined:

1. "Immediate" messages—In keeping with Type 2 timing above, there should be provision for single ASCII character messages. These should be generic in nature so that each detector can specify the exact response it expects. For reasons explained below, these single character instructions will be restricted to lowercase alpha, excluding "a,b,c,d,e,f". This provides 20 total immediate instructions (which can be reassigned for different interfaces).
2. Normal messages—These messages are four or six ASCII characters in length, and the last character is a checksum. These messages would be available for the near real-time activities, with no acknowledge expected. As detailed

Salient Systems, Inc., 4140 Tuller Road, Dublin, Ohio 43017.

below, these begin with an uppercase alpha followed by two or four data characters that are tentatively restricted to decimal or hex values (using lowercase "a" through "f" for upper hex numbers). Two leading opcode characters could be allowed at the expense of maximum data value.

3. "Extended" messages—This category allows a verbose exchange of information during the posttrain period using either begin and end braces "{,}" (printable ASCII) or "STX,ETX" (unprintable ASCII) to enclose the message string. The internal format of the string can be flexible depending on the mutual agreement of the parties involved. Ordinarily, it should have a header defining the type of message (and possibly message length), a body allowing most printable characters, and a tail with provision for a checksum.

The main purpose in configuring each type of message differently is to allow a "smart" device driver servicing the serial port to determine the type of action needed by interpreting the lead character of each message. In particular, a leading lowercase alpha could cause an immediate interrupt that, for instance, may not be desirable for an extended message.

DETAILED DISCUSSION

Real-Time/Immediate Messages

The greatest danger of applying the single-character messages to achieve the 1-msec response time lies in the compromised reliability. Not only is there no verification mechanism, but many times serial ports may behave less reliably on the transmission of a first or single character (due to flushing of buffers, etc.). The key issue is the trade-off between the actual reliability between two specific devices and the alternative of incorporating another interface that can meet the 1-msec timing limit. Increasing the RS232 baud rates to 19.2 or 38.4 kbaud would certainly be one possibility given that the detection devices are probably in the same or adjacent racks in the wayside enclosure. This would allow moving up from the "immediate" to the "normal" message type while still staying within the 1-msec response domain.

Another approach to meet the 1-msec criterion is to incorporate a faster interface. Many single-chip micros have high-speed ports for local (on-board) connections that can be interfaced to other standard serial interfaces such as RS-422, RS-423, and RS-485. These standards allow for an order of magnitude or more of a speed increase. Obviously, this approach works only if both devices are able to incorporate the higher-speed port.

Near Real-Time/Normal Messages

As summarized above, these messages would be four or six ASCII characters in length, with the last character being a checksum, and no acknowledge expected. Normal messages begin with an uppercase alpha followed by two or four data characters that are tentatively restricted to decimal or hex values (using lowercase "a" through "f" for upper hex numbers). There are plenty of opportunities for subsets in this category that implicitly use decimal or hex exclusively, as well as being implicitly four or six characters in length. Two leading opcode characters could be allowed at the expense of maximum data value.

The details of this message format are not particularly important up to the point that they become a de facto standard after someone has implemented them. The main point is that with a little care there can be some inherent redundancy in the message structure that would improve its reliability without overly taxing the link.

Some devices, such as automatic equipment identification (AEI) tag readers, operate within this area of near real time, currently without any benefit of standardization other than their own. Generally, this means that whatever device receives the tag reader's message string also must recognize the format and make its own assumption about the timing of the message arrival and the corresponding moment when the tag was correctly read. This happens to work reasonably well because there is a fair amount of uncertainty as to when the tag gets read within the field of its antenna. This is also a case of a relatively long ASCII string that must be treated as a near real-time event.

Posttrain/Extended Messages

This area of the proposed standard has the most flexibility, and coincidently it also has the most work already invested in it. Because most detection devices are ordinarily designed to stand alone, it is understandably easier for most devices to handle their own real-time jobs and then transfer the results to other devices after train passage.

Because of the added time available, the extended messages can afford to be (and are recommended to be) acknowledged by the receiving device. ASCII "ACK" and "NAK" characters are proposed. Additionally, busy and error conditions should be returned when appropriate.

The means of properly correlating events during train passage are usually related to time since train arrival (signaled between designated sensor device and all others needing the sync signal) and/or to the axle number associated with a particular event. Some devices can also sense car count or car position in the train, but the lower reliability of this capability suggests the added redundancy of time and axle count is desirable. Consistent with the rest of the standard, the reported time resolution is generally to 1 msec.

Abbreviated Form of Extended Message

For devices with limited ability to create elaborate messages because of code and data space limitations, a simplified version of the extended message is proposed. The message string would begin with two uppercase alpha opcode characters after the beginning brace or "STX". Any number of following data characters (up to some practical limit such as 72) would be followed by an end brace or "ETX" and two hex checksum characters. The data field can include any uppercase alphas for text and hex or decimal numbers. Space, tab, and basic punctuation characters are allowed for separation and visual convenience. In this case, the checksum is suggested to occur after the end brace. This allows simpler devices to strip the

checksum off without much difficulty if they are incapable of using the integrity check.

Given the limited capability of the devices that might use this version of the extended message, it is likely that message acknowledgement would not be incorporated here.

Note that regarding the choice between braces and STX, ETX characters, braces have the advantage of being printed to a terminal screen, thus allowing for ease of development and debugging with simple techniques.

One of the fringe benefits of the proposed standard is the potential for separate contractors to exchange simple text files in order to simulate each other's equipment.

Fully Developed Form of Extended Message

The appendix presents a more complete form of extended message. This message structure is consistent with the abilities of larger detection devices that have some reserve capacity to handle more complete communication sessions. The details of this message format represent the fruits of much effort on the part of the Video Masters staff after numerous sessions with Salient Systems' crew.

CONCLUSIONS

There is a demonstrated need for a set of ground rules to cover communications between various wayside detection devices. This paper is an attempt to define some of the inherent limitations that are present in the wayside detection process and with the current equipment available to perform these tasks. A set of guidelines were then outlined and the framework for a communication standard presented. It is hoped that this effort will provide an easy starting point from which to adopt a formal standard for acceptance by the railroad industry as a whole.

ACKNOWLEDGMENTS

The author wishes to thank the Burlington Northern and Union Pacific Railroads for providing the encouragement to resolve this issue. In addition, the staff at Video Masters, Servo, and Salient Systems have put in many long hours to hammer out many of the details and start putting them to use in the field.

APPENDIX
Extended Message Formats (Version 1)

General Format: [header] { message stream } [tail]

All data are expressed in ASCII. All transmitted hexadecimal data are the ASCII representations of the hexidecimal values.

Header Format: [header] = [STX] [version] [nchar]

field name	[STX]
size (char)	1
type	constant
description	02H
field name	[version]
size (char)	1
type	character
description	Provides interface version number for future expansion and compatibility, starting with "1".
field name	[nchar]
size (char)	4
type	hexchar
description	Total characters in message stream (hex count)
special values	**ffff** = test data (of undetermined length) in message, look for special tail indication for end of data.

Tail Format: [Tail] = [cksum switch] [cksum] [ETX]

special values **ffff,ETX** indicate end of data for text files.

field name	[cksum switch]
size (char)	2
type	hexchar
description	indicates that cksum field is active
special value	**ff** = cksum *does not* have to be checked.
field name	[cksum]
size (char)	2
type	hexchar
description	2's complement of the 8-bit sum of all the binary equivalents of ASCII characters in message stream
special value	**ff** when combined with **ff** above closes the undetermined length, text file declared in header.
field name	[ETX]
size (char)	1
type	constant
description	03H

General format of message streams:
{ message stream } = [mtype]{ message fields }

Field Descriptions found among different message types:

field name	[mtype]
size (char)	1
type	upper case alpha
description	Identifies message type, establishes record structure.
field name	[d]
size (char)	1
type	char
description	Delimiter to separate data within a record (e.g. comma, dash, space, or tab).
field name	[d2]
size (char)	2
type	char
description	Delimiter to separate data within a record (normally CR/LF characters).
type description	timestamp [MSB] [ ] [ ] [ ] [LSB] Data are sent in order, from MSB to

LSB. All values are the ASCII representations of the hexidecimal. Assume leading zero's to pad message.

Message 'T': Train Data

The train data consists of the following fields: [mtype][d2][date/time][d][TrnOD][dir][d]-[naxles][d] [alarms][selft][d][TotTime]

field name	[mtype]
size (char)	1
type	constant
description	'T'
field name	[date/time]
size (char)	10
type	decimal
description	date and time train arrived at site (24 hr. format)
field name	[TrnOD]
size (char)	2
type	hexchar
description	(train of day) train number after midnight
field name	[dir]
size (char)	1
type	char
description	1 = left to right; 0 = right to left.
field name	[naxles]
size (char)	3
type	hexchar
description	total number of axles in train
field name	[alarms]
size (char)	2
type	hexchar
description	alarm count (' ' = no alarms)
field name	[selft]
size (char)	2
type	char
description	2 characters indicating system status
field name	[TotTime]
size (char)	6
type	timestamp (hexidecimal)
description	total time (in msecs) TRAIN PRESENT signal was engaged for train.

Message 'A': Axle Data

The axle data consist of the following fields: [mtype] { axle records }

field name	[mtype]
size (char)	1
type	constant
description	'A'
record name	axle record
size (char)	12 per record, up to 1023 records
structure	[d2] [axle number] [d] [axle timestamp]
description	provides the timestamp for axle events
field name	[axle number]
size (char)	3
type	hexchar
description	assigns a number value to an axle event
field name	[axle timestamp]
size (char)	6
type	timestamp (hexidecimal)
description	Time of axle event since TRAIN PRESENT signal was engaged.

Message 'E': Event (or Exception) Data

The event data consist of the following fields: [mtype] { Event records }

field name	[mtype]
size (char)	1
type	constant
description	'E'
record name	Event record
structure	[d2][Event axle number][d][Event type] [d][Event label][d][event time]
size (char)	31 per record, up to approx. 100 records
description	Provides a description of significant events (typically, an alarming axle).
field name	[Event axle number]
size (char)	3
type	hexchar
description	Assigns an axle number value to the event or alarm.
field name	[Event type]
size (char)	1
type	char
description	Describes type of event or alarm.
field name	[Event label]
size (char)	16
type	character
description	A string of characters used to describe the event on the screen.
field name	[Event time]
size (char)	6
type	timestamp (hexidecimal)
description	Time of event since TRAIN PRESENT signal was engaged.

NOTE: This message may require further revision to allow more complete care information to be conveyed with the basic axle alarm data. The device creating the alarm may also be one reading the car tag or counting the cars.

Message 'S': Train Summary

The Train Summary data message provides a means for one detector system to send text data through another system for display at a remote location without the intermediate system taking an active part in processing or handling the information. This data message consists of the following fields:

[mtype] { summary text }

field name	[mtype]
size (char)	1

type	constant
description	'S'
field name	{ summary text }
size (char)	undetermined
type	character
description	Train data to be stored and displayed upon command from the remote destination.

Message 'D': Detail Train Data

The Detail Train data message provides a means for one detector system to send text data through another system for display at a remote location without the intermediate system taking an active part in processing or handling the information. This data message consists of the following fields:

[mtype] { detail text }

field name	[mtype]
size (char)	1
type	constant
description	'D'
field name	{ detail text }
size (char)	undetermined
type	character
description	Train data to be stored and displayed upon command from the remote destination.

Message 'N': No Train

The No Train command consists of one field:
[mtype]

field name	[mtype]
size (char)	1
type	constant
description	'N'

Message 'Q': Status Request

The Status Request command consists of one field:
[mtype]

field name	[mtype]
size (char)	1
type	constant
description	'Q'

Message 'L': Last Message

The Last Message command consists of one field:
[mtype]

field name	[mtype]
size (char)	1
type	constant
description	'L'

Improvement of ATCS Operational Safety and Efficiency Through Human Factors Applications

GIDEON BEN-YAACOV

Successful implementation of any advanced computer system depends on its usage. Users—in this case, enginemen—will not be enticed to use advanced train control systems because these systems have sophisticated software structures and complex hardware configurations. Rather, they will use the systems only if useful information can be obtained through simple man-machine interface (MMI) procedures. Human factoring methods can be used to develop MMI and related displays for the locomotive on-board terminal and to promote safe, efficient, and reliable train operation.

North American railroad companies, in a joint venture, have developed specifications for an advanced train control system (ATCS). These specifications emphasize standardization of ATCS structure and functions while allowing railroad companies and vendors to decide how various functions will be provided at the on-board terminals installed on locomotives.

Several railroad companies have developed subsystems and components with which they can test, on a limited scale, some ATCS functions. These railroad companies are satisfied with the results of their experimental equipment and now want to build pilot ATCSs. The pilot ATCSs are expected to have most or all of the functions and possibly some additional functions or features that individual railroad companies would like. A typical pilot ATCS would be installed on 50 to 100 locomotives and used over a few hundred miles of track.

The pilot ATCS's on-board terminal equipment and functions will probably be designed according to the North American railroad industry's ATCS Specification 320, Locomotive Displays and Control. This specification lists the information that should be available for display at the terminal's screen. The specification also identifies numerous functions—each of which can be requested through the terminal's keyboard. Furthermore, individual railroad companies will probably require additional information and functions, for example, operational instructions ("take siding" and "hold main"), locomotive health monitoring (locomotive load and brake pressure), and operational information (brake and throttle settings, time, direction, and distance to stop). The large amount of information expected to be available at the terminal and the many functions provided for selection at the terminal will pose a serious challenge to the designer—how to design an on-board terminal that is capable of displaying desired information presented in a useful format and how to design man-machine interface (MMI) procedures in a manner that makes them simple and easy to use.

Automated Monitoring and Control International, Inc., 11819 Miami Street, Omaha, Nebr. 68164.

On-board terminals are supposed to help enginemen operate trains. If an on-board terminal is difficult to use or has displays that do not provide useful information, or the information cannot be easily read, then enginemen will have difficulty using the terminals. In extreme cases, enginemen may be overwhelmed by the complexity of the MMI procedures or by the amount of irrelevant information shown. This could result in operational errors.

Useful information is available only when different displays that support different operational needs or different ATCS functions are provided. The format and content of each display must be geared toward a specific operational need. Each display must be quickly and easily accessible. For this to happen, simple and efficient MMI procedures are needed.

Proper design and successful implementation of the ATCS on-board terminals can be achieved only if the design is supported by suitable human factors activities.

Successful implementation of human factors activities will depend on the following:

- Use of a detailed human factors plan for the design, development, and evaluation of the enginemen's terminals;
- Use of suitable human factors guidelines;
- Careful evaluation of proposed design; and
- Use of human factors specialists who are experienced in design, design evaluation, and follow-up evaluation of MMI equipment and related displays for a variety of advanced computer systems.

The following sections include recommended human factors activities for the development of suitable human factors plans and guidelines and for the evaluation of the proposed design of the on-board terminal.

HUMAN FACTORS PLAN

Human factors tasks for a specific ATCS project are defined in an ATCS human factors plan document. These tasks are intended to do the following:

- Analyze engineman role, engineman information needs, and the new operational procedures,
- Develop a human factors guidelines document, and
- Evaluate a prototype terminal design and perform a long-term operational evaluation.

Figure 1 shows a typical outline of the scope and contents of an ATCS human factors plan. The following paragraphs describe the scope and contents of each section of the ATCS human factors document outlined in Figure 1.

Purpose. This section describes the purpose and scope of the proposed plan and the project for which the plan is designed.

Reference documents. This section identifies documents that support plan implementation.

Plan overview. This section describes the organization, schedule, resources, and responsibilities necessary to perform the ATCS human factors tasks.

Analysis efforts. This section describes the specific human factors analysis tasks for the ATCS project. The following types of analyses are performed:

- Definition of engineman role, including definition of data needed by the enginemen to support their duties, identification of data critical to train operation safety, and identification of data of secondary importance to train operation safety;
- Preparation of specifications, including descriptions of those functions used by the enginemen in a manner that shows the intended train operational procedures;
- Analysis of engineman informational needs, including identification of data items needed to support each major operational task, definition of groups of data items to be available for display at the terminal, and evaluation of the available backup information for use when the ATCS or the terminal fails; and
- Integration of operational procedures, including evaluation of the new train operational procedures proposed to support the ATCS system, evaluation of whether the new operational procedures are effectively integrated into the existing operational procedures, and verification that the proposed new operational procedures are accepted by the enginemen.

1. PURPOSE
2. REFERENCE DOCUMENTS
3. PLAN OVERVIEW
   - 3.1 Organization
   - 3.2 Schedule
   - 3.3 Resources
   - 3.4 Responsibilities
4. ANALYSIS EFFORTS
   - 4.1 Definition of Engineman Role
   - 4.2 Participation in Specification Preparation
   - 4.3 Analyzing Engineman's Informational Needs
   - 4.4 Evaluating New Operational Procedures
5. DESIGN EFFORTS
   - 5.1 Developing a Human Factors Guidelines Document
   - 5.2 Design Support
6. EVALUATION EFFORTS
   - 6.1 Design Evaluation
   - 6.2 Long Term Operational Evaluation
7. ADMINISTRATIVE PROCEDURES
   - 7.1 Deficiency Reporting and Resolution
   - 7.2 Configuration Management Procedures
   - 7.3 Standards, Practices, and Policies

FIGURE 1 Scope and contents of the human factors plan document.

Design efforts. This section provides a detailed plan of human factors tasks associated with the terminal design. The tasks are developing a human factors guidelines document and human factors design support.

Evaluation efforts. This section provides a detailed plan of the human factors evaluation tasks associated with the terminal development. These tasks are human factor design evaluation of a prototype terminal specifically installed for that purpose in a locomotive and long-term operational evaluation of terminals associated with a pilot ATCS installation.

Administrative procedures. This section describes the following human factors administrative procedures: deficiency reporting and resolution; configuration management procedures; and standards, practices, and policies.

HUMAN FACTORS ANALYSIS EFFORTS

Human factors analysis activities that are performed during the design of the on-board terminal's MMI equipment and displays are defined in a human factors plan document. There are four categories of these activities:

1. Definition of engineman's role. Definition of the engineman's responsibilities and the manner in which the new ATCS can support these responsibilities is an important part of the initial ATCS design activity. The engineman's train-operating responsibilities are identified concurrent with the crew's information needs. The following must be defined: train-operating crew's role and responsibilities with respect to ATCS use, type of information enginemen need to support their duties, information critical to train-operation safety, and information of secondary importance to train-operation safety.

2. Participation in the functional specification preparation. Functions used by enginemen should be described in terms of intended train operational procedures.

3. Engineman's interface analysis. MMI equipment and displays that are poorly designed increase the probability of engineman errors and thus increase the probability of negative operational consequences. To ensure effective user interface, the following types of analysis are suggested: definition of information that would be displayed on the locomotive's on-board terminal, identification of information that meets engineman-required or task-required needs, and evaluation of available back-up information when the ATCS or terminal fails.

4. Procedures integration. Integration of the ATCS's related operational procedures with established train operational procedures reduces the potential for engineman errors. Furthermore, procedure integration reduces training requirements and improves the probability of the enginemen's acceptance of ATCS. The following activities are to be performed: identifying movement authority and speed restriction procedures supported by the new terminals on board the locomotives, integrating the new ATCS operating procedures with the existing train operation procedures, and verifying that the proposed new operating procedures associated with the ATCS's on-board terminals are accepted, from a human factors point of view, by the enginemen who evaluate them.

HUMAN FACTORS GUIDELINES DOCUMENT

A human factors guidelines document for the on-board terminal's MMI equipment and related displays must be developed by human factors specialists. The guidelines should include a description of human factors principles and criteria. The guidelines should address areas where human factors principles should be taken into consideration when designing terminals. These guidelines will support the design and the design review process. Figure 2 shows an example of a table of contents for a human factors guideline document developed for the ATCS on-board terminal. As can be observed, this document addresses large numbers of human factors aspects associated with the terminals.

Human factors guidelines will be used in: (a) assessing what hardware will be needed to support the terminal; (b) designing display screen, keyboard, and data displays for the prototype terminal; (c) reviewing the terminal design specification document against human factors criteria; (d) reviewing the human engineering suitability of the prototype terminal installed on a locomotive; and (e) evaluating terminal usefulness based on long-term operational experience.

For the human factors guidelines to be effective in supporting the design review process, each human factor criterion should be supplemented by an evaluation questionnaire arranged in a checklist format. The questionnaire will help ascertain whether the aspects of human factors principles associated with each criterion have been adhered to. Figure 3 shows an example of a section taken from a human factors guidelines document. This example contains an evaluation questionnaire presented in a checklist format.

REVIEW PROCESS

The human factors review process consists of three phases: the design, the design evaluation, and the long-term evaluation. The review process, especially during the design and the design evaluation phases, is interactive. It provides an opportunity to resolve problems before the manufacture and installation of the pilot ATCS on-board terminals begin. The design and design evaluation phases are performed in a multipass process. (See the feedback pass in Figure 4.)

The first phase of the review process, the design phase, begins by defining the ATCS functions that will be provided at the enginemen's terminals. From this information, the onboard terminal's physical characteristics are specified and candidate displays are defined in sketches or drawings.

During the second phase of the review process, the design evaluation phase, the terminal design specification as well as a prototype terminal installed on a locomotive are evaluated for human engineering suitability. The evaluation proceeds from a review of the suitability of the on-board terminal's MMI procedures for simple and effective use; an evaluation of the display's formats and contents for compatibility with the enginemen's assigned duties; and a determination of the understanding of the displayed information by the enginemen to an empirical assessment of the effectiveness of the displays in helping the enginemen operate the train in a safe and efficient manner.

The third phase of the review process is the long-term operational evaluation of the on-board terminals installed on locomotives selected for the pilot ATCS. These terminals are evaluated by their initial operational experience.

The following sections provide additional details about the specific human factors activities recommended for each of the three phases of the review process.

Design Phase

During the design phase, candidate displays for the on-board terminal are defined, and the hardware components of the terminal's MMI equipment are listed. During this phase, display objectives and MMI procedures should be considered.

The specific human factors design activities are as follows:

- Recommending specific display format(s) and display contents. When designing a locomotive's displays, it is necessary to establish a format for data presentation that allows the enginemen to rapidly access the point to which they must direct their attention. The format can be textual or graphic. The display format should contain vital information permanently located in a specific area of the screen, thus allowing enginemen to rapidly view the train's primary or most pressing operational needs. Additional operational data may be presented outside this area. This additional data should supply detailed information and explanations for specific operational situations.
- Recommend MMI equipment layout and user procedures.

The display objectives should emphasize their required information content and the choice of optimal techniques to enhance information displayed (e.g., character size, inverse video, blinking, color, intensity, etc.).

MMI equipment must be specified in terms of its physical dimensions, performance characteristics, and environmental suitability. MMI procedures deal with the methods enginemen use to dialogue with the terminal, the on-board computer's response time, and ergonomic considerations.

The candidate displays and MMI equipment should be described in detail in the terminal design specification document. This document should include information about each of the proposed displays (e.g., the display purpose, list of applicable train operating data, and schematic illustration of the proposed display).

The on-board terminal design could be outlined by the railroad's engineering group, the railroad's operating group, or by the on-board terminal supplier's engineering group. The actual design efforts will probably be performed by members of the supplier's engineering group. When, and if needed, these individuals can be assisted by the railroad's engineering or operation personnel.

Design Evaluation Phase

The terminal design specification document should be reviewed against human factors criteria and design objectives

1.0 DATA ENTRY-KEYBOARD
- 1.1 Numeric Keys Arrangement
- 1.2 "Hard" Function Keys Arrangement
- 1.3 "Soft" Function Keys Arrangement
- 1.4 Key Dimensions
- 1.5 Labeling and Nomenclature
- 1.6 Key Displacement and Resistance
- 1.7 Positive Key Actuation Indication
- 1.8 Keyboard Slope
- 1.9 Separation of Keys
- 1.10 Key Type

2.0 TERMINAL'S RESPONSE TIME TO ENGINEMAN ENTRIES
- 2.1 Key Activation
- 2.2 Error Feedback
- 2.3 Response to Switch Status Inquiry
- 2.4 Request for Switch Control Function
- 2.5 Response to "Execute" of Switch Control Function

3.0 ENGINEMAN USER'S GUIDES
- 3.1 Terminal Operating Procedures
- 3.2 Train Operating Procedures

4.0 DISPLAY SCREENS
- 4.1 Display Readability
- 4.2 Reflected Glare
- 4.3 Screen Luminance and/or Contrast
- 4.4 Display Resolution
- 4.5 Refresh Rate
- 4.6 Display Controls
- 4.7 Flicker/Movement
- 4.8 Viewing Angle
- 4.9 Gray Scale/Color Ability
- 4.10 Image Quality
- 4.11 Color Quality
- 4.12 Low Light Conditions
- 4.13 Sunlight Readability

5.0 ENGINEMAN-DISPLAY RELATIONSHIP
- 5.1 Viewing Distance
- 5.2 Viewing Angle
- 5.3 Screen Location
- 5.4 Visibility of Data

6.0 DATA PRESENTATION FORMAT
- 6.1 Usefulness of Data
- 6.2 Lists
- 6.3 Tabular Information
- 6.4 Headings

7.0 SCREEN LAYOUT AND STRUCTURE
- 7.1 Location of Data Groups
- 7.2 Location of Instructions
- 7.3 Messages Area
- 7.4 Alarms Area
- 7.5 Screen Loading

8.0 MESSAGES AND WARNINGS
- 8.1 Types Needed
- 8.2 Error Messages
- 8.3 Feedback Messages
- 8.4 User Guide Messages
- 8.5 Message Contents
- 8.6 Visual Versus Auditory Presentation
- 8.7 Intensity and Frequency

9.0 GRAPHIC SYMBOLS AND HIGHLIGHTING
- 9.1 Use of Graphic Symbols
- 9.2 Use of Color
- 9.3 Use of Highlighting
- 9.4 Blinking of Symbols
- 9.5 Inverse Video

10.0 DISPLAY CONTENTS
- 10.1 Content Density
- 10.2 Content Integration
- 10.3 Format Orientation
- 10.4 Cognitive Fidelity
- 10.5 Information Needs
- 10.6 Display Layout and Contents

11.0 USER CHARACTERISTICS
- 11.1 Anthropometrics
- 11.2 Visual Activity and Color Perception
- 11.3 User Experience with Respect to Train Operation
- 11.4 Age (Cognitive Capacity, Visual System Degradation)
- 11.5 Time Available for Viewing the Displays (Work Load)
- 11.6 User Response Time

12.0 USER TASKS
- 12.1 Entering Data into the Terminal
- 12.2 Train Operation
- 12.3 Voice Communication
- 12.4 Train and Track Surveillance
- 12.5 Managing Problems

13.0 TRACK AND ENVIRONMENTAL FACTORS
- 13.1 Day and Night Operation
- 13.2 Sunlight Readability
- 13.3 Ambient Noise Levels
- 13.4 Existing Alarms and Warnings
- 13.5 Temperature
- 13.6 Humidity
- 13.7 Vibration
- 13.8 Abrasive Environments

14.0 LOCOMOTIVE CHARACTERISTICS
- 14.1 Available Mounting Locations (i.e., Console Cabs and Traditional Cabs)
- 14.2 Cab Seat Feature and Location
- 14.3 Location of Windows, Doors, Walkways in Cab
- 14.4 Foot Rests
- 14.5 Other Display and Controls Used by the Enginemen (e.g., Auxiliary Speedometer, Emergency Brake Valve, Radio, etc.)

FIGURE 2 Table of contents for a human factors guidelines document.

	COMPLIANCE CHECKLIST			
	N/A	YES	NO	Reference/ Comment
HUMAN FACTOR PRINCIPLE: Section 4.0: THE DISPLAY SCREEN - In advanced train control system, video display devices (for example, flat panel display screens) comprise the principle interface between the system and the enginemen. It is, therefore, important that the characteristics of the display device promote transfer of system data to enginemen. The quality of the displayed image must be consistent with enginemen needs.				
EVALUATION CRITERIA: Section 4.1: DISPLAY READABILITY - Alpha-numeric and graphic characters should be easily readable by the enginemen under all lighting conditions expected in a railroad environment.				
Section 4.2: REFLECTED GLARE - An anti-glare protection filter, a display hood, and/or display tilt and swivel feature should be installed to minimize or eliminate reflected glare at normal enginemen viewing angles.				
Section 4.3: SCREEN LUMINANCE AND/OR CONTRAST - (1) For an EL display, when adjusted to full brightness, the minimum pixel luminance should be 20 ft. lamberts. (2) For an LCD display, the contrast ratio for low and high ambient levels should be between 3:1 to 40:1.				
Section 4.4: DISPLAY RESOLUTION - Discrimination of fine details is a function of the number of scan lines or addressable points ("pixels") per unit length. (1) Video displays for displaying alpha-numeric text should have a minimum of 50 pixels per inch.				
(2) Video displays for displaying complex symbols and graphic details should have a minimum of 70 pixels per inch.				
Section 4.5 REFRESH RATE - The refresh rate for a particular video should be above the critical frequency at fusion so that the occurrence of disturbing flicker is not perceptible.				
Section 4.6: DISPLAY CONTROLS - Parameters such as a luminance (brightness), contrast, or color should be able to be adjusted by the user.				
Section 4.7: FLICKER/MOVEMENT - The display screen should have no discernible flicker and be free of any movement (change in position) throughout the specified range of vibration, shock, temperature, and humidity.				
Section 4.8: VIEWING ANGLE - The viewing angle of the display screen should be at least 30 degrees from the perpendicular.				
Section 4.9: GREY SCALE/COLOR ABILITY (1) Changing the intensity of the pixels provides for grey scale capability. A monochrome display screen should have at least 8 levels of grey scale. These grey scales should be distinguished from each other.				
(2) A color display screen should have at least 8 colors. These colors should be distinguished from each other.				
Section 4.10: IMAGE QUALITY - The display screen should have a good image quality, i.e. the pixel should be uniformly shaped, focused and have an excellent contrast ratio and viewing angle.				
Section 4.11: COLOR QUALITY - The monochrome display screen should have a color which is easy for the eye to see (for example, blue/black on white or amber on black).				
Section 4.12: LOW LIGHT CONDITIONS - The display screen should be able to produce a bright display under low light conditions.				
Section 4.13: SUNLIGHT READABILITY - The display screen should be capable of producing readable characters under direct sunlight (readable characters are defined as being capable of exhibiting a 3:1 contrast ratio in a 10,000 candle power ambient light).				

FIGURE 3 Example section taken from a human factors guidelines document.

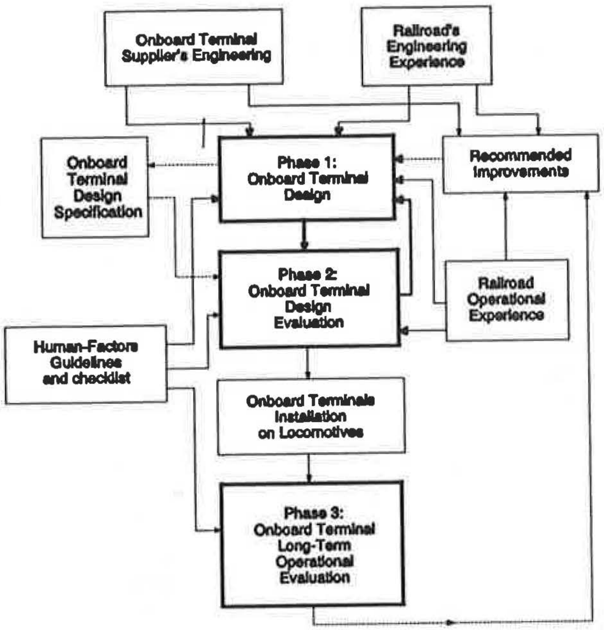


FIGURE 4 On-board terminal design review process.

(i.e., Have the operational needs been adequately addressed? Is the design practical from the enginemen's point of view? Are unforeseen problems in seemingly unrelated aspects of train operation produced by the information included in the enginemen's terminal?) Human factors guidelines and the evaluation checklist should be used to facilitate systematic and consistent evaluation of the proposed on-board terminal.

The design evaluation process should include a review of a prototype on-board terminal, preferably a terminal that was installed for this purpose on a locomotive. The on-board terminal could then be evaluated under "real" conditions.

At least four separate reviews of the on-board terminal design specification document and the prototype terminal should be performed. These reviews could be performed by the railroad's operational representatives. Enginemen, the ultimate users of ATCS, should comment on the proposed terminal's design and evaluate the MMI equipment and related displays by the simplicity of its procedures and by the usefulness of the information accessed, the railroad's engineering representatives, independent individuals from the supplier's engineering group (i.e., individuals other than those who created the terminal design specification document), and human factors specialists.

Design evaluation is vital to the development of an acceptable on-board terminal because it provides an opportunity to detect and correct deficiencies in the proposed design before the pilot on-board terminals are built and installed on locomotives. The design evaluation may indicate a need to modify the keyboard layout, keyboard arrangement, number of keys, screen size, screen type, character size, display contents, displays formats, terminal location, etc. If changes are made, the review process returns to the first phase—the design phase. When the terminal design specification document and the prototype terminal are successfully evaluated, manufacturing and installation of pilot terminals can begin.

The final version of the "as-built" pilot terminals and their displays should be documented in the final version of the terminal design specification document. This document should include a cover sheet that contains the approval signatures, the design specification, and the appropriate technical information showing all relevant data being presented at the on-board terminals.

Long-Term Operational Evaluation Phase

There should be more than one evaluation. The first evaluation should be done upon completion of the design phase (design evaluation). The second evaluation should be done following installation and initial operation of the pilot ATCS on-board terminals (long-term evaluation).

Long-term acceptance by the enginemen and usefulness in long-term routine operation cannot be addressed during the limited time available in the earlier design evaluation phase. The feedback loop from the "long-term operational evaluation" of the pilot system to "recommended improvements" (see Figure 4) indicates that long-term operational experience is part of the total on-board terminal evaluation. The long-term evaluation may precipitate changes that could then be implemented in a review process similar to the one used for the initial design and implementation of the on-board terminals. Change originators may be the railroad's operational group, the railroad's engineering group, or the supplier's engineering group. Change implementation will be performed by either the supplier or the railroad's engineering group.

CONCLUSION

During the terminal development phase, several different methodologies should be employed to define the human factors plan; to formulate design requirements; to evaluate terminals based on human factors criteria; and to resolve human factors deficiencies. Among those methods are these four:

- Use of existing human factors literature (this provides data on past research efforts);
- Use of experienced human factors and railroad industry specialists;
- Evaluation of the terminal's design specification document and a prototype terminal installed on a locomotive [this evaluation will be performed by human factors specialists and candidate users (enginemen)]; and
- Long-term operational evaluation of pilot ATCS terminals installed on locomotives (this evaluation will be conducted by human factors specialists, railroad industry specialists, and enginemen).

The performance of ATCSs and, in particular, their on-board terminals, will depend on matching terminal design with the capabilities and limitations of the human users (i.e., enginemen). Users' performance levels will be measured in terms of their ability to meet their task demand. The readability of displays, usefulness of information, ease of locating desired information, proper highlighting of emergency situations, and simplification of MMI procedures are all important perfor-

mance factors. If ATCS are to be viable operational tools, then task demand must match the ability of the enginemen. Human factors considerations influence users' ability to meet task demand. They are, therefore, an important part of the ATCS design.

Systematic evaluation of human factors issues in relation to ATCS requirements is essential in terminal design. When human factors are considered early in the design process, it is probable that design deficiencies will be corrected before the ATCS is operational. In contrast, when attention is focused only on some other components of the ATCS design (e.g., verification and validation and extensive testing) it is possible that operational problems could arise in human factors areas not considered.

SUMMARY

The human factors engineering process described in this paper is provided for the following:

- Planning human factors methods that support the design, design evaluation, and long-term operational evaluation phases;
- Use of human factors principles and criteria when designing the prototype on-board terminal;
- Evaluation of the prototype terminals based on human factors principles and criteria using a questionnaire arranged in a checklist format; and
- Use of evaluation results to correct human factors deficiencies.

Keyboard layout and key arrangement, display format and contents, legibility of the display device, and simplification of MMI procedures are all important design factors. Furthermore, the information shown at the on-board terminal's displays must be integrated into the train operation procedures.

The importance of human factors activities in ATCS design cannot be underestimated. Emphasis on effective on-board terminal equipment and displays is expected to increase partially because of the frequency and the severity of the accidents experienced within the railroad industry and partially because of the awareness within the railroad industry of the additional costs incurred as a result of human errors.

A human factors plan should be developed and then implemented during the design and development of on-board terminals. The plan should provide the framework for the application of human factors principles and criteria during the progressive stages of ATCS on-board terminal development. It should summarize present state-of-the-art methods of human factors planning, analysis, design, and evaluation efforts related to ATCS on-board terminal development.

The proposed methods for applying human factors principles should be geared to meet the specific needs of the railroad industry and should provide railroad management, consulting engineers, system suppliers, and the railroads' design teams with procedures and criteria that will help improve train operation safety and efficiency through human factors applications.

The human factoring knowledge used when designing on-board terminals is as important as that used by military, aerospace, and nuclear industries. Although the railroad industry can take pride in its improving safety record, the adverse public reaction to accidents caused by human error makes it essential that on-board terminals be optimized from a human factors point of view.

Performance and Capacity Analysis of an Operating ATCS Communications System

EDWARD L. FURMAN

Automated Monitoring and Control International has been working with Union Pacific Railroad to implement a communication network based on Advanced Train Control Systems (ATCS) Specification 200. This is the first large-scale implementation of the ATCS communications systems. Computer performance predictions and data from the installed system give insight into the capabilities of this type of mobile data network. Expected throughput of the network is estimated on the basis of analytical models. The successful large-scale implementation of a Specification 200 network on Union Pacific indicates that the ATCS specifications provide the basis for a viable communication network.

The use of mobile data technology is becoming more important in the transportation industry. In the railroad industry, business and safety systems have extended the reach of host computer networks and dispatch operations to the locomotive through the use of data radio communications. Business applications such as Automated Monitoring and Control International's (AMCI's) Work Order Reporting System (WORS) give railroads a distinct competitive advantage through improved productivity and customer service. WORS is designed to operate within the architecture established by North American railroads for the Advanced Train Control System (ATCS) (*1*).

Figure 1 shows the components required to complete an end-to-end data connection in ATCS. The dispatch computer determines an appropriate movement authority and creates ATCS messages to the locomotive. This movement authority is delivered and displayed to the engineman in the locomotive. The communication process begins as the dispatch computer delivers the message to the front-end processor (FEP), which then sends the message to the appropriate cluster controller (CC). The CC selects the optimum base communication package (BCP) to deliver the message to a mobile communication package (MCP) located on the locomotive. The BCP transmits the message to the MCP over a 900 MHz radio frequency (RF) path. When the message is received by the MCP it is then routed to the on-board computer (OBC), which drives a display that shows the movement authority to the engineman.

AMCI has been working with Union Pacific Railroad to implement a communication system based on ATCS Specification 200-compliant components. A locomotive in WORS communicates over the same network as an ATCS locomotive, but the messages are sent between a conductor's terminal on the locomotive and the railroad's management information system (MIS) host computer system. The FEP routes WORS traffic to the MIS host instead of the dispatch computer. On a locomotive in WORS the MCP connects directly to an on-board terminal (OBT) instead of the on-board computer. With the addition of such an on-board computer, a WORS locomotive can be expanded to full ATCS.

WORS is a system that enables crewmen to report car movement information directly into the railroad's central computer system from terminals on board the locomotives. Work is reported in a real-time manner rather than through after-the-fact clerical input. Train operations, such as train movement data, delivery and collection of cars, car status updates, and interchange of cars with connecting carriers, are entered at the conductor's on-board terminal. If the planned train work needs to be modified, the modifications are transmitted to the train. This approach avoids after-the-fact clerical input and manual data entry tasks, which in turn increases productivity and dramatically improves operational data reliability. A reduction in car-hire costs and improved customer service are additional benefits.

The communication network operates according to specifications established by the ATCS project of the Association of American Railroads (AAR) and Railway Association of Canada (RAC). These specifications are contained in the ATCS Specification 200 series. The RF link uses a pair of radio channels in the 900 MHz band. Each channel operates at a data rate of 4,800 bps. One channel (the inbound channel) receives data from the locomotives operating within the coverage of a base station. The second channel (the outbound channel) carries transmitted data from the base station to the locomotives.

Messages sent over the link are separated into individually addressed packets of data. Messages longer than 251 bytes require multiple packets for transmission. Two primary modes of packet transmission are used. Mode 1 requires an acknowledgment packet upon receipt of each packet. Mode 2 uses an acknowledgment or resend request for groups of 10 packets rather than for each packet. The last two packets of each group of 10 request acknowledgment.

WORS is the first large-scale implementation of ATCS technology in a revenue-producing environment. In terms of how this Specification 200 communication network performs, such evaluations are generally made by considering throughput and response time. The WORS network at its current load is not operating near the expected capacity. Throughput and response time under heavily loaded conditions have only been experienced during laboratory testing and through computer simulation of the network. But it is possible to estimate the expected throughput of the network and to review the

Automated Monitoring and Control International, Inc., 11819 Miami Street, Omaha, Nebr. 68164.

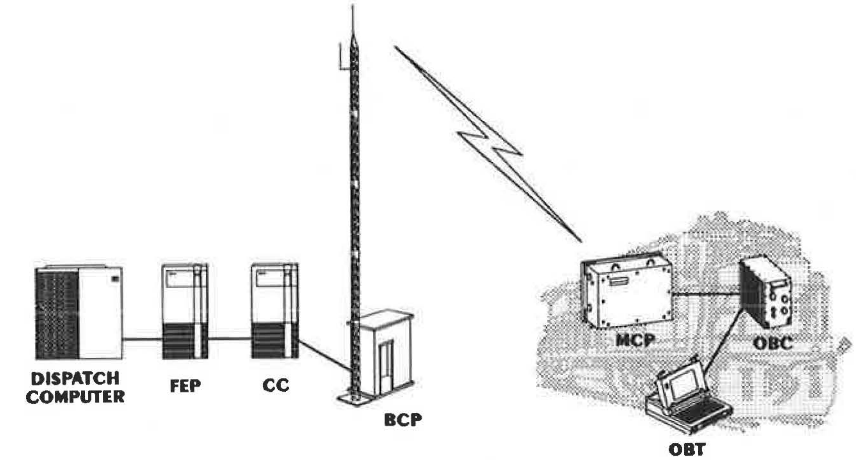


FIGURE 1 ATCS communication system.

current loading of the network and to examine mechanisms for adjusting the capacity of this type of network.

NETWORK CONFIGURATION

As of March 15, 1991, 67 base communication packages were installed at Union Pacific sites providing radio coverage in the states of Oregon, Idaho, Washington, Utah, Wyoming, and Nebraska. Figure 2 shows the Union Pacific Railroad and the areas currently using this ATCS-compliant data radio system. More than 500 locomotives are equipped with mobile communication packages. The communication network is controlled from a centralized fault-tolerant combination front-end processor/cluster controller. On average, 90 jobs use this network every day for WORS. The base communication packages are connected to the front-end processor/cluster controller on a point-to-point basis using the ATCS specification balanced HDLC protocol. Currently all the base communication packages are operating on ATCS Channel 1. The mobile communication packages are able to operate on any of the six ATCS channels.

The traffic currently on the network is of three types: basic operation of the network, WORS, and 3270 pass-through. All messages routed through the network conform to ATCS Specification 250. The messages result in packets of varying sizes. Outbound messages for WORS and 3270 pass-through utilize the extended ATCS packet length of 256 bytes. Most inbound messages in WORS and 3270 pass-through are less than 128 bytes and utilize the standard ATCS packet length of 128 bytes.

The number of locomotives using a base communication package in the operating WORS network was evaluated using the front-end processor/cluster controller's diagnostic function, "LIST DEVICE CONTACTS." This function produces a report for selected base communication packages showing a list of locomotive unit numbers and the time when each of those locomotives last used the base communication package. The reports are evaluated by considering a 1-hr window at and before the sample time and counting the number of locomotives reported during that time. This is based on the assumption that sometime during the hour all of these locomotives may have been using the base station concurrently. Based on these reports, as many as 29 locomotives are operating through a given base communication package.

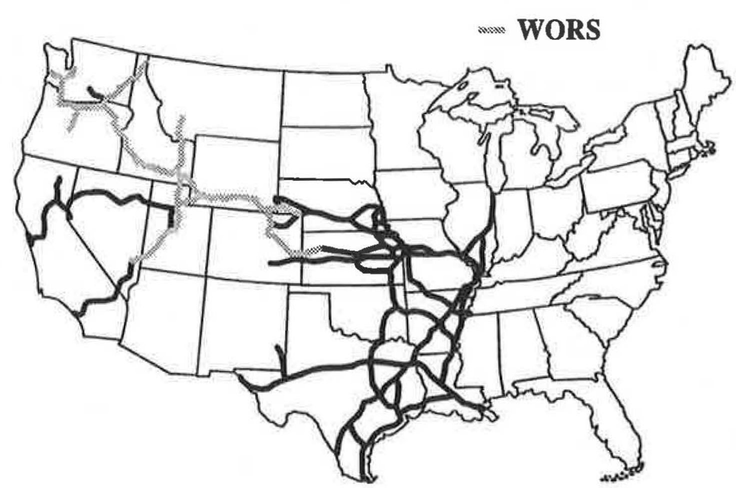


FIGURE 2 Union Pacific Railroad system map.

EXPECTED CAPACITY

RF Channel Capacity

The capacity or throughput capability of the network is primarily determined by the capacity of the RF link. The link must be further separated into the inbound and outbound directions because different performance considerations are involved for each direction of traffic. Capacity can be considered in packets or bytes. Because the available performance statistics from the front-end processor are in packets, most of the capacity discussion will be in terms of packets. This introduces a level of inaccuracy because the system uses packets ranging in size from 8 to 256 bytes. A lower limit for packet capacity can be estimated by using the largest packet size.

In the outbound direction, the capacity is determined by the data rate over the RF. Because a single base is used to reach multiple mobiles under the control of the front-end processor/cluster controller, it can be assumed that the base transmissions can be scheduled to utilize fully the available capacity of the base. Throughput will be estimated in packets per hour. For the largest packet size—256 bytes—approximately 5,000 packets can be sent from a base station in 1 hr. If smaller-size packets are transmitted then, of course, more packets per hour can be sent over the base.

The inbound channel is characterized by multiple mobiles trying to communicate with a single base. When multiple mobiles attempt to use the same channel more than one mobile may attempt to transmit at the same time. This can result in a "collision" and both mobiles will have to retransmit the packets. The inbound capacity is determined by the effectiveness of the mechanism used to minimize inbound collisions.

There are two possible modes of inbound channel operation, both of which are used in the network. One mode, further described below, is by "busy bit" control. The other mode, called ALOHA, is where mobiles simply access the channel and transmit a packet. If the packet fails, it is retried. From Figure 3, for $a = 1$, the inbound capacity of ALOHA for 256-byte packets is approximately 1,000 packets per hour.

The ATCS radio link uses digital sense multiple access (DSMA) to control inbound channel access (*2*). When a base station is transmitting data outbound, it will insert busy bits in the data stream upon detection of inbound traffic. The busy bits are monitored by mobiles wanting to use the channel.

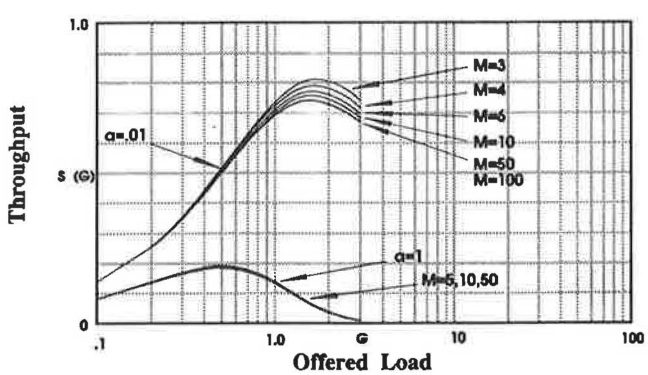


FIGURE 3 Normalized throughput for inbound channel.

When a mobile communication package checks the channel and finds that some or all of the last three busy bits were set to busy, it will wait a random time from 10 to 2,000 ms before checking the channel again. If the mobile finds the last three busy bits set to idle when it checks the channel, it will access the channel and transmit a packet. If a sync pattern is detected, the mobile will wait a random time from 10 to 800 ms and check the channel again. If less than three busy bits have been received since the sync pattern, the mobile will wait a random delay of 10 to 50 ms and check the channel again.

AMCI developed an analytical model to predict the capacity of the ATCS inbound channel (*3*). Figure 3 presents the performance prediction of this model in the form of normalized throughput curves. This graph considers throughput in the manner used by Kleinrock and Tobagi (*4*) and assumes that all the packets are the same length. This method of looking at the inbound channel removes the throughput differences related to packet length by normalizing all time to the time required to transmit a packet.

The y-axis is the throughput normalized by dividing the number of packets that got through by the time required to transmit one packet. The x-axis is the load presented to the channel in packets, normalized by dividing by the time required to transmit one packet. The throughput is a function of the parameter a, which represents the portion of time when each mobile has no information about other inbound traffic on the link. When $a = 1$, mobiles have no information about other traffic and access the inbound channel randomly. This results in the lower throughput curves in Figure 3, which represent the well-known performance of ALOHA channel access (*4*). For a carrier sense multiple access (CSMA) system, a is primarily a function of propagation delay. In the AMCI model for the DSMA inbound channel used by ATCS, the parameter a is determined by modem bit synchronization time and busy bit insertion time in addition to propagation delay and other factors. Figure 3 shows the changing throughput characteristics for various numbers of mobiles (M) for an assumed a of 0.01. The curves in Figure 3 show that the throughput of the channel increases as more load is offered to the channel. For very high offered loads the throughput will decrease as more collisions occur. The family of curves also shows that with fewer mobiles there are fewer collisions and greater throughput is possible.

For any given packet length the normalized throughput curve can be converted to a packet throughput for a specific time period. Assuming a 256-byte packet, a lower limit of inbound capacity of a base station is estimated from Figure 3 to be approximately 4,200 packets per hour.

Required Capacity for ATCS

The expected operating scenario for a railroad's ATCS implementation determines the characteristics of the required data traffic. The key parameters of this traffic are not only data volume but also the geographic location of the mobiles sending and receiving the traffic as well as the distribution of wayside and other devices in the system.

The RF channel requirements for ATCS were estimated by Dr. Sheikh of Lapp Hancock (memorandum from R. Ayers to Component Specification Drafting Committee members,

Oct. 9, 1987). This requirement estimate was in the form of multiple curves produced for various situations. As an example from this estimate, three radio channels were sufficient to support the following load:

- 20 trains,
- Location reporting every 60 sec,
- 70 data base downloads per hour,
- 350 wayside units, and
- Status updates every 90 sec.

Many of the control flows and operating assumptions of ATCS are still under development, so the exact communication requirements are not known at this time.

RF Capacity Expansion

The amount of traffic that a base communication package receives is determined by the number of locomotives and other devices within its coverage area. The traffic load on the system further increases when base communication packages are located to provide overlapping coverage. Generally, the larger the coverage area of a base communication package, the more locomotives and other devices it will be communicating with. The coverage area of a base station can be controlled by antenna selection and mounting height.

When a base communication package is installed at a specific location with an antenna of known pattern at a given height, a coverage area can be calculated. Each base communication package has a known capacity for inbound and outbound traffic and will provide this capacity for all the trains, wayside units, and other RF users within its coverage area. If the traffic demand within the coverage area is expected to exceed the capacity of the base station, the network capacity can be expanded by installing additional base communication packages on other RF channels. Six channels have been licensed to the AAR for ATCS. If more capacity is required in an area than can be provided by base communication packages on six channels, the network capacity can be expanded by reducing the coverage area of each base station and installing additional base communication packages with smaller coverage areas.

Known values of transmitter power and receiver noise figure are used to compute a carrier-to-noise ratio (C/No) for RF data reception. AMCI uses a computer model in conjunction with digitized topographical data to calculate contours where the C/No is acceptable for a given level of service. This predicted coverage area identifies the physical locations where users can expect to obtain normal performance from the data radio network. Field tests have established the accuracy of the coverage prediction (*5*). Coverage analysis is used in conjunction with railroad operating plans to determine the traffic volume expected for a base station. Base station locations and antenna configurations are selected for acceptable system loading and to provide the required level of service to mobiles.

An ATCS communication network can be engineered to meet the capacity required by the railroad. The areas in coverage and the capacity of the RF network are controlled by engineering the coverage of base communication packages. The remaining components in the network can be expanded as needed to support the base stations installed.

SYSTEM PERFORMANCE

Total Traffic Volume

A macro view of the WORS ATCS Specification 200 network in operation is available from a statistical report produced by the front-end processor/cluster controller. This report counts the packets received and transmitted for each hour by each base communication package. Data from this report for March 15, 1991, a randomly selected day, were used to graph the inbound and outbound traffic per base station per hour. The data from each base station are used to calculate the average and standard deviation of packets per hour during each hour of the day. The maximum traffic volume in packets for any base communication package for each hour is also graphed.

The traffic for each base communication package was added during each hour to determine the total traffic volume through the network. During the heaviest hour for total inbound traffic, Hour 0, the front-end processor/cluster controller handled 28,995 packets. The heaviest hour for outbound traffic was Hour 20 when 18,511 packets were routed by the controller. The graph in Figure 4 shows inbound and outbound total traffic by hour.

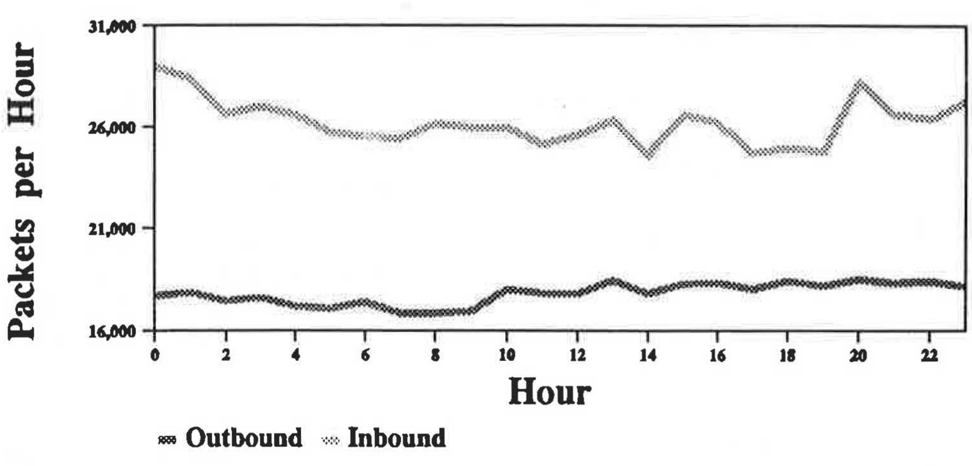


FIGURE 4 Total traffic.

Outbound traffic from each base station was a fairly constant 240 packets per hour. The maximum outbound traffic was in Hour 6 from Rocky Point, Oregon (near Portland), when 693 packets were transmitted. The second highest figure was for the same hour through North Platte, Nebraska, when 568 packets were transmitted. Figure 5 shows the average, standard deviation, and the maximum packets per hour outbound for each hour of the day.

The variation of inbound traffic was much greater than outbound. The busiest inbound base station was Trail Creek, Idaho (near Pocatello), during Hour 23 with 2,336 packets received. The second busiest was Black Mountain, Idaho (near Bonners Ferry), during Hour 20 when 2,238 packets were received. Figure 6 shows the average, standard deviation, and the maximum packets per hour inbound for each hour of the day.

The average outbound traffic volume of 240 packets per hour shown in Figure 5 is easily explained. ATCS specifications require each base communication package to transmit a cluster controller ID every 15 sec if there is no other traffic. This results in a baseline outbound load of 240 packets per hour.

The inbound traffic volume shown in Figure 6 varies greatly because locomotives move between base stations and base stations that provide coverage for yards have more locomotives within their coverage area. The inbound traffic is primarily from packets that do not directly support WORS or 3270 pass-through. ATCS specifications require each mobile communication package to transmit once every 5 min if it has no other traffic. This results in 12 packets per hour from every locomotive equipped with a mobile communication package. On-board terminals in WORS send a "WAMI" message every 2 min while in RF coverage, which results in 90 packets per hour from each locomotive with a terminal installed. These automatic messages are only sent if there is no other traffic during the time interval.

Traffic for a Job

The traffic to and from a "typical" WORS job is determined by monitoring randomly selected jobs. The AMCI front-end processor provides a function called TRACE, which monitors traffic to and from a selected locomotive. TRACE is started on a specific locomotive to record the traffic to and from that

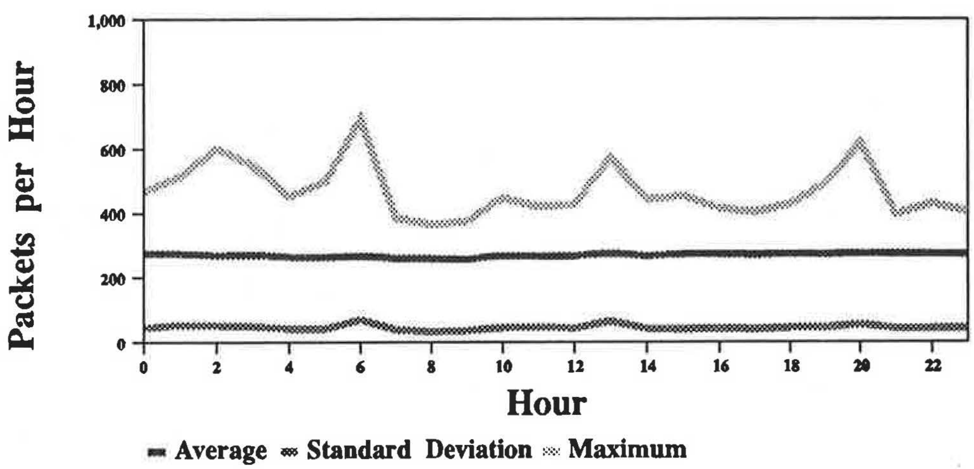


FIGURE 5 Outbound traffic per base communication package.

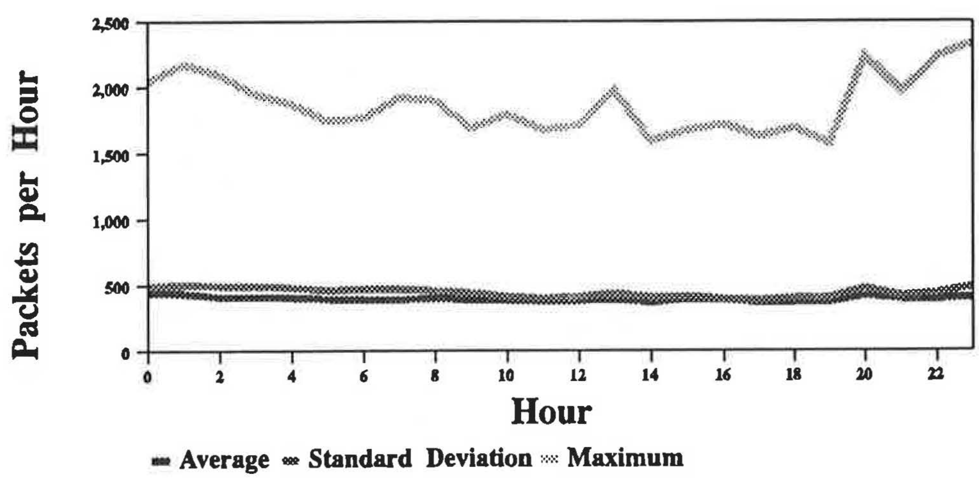


FIGURE 6 Inbound traffic per base communication package.

locomotive. The TRACE file is then converted to dBase for further analysis. This permits locomotive traffic to be processed and counted. When multiple packets are received at the front-end processor/cluster controller due to overlapping RF coverage, duplicate packets are removed when counting mobile packets per hour. Figure 7 summarizes the inbound traffic characteristics from 15 jobs. Figure 8 summarizes the outbound traffic characteristics from the same 15 jobs. The graphs show the minimum, maximum, and average packets per hour from the sampled jobs.

ANALYSIS

The average WORS traffic outbound for each base station uses less than 5 percent of the capacity of the base communication package. The busiest base used only 12 percent of the available outbound capacity. The heaviest load is on the inbound RF channel. The average inbound load is 500 packets per hour, conservatively 12 percent of the base communication packages' minimum inbound capacity of 4,200 packets per hour based on 256-byte packets. The maximum inbound traffic volume in an hour was 2,336 packets. Even during peak loads for WORS, almost half of the capacity of each base station is still available for operation of ATCS. Smaller-sized packets will result in a higher capacity in packets per hour. Because the actual traffic is a mixture of small and large packets, this analysis of capacity is conservative.

The equipment in an ATCS communication system automatically generates traffic to verify that it is in communication with the central system. Base stations send a packet every 15 sec. On-board terminals originate a message every 2 min. At the present time, this type of traffic represents the majority of the load on the network. When "real" messages are sent during a time interval, the automatic message is not sent. This automatically generated traffic establishes a minimum traffic level for the network. Initial application traffic merely replaces the automatically generated traffic and does not result

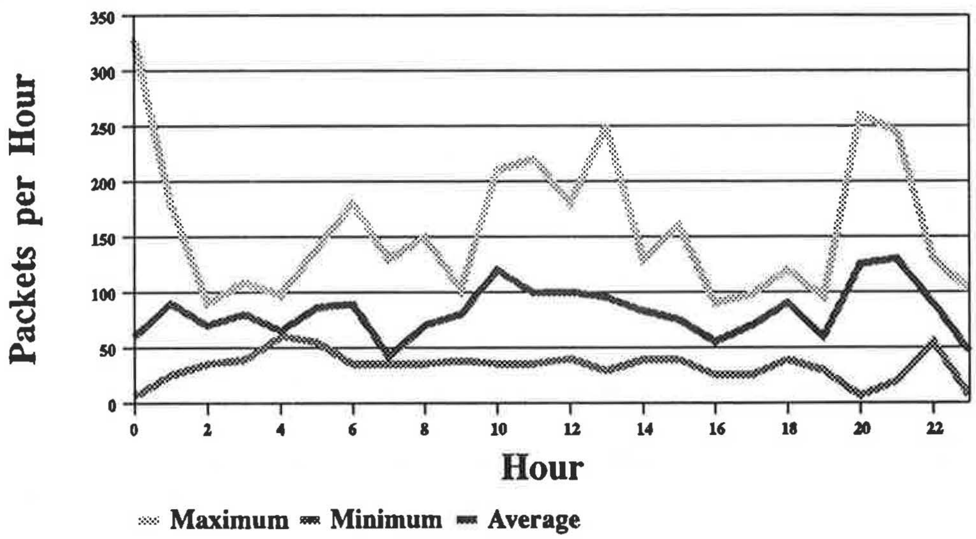


FIGURE 7 Inbound traffic per job.

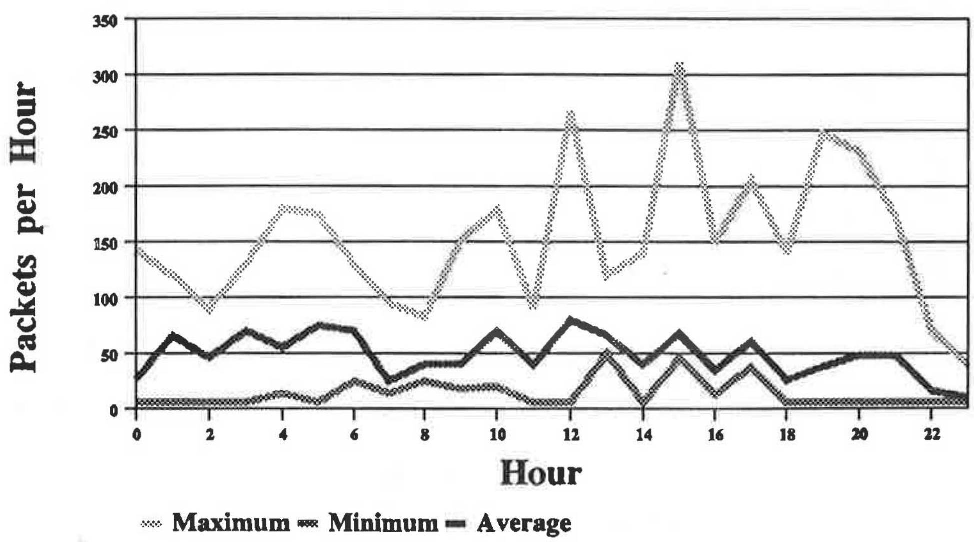


FIGURE 8 Outbound traffic per job.

in much increase in network traffic. Beyond this minimum traffic level, additional application traffic will result in increased loads on the network.

The successful large-scale implementation of Specification 200 on Union Pacific indicates that the ATCS specifications provide the basis for a viable communication network. A Specification 200 communication system can provide reliable communications for a variety of applications including ATCS, work order reporting, locomotive health monitoring, code line replacement, and maintenance-of-way management. The RF network can be engineered through site selection and use of multiple channels to provide the needed capacity for a railroad's data communication requirements.

The ATCS communication system is designed to support other applications in addition to train control. To ensure that critical traffic and low-priority traffic can share the same communication system, a priority value is assigned to each message. The priority determines the retry interval for the message and causes the message to pass lower-priority messages in any queue. The ATCS protocol suite establishes eight priority levels with two subpriority levels (RF ARQ mechanism disabled or enabled) per priority for a total of 16 "channel groups." The eight priority levels permit traffic to move through the network according to its relative importance. Packets for emergency messages have the highest priority. Packets involving train control have a higher priority than WORS packets or data base download packets.

A network handling multiple types of traffic will be engineered so that the delay requirements established in Section 3.2.1.1 of Specification 200 are met. According to these guidelines, the delay for 99 percent of emergency traffic between vehicle and wayside will be 4 sec or less, and dispatch system to a vehicle will be 10 sec or less. Normal traffic will have an average delay of 30 sec and a 99 percent delay of 225 sec. Operational traffic will have an average delay of 10 sec and a 99 percent delay of 75 sec.

SUMMARY

AMCI has been working with Union Pacific Railroad to implement a communication network based on ATCS Specification 200. This paper presents an estimate of the expected capacity (throughput) of the network based on analytical models. For convenience, throughput is estimated in packets per hour. For the largest packet size, 256 bytes, approximately 5,600 packets can be sent from a base station in 1 hr. Assuming a 256-byte packet, a lower limit of inbound capacity of a base station is estimated to be approximately 4,200 packets per hour.

The current loading level of the installed network is compared with the expected capacity. As many as 29 locomotives are operating through a given base communication package. The maximum outbound traffic from a base station was Rocky Point, Oregon, when 693 packets were transmitted during 1 hr. The busiest inbound base communication package was Trail Creek, Idaho, with 2,336 packets received in 1 hr. The average WORS traffic outbound for each base station uses less than 5 percent of the capacity. The average WORS inbound load uses, conservatively, 12 percent of the inbound capacity.

A Specification 200 communication system can provide reliable communications for a variety of applications including ATCS, work order reporting, locomotive health monitoring, code line replacement, maintenance-of-way management, and other future applications. To ensure that critical traffic and low-priority traffic can share the same communications system, a priority value is assigned to each message. The eight priority levels permit traffic to move through the network according to its relative importance. The RF network can be engineered through site selection and use of multiple channels to provide the needed capacity for a railroad's data communication requirements.

The successful large scale implementation of a Specification 200 network on Union Pacific indicates that the ATCS specifications provide the basis for a viable system capable of supporting ATCS, WORS, locomotive health, and other applications.

REFERENCES

1. Coll, Sheikh, Ayers, and Baily. The Communication System Architecture of the North American Advanced Train Control System. *IEEE Transactions on Vehicular Technology,* Vol. 39, No. 3, August 1990.
2. Sheikh, Coll, Ayers, and Baily. ATCS: Advanced Train Control System Radio Data Link Design Considerations. *IEEE Transactions on Vehicular Technology,* Vol. 39, No. 3, August 1990.
3. Sharif and Furman. Analytical Model for ATCS Inbound RF Channel Throughput. *Proceedings of the 1991 IEEE Vehicular Technology Conference.*
4. Kleinrock and Tobagi. Packet Switching in Radio Channels: Part I—Carrier Sense Multiple-Access Modes and Their Throughput Delay Characteristics. *IEEE Transactions on Communications,* Vol. COM-23, December 1975.
5. Mautschke, Furman, and Decker. Mobile Data Transmission in a Railroad Environment. *Proceedings of the 1990 IEEE Vehicular Technology Conference.*

Railroad Operation Using the Advanced Train Control System

David A. Poltorak and John H. Bailey

The basic approach taken in the development of the Advanced Train Control System (ATCS) has been that railroad operations can be made safer and more efficient by applying modern command, control, and communications technology. By using precise speed and location information, the system is able to provide more timely and precise traffic information than traditional train control systems. This allows ATCS to benefit from "moving block" operation, where the separation necessary for safe operation is dynamically determined from traffic levels and capabilities of trains (e.g., braking distances and location updates). Smaller train separation is made possible with a resultant increase in line capacity. Major design elements employed by ATCS are the data communications system and information-processing nodes that reside at the central dispatch office, on board locomotives, on board work vehicles, and in field devices. The use of these elements will have a significant impact on the manner in which dispatchers, enginemen, and foremen conduct their daily operations. These elements also provide for numerous applications besides train control.

Insight is provided into railroad operation with the Advanced Train Control System (ATCS). The basic architecture and operation of ATCS as currently specified are described. Alternate applications of ATCS will also be explored. Those who wish to investigate the operations and operating logic of ATCS more thoroughly are referred to *ATCS Concept of Operations* issued by ARINC Research Corporation in 1991.

The ATCS concepts had their origin in Canada. In the late 1970s, several Canadian railroads, including the Canadian National, British Columbia, and Canadian Pacific, studied the potential for using technologically advanced computer and communications systems to provide a new system of train control. In the fall of 1983, the Canadian railroads were joined by several United States railroads, among them the Burlington Northern, Norfolk Southern, and Seaboard System (now a part of CSX Rail Transportation), soon followed by the Union Pacific and Southern Pacific. By 1984, through an agreement with the Association of American Railroads and the Railway Association of Canada, a central project office was established.

In 1985, ARINC Research Corporation was retained to design the system architecture with oversight provided by railroad officials. Railroad operating officers, equipment suppliers, and ARINC then participated in component specification drafting committees to discuss requirements and architectural elements. By 1988, the ATCS specifications were written and distributed to interested parties. These specifications are dynamic documents that will evolve with technology and railroad implementation of ATCS. The ATCS project is currently managed by the Association of American Railroads with oversight provided by a railroad industry steering committee and an executive committee.

The overall architecture of ATCS, basic railroad operation with ATCS, and potential spin-offs from the ATCS architecture are discussed. The paper concludes with some challenges for the application of advanced technology to rail operations.

SYSTEM DESCRIPTION

ATCS has five major elements. Four of these elements are the various information-processing nodes that reside at the central dispatch office (the central dispatch computer), on board locomotives (the on-board computer), on board work vehicles (the track forces terminal), and in the field (the wayside interface unit). These nodes are designed to replace most of the voice communications that are required in today's operations and to determine vehicle location and speed and the status of wayside devices. The nodes collect, process, and distribute data with minimal input from dispatchers, enginemen, and foremen.

Dispatchers are constantly updated by computers on what is happening on the railroad. Enginemen and foremen are constantly updated by their computers on what their vehicles are doing and are prompted to take required actions. The fifth element, and system keystone, is the modern data communications network that ties the various information-processing nodes together (Figure 1).

Data Communications System

The ATCS data communications system is responsible for the interchange of data between the dispatch system, locomotives, gangs, and field devices. The communications system consists of three major pieces: a communications handler, which behaves like a telephone exchange in that it routes messages to the appropriate point; a set of radio base stations; and mobile data radios (Figure 2).

The data communications system (in conjunction with the information-processing nodes) significantly reduces the need for voice communications. This eliminates the time and effort required to raise someone on the radio, read the information, and write the information down—all of which can be very time consuming. Instead, once orders and requests are entered into computers, the users can move on to other activities. The majority of information flows to and from the cen-

ARINC Research Corp., MS 5-371, 2551 Riva Road, Annapolis, Md. 21401.

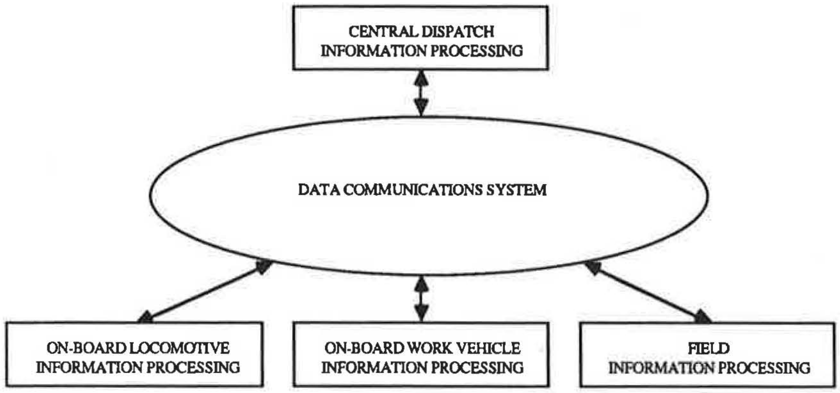


FIGURE 1 A modern data communications system ties the various information-processing nodes together.

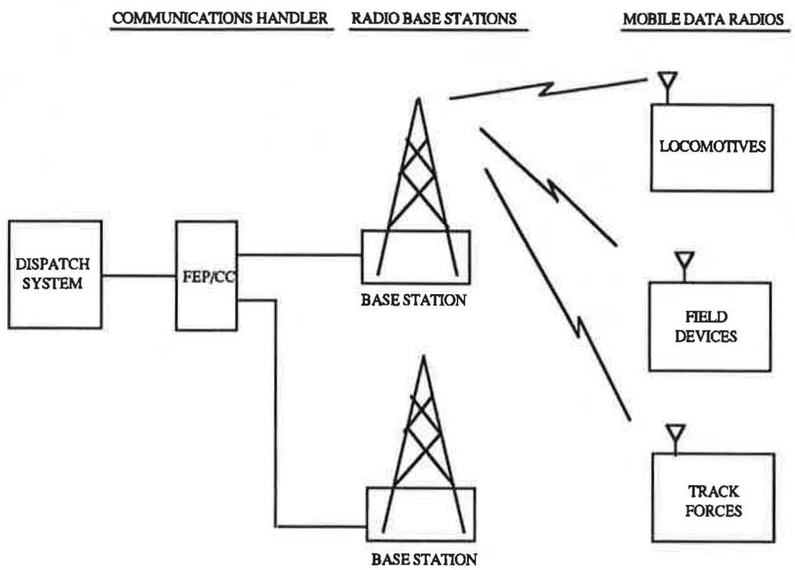


FIGURE 2 ATCS data communications system.

tral dispatch system, although the data communications system is not limited to such transfers.

Dispatch System

The function of the dispatch system is to manage the movement of trains throughout the rail network with the objective of guaranteeing safe operations without incurring train delays. The central dispatch system determines whether a particular movement is safe by checking for human and machine errors.

Key elements of the ATCS dispatch system are the dispatching consoles, the management system, the safety system, and the corporate management information system (MIS) interface. Dispatching consoles provide the interface between the dispatchers and the rest of the ATCS dispatch system. Consoles are not part of the ATCS design, although certain information that is exchanged between the consoles and the remainder of the dispatch system is specified in the design.

The management system handles all the operations planning aspects of dispatching that do not directly affect safety. This element of the dispatch system does not need to be implemented identically on each railroad to ensure interoperability.

The safety system handles all safety-related aspects of dispatching and is functionally standardized. Although the exact design of the hardware for the central safety system is not specified, the system must be capable of checking for errors and must deal with detected errors in a safe manner. The reaction of the system to such errors ranges from refusal to grant an authority (in the case of a dispatcher requesting an unsafe authority) to shutting down the dispatch computer and reverting to manual operation in the case of a hardware failure. In this fashion, the system has been designed to handle faults gracefully.

The final element in the ATCS dispatch system is an interface to various MISs that might be employed by the railroads to coordinate schedule tracking, work order reporting, locomotive utilization, maintenance planning, crew scheduling, or other applications. The form of the data interchange between the dispatch system and MISs is not part of the ATCS

design, because the form and existence of these data systems is railroad-specific.

Locomotive Systems

The function of the locomotive system is to provide automatic location tracking and reporting, predictive enforcement, and automated transmission of movement authorities, and switch monitor and control information via the data communications system. A mobile communications package, an on-board computer, and a display are the only equipment requirements for automated transmission of movement authorities and switch monitor and control information. The mobile communications package, which is used to transmit and receive information, is a mobile data radio.

Automatic location tracking and reporting is accomplished with the use of locomotive-mounted interrogators and track bed transponders. Interrogators activate the passive transponders by emitting a signal that is coupled to the transponders. Each transponder contains unique, coded identification information that is transmitted back to the locomotive interrogator while the interrogator and transponder are still near each other. The identification information is sent to the on-board computer where the coded identification information is matched to an exact location. The on-board computer then communicates this location information to other ATCS users such as the dispatch system. Odometer signals are used by the on-board computer to interpolate between transponders. Odometer error is zeroed each time a transponder is encountered.

Predictive enforcement ensures train compliance with speed and authority limits. Information provided by brake, throttle, and axle sensors is fed to the on-board computer to calculate braking distance and eventual speed.

Work Vehicle Systems

The primary function of the work vehicle system is to provide the capability for a track maintenance foreman to communicate with the central dispatch system and other vehicles via the data communications system. The system is designed to support various data input and output functions, including location tracking and reporting; switch monitoring and control; requesting, securing, and releasing authority for track occupancy; requesting and receiving slow orders for designated mileage limits; transmitting slow order requirements to the central dispatch system; transmitting track work reporting to the central dispatch system; releasing slow orders; and requesting and receiving advisories. A mobile communications package, a track forces terminal (a scaled-down on-board computer) and a display are required to provide these functions.

Field System

The primary function of the ATCS field system is to provide remote monitor and control of wayside devices. Communication between the field system and the dispatch system is conducted via the data communications system as well as land lines (e.g., pole line, telephone company line, fiber-optic line). In those cases in which devices are not able to communicate directly with the dispatch system, information is relayed to the dispatch system by locomotives. Locomotives and work vehicles also communicate with the field system for their own purposes.

An ATCS field system is made up of a wayside interface unit and a mobile communications package. The wayside unit provides the interface between existing field devices and the mobile communications package. Many field devices can be connected to the ATCS communications network via a single wayside unit and mobile communication package.

Within ATCS, field devices are classified as either controllable or noncontrollable. Controllable devices are all field devices that can change a route, either by providing alternate paths or by physically interrupting the path. These devices include power-operated switches, hand-operated switches, movable bridges, and railway crossings at grade. If the devices are appropriately equipped, they can be controlled by ATCS (i.e., from locomotives, work vehicles, or central dispatch). If the devices are not equipped for remote control they can at least be equipped for remote monitoring, which provides for efficient determination of the route through the device.

Noncontrollable devices are all field devices that do not provide alternate paths or physically interrupt a path. ATCS is designed to monitor these devices remotely for status indications. Noncontrollable devices include occupancy detectors, intrusion detectors, train defect detectors, track integrity indicators, and route integrity indicators.

Specifications

The field system, dispatch system, locomotive system, and work vehicle system all use the communications system to exchange information in a timely manner. Specifications that emphasize functionality rather than dictate design have been developed for these systems. The ATCS specifications represent the minimum necessary to provide for component interoperability. They define a system architecture that is open in the sense that interfaces (but not the contents of any computer or radio or any other box) are explicitly specified, which should reduce user costs by providing for expanded sources of supply. There are, however, aspects of system operation that do not impinge on component interoperability and must therefore be specified by the procuring railroad. Items that must be specified by the procuring railroad include the specifics of man-machine interfaces (e.g., use of a mouse versus command-line interfaces at dispatch consoles); procedures for dispatcher position changes; the form and format of data bases; the specifics of interfaces to corporate data systems (e.g., form and format of data queries and responses); and any special add-on devices or computer programs that adapt "standard" ATCS to individual railroad operating practices.

SYSTEM OPERATION

Railroad operation under ATCS can be considered in five categories: start-up, general operations, emergency handling,

shutdown, and impact on people using the system. Start-up refers to activities related to putting trains and gangs together and preparing them for departure. General operations concerns activities related to the normal operation of trains and gangs, such as setting routes, issuing authorities, and traveling toward destinations. Emergency handling refers to activities taken to continue safe railroad operation in the presence of unusual circumstances such as emergency stops and route defects. Shutdown refers to activities related to terminating trains and gangs. ATCS will have a significant impact on the manner in which people run the railroad. This impact is discussed at the end of this section.

Start-Up

Activities related to putting trains and gangs together and preparing them for departure include initialization, notification of conditions along intended routes, and route data download. Initialization requires the dispatcher to identify a controlling locomotive or a work vehicle that is valid and available. Just how this is done is railroad-specific, although one can imagine some sort of motive power MIS that the dispatcher might access. The dispatcher must also inform the system of car consist, power consist, crew, intended route, and destination; when this has been done, the initialization process is complete.

ATCS then checks conditions along the intended route. Conditions might include track condition notices, track work permits, and advisories. Track condition notices convey temporary conditions such as temporary slow orders and stop and inspect orders. A track work permit is an authority for a gang foreman to conduct work on or along track that may render the track temporarily impassable. Advisories include any other conditions that warrant special attention such as severe weather.

The next step in the start-up process is the route data download (via the data communications system) to the vehicle. The data download, which does not require dispatcher involvement, consists of initialization data, condition data, and physical plant data for the intended route. Physical plant data include speed limits, transponder locations, communication coverage limits, highway crossing locations, switch locations, and information on many other factors that are required by on-board computers for general operation of trains and gangs.

General Operations

General operations concerns activities related to the normal operation of trains and gangs. These activities include requesting authority, setting switches along the intended route, issuing authority, and tracking the progress of trains and gangs toward their destinations.

After start-up and notification (via the data communications system) that the train crew or gang is ready to leave, the dispatcher decides the type of authority a train or gang should have and informs the central dispatch computer. The dispatch computer then decides whether the authority is appropriate and permissible; the requested authority is permissible if it does not conflict with other authorities already held by trains and gangs. The dispatch computer then "reserves" and sets the route the dispatcher has requested for the authority. Once a route is reserved any subsequent requests for conflicting routes will not be granted.

ATCS sets the route without dispatcher interaction by commanding all necessary switches into the position needed for the requested route. ATCS also verifies that the field logic agrees that the requested position is safe and awaits confirmation from the field devices that they are in the correct position before issuing the requested authority. A train or gang that has been issued an authority is ready to proceed toward its destination.

If the requested route conflicts with another authority, the system will request a decision from the dispatcher as to the disposition of the request. The dispatcher also has the ability to restrict (that is, reduce) the limits of authority after an authority has been issued. Authorities are cleared (released) by the system as a result of train location, control-point occupancy indications, report at indications, or dispatcher request.

Because train location is automatically reported to the dispatch system (at a rate determined by the central dispatch system rather than at some preengineered block-length interval), track can be released for use by other trains more quickly than with today's fixed block systems, thereby increasing system capacity. The progress of equipped trains and gangs is tracked by the system based on messages sent automatically over the data link from locomotives and gangs; unequipped trains and gangs require voice communication between the engineman or foreman and the dispatcher. Although the dispatch computer tracks the location of every train and gang, the reverse is not true. Each train or gang computer is only aware of its location relative to its issued authority. An enforcement system is used to keep trains within authority and speed limits.

Train-Specific

ATCS automatically enforces train speed and movement authority limits through automatic brake applications. Enforcement is designed to stop a train within its limits of authority if it is expected to exceed its authority limits. The enforcement system is also designed to allow trains to operate within a speed-performance envelope defined by the on-board computer but will stop a train if a speed limit would otherwise be exceeded. The system is designed to allow the engineman maximum opportunity to retain full control of the train including sufficient capability to execute an emergency brake application at any time.

Gang-Specific

A track work permit may be issued only to work crews. A permit is a bidirectional authority that may have other authorities within its boundaries; however, all movements within boundaries set by the permit require the foreman's explicit permission in addition to a movement authority.

A foreman requests a track work permit from the computer terminal in his vehicle. If there are no conflicts with other trains and track forces, the permit will be issued by the system.

The foreman may also request that a track condition notice be issued to take track out of service or to inform other trains of temporary changes in track conditions. The foreman is also capable of removing a track condition notice or releasing a track work permit that he has been granted.

It is important to note that these permits and notices can be issued and released by the system without any input from the dispatcher; it is, however, a simple matter for the dispatcher to issue a permit or notice if so required. In either case, the system will preclude the issuance of authority for track that is out of service and will notify any trains or track forces that will be affected.

ATCS will prompt an engineman on how to conduct his operation as a track work permit or track condition notice is approached and traversed. The need to record and track such information is handled by the system as are any communications between the train and track force.

Emergency Handling

Emergency handling refers to activities taken to continue safe railroad operation in the presence of unusual circumstances, such as emergency stops, rogues, route defects, and other potentially dangerous situations. ATCS is designed to advise all affected system users of emergency brake applications or other emergencies upon notification from the train or the dispatcher. The system is also designed to detect and handle unexplained track occupancy indications (rogues) and trains exceeding or predicted to exceed their authority limits. ATCS will notify the dispatcher of any reported route defects and advise the dispatcher to notify any affected trains that it could not contact via the data communications system. Reported route defects might include high water, broken rail, or a switch becoming unlocked. Finally, the system will notify gangs to clear the track promptly when potentially dangerous situations arise.

Shutdown

Shutdown refers to activities related to terminating trains and gangs. A train is cleared from the on-board computer and the central dispatch computer upon reaching its final destination. The dispatcher has the capability to cancel a train that is being created. Similar capabilities exist for gangs.

Impact on People

The technology used to implement ATCS will have a significant impact on dispatchers, enginemen, and foremen. ATCS should relieve the dispatchers from most of the mechanics of dealing with trains and work gangs. This will allow more time for dealing with the broad, tactical requirements of running the railroad. The potential for improved traffic throughput is clear.

Enginemen should find that the more timely flow of information will allow for more efficient operation of their trains. Under current signal system operations, the engineman only receives information at fixed points and may be forced by rule to operate at a lower speed than necessary until the next signal. Under ATCS, information is provided to the train as soon as it is available.

The enginemen should also benefit by having all information germane to current operating conditions integrated on a single display. It will no longer be necessary to keep a mental picture of allowed speed, authority limits, applicable bulletins, slow orders, and work orders. Finally, the ability to determine and properly control switch conditions from the cab should make switching operations much more efficient.

Work gangs should find operation under ATCS to be more efficient, because most of the day-to-day information exchange will take place without the need to gain a dispatcher's attention. In addition, because track work permit and track condition notice speed restrictions are enforced on the locomotives, there is greater assurance that special instructions will be adhered to.

Many applications besides train control are made possible by the data communications system and the various information-processing nodes that are used by ATCS. Some of these applications are discussed in the following section.

SPIN-OFFS

The five ATCS elements provide for numerous applications besides train control. Many railroads may find it beneficial to implement some of these applications, which include work order reporting, locomotive health reporting, code line replacement, and track forces management, prior to implementing train control.

It is reasonable to expect that the work order reporting process could be significantly streamlined by using the data communications system to provide a timely exchange of information on pickups and setouts between a locomotive and a central dispatch system. Work orders could be sent to the on-board computer as part of the data download and completed work could be reported immediately with the on-board computer. Work order changes could also be sent to the locomotive while it was en route or doing work. The potential for improved efficiency is obvious.

Locomotive health monitoring can be used to evaluate locomotive performance and to diagnose failures. On-board sensors would be used to monitor important mechanical and electrical parameters continuously, feeding information to the central dispatch system or another central facility via the data communications system. Remote monitoring at the central system might be used to optimize fueling procedures, detect incipient failures, or to diagnose transient or intermittent failures. Such a system could eventually lead to scheduling maintenance when necessary rather than on a fixed schedule.

The ATCS design elements could also be used for code line replacement. Code messages would be generated as they are today and then run through a converter to format them into an ATCS data system message. The code messages could then be exchanged between locations and vehicles by using the data communications system. Another potential application of the ATCS design elements is track forces management. Voice communication between dispatcher and foreman might be significantly reduced by using the data communications system to transmit work requests and to report work com-

pleted. Paperwork might also be minimized, because the computers could be used to record and retrieve pertinent information.

These are but a few of the many potential applications that could be implemented with the ATCS design elements. Improved performance and reduced costs that might result from these applications could be used to pave the way for advanced train control.

CONCLUSION

ATCS has been designed to improve railroad efficiency and safety. Form, fit, and function specifications have been developed for the system and the system design elements. These specifications provide for standard data interchange among components built by various vendors for a single railroad and among various railroads.

Specification of hardware and software requirements has made it possible for suppliers to initiate preliminary and then detailed design efforts. Hardware and software integration and testing is an ongoing process being conducted at a number of test beds such as the Canadian National Toronto ATCS Test Bed and the Canadian National B.C. North Line installation.

Successful completion of system integration and testing should lead to full-scale production and deployment. Deployment need not be an "all at once" occurrence as a result of the modular system design.

Although ATCS has been designed for train control, many other applications of the system design elements could improve railroad efficiency, service reliability, and profitability. These include work order reporting, locomotive health monitoring, code line replacement, and track forces management. The challenge facing railroads and their integrators will be to integrate the vast amount of information flowing to decision nodes properly so that the people charged with making the tactical operating decisions are not overwhelmed, and the full potential of ATCS is realized.

Essential Elements of System Integration

AGU R. ETS AND KEN KOZIOL

The inclusion of microprocessors and their concomitant software in railroad control systems such as the Advanced Train Control System (ATCS) marks a significant change from the previous integration practices of railroad equipment and also introduces issues and concerns that have not been encountered previously with the integration of analog systems. Systems integration is an engineering discipline that works to integrate many diverse parts, with independent operating characteristics, into an entity that functions as a system. The three basic elements of systems integration as they relate to the integration of ATCS by the railroads are goals, planning, and execution. The goals serve as a tangible framework for the many design decisions that must be made in integrating a complex system. Goals must identify the customer, provide a basis for an implementation strategy, and provide a manner for measuring success. Planning is needed to achieve these goals. Proactive project managers engage in thorough and extensive planning to ensure the success of the project. Experience has confirmed repeatedly that efforts spent in up-front planning have a 10-fold payback during the project execution phase. Execution is managing the project to the plan, modifying it when necessary to accommodate the changing environment. The systems integration process is dynamic and design decisions must be made constantly. Meticulous attention to these activities ensures success, reduces life-cycle cost, and reduces the project schedule. Failure to attend to these activities guarantees increased cost, extended schedule, and reduced technical performance.

Systems integration is more than just cabling black boxes together, turning them on as a unit, testing them, and then installing them in the field. Systems integration is a managed process that combines diverse elements into a single entity to fulfill a specific need. This engineering discipline was developed in the 1960s by the National Aeronautics and Space Administration (NASA) to manage mission-critical engineering efforts. It is a complex, nontrivial process.

Although systems integration is sometimes equated with software development, it is in fact broader in scope. For digitally based systems like the Advanced Train Control System (ATCS), systems integration includes the development of software as a subordinate task. Discipline is needed to successfully integrate many diverse parts, with independent operating characteristics, into an entity that functions as a system.

Systems integration is the resolution of design disconnects that occur when many diverse components are brought together, often for the first time, to solve a problem or achieve a goal. Each resolution requires engineering and management decisions, and each decision must consider the impact on the system as a whole. The impact of these design decisions on the components (downward) and on the system (upward) must be considered.

Smartware Associates, Inc., 4805 Dayron Street, Laurel, Md. 20707-3105.

Systems integration has been an evolving discipline over the past 30 years. It entails engineering and management aspects that have been developed as a body of knowledge for implementing complex systems. The weapons of Desert Storm are examples of successful integration projects. Although NASA invented systems integration, it has also provided examples of what happens when the planning and management guidelines are sidestepped or ignored: the Challenger in 1986 and the Hubble Space Telescope in 1990. In both cases, established procedures were sidestepped. Engineering findings that indicated seemingly trivial problems were ignored. But NASA has also provided examples of proper management of technical failures—for example, the safe return of the Apollo 13 astronauts after an explosion in the command module en route to the moon.

The inclusion of microprocessors and their concomitant software in railroad control systems marks a significant change from the previous integration practices of railroad equipment. Although the use of digital systems can provide better control and other advantages, it also introduces issues and concerns that have not been encountered previously with the integration of analog systems. In digital-based system, the software implements the required functionality and logic for train control and management information.

Changing requirements is a prevalent problem, especially in the design and implementation phases of software-based systems. Although software is flexible, it adversely affects the structure and coherence of the design when basic architectural structures are modified. Changes to the system required to accommodate these design decisions must be tracked, and their impacts on the system as well as on the schedule must be considered. This in turn affects design documentation and further implementation.

Systems integration is a cyclical process. Engineering and management must be involved in making design decisions to resolve ambiguities in the specification of the system and to overcome design problems. Projects rarely fail for technical reasons. Failures can generally be traced to the management of the project. The essential elements of systems integration can be distilled into three seemingly trivial and self-evident elements:

- Understand the goal.
- Plan the project.
- Execute the plan.

The commitment to these elements will define the degree of ATCS success.

Integration of a system such as the ATCS can be accomplished in many ways. A railroad may serve as its own systems integrator, contracting out the many parts of hardware and software needed for the entire system. This puts a significant

engineering and management burden on the railroad. Mechanisms and procedures must be in place to resolve the inevitable design ambiguities and disconnects that will occur. These decisions must be tracked with regard to implementation and analyzed with regard to system impact.

At the other end of the spectrum, a railroad can hire a systems integrator to produce a turnkey system. Although greatly reducing the number of engineering and management design decisions to be made, this approach does not reduce their importance. In fact, with a systems integrator, design, engineering, and management decisions made by a railroad will be more important because they will be made at a higher level of system design and thus carry greater leverage.

A third approach is somewhere in between these two extremes, with a railroad using a carefully crafted strategy of implementation in conjunction with a systems integrator.

UNDERSTAND THE GOALS

Although it seems trivial and self-evident, the importance of understanding the goals of the project cannot be emphasized too much. "Understanding" means an intrinsic knowledge of what benefits are to be realized when the goals are achieved and how the goals interact among themselves. It is not sufficient to write a list of goals, then nod approvingly and relegate the list to the shelf. The stated goals must be articulated and made part of everyone's thinking.

The goals serve as a tangible framework for the many design decisions that must be made in integrating a complex system such as the ATCS. The basic goals of ATCS implementation are essentially the same for all the railroads: better operations, economy, safety, and customer service. However, in implementing the ATCS to support these broad, generally accepted goals, a diversity of opinion exists among the railroads.

What the ATCS is seen to provide for each railroad depends on that railroad's existing operations, market, economic viability, and labor relations. In the case of the ATCS, each railroad has stated different goals for its implementation. One sees the ATCS as an extension of the management information system (MIS) for dispatch orders. Another sees the ATCS as a communications system linking the operational elements of a railroad. A third sees the ATCS as a safety system. A fourth sees the ATCS as an active train control element, whereas a fifth views it as a system to help monitor locomotive health. In all cases, the ATCS is seen as a system to help the railroad as a body to become more efficient and competitive in a rapidly changing transportation marketplace.

These goals are manifested as the order in which the various ATCS functions and applications are executed and implemented by each of the railroads. But what of the process by which understandable goals are defined for a systems integration project? A starting point is to pose and answer a series of questions.

At the strategic level we must ask, What is the desired benefit for the business? Is it increased customer service, reduced operating expenses, more efficient operations, reducing accidents, or enhancing train control? The answers must include the specific problems to be solved and the operational concept envisioned to solve the problems. The goal must be stated specifically enough so that a strategy for implementation can be devised.

Second, we must identify who the customer is for the ATCS implementation. Is it the operating divisions, the information division, the maintenance division, or the R&D division? Some subsidiary questions include who must be satisfied—that is, Who is the system being built for?

Third, we must ask when the system needs to be operational. Now? next year? or 10 years from now? Some subsidiary questions cover identifying the schedule drivers, the timetable for realizing the benefits from the system, and any other schedule-related implementation requirements.

Fourth, we must define the commitment in terms of money and resources for the integration. How much is it going to cost? How will it be paid for? How will cash flow and operating costs be affected by the investment outlay for ATCS? What specialized external resources are required?

Fifth, we must specify what exactly we are going to build. Is it going to be ATCS specification components, an ATCS-compatible communications system, a train locator, or an enhanced MIS? We must ask what the operational elements of this system are and whether they satisfy the envisioned operational concept.

Finally, we must determine how we will measure success. Is it to be profitability, market share, customer response, or some other performance measures? In other words, how will we know when we have achieved our goals? The goals must be articulated clearly and explicitly linked to measurable performance parameters. If the goals cannot be explained and formulated so they are amenable to scheduling, staffing, and planning, they are too complex or too ambiguous or both.

Goals must be clear and well defined. Only then does everybody understand the same concept of what is to be implemented or realized. Only then can the team work in concert to achieve the goal. Systems integration of the ATCS is a complex undertaking, but the complexities can be reduced to manageable pieces in the context of clearly defined goals. With clear goals, we will know which path to take to reach our destination.

PLAN THE PROJECT

Once we thoroughly understand the goals, we must plan to achieve them. The second element in successful systems integration is to plan the project. Successful projects are managed by project managers who are proactive—or very lucky. Although luck is capricious, planning is not.

Proactive project managers engage in thorough and extensive planning to ensure the success of the project. The ATCS is a complex and sophisticated system. This makes planning all the more essential, because people will change and a realistic project plan will provide the only continuity. The importance of continuity is that it provides a baseline against which the steps toward achieving the goal can be measured.

Some believe that extensive planning is a waste of time. Managers who ascribe to that philosophy are demonstrating their own lack of proficiency. Experience has confirmed, repeatedly, that efforts spent up front planning have a 10-fold payback during the project execution phase. The dividends paid by proper planning always exceed their cost.

It is, for example, very inefficient to staff a software development team to complete a task in 6 months when the

supporting hardware cannot be fully deployed for a year. The software development team could have done the same job, within the overall schedule, with fewer people and better coordination if it had been allowed the year. Overall, project planning allows the project manager to establish appropriate staffing levels to increase staff efficiency.

Proper planning also allows a manager to ensure that each required task is being addressed. How many times have we heard, "I thought Joe was doing it." Of course both Joe and the project manager thought Judy was responsible. When the oversight is discovered, Judy and her now understaffed and underfunded organization must do the work under extreme pressure. The project will probably be poorly executed and costs will spiral. Is it Judy's fault? Of course not. It is the manager's mistake for not properly planning and delegating the work.

The ATCS is a top-down design. This makes it very important that the integration be controlled by a series of plans so the resultant product does in fact meet the top-level goals. What we want to avoid is the electronics equivalent of the classic cartoon of two railroads meeting—but only one rail is connected. Carefully managed systems integration will avoid such a faux pas.

Based on the nature of a systems integration project and the supporting organizations, several management plans are required. One aspect of proper planning is to determine the plans required for a particular project. Some of the most critical and most often overlooked management plans should be developed very early in the project during the system requirements definition phase. Even the simplest turnkey projects require these planning steps. More complicated in-house integration efforts require more extensive planning.

Project Management Plan

The most important plan—and a key element in every successful project—is the project management plan. It is also the most often overlooked, and when not overlooked, the most often deemphasized. Every experienced systems integration project manager will agree that a thorough, well-thought-out project management plan will virtually guarantee success. The lack of one portends failure.

A comprehensive project management plan states, in a single place, the goals of the project, how those goals will be achieved, and how success will be measured. It identifies the customer for whom the system is being built. Even in-house development efforts have a customer, and it is critical to identify that customer and thoroughly understand the customer's expectations—that is, the project's goals.

The project management plan states the roles, responsibilities, and authorities of key members of the project team. It includes a project work breakdown structure and assigns a budget and responsibility for each element in the work breakdown structure. It identifies project and nonproject resource requirements. It specifies project constraints and establishes an overall project schedule. The plan clearly specifies the requirements for risk management, change management, product assurance, and vendor and contractor management. It establishes the mechanics of project performance monitoring and analysis. It specifies testing and acceptance criteria. It defines reporting requirements and establishes the administrative details for the project management office. It is impossible to manage successfully a complex systems integration project without having considered each of these subjects as they apply to the particular project.

Risk Management Plan

A second, often overlooked, project management plan is the risk management plan. This plan describes the process for evaluating implementation risks and mitigating their impact. Risk management includes both risk assessment and risk control. Risk assessment is identifying, ranking, and analyzing the probability of occurrence and the impact of risks. Risk control is mitigation planning, resolution, and monitoring.

The risk management plan specifies the procedures used to implement the risk management requirements from the project management plan. Some risks can be identified at the beginning of the project, others are more subtle and invisible until the project is under way. The risk management plan identifies each anticipated risk and is continually updated to keep it current. For each risk, the plan specifies a course of action to be taken when a risk tolerance threshold is exceeded. Some predictable risks should always be addressed:

- Productivity and sizing estimates,
- Contractor and vendor performance,
- Vendor decision to abandon the business area,
- Requirements and specification interpretation,
- Availability and attrition of key personnel,
- Assumptions for planning decisions and technical specifications,
- Changing requirements and technical specifications,
- Changing priorities,
- Unnecessary functions, and
- Disruptions to operations.

In addition, each project has a set of unique risks. Overall project cost, schedule, and uncertainty are reduced when each of these risks is identified and monitored from the beginning of the project.

Change Management Plan

The third plan that must be addressed from the beginning is the change management plan. Requirements will change, and it is only prudent to recognize this in the beginning. The change management plan specifies the procedures for managing the change. It must cover, at a minimum, (a) technical, cost, and schedule impact analysis of proposed changes; (b) review procedure for proposed chances; (c) classification and prioritization of proposed changes; (d) approval authority delegation for proposed changes; (e) notification to affected organizations of the approval of changes; and (f) performance monitoring of changes.

The importance of understanding the impact of changes on the total project is critical. A seemingly trivial software change may affect the system, hardware, and installation engineering, as well as training and product assurance. "Inconsequential" efforts may end up costing the project tens or even hundreds of thousands of dollars.

Once the true impact is known, who has the authority to approve such changes? Does the customer really want the change? How much is the customer willing to pay for it? Does the development staff understand the proposed change's broader implications—to operations, for example? What is the impact of delaying the change to a later build? The change management plan must include procedures for answering each of these questions. An effective change management plan has the added benefit of obtaining customer buy-in for each change and for the overall project effort.

Product Assurance Plan

The product assurance plan should address the need for and the methodology of implementing independent verification and validation, quality assurance, configuration management, data management, and independent testing. It should specify the methodology for certifying individual configuration items, subsystems, builds, and the completed system. It must address the need for project standards, organizations responsible for developing project standards, and the review and approval cycle for project standards. It should define (generically) configuration items and establish certification and audit requirements.

A key part of product assurance for software-based systems is the verification and validation plan. This should describe an overall design verification program, covering development of the functional requirements, criteria, specifications, test, and qualification methods and procedures; this should include a plan for software design verification and validation. Specific methods of conformance to approved specifications and accepted guidelines must be demonstrated in detail as well as how the independence of the verification team is achieved.

Effective verification and validation is more than just checking off blocks on a test sheet and generating reports and other documentation. The test sheets and reports are evidence that the software product was examined and reviewed in detail and that it was found to be correct with regard to the requirements. The emphasis of the verification and validation plan should be on tools and techniques for finding problems and the process for resolving the problems identified.

Deployment Plan

Although often delayed until later in the project development cycle, a deployment plan should be considered early in the development cycle. Often deployment considerations affect development sequencing. Long lead-time procurement items or extensive installation efforts may require that system architectural decisions be made early in development. The deployment plan can uncover new risks that need to be addressed. Very frequently, deployment considerations will drive the design. It is always less expensive to do it right the first time than to do it over.

Operation and Maintenance Plans

Like deployment plans, operation and maintenance plans are often delayed until late in the development effort. As with deployment planning, operation and maintenance planning often uncovers design, scheduling, and risk considerations that should be addressed early in the project. The operation and maintenance planning also provides a vehicle for input from those people who must work daily with the system being developed, in this case, the ATCS.

Other Elements

These six plans, once developed, are not unchangeable. They are living documents that are modified to reflect changing conditions with respect to the project. They do, however, contribute to a life-cycle cost baseline against which proposed changes may be evaluated.

To complete the list of essential planning elements for systems integration, the system engineering plan and the system development plan require mention. The necessity for additional implementation-type plans will grow out of the requirements established in these high-level plans. Training plans, staffing plans, configuration management plans, and other detailed planning documents will evolve from the requirements established in the top-level plans.

In addition to these plans, each project should have a project bill of materials, a configuration item list, and a project critical path network. The bill of materials, of course, is necessary to track the materials required, ordered, and received. It is often expanded to include storage and need locations. The quantity portions of the bill of materials are usually configuration-managed.

The configuration item list consists of all deliverable software, procedures, plans, and documentation. It is sometimes expanded to include services such as training.

A critical path network is necessary to understand project dependencies, resource utilization, and cash flow. It is essential to understand project performance and the impact of changes. To be effective, a critical path network must show all internal and external project dependencies. A good rule of thumb is that the critical path network should contain one task for each man-week of labor. For example, the critical path network for a 20 man-year effort should contain about a thousand tasks.

The experienced project manager will involve key members of the project team, his superior, support organizations, and the customer's organization in developing these plans. Once developed, the project manager will secure acceptance and approval from these same people. Before finalizing these plans, the project manager will likely have the drafts audited by an independent third party—a coworker or a peer—to make sure that everything has been considered and a solution set has been established for managing a complex systems integration project.

EXECUTE THE PLAN

Proper planning is reduced to an academic exercise if the plan is not maintained and followed. For the plans to be responsive to the needs of the project they must be maintained, reviewed, and updated to reflect the changing project priorities and requirements. The plans must be thoroughly understood and followed by everyone on the project team, and customer, and support organizations.

The management of a complex systems integration effort, such as the ATCS, is beyond the capabilities of any one individual. Such an effort is generally managed by an organization established for that purpose. The project manager heads the organization, generally called the program management office (PMO). The PMO is chartered to be responsible for the following:

- Ensure that project goals are met;
- Develop, implement, and refine the plans;
- Define the framework for project design decisions; and
- Document, coordinate, and communicate the decisions.

In practice, the PMO performs the following functions, as a minimum (additional responsibilities may be included to satisfy the organizational needs of a particular project):

- Manage the project operations;
- Track, analyze, and audit technical performance;
- Track, analyze, audit, and improve cost and schedule performance;
- Manage vendors and subcontractors;
- Manage risk;
- Manage changes;
- Coordinate with users, customers, and executives;
- Coordinate designs and activities within the project organization;
- Plan project activities;
- Administer the project; and
- Report on project progress and issues.

Historically, projects of the complexity of ATCS invest 8 to 10 percent of the total project budget in the operation of the PMO. The PMO will save many times that amount in project cost and schedule.

CONCLUSION

Planning is essential for success. Each system integration of the ATCS must focus on success. This means that the goal is clearly in mind when each of the intermediate steps is taken and design decisions are made. There is no magic "cookbook" approach to systems integration. Prepackaged solutions are applicable only to prepackaged problems.

The systems integration process is dynamic; design decisions must be made constantly, whether by the railroad or the turnkey systems integrator. Keeping the goals in sight and the plans updated, the systems integrator must always consider ways to reduce risks of implementation and to enhance the safety of the overall system, all the while working within the cost and schedule constraints imposed by top management.

Each railroad must make an assessment within the context of its own business operations and ATCS goals how these essential elements of system integration are to be satisfied. Although having each of these elements in place does not guarantee success in integration, the potential for failure is increased when these elements are overlooked.

ATCS at CP Rail—Steady and Measured

R. J. HIPFNER AND R. A. SHEA

At CP Rail, an extension of an installed computer system for a radio block dispatching method provides the features of advanced Train Control Systems (ATCS). The design strategy followed calls for a progressive installation—both in terms of equipment and functions. The base system was installed in 1985 and has been upgraded several times since inception. It now is used on every CP Rail dispatching desk and has been installed on three other railways. This new implementation adds a local area network and connection to ATCS Specification 200 data communications. Extended principles of operation were developed to allow for mixed-mode issuance of train operating clearances and track occupancy permits. The system works better with more equipment installed, rather than requiring a threshold for ATCS operation. Locomotive engineers receive their operating authority on a flat panel monochrome screen by data radio in the same format as if issued by the rail traffic controller (RTC) by voice. When data communication is not available, the crew enters the movement authority into the on-board computer with point-and-select menuing. A check word is calculated and exchanged between the locomotive crew and the RTC to ensure accurate transcription into the on-board computer. This mode allows smooth interaction across areas and subdivisions with and without radio data communications. A phase 1 pilot between Calgary and Edmonton is under way, having been installed in the fourth quarter of 1991. CP Rail is now at work on its future plans for the system.

CP Rail's involvement with the Advanced Train Control Systems (ATCS) dates back to the initial concept stage conducted by Canadian Railways in the early 1980s. At this point, it is possible to summarize CP Rail's internal developments in ATCS to the present time and to outline what developments the railroad is considering in the next several years.

To date, CP Rail has accomplished the following:

- Developed and installed a computer-assisted dispatch system on all territory not under centralized traffic control (CTC);
- Developed and installed its own CTC office system with an eventual goal of a single parameter-driven office system;
- Demonstrated ATCS concepts with "in track" installations; and
- Progressed from the development of an "ATCS-like" system to a pilot production system.

Before expanding on the above it is useful to outline some of CP Rail's policies regarding its approach to ATCS development:

- CP Rail is its own systems integrator.
- CP Rail uses systems and components built to Association of American of Railroads and Railroad Association of Canada specifications wherever possible. Decisions to deviate from these specifications are made only after an analysis of the impact of that deviation. Equipment availability, implementation schedule, or testing ability are some of the considerations permitting deviations.
- CP Rail directs its efforts to systems intended for revenue service. This means that development must address such issues as robustness, serviceability, reliability, and above all, safety.
- CP Rail recognizes that implementation requires many years. Consequently, development and implementation proceed in incremental steps. Development must adapt and work with existing systems that cannot be replaced all at once.

R. J. Hipfner, CP Rail, 40 University Avenue, Suite 1200, Toronto, Ontario, Canada, M5J 1T1. R. A. Shea, CP Rail, P.O. Box 6042, Station A, Montreal, Quebec, Canada, H3C 3E4.

COMPUTER-ASSISTED DISPATCH

In 1985, CP Rail began production installations of a computer-aided manual block system (now called the occupancy control system, or OCS). Each rail traffic controller (RTC) has an OCS work station with the following features:

- Simultaneous entry and issuance of train and maintenance-of-way authorities by the RTCs. Advanced rules checking prevalidates menu choices and presents only valid options for selection. Forced rules items are inserted automatically on authorities during issuance;
- A graphics monitor that shows the subdivision in a train sheet format and individual locations in a station format like CTC;
- Minimal data entry and familiar formats, which means training can usually be accomplished within hours;
- Full backup and recovery of authorities and train supplies is provided. Information is automatically logged to a central computer in real time for long-term storage and second backup, and to feed other applications with train activity and location information;
- Data-driven rules and graphics. Mainframe data bases are updated by local personnel and automatically fed to the individual RTC work stations at shift transfer, which immediately reflect the table changes;
- Rules logic resides in the RTC work station. The RTC may continue dispatching without regard to the status of communication links to other computers;
- Tracking of manually operated switches permitted to be left out of correspondence. Subsequent train authorities over the switch are warned to normal the switch. OCS also tracks the reporting of normalling of switches;
- Compatibility with IBM PC family of computers; and
- Bilingual operation (French/English) where necessary.

CP Rail developed a version tailored to track warrant control rules for the Burlington Northern. Installations began in

1988. OCS or its derivative, computer track warrant control, is now in use on 85 percent of CP Rail track, as well as on Burlington Northern, SOO Line, Algoma Central, BC Rail, and Chicago Central and Pacific.

COMPUTER CTC SYSTEM

In 1987, CP Rail began work on an office system to communicate with and control the CTC field devices. It has the same design goals as OCS: table-driven graphics and rules, quick response, high-quality graphics, interactive rules on forms processing, mainframe logging, IBM-PC compatibility, low failure rate, and low mean time to recover.

Revenue service began in 1989. Five installations are now in use and several more in development. A second-generation CTC system is underway to attain the design goals more completely.

GENERIC DISPATCH SYSTEM

Since 1987, CP Rail has been steering its development efforts to a single integrated train-dispatching system. The goals are as follows:

- Reduce RTC workload by providing for single recording of train operating, status, or delay information;
- Reduce RTC stress by improving hand-off among the varying protection methods and rules;
- Improve system assurance and availability by reducing the variety of hardware components and connections;
- Improve information by capturing more data directly at the source and time of creation;
- Improve traffic control by consolidating all information faster and improving the quality of the decision making;
- Reduce installation and upgrade costs by building a single data-driven system; and
- Improve safety by reducing and streamlining the rules base.

CP Rail has some way to go before the ultimate goal is reached, but in increments and by careful design considerable progress is being made.

ATCS DEMONSTRATION AND PROOF OF CONCEPT

In 1980 and 1987, CP Rail installed and demonstrated remote switch control from a moving locomotive. Two different control systems were installed. Both interrogated switch position and allowed switches to be changed from the locomotive. Radio communication was used between the switch and the locomotive. Because only one locomotive was equipped for a particular switch, this demonstrated the principle but did not address the more complex problem—having multiple locomotives equipped with the ability to throw a particular switch and having only one locomotive at a time in possession of this control and all other locomotives locked out.

Automatic enforcement of authority limits and speeds was clearly demonstrated as being feasible. Apparatus was installed on one locomotive and proved to be effective in both slowing and stopping a train. The demonstration revealed that there was significant variation in braking performance and that considerable technical effort will be required to develop a means of predicting braking distance. Providing a braking algorithm that is selfcorrecting for the many variables that affect braking performance is now under development. Significant simulation and field-testing work has been carried out with additional testing scheduled for 1991.

ATCS PILOT AREA

In late 1987, a CP Rail group was given the mandate to implement an ATCS pilot project. From that time to about 1989, a detailed analysis was carried out to do the following:

- Define the expected outputs from the pilot project;
- Understand and interpret the industry specifications and identify the issues that needed to be addressed, particularly with respect to integration;
- Prepare a design and implementation sequence; and
- Evaluate apparatus being developed by various suppliers.

PRIMARY RADIO TESTS

During the initial stages to develop a design for the pilot project, an early conclusion was reached that communication in the 900 MHz bands could present a problem if CP Rail were to use existing base station sites. Consequently, CP Rail carried out a set of tests, using message success rates as a criteria for coverage and using radios and modems that were as close to ATCS specifications as were available at that time. These tests provided sufficient information to make CP Rail confident that the use of the UHF bands would not require a substantial increase in base station infrastructure.

The remainder of this paper will describe the purpose, approach, and implementation of the pilot project.

PILOT PROJECT

The pilot project can be best discussed under these headings:

- Policy,
- Purpose,
- Phased implementation,
- ATCS office components,
- ATCS on-board components,
- Software quality,
- Safety assurance, and
- Future phases.

For the pilot project, these specific policy decisions were made:

1. The primary use of the pilot project would be for train control. Secondary uses of the communication system could be examined in later phases of the project.
2. The installation would provide for a production system to be used in normal revenue service.

3. The system would provide for the simultaneous operation of trains that were equipped with ATCS apparatus and those that were not.
4. Integration and application programming would be carried out by CP Rail.
5. The ATCS communication 900 MHz frequencies and protocol would be used.
6. The design of the pilot would allow for installation, testing, and production implementation in discrete steps.

These specific policy decisions are consistent with the policy laid down in the earliest stages of CP Rail's development cycle.

PURPOSE

The decision to proceed with a pilot project was predicated on satisfying four major purposes:

1. To prove the ability to run a mix of equipped and unequipped trains (this was seen as the implementation scenario CP Rail would be faced with for some time);
2. To prove the ability to install functions in a progressive manner without interfering with railway operations;
3. To provide a platform for carrying out operational integration of components and subsystems from different suppliers; and
4. To provide a concrete means of establishing the costs and benefits of ATCS functions.

PHASED IMPLEMENTATION

It was determined during the analysis phase of CP Rail's development cycle that pilot project implementation was best thought about in terms of these specific functions:

1. Electronic transfer of train movement authorities from office to train;
2. Electronic transfer of crew acknowledgments or rejections of these authorities to office;
3. Provision for insertion into on-board devices of the train movement authorities that are issued to a crew by voice communication from the RTC; and
4. Reports of train location and speed using tachometer and transponder technology transferred electronically to the RTC work station displays.

Another consequence of the phased implementation policy was the decision to maintain existing operating rules. (Canadian railways all operate under the authority of a single rules set, the Canadian Rail Operating Rules, unlike U.S. railways, which have a proliferation of operating rules and practices.) ATCS specifications require adoption of another set of operating rules, one that is not in use on any existing railway. CP Rail decided that maintaining an operation with two differing rules sets was not an expense or risk that it was prepared to undertake.

Given this decision, CP Rail has maintained its use of OCS logic and rules. This led to the use of OCS data bases, data elements, and data messages that do not conform to ATCS Specification 250. However, where new data elements were needed, for instance transponder serial numbers, CP Rail adopted the specified ATCS format.

With the equipment installed in Phase 1, CP Rail is in a position to add and test additional functions:

- Switch control;
- Automatic track release using train tracking;
- Automatic brake enforcement;
- Issuing track occupancy permits to track maintenance forces.

As a major consequence of CP Rail's requirement and design philosophy of adding functionality in incremental steps, the system also allows for incremental fallback when parts of the system either fail or are unavailable. For example, if electronic data communication fails, an ATCS-equipped train does not inevitably revert to manual, paper-based authorities. If voice communication is operational, the train crew continues to use the on-board computer, entering the authorities into the computer as they are voiced by the RTC. An exchange of computer-generated cyclic redundancy check (CRC) codes between the RTC and the train crew ensures that authorities have been entered correctly on each of the computers. CP Rail has named this "voice/on-board terminal mode." When data communication is restored, the two computers immediately begin to use the data radio (DNET), by preference.

OFFICE COMPONENTS

The base for the new office system is OCS, but several changes and enhancements have been made for the ATCS pilot (see Figure 1):

- Split the existing OCS system onto two machines. The RTC interaction and rules verification modules are isolated on their own computer, separate from the mainframe communications modules on the server.
- Connect the RTC work stations to the server with a local area network. Ensure that a failed RTC work station can be rebuilt from the server, and the server can be rebuilt from the RTC work stations.
- Connect the ATCS radio data network interface to the server. Implement ATCS Specification 200, "wrapping" and "unwrapping" OCS clearances and messages to and from the mobiles.
- Design and build a DNET manager to handle and report on the messages, queues, trains, base stations, and other DNET components, both hardware and software.
- Enhance the RTC work station logic to handle multiple unsolicited updates, in this case from the RTC's keyboard and, through the server, the corporate mainframe computer, other RTC work stations, and the DNET.
- Add a larger, superresolution monitor to the RTC work station with simultaneous OCS- and CTC-style graphics. New icons would show the status of the new data available from DNET, such as a train's head-end and tail-end location.
- Enhance the RTC work station to implement new synchronization logic to ensure that both the train onboard com-

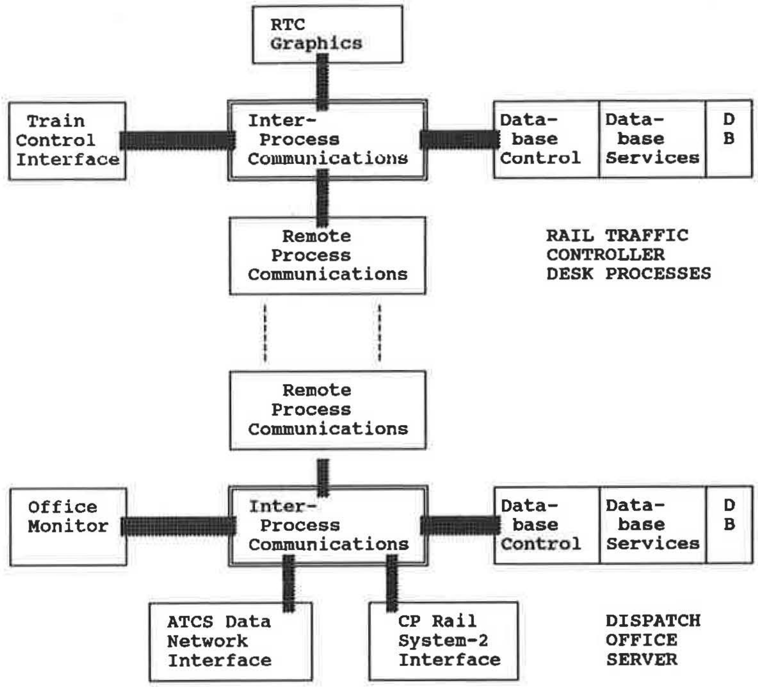


FIGURE 1 The office architecture.

puters and the office have protected their authorities. Calculate and test the check words for voice/on-board terminal mode.

- Add new logic on the RTC work stations and server to ensure that corrupted messages and data bases are detected immediately.

OFFICE COMPONENTS FUNCTIONAL BREAKDOWN

1. Train control interface

RTC functions

Train supply:	create, edit, delete, change lead unit, display
Authority:	issue, track release, cancel, acknowledge, change call time, display
Bulletins:	edit
Report times:	over main track switch
Switch report:	normalized report
Transfer:	same desk, new desk, mainframe, diskette
Sign-on:	recover, receive territory

Service functions

Switch status:	tracks the known position of manually thrown switches
First rules calculator:	calculates the valid options for authorities and supply choice, by display field
Second rules checker:	postvalidates the selected and generated authority items prior to generating a complete time for the RTC
Receive MTP supplies:	automatically introduces new train supplies, and updates to them, as received from CP Rail's master train plan (MTP)
Update graphics:	updates the RTC's graphics monitors when information is received, deleted, or modified
Local backup/recovery:	maintains a data image on local hard disk; rebuild random-access memory (RAM) data image from hard disk after power restored
Router backup/recovery:	maintain a data image on router disk; rebuild RAM data image from router disk after failed RTC work station replaced
RTC printouts:	print authorities, supplies, transfers, status reports on RTC local printer
Activity logging:	maintain an audit trail of RTC activity for problem analysis.

2. Router/server

Host management functions

Security:	validate RTC access, force password changes, control territory ownership
Session management:	maintain communications session with host, including error recovery and automatic application sign-on
Incremental download:	detects updates to MTP supplies and loads them onto the router
Initial download:	detects the out-of-sync condition for a territory's track tables, MTP trains, or complete operating sit-

	uation and forces a full refresh of the necessary components
Incremental upload:	logs any change in the operating situation (supplies, authorities, etc.) for backup and audit

Office functions

Distribution:	queues and switches messages between the connected entities, the DNET, the RTC work stations, CP Rail's host
Backup/recovery:	provides a second-level situation backup for the RTC territories, recovers an RTC work station on demand

DNET functions

Locomotive maintenance:	establish a communications session between a locomotive and an RTC, ensure that on-board terminal/computer software is of the correct release/version, update the on-board track tables if necessary, terminate a session
Message tracking:	transmit field-bound messages, track message status and retransmit as necessary; acknowledge messages arriving in assurance mode; track health of on-board equipment.

ATCS ON-BOARD COMPONENTS

The locomotive on-board configuration consists of two computers: (*a*) An input/output processor, which links the mobile communications package, the transponder interrogator, and the axle generators; and (*b*) an MS-DOS–compatible 80286 computer, which also provides a flat-panel 24 × 80 VGA display and a limited function keypad.

Applications programs for receipt and acknowledgment of clearances, creation of track releases, cancellation of clearances are resident on the second machine (the on-board terminal) and have been written by CP Rail. Internal logic is based on the logic of the office OCS process.

ON-BOARD COMPONENTS FUNCTIONAL BREAKDOWN

1. On-board computer

Initialization:	establish the existence and status of the connected devices (transponder interrogator, dual tachometers, locomotive I.D. unit, on-board terminal [OBT], mobile communication package); obtain ground address table and send to OBT; report on-line status to OBT and office; set distance from zero for movement tracking
Message routing:	routes ATCS Spec 200 messages to/from the connected devices; processes vital messages addressed to OBT as follows: on input, the on-board computer intercepts the message and does a CRC check before routing it to OBT; on output, the on-board computer intercepts the message and appends the CRC value before passing it to the mobile communication package; the onboard computer implements a 31-bit CRC failure check and sends message to office
Movement tracking:	acts on a transponder hit and sends to OBT the serial number of the transponder encountered; each second reports (distance from last transponder encountered, tacho [i.e., wheel] direction, speed
Diagnostics:	monitors status of connected devices and reports to OBT and office; on-board computer will not shut down if OBT link, or other device fails

2. OBT

Locomotive engineer functions

Train supply:	display, edit train characteristics
Clearance:	acknowledge, copy, cancel, change call RTC time, reject, track release, display
Coupling:	report train coupled/uncoupled
Transfer:	off duty, on duty
Time/date:	manual correction
Sign-on:	voice-respond to version request, mark track table as valid

Service functions

Power-up/sign-on:	establish connection with on-board computer, initialize, respond to version request, check validity of track tables and software
Clearance:	maintain synchronization with office—receive or reject clearances from DNET, allow or reject crew actions
Location tracking:	report speed and location to office
Diagnostics:	tests the state of the on-board computer link, manipulates mobile communications package via the onboard computer link
Message testing:	generates a CRC for all application messages, ensures that both electronic and keyboarded messages pass the CRC check

SOFTWARE QUALITY

The largest problem CP Rail has encountered in building control system software has stemmed from the lack of under-

standing between the developers and the clients. For ATCS, CP Rail has instituted several types of documentation that help the designers clarify their thoughts and provide logical, complete guidance for the developers:

- *Principles of operation*—Describes the root concepts of the control system, including maintaining synchronization, effects of failure, recovery methods, detection of errors, etc. A small set of flow charts illustrates the more complex authority protection mechanisms.
- *OBT software, high- and low-level design*—Documents the functions, screen layouts, edits of each field, data flows, logic flows, and internal data bases.
- *Office architecture, high- and low-level design*—Describes the extensions to the OCS product, including the modified logic, data structures, interprocess communications, data flows, and logic flows.

Changes to the design documents are passed through a change control board, which includes both the developers and the client designers. Software is checked in and rebuilt by a control officer, who is independent of the development team. This officer checks coding standards, checks that the modules conform to standard quality metrics of complexity, ensures that documentation is complete and matches the code, and protects the module by submitting it to version-controlled backup.

The control officer is also responsible for testing, documentation of the tests, tracking of defects, and final resolution. Testing is carried out by experienced independents in a site off the development floor using both white box and black box principles.

Much of the regression testing has been automated with programs and computer files of previously generated tests. Test machines have been set up to generate bad data into the on-board computer, respond incorrectly to good queries from the OBT, and pretend to be a misbehaving data network—to give some examples.

CP Rail decided as well to include the radio system components as part of the test facility. Thus, several levels of stepwise testing and integration have been possible without resorting to field work.

SAFETY ASSURANCE

The OCS office program is the basic protection against the generation of authority overlaps. It presently has two separate and distinct algorithms for checking and rejecting potential overlap conditions, as well as ensuring that all dispatch rules have been complied with in the generation and issuance of an authority. With the development for the pilot project, some fundamental considerations have been established as a result of the design process and the fault analysis contained therein. Examples of this are as follows:

- The crew will remain in the control loop for all processes associated with the granting, altering, or cancelling of movement authorities.
- Basic functions are protected by at least two checks, e.g., OCS has two separate algorithms for overlap checking. Assurance that a clearance as generated by the RTC on the OCS system is replicated exactly on the locomotive is provided by having an independent mathematical check carried out on the complete text of the authority as displayed on the OCS machine and on the locomotive OBT. The mathematical check is carried out by a different algorithm on the OCS machine than that used on the locomotive. Note that this check is over and above the cyclic redundancy checks being carried out on the messages moving between the office and the locomotive.
- The loss of a message never results in the potential for the generation of an unsafe condition.
- Messages arriving in a different sequence than that in which they were generated will not result in an unsafe condition.
- Messages from the office crossing messages from the locomotive cannot result in an unsafe condition.
- Temporary failures, however short, of any component (for example, loss of power to an on-board computer or OBT) will require complete consistency checking between the office and the locomotive.
- Software and data base version checks are inherent in all activities.

Future Phases

The 1991 pilot between Calgary and Edmonton will be incrementally expanded in future phases. Once confidence in location tracking has been established, we will move to automatically release track behind trains operating with proceed clearances. Both portions and entire clearances could be released by the OBT logic, using the identical logic now used in the OCS system.

As OCS currently tracks the reported position of manually operated switches, it is a small step to centrally control the position of powered switches, incorporating the CTC control logic already developed. However, CP Rail is still studying the merits of direct locomotive control of the switches through the on-board computer and the operational considerations before committing to either central, distributed, or dual-switch control.

Considerable effort has been expended on the development of a braking algorithm. CP Rail's approach is to develop a self-correcting algorithm that will converge from an initial, conservative, braking curve for a given consist toward an algorithm tailored for the actual operating characteristics of the train, track, and conditions. Each braking action taken by the locomotive engineer would add to this data base and improve the situational algorithm. Pilot implementation of such a predictive algorithm is a priority item but will require significant additional effort. Testing and development is continuing.

In a recent RTC workload study it was noted that, where train lineups have been abolished, more than 50 percent of the RTC's time is spent dealing with maintenance-of-way forces. CP Rail is looking at automating both the delivery of real-time train location and speed, and the issuing of the authority to occupy the track. These functions are likely to be deployed

on several technologies using both fixed and mobile communications.

CONCLUSION

Compared to some other railroads, progress on ATCS at CP Rail may not have been as observable. The deployment of the live pilot this year in the Calgary–Edmonton corridor should go some way toward changing this.

The unique element, an incremental approach to deployment, matches more closely CP Rail's tradition of installing new technology on a steady timetable. This approach allows safety to be maintained and new methods adopted in small, manageable increments where affordable and where benefits are obtainable.

Operation Control and Signaling System for High-Speed Lines

KLAUS H. HÜMMER

Increasing traffic demand combined with the consequent congestions of road and air traffic has led several European countries to improve passenger rail service by introducing fast, intercity links and new high-speed lines. The signaling for the German Intercity-Express high-speed trains consists of continuous automatic train control (CATC), decentralized microcomputer interlockings, and operation control centers for automatic train supervision. The CATC is an automatic cab-signaling system transmitting, via an induction cable installed along the track, an array of data from track to train and vice versa. The most important signals are maximum speed, target speed, and distance-to-go. On-board fail-safe computers calculate the required speed, control the train, and supervise the performance. Microcomputer interlockings work on a fail-safe, 2-out-of-2 technique with full redundancy. Control range for each central interlocking can reach 40 to 60 mi. The software consists of a basic operating system and special railway applications. Track monitoring is achieved by jointless audio track circuits with a frequency range of 4 to 6 kHz and 9 to 17 kHz, and a frequency shift keying (FSK) code bit-pattern to make the circuits immune to interference. Train movements are operated from control centers spaced at approximately 200-mi intervals along the lines. Train routing is implemented by means of train numbers. The progress of trains is monitored on indication panels, track lay-out monitors, and train graphs that allow dispatchers to react quickly. Links between control centers and the line is via a fiber-optic PCM high-speed communication system. Automatic train supervision, decentralized microcomputer interlockings, and continuous automatic train control are interconnected by well-defined data interfaces.

In Europe, but not only there, railways have been enjoying a renaissance during the past 10 years. There are clearly two reasons for this: the increasingly chaotic conditions on highways and the congested airspace with its accompanying flight delays.

For some years now, in the major European countries, high-speed rail lines have been built and, step-by-step, put into service. A European network for high-speed lines is planned for the future. Same branches of it have already been realized (Figure 1). In France, two lines are operating at present. In Spain, the high-speed Madrid–Sevilla line is under construction.

As long ago as the early 1970s, the German Federal Railway started an Intercity-Express service linking 50 major towns in Germany. Speeds of 120 mph were achieved on existing lines. Running parallel to faster service on existing lines in Germany are newly constructed lines and the modern, high-speed train known as "ICE," the acronym of Intercity-Express. In June 1991, the German Federal Railway put high-speed traffic into service along the first 270 mi of the newly constructed Hannover–Würzburg and Mannheim–Stuttgart lines. The current line used by high-speed trains from Hamburg in the north of Germany to Munich in the south is 570 mi long. The ICE network will be extended next year.

Siemens AG, Transportation Systems Group, Ackerstrasse 22, W-3300 Braunschweig, Germany.

Today, the new ICE trains travel to schedule at a speed of 160 mph. The vehicles are authorized to travel at 175 mph, which they do when there are train delays. The signaling system is designed for 200 mph.

Because of the hilly countryside, many bridges and tunnels had to be built. To make that large investment yield a better profit, freight trains also travel on high-speed lines. With this in mind, line gradients have been limited to 1.25 percent.

The chosen signaling system permits optimal traffic flow even when there is mixed traffic of high-speed trains and freight trains. The high-speed ICE is the latest technology with a three-phase motor, power recovery in network when braking, data transmission within the trains using fiber-optic cable, and on-board diagnostic systems. The cabs and passenger cars are pressure-proof because of the strong changes of air pressure at the entrances and exits of tunnels.

An ICE train set is made up of a traction unit at each end and 14 passenger cars. First and second class provide a seating capacity of 759. In 1988, a prototype of this train reached a speed of 253 mph, a world record. But however fast the trains themselves may be, the efficiency of high-speed lines is largely determined by the capability of the signaling and operation control system employed.

THE COMPLETE SYSTEM

Railway signaling equipment must guarantee the protection of train movements and permit an efficient use of the line by keeping headways short. The equipment for signaling and operation control on the new German high-speed lines is characterized by the application of electronic technology and by design and testing done in accordance with the high-safety level of German Railways specifications.

The complete system is made up of continuous automatic train control (CATC) for protecting trains and for cab signaling, on-board automatic speed control (ASC), decentralized microcomputer interlockings for safeguarding routes, and the operation control center's automatic train supervision (ATS) (Figure 2).

The CATC System

The basic requirements for the signaling system result from the fact that the maximum speed is 175 mph. At this high

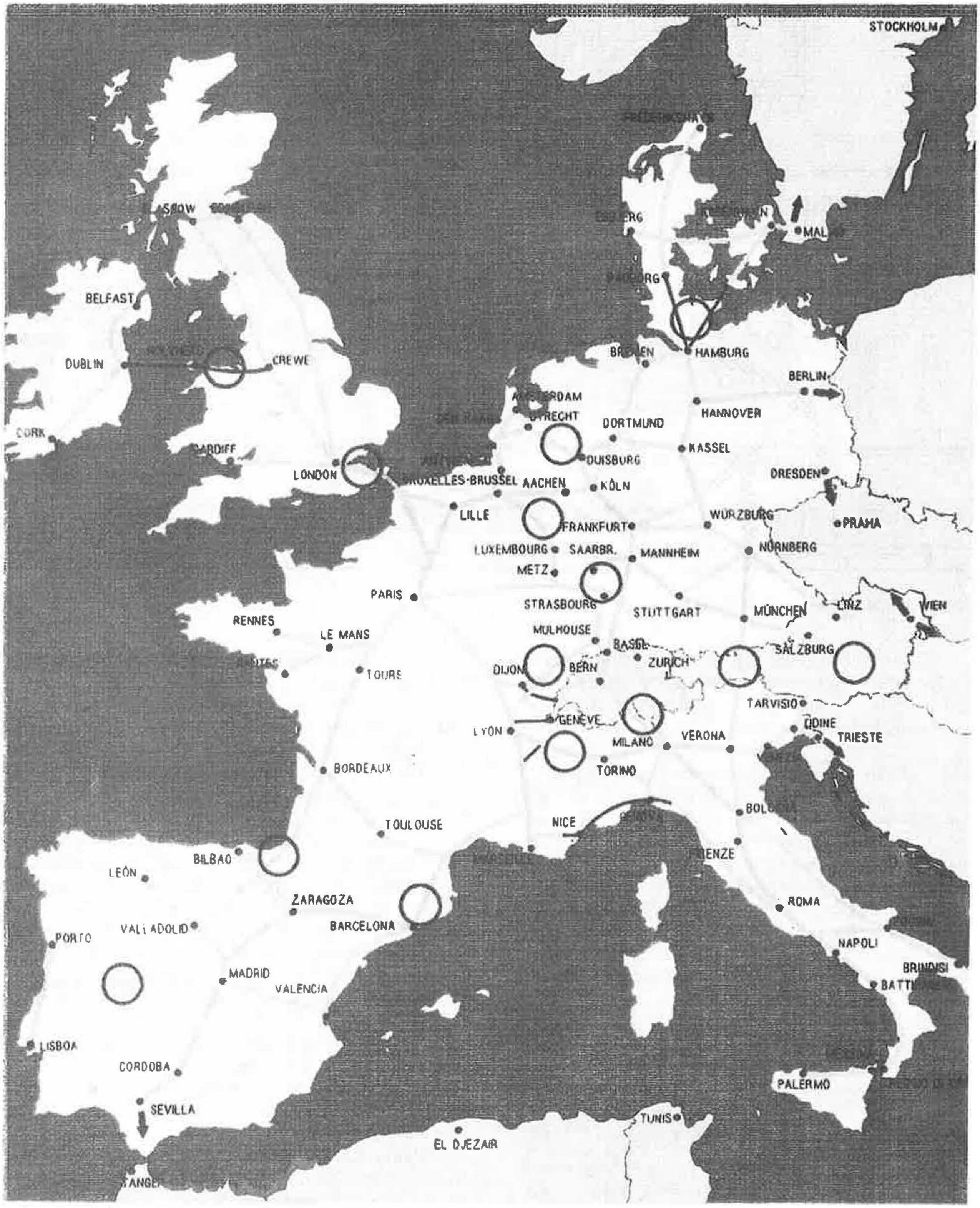


FIGURE 1 Planned European high-speed network.

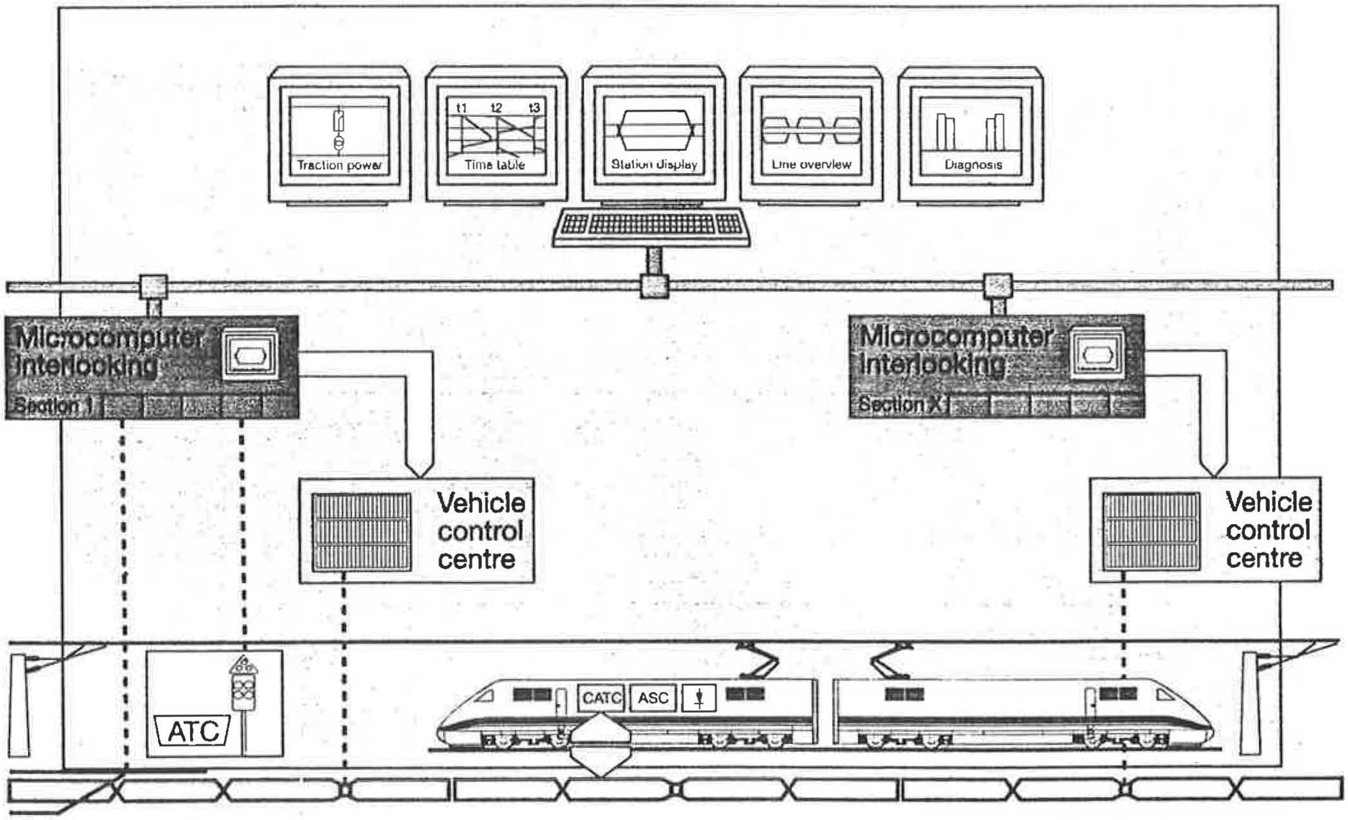


FIGURE 2 Train control for high-speed lines.

speed, the engine driver is not able to observe optical signal aspects. The braking distance of a train traveling at this speed at a braking rate of 0.5 m/s is 3.7 mi. For this reason, it is necessary to install a continuous automatic train control system.

On the German high-speed lines, trains are controlled by the CATC system known as the LZB 72-80, which permits either automatic or driver operation. Among the principal features of the CATC (Figure 3) are the unmanned vehicle control centers along the line at great distances. Each vehicle control center has a 2-out-of-3 computer system for handling all data processing. This configuration guarantees fail-safe and reliable operation. Two computers are needed for the vital operation by cross-comparing the outputs, the third computer serves as a hot standby. The vehicle control center also includes the corresponding equipment for transmitting data to the vehicles, to the other vehicle control centers, and for receiving data from the interlockings.

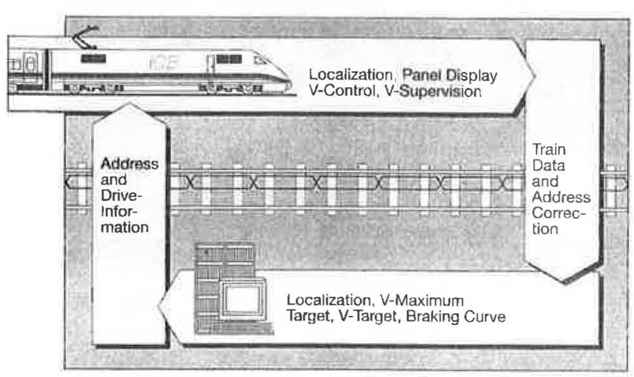


FIGURE 3 CATC basic functions.

Another feature of the CATC is storage of permanent controlled-area data in the computer system of the vehicle control center, along with data on line gradients, currently permitted line speeds, and, if applicable, signal locations. Variable data such as the location of the trains, individual braking efficiency of trains, and positions are transmitted to the vehicle control centers by the trains and interlockings.

The vehicle control center computer calculates the running orders to be transmitted to the on-board computers: target speed, available distance to the next target, and the number of the braking curve to be used in the train-borne computer. The trains calculate their driving and braking profile and drive in accordance with this limitation at maximum permissible speed.

Changes in line data, such as permitted maximum speed along track construction sections, can be input into a vehicle control center. The data transmission functions exactly as a closed-loop data transmission between train and track, and track and train (Figure 4).

A double-track section of up to 60 mi in length can be monitored and controlled from one CATC control center. In practice, however, on the new high-speed lines, section of 25 to 40 mi is supervised and controlled by a single vehicle control center.

Trains are traveling nearly at moving block intervals—the shortest possible headway. At a speed of 200 mph, trains can travel at 2-min intervals. The vehicle control center provides

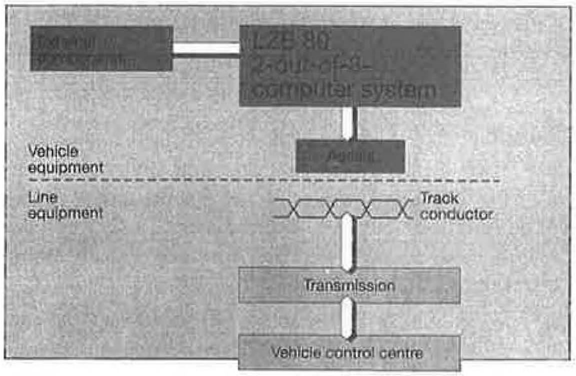


FIGURE 4 LZB 80 continuous automatic train control.

all trains within its control range with information for observing the distance between trains and with the permitted speed. The traction unit is addressed by its location when the vehicle control center is sending data. The on-board computer derives the relative location of the train from the odometer and crossings of a wire-loop in the track and calculates the current permitted speed on the basis of the distance to the next target and the target speed transmitted from the vehicle control center. The vehicle is automatically braked if the permitted speed is exceeded (Figure 5).

The vehicle computer also calculates the data for automatic speed control as well as the permitted speed from the intermittent automatic train control (ATC) system, when this is employed as a backup system.

Data exchange between vehicle control center and train is bidirectional. Transmission is via a pair of wires laid between the rails. The wires are crossed at intervals of 110 yards. The transpositions give the exact location of train. The receiving antennas are connected with the two channels of the receiver, which amplifies and demodulates the call-telegrams sent from the vehicle control center. The receiver forms the information "transposition point" from the phase change of the receiving signals at transposition points. The location and the driving information is evaluated by the train-borne 2-out-of-3 computer and processed for controlling speed and activating the brakes. The train answers the call of the vehicle control center via its sending antennas, transmitting its location, locomotive number, train length, and its actual speed (Figure 6).

Several trains can be in the area of one loop at the same time with a theoretical length of 8 mi. They are individually addressed by the LZB control center and provided with information in respect of their run. For availability reasons, the loop is split into sections of 300 m, which are individually fed by transmitters and receivers with identical information. The data transmission frequency from vehicle control center to train is 36 kilocycles/sec, the data-signaling rate is 1.2 k/bit per second (bps). In the other direction, from train to track, the frequency 56 kHz is used with a data rate of 600 bps. The carrier frequencies are frequency modulated. The transmitted data are coded for hamming distance 4.

Future modifications are planned that will allow the LZB data transmission to be via radio instead of via cable loop. A point to be emphasized here is that the LZB 72-80 system provides bidirectional transmission of data. This ensures maximum safety and efficiency of control. The system takes into consideration the individual train characteristics and topography of the line, and so achieves close train headways and maximum line-carrying capacity. The method of transmission of target information (instead of speed information) is a major advantage over coded track circuit systems in mixed traffic applications.

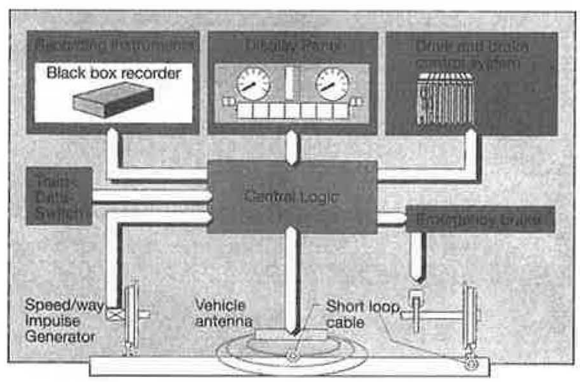


FIGURE 6 CATC train-carried equipment.

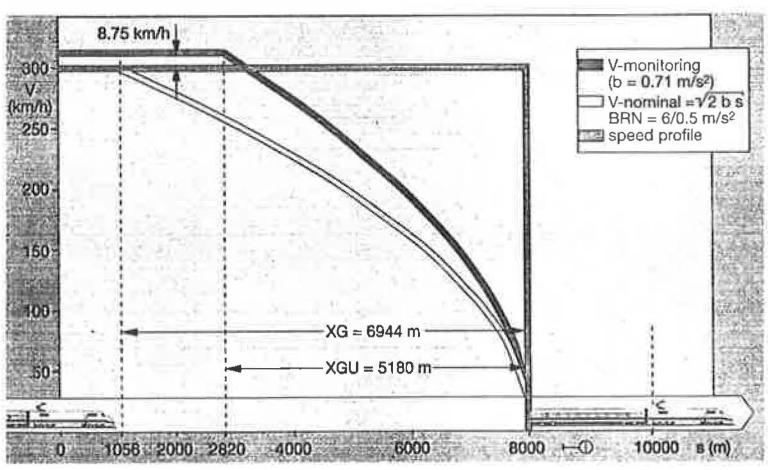


FIGURE 5 CATC determination of speed.

Figure 7 shows the cab of the ICE high-speed train with the display of the CATC system LZB 72-80. The German Federal Railway has more than 750 mi of line equipped with this system. The CATC system has been in operation for 15 yrs on high-speed lines catering to speeds of up to 120 mph. The second generation of the LZB 72-80 is working now very successfully and reliably for trains with speeds of 175 mph. About 400 locomotives and ICE high-speed trains are equipped with this system. It is in accordance with the ORE-A46 recommendation of the International Union of Railroads (UIC). Therefore, it will be introduced also on other new lines in Europe, such as the future high-speed line in Spain.

Interlockings

Route setting and locking, points, track vacancy detection and equipment (and, if applicable, line signals) are to be linked by interlocking. Germany's new high-speed lines are equipped mainly with electronic interlocking. These interlockings provide safeguarded routes and also give information to vehicle control centers. The microcomputer (MC) interlockings, with solid-state technology, have a very high level of safety. Each type of electronic interlocking is type-tested for safety by the German Federal Railway. These MC interlockings work to the principle of geographical circuitry logic. Therefore, the entire safety logic is programmed and tested just once for one railway authority. The general layout of MC interlocking is shown in Figure 8.

FIGURE 7 ICE-cab with CATC system LZB 80.

The interlocking software is divided into basic operating system, special railway applications, and station-specific software. The individual interlockings are project-planned and only the function is tested, as the system's safety has already been approved. This combination provides the safest and most economic procedure.

Interlocking fail-safe function is provided by cross-checking results of the two-channel microcomputer system, SIMIS (shown on the right side of Figure 9). The results of the two parallel computer channels are continuously cross-checked. Only if the results are equal will the command be executed. Availability is made by employing full redundancy, using two computer systems each with a SIMIS dual-channel computer system. In the dual computers the same software is used. Figure 11 shows the hardware for fail-safe microcomputer interlocking.

Via a data transmission line, each main interlocking is linked to several extended interlocking components with decentralized section computers. The maximum distance between a main interlocking and the extended interlocking components is 7.5 mi. Up to nine decentralized interlockings can be connected to one main interlocking, which has display and operating facilities. Therefore, at maximum, up to 60 mi of line can be controlled and safeguarded by one main MC interlocking. Figure 10 shows, as an example, the installed MC interlocking Orxhausen, controlling a long section of the ICE high-speed line from Hannover to Würzburg.

Track Monitoring

Special track vacancy detection equipment has been developed for use with the most modern high-speed trains. This equipment is immune from electromagnetic interference and is fail-safe. It functions with a frequency shift keyed (FSK) coded 8-bit pattern at an audio frequency range of 4 kilocycles/s to 6 kilocycles for block sections and 9 to 17 kilocycles/s in stations.

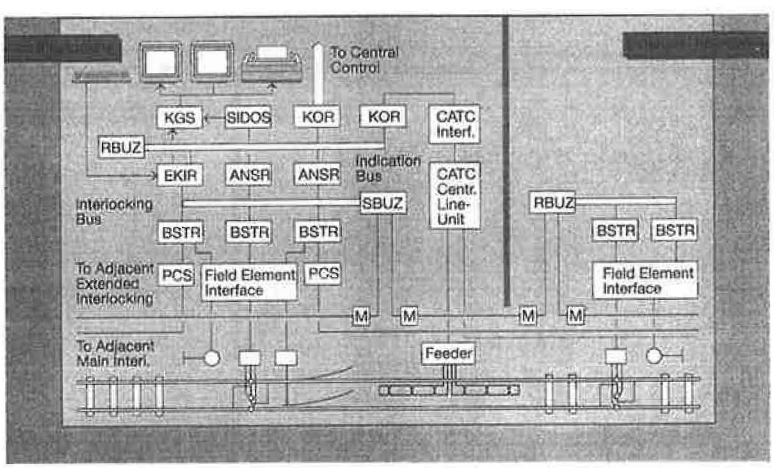


FIGURE 8 Microcomputer interlocking general layout.

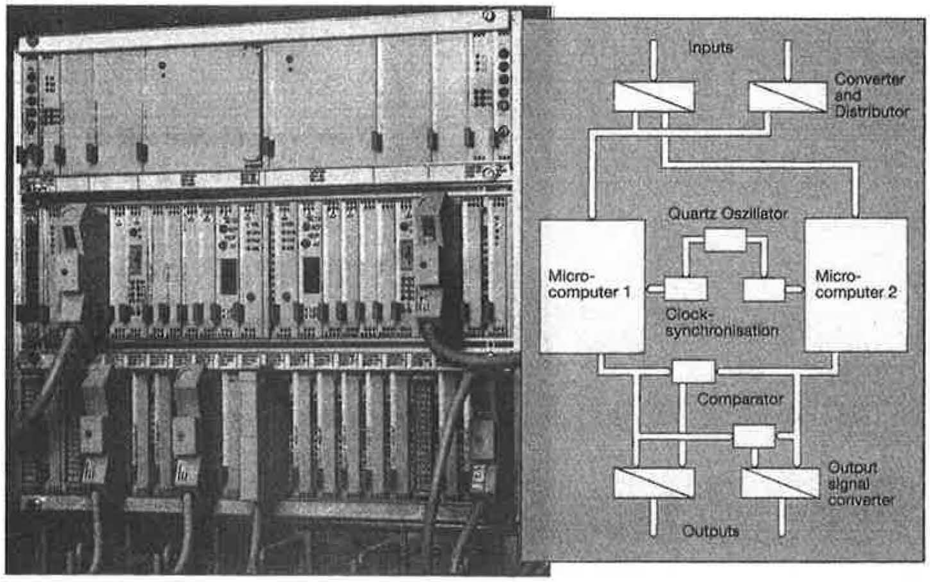


FIGURE 9 Fail-safe microcomputer interlocking.

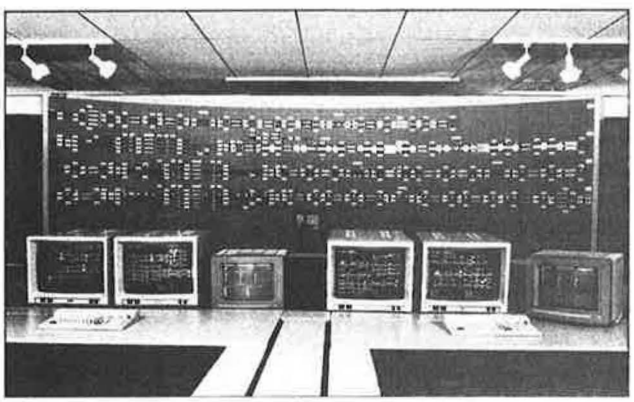


FIGURE 10 Microcomputer interlocking Orxhausen.

The track circuits naturally have electric joints, so that the rails are not required to be cut. This enables an optimal traction current return to the power station. The length of a center-fed circuit for blocks is up to 1.2 mi. This type of track circuit uses the most modern electronic components. They are remote-controlled; that is, transmitter and receiver are accommodated in signal cabins. As a result, reliability and availability are excellent.

Line Operation

Line operation functions as follows: Computer-aided traffic control centers monitor line sections of approximately 200 mi on high-speed lines. They supervise traffic movements for adherence to timetable and to manage traffic control in exceptional cases.

Route protection and control is automatically performed. The CATC system provides for the protection and control of trains. A train can be controlled by the driver, who refers to a cab display, or be controlled automatically in accordance with the current calculated maximum permitted speed.

For train operation automatic train supervision is installed. The high-speed trains set their routes by themselves. Each train has its own identification number. As the train travels along the high-speed line, its number is switched from interlocking to interlocking. This identification permits individual train tracking throughout the controlled area. The safe-guarded route is automatically established as the train proceeds over the line. There is no operational control by a dispatcher. Location and description of the train are transmitted from each process area to the operation control center. Here, the computer updates the train graphs that are shown on color monitor·

The stored timetable data are compared with the current train run data. Train delays can be immediately perceived and decisions made by the dispatcher. Traffic supervisors are also aided by complete line displays, with points and track sections. Full line displays can be either shown on monitors or on a panorama panel. The operation control and signaling system is part of the entire transportation control system for high-speed lines.

The different operating areas, such as train control, train monitoring, power supply, passenger information, station supervision, and so on, are connected by local area network to a multifunctional, integrated telecontrol system (Figure 11). Standard operating systems such as UNIX and standard hardware are used for the work stations.

The complete system handles all controlling and supervising functions (Figure 12). First, there is the already described train operation area for supervising train movements, including train operation and line overview. Other features of the transportation control system are passenger information, the public address system, track-to-train radio, and closed circuit television for monitoring platforms as well as power management for electric traction and building control. Building control includes fire protection and the supervision of elevators.

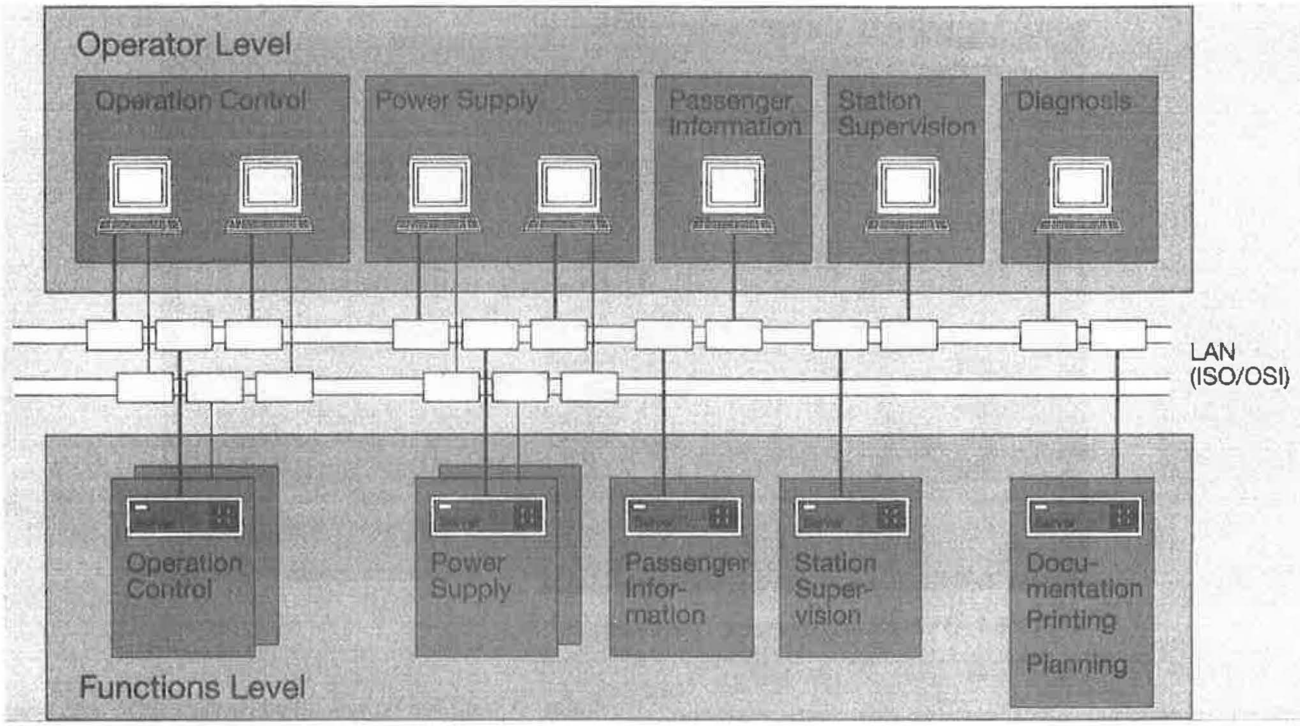


FIGURE 11 Multifunction integrated operations control center.

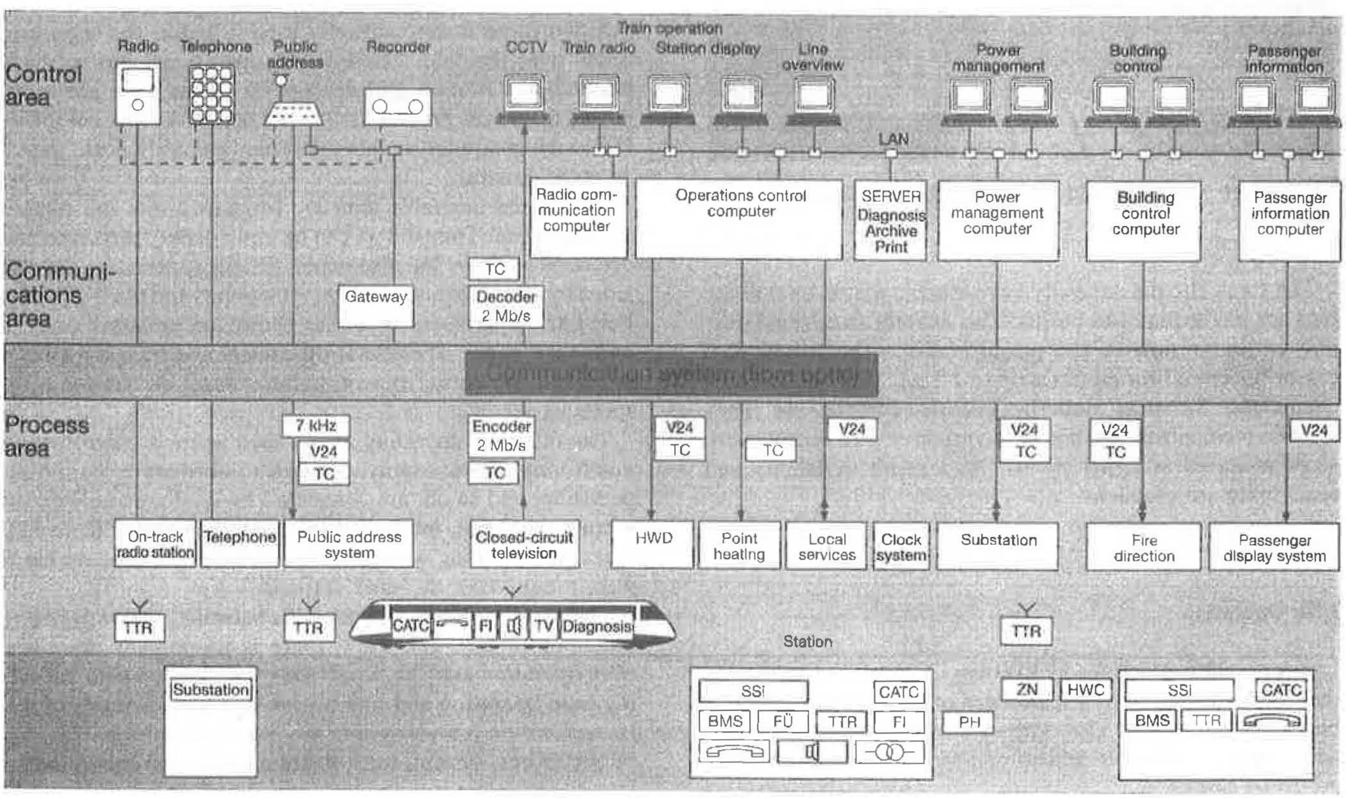


FIGURE 12 Transportation control system for high-speed lines.

Information to and from the control center is transmitted by metropolitan area network via a fiber-optic pulse-code modulated high-speed communication system.

SUMMARY

The operation control and signaling systems for the ICE high-speed lines are clearly structured. Automatic train supervision, decentralized microcomputer interlockings, and continuous automatic train control are interconnected by well-defined data interfaces. All equipment is well proved.

The system is characterized by a very high level of safety and availability. It is flexible and can be adapted very easily to new requirements. Modern electronic hardware and a straightforward structure of software guarantee successful application both now and in the future.

ACKNOWLEDGMENT

The author would like to thank the director of the technology, quality assurance, and standards department of Siemens Transportation Systems Group for permission to publish this paper.

Migration from Conventional Signaling to Next-Generation Train Control

JEFF TWOMBLY

Migration from present-day signaling systems to next-generation train control, such as Advanced Train Control Systems (ATCS), will need to be an evolutionary process. Interoperability and compatibility with existing signal systems will be essential if ATCS, integrated with management information systems (MIS), are to provide suitable returns on investment. Some of the benefits of ATCS can be realized now, some later, and probably the "train control" feature will be the last implemented, certainly where signaling is now in service. Some of the benefits sought are already being realized through developments applied to existing, traditional technology.

The introduction of computer-based equipment and high-speed communication links brought to the railroad industry promises of timely transfer of information, which could benefit operations. It promised that not only could nonvital functions be implemented, such as train health monitoring, but also vital control data could be transmitted over the communication links between the central control and the trains. It promised to eliminate the need in many areas for wayside equipment such as track circuits, signals, and line wires. But with these promises come many questions about how to implement a smooth migration path from existing signaling to next-generation train control.

There are several key issues regarding both the ease of migration and the overall total cost/benefit of converting from conventional signaling to next-generation train control. These issues include compatibility with existing signal systems, movement of unequipped trains over controlled territory, ability to implement in a piecemeal fashion, and the possibility of significant training costs for dispatchers, locomotive maintenance personnel, and train crews.

DEFINING ATCS

Definitions are a central problem in the current discussion about next-generation train control systems. Terms need to be carefully defined to avoid any confusion. One definition of an Advanced Train Control Systems (ATCS) is that proposed by the Association of American Railroads (AAR) and the Railway Association of Canada (RAC) and described in specifications generated by ARINC Research Corporation. Frequently "ATCS" in capital letters is also used to describe any next-generation train control and train management system, regardless of its resemblance (or lack thereof) to the AAR/RAC specifications. It is worth noting that, so far as is known, no existing, functioning system fully implements the AAR/RAC specifications.

For clarity's sake, in this paper "ATCS" refers only to the system defined by the AAR/RAC specifications. The term "next-generation train control" refers to other types of advanced train control systems. (Others define these systems using lowercase "atcs.") We will further define "next-generation train control" to refer only to systems that provide *actual control of train movement*, as opposed to management information systems (MIS).

Management information is one of the two principal functions of advanced train control systems: issuing work orders, locomotive health monitoring, crew calling, event recording, and planning dispatching strategy, to mention a few. The second principal function is vital and nonvital train control: throwing switches, moving trains, and stopping trains.

The railroads must weigh the expense of tying both management information and *train control* functions into a single package. Next-generation train control systems can be designed to be compatible with MIS just as existing train control systems are compatible with these systems (e.g., crew calling, event recording, and dispatcher assistance).

BENEFITS OF ADVANCED SYSTEMS

Although not everyone in the railroad industry agrees on the choice of a single advanced system, probably everyone agrees on the desirability of the benefits an advanced system can provide: the ability to move more trains more efficiently on existing tracks, to maintain the existing high standards of safety, and to reduce the costs associated with the signal and control system, especially the costs of wayside equipment. In addition to these an important practical consideration can be added: interoperability/interchangeability with existing signal and control systems, facilities, and personnel. Beyond these, there are additional benefits that can be realized when an MIS is added to a next-generation train control system.

With ATCS, for example, one of the key features will be the integration of MIS. Much of the information needed to provide more efficient train handling (hence reduced fuel consumption and less damage to lading) comes from MIS waybill and car equipment data on the type and contents of each car (Universal Machine Language Equipment Register, or UMLER). In addition, MIS provides the number and identity of cars in the train.

General Railway Signal Corporation, P.O. Box 20600, Rochester, N.Y. 14602-0600.

Traffic Capacity

Some proponents of ATCS assert that the "moving-block" feature could increase traffic capacity by allowing closer headways between trains. Although this may be true for the traffic mix on European railroads, it does not appear to have been demonstrated for the traffic typically encountered on U.S. and Canadian railroads—a large number of long freight trains and a small number of passenger trains. For the U.S./ Canadian situation, fixed-block systems with blocks of suitable length can provide comparable traffic capacity. To increase capacity, next generation train control systems will allow shortening block length easily and at a reasonable cost.

Interoperability/Interchangeability

The definition of interoperability/interchangeability needs to be expanded to include interoperability with existing systems. Many of the benefits of interoperability/interchangeability are obvious, especially as they relate to providing a smooth transition from existing systems to advanced ones without the problems of the "D-Day" cut-over required when changing to a completely different type of system. These benefits are discussed in greater detail below.

Training/Facilities

There are also some less visible benefits of interoperability/ interchangeability. When a new system closely follows the framework of the existing one, very little retraining is needed for dispatchers and engine operators, and existing centralized traffic control (CTC) offices can still be used.

If the benefits provided by a completely different system were significant enough, a radical change might be justified. However, other advanced systems offer similar benefits to ATCSs, plus interoperability/interchangeability. Indeed, many of the benefits can be obtained by upgrading conventional systems. Many products from signal suppliers are designed for compatibility with train management systems, including the train management functions of ATCS, based on meeting the ATCS specification for communications. (Note that following the ATCS specification may result in some degradation of performance from conventional systems because of factors such as message response time.)

THE TRANSITION TO ATCS

Train operations and control can be classified as follows:

- Dark territory (no signaling) operated by timetable and train orders, direct traffic control, or track warrants;
- Manual block signaling;
- Automatic block signaling with timetable and train orders or track warrant system;
- CTC or other traffic control system; and
- Cab signaling with or without automatic train control or automatic train stop.

To overlay or apply ATCS to any of these situations will require a smooth transition period, not only to install the digital communications network and the computers on the locomotives and at dispatching headquarters, but also for training dispatchers and operating and maintenance personnel.

Probably the easiest transition will be that of installing ATCS in dark territory. Next easiest will probably be in manual block signaling territory. Next would be in automatic block signaling territory, then in CTC territory, and finally in territory with cab signaling, whether with or without automatic train control or train stop.

In all cases a transition path must be determined that will allow a smooth, orderly change with no disruption in operation, including the ability to handle non-ATCS–equipped trains. This is important not only because of the time required to install wayside and motive power equipment, but also because of the time required to train personnel, especially for handling maintenance of motive power units.

Note that the Federal Railroad Administration will probably not allow ATCS to be installed to replace present signal systems without a period of parallel operation, until ATCS can provide proof that they can be as safe as the signal system currently in service.

Also note that although ATCS provide the movement authority to the train, similar to a train order, the train will still be guided by any wayside signals in regard to safety of the movement. Obviously the engineer would stop the train when it encounters a red signal. Full ATCS would also provide a display on the locomotive computer, telling the engineer to stop in approach to a red signal.

Carrying this further, to the overlay of ATCS on CTC, one school of thought is that the safety features of CTC would be retained, though possibly wayside signals could be retired because the ATCS on-board computer would provide signal displays in the cab.

As for interlockings, the local control would probably be turned over to the dispatching center, but again the interlocking's safety features would be retained. The engineer's cab display would allow the train to move through the interlocking only if it were safe to do so. In many instances the interlocking might never be taken under ATCS control because of its complexity.

Due to the tremendous investment in traditional signaling, the ability to address these transition issues will have significant impact on the acceptance and implementation of future train control systems.

TRANSITION TO A NON-ATCS ADVANCED SYSTEM

Past experience in upgrading systems suggests there will be a relatively long period during which the old and new must coexist, with the old system remaining on less-used routes for several years or longer.

The most likely scenario would be that a currently dark section would be added to an existing signaled section, where this had not been cost-effective using traditional technology. This highlights the need for next-generation train control systems that maintain interoperability with conventional systems.

Another scenario that highlights this requirement for compatibility is overlaying a new technology over existing signaled territory. This could occur when portions of the existing systems are in need of major repair or replacement, or when it is desired to keep a major portion of an existing system, such as a large interlocking. The differences in implementation and operation between currently proposed schemes and existing signaling approaches do not allow for the coexistence of the two systems on the same territory; however, there are some very practical reasons why coexistence should be provided for. These include installed investment, cost of change greater than benefit, existing practices, and many other practical considerations. The compatibility issue is critical to the success of any future control system.

One alternative approach to delivering the benefits sought by our industry—one that addresses the compatibility issues—is based more on existing signaling techniques and practices (Figure 1). It includes a nonvital office work station (or panel board) handling the train movement controls. This computer then feeds control information to centrally located vital Boolean expression evaluators. These vital controllers contain all of the same vital field logic that would formerly have been stored in local controllers located in wayside bungalows. These vital controllers communicate by radio links directly with both the car-borne equipment carried in the locomotives, and the wayside units that handle the physical control of the ground equipment.

Position locating is achieved through the use of wayside transponders that mark block boundaries. These blocks can be short, train-length blocks. The equipment required is very inexpensive. For overlaying on existing systems, the transponders are located at existing block boundaries (i.e., where signals are now located.) Equivalent logic will be provided in the office to allow wayside signals and cab signals to be in

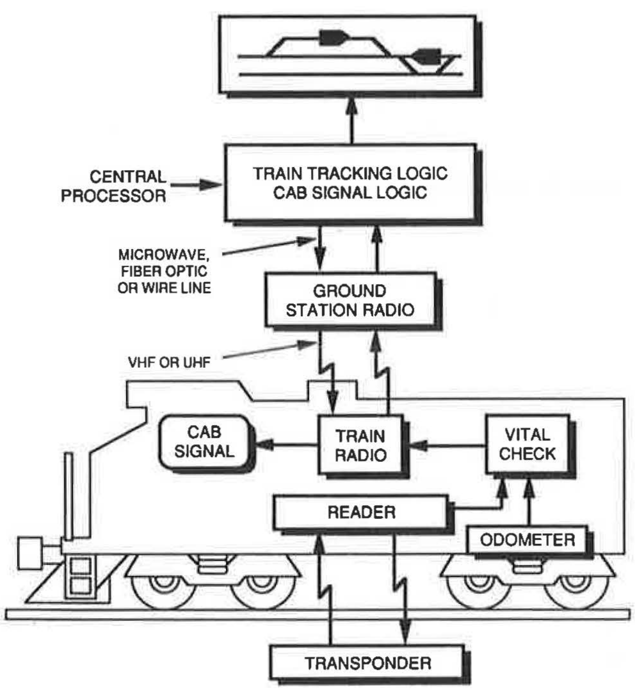


FIGURE 1 Alternative approach.

agreement. Eventually, wayside signals could be removed or block lengths redesigned very economically.

This approach makes use of existing, proven practices and technologies. The office controller is a traditional CTC device, the vital controllers perform the same functions as existing microprocessor-based interlocking controllers, and the fixed-block approach provides a very economical element of compatibility with existing systems.

Because the office computer is the same as today's, and a fixed-block design is used, this system can easily be added to an existing signal system. It would require adding a separate code line (or radio-based code transmission) to the office computer, its implementation technique transparent to the rest of the system.

This approach to next-generation train control offers many practical advantages, particularly in the area of compatibility. In this system, an existing office computer, or any CTC computer (readily available from a number of suppliers) could be used. Because the existing basic principles of office control are used, and every component of the system is based on proven, safe technological and operational practices, the confidence curve in demonstrating the safety of the system will be much shorter than if new approaches are used. This evolutionary approach to next-generation train control provides benefits comparable to revolutionary approaches, but far fewer risks.

Operational compatibility with existing systems is another benefit with practical implications. The block approach presents the same operational information to the engineer regardless of whether this information is presented on the cab display or on wayside signals. This compatibility allows for smooth travel of trains over territory controlled under existing technology or under new technology. It also allows the simultaneous operation of both systems during the interim phase when cab and wayside signals are to be used in tandem. And it minimizes the extent of training required when installing the new system. Signal aspects or speed limits, displayed in the cab with meanings similar to existing wayside signal aspects, are an efficient way to convey information to the operator and do not require extensive training to learn how to understand new, more complex material.

This coexistence of existing and next-generation train control provides a practical scenario for migrating from an existing system to a next-generation system. This allows for piecemeal migration over time as opposed to the alternative, which calls for the instant abandonment of existing systems and cut-in of new systems. Thus, initial installation of next-generation systems can be made in those segments of a territory where it is most cost-effective. And such train control technology is compatible with developing management information technologies, whatever form they may eventually take.

CONCLUSION

The transition to a next-generation system must be evolutionary. Many benefits can be obtained before the last piece is installed—actual train control.

There are three requirements for a smooth transition to a next-generation system:

1. Compatibility between new and existing equipment and systems during the transition period, which may take several years;
2. Extensive training of dispatchers, operating personnel, and maintenance personnel; and
3. Investment and benefits realized incrementally.

A fourth point might be added: educating and working with governmental regulatory agencies to foster cooperation with the railroad industry and suppliers in making the transition to the new systems.

In closing, it should not be overlooked that many of the benefits promised by ATCS are being realized today. Through the application of vital electronic interlocking controls, vital electronic track circuits, vital electronic cab signals, radio transmission of control data, desktop central offices, and remote diagnostic systems, traditional technology is being upgraded economically. Maintenance requirements are being reduced, training simplified, line wires are coming down, and information is being exchanged electronically, making railroads more competitive, more profitable, and safer than ever.

AUSTRAC: The Australian ATCS

KEITH LUGSDEN AND RON DAVISON

AUSTRAC is the Australian implementation of the North American Train Control Systems (ATCS) series of signaling specifications. AUSTRAC is being installed on Australian National's Trans-Australian and Central Australian lines. AUSTRAC differs from ATCS in several respects. The development of AUSTRAC pinpointed some problems found with the ATCS specifications and also led to the incorporation of unique features into AUSTRAC.

Australia's present rail network is operated by individual state governments, the exception being the federal government-controlled Australian National (AN) infrastructure across what was the old South Australian country network. Not only are there five independent systems involved in cross-country and interstate operations, the problems and efficiencies are further frustrated by a lack of common gauge. Queensland and Westrail have narrow and standard gauge, New South Wales and AN have standard guage, and Victoria has broad and standard gauge. All of the systems have independent operating rules, different communications and signaling systems, and a variety of management information systems (MIS). With this mismatch of infrastructure and duplication of management and organization overhead, rail transport finds it very difficult to compete with its road and marine competitors. Railways have to maintain their own right-of-way, whereas the competition operates through publicly funded roads and ports allowing them the benefit of moving in and out of transportation corridors at will without exposure to loss of substantial capital investment.

In late 1989, the National Freight Initiative (NFI) was established with the participation and support of the individual railroad systems, major rail users, and the federal government to determine the feasibility of a national rail freight organization to perform the interstate rail transport task in Australia. This organization will be an incorporated company integrated across state borders and operating at cost levels significantly below those now prevalent in the rail system.

In 1989, AN was looking to introduce some form of control and communications infrastructure on the Trans-Australian and Central Australian railways (TAR and CAR respectively). Fortunately, AN's managing director, Russel King, had the forethought to consider the future integrated network. From this coast-to-coast integrated freight network requirement the concepts of AUSTRAC were formulated. Consequently, AUSTRAC has been designed to meet AN requirements for an efficient signaling system on the TAR and CAR. The design is based on the principles of Advanced Train Control Systems (ATCS) specifications. The TAR and CAR cover about 2 000 km of single track east from Port Augusta to Kalgoorlie and north from Tarcoola to Alice Springs, respectively. AUSTRAC is being installed at 41 of the crossing loops, mostly on TAR. It is the first system installed in the world to incorporate the facilities described as ATCS Level 30.

International Railroad Systems Pty, Ltd., 209-217 Wakefield Street, Adelaide, South Australia, 5000.

OVERVIEW OF ATCS

ATCS will not be described in detail here because this topic has been covered elsewhere (*1*). However, there are some important differences between ATCS and AUSTRAC, and some of these will be highlighted in this paper. In this section, a brief overview of ATCS is presented.

ATCS specifications were produced by ARINC Research Corporation under the control of a steering committee formed by the American Association of Railroads (AAR) and Railway Association of Canada (RAC). The specifications have been designed "to facilitate compatibility and standardisation without limiting the individual design approaches of individual suppliers" (*2*). As a practical consequence of this, the specifications as they stand to date are overly constraining in some areas and too loose to achieve interoperability in others. In addition, the specifications are generally in a form that, although readily understandable by signaling engineers, are not very amenable to implementation in software. Therefore, the design of AUSTRAC has highlighted the need to clearly define many concepts that are not covered adequately in the ATCS specifications.

Features

ATCS can be broadly characterized by the following features:

- Management of moving-block track occupancies by a centrally located dispatch center;
- Automatic tracking of the location of trains using a data communications network;
- Issuance of movement authorities over the data network;
- Enforcement of speed and authority limits;
- Automatic control of a diverse range of wayside devices such as motor-operated points and highway crossing devices;
- Reporting of train, switch points, and ATCS component health; and
- Three different modes of operation already defined (known as ATCS Levels 10, 20, and 30) and a fourth envisaged (ATCS Level 40).

ATCS Specification 100 (*2*) defines the three operating modes as representing different levels of sophistication:

- Level 10—Centralized route and block interlocking logic in a computer-aided dispatch system;
- Level 20—Automated transmission of movement authorities and other instructions via a data communications system;
- Level 30—Automatic location reports and full train tracking within the dispatch system; and
- Level 40—Full field locking.

Control Flows

One of the major obstacles to be overcome when attempting to implement the ATCS specifications is caused by its functional description by control flows (*2*). These were still in early draft form in 1988 when the AUSTRAC project commenced. Once a detailed analysis was performed on the control flows, several shortcomings were uncovered: First, there are numerous errors, inconsistencies, omissions, and ambiguities present in the control flows. These include "dangling control" (where a control flow has a path with no defined end point), "hanging control" (where a control flow may be suspended pending some event that may never occur), and conflict with other ATCS specifications (such as messages that do not have the required information, or messages that, according to the control flows, are never sent). Second, the lack of a coherent structure (e.g., hierarchical) makes many of the control flows impossible to comprehend. For example, a control flow may invoke another control flow that may eventually invoke the first one. Third, there are no textual descriptions of any of the control flow functions. This means that the implementor must comprehend the functionality of ATCS just from three or four words written in control flow boxes and by correlation with message definitions in another ATCS specification. Finally, there is no recognition in the control flows that they are to be implemented in software on distributed processing systems (central office, waysides, and locomotives) as defined by ATCS itself. For example one control flow may require action by three physically different devices and hence the control flow cannot be attributed to any one device.

ARINC and AAR have acknowledged that there are problems with the control flows and are currently trying to produce a more coherent set that should facilitate easier implementation in software and hardware. In fact, ATCS specifications have been under constant development since the first drafts appeared a few years ago. This has meant that the specification of AUSTRAC has had numerous changes to its original baselines and has required keeping ATCS specifications under constant review.

OVERVIEW OF AUSTRAC

AUSTRAC has nearly all of the features of ATCS described in the preceding (see Figure 1), but with some important differences that generally relate to AN's implementation requirements. In particular, traffic density was considered sufficiently low on TAR and CAR to allow the removal of some of ATCS' more complex (and hence more costly) features. Consequently, AUSTRAC now has the following characteristics:

- Fixed block only (i.e., an entire section of track is marked as occupied until the entire train has vacated it);
- The communications backbone is based upon an optical-fiber cable that connects a central communications controller (known as the FEP/CC) on a point-to-point basis with a series of radio base stations located near to the track;
- All wayside installations and equipped locomotives are connected to the network by data radio;
- Fail-safe hardware is employed in all safety-critical areas—on-board computers (OBCs), wayside interface units (WIUs), and office interlocking computers (OICs);
- Formal specification and analysis of all safety-critical software;
- Vital interlocking of safety-critical commands from the central office using a pair of safety computers;
- Vital relays inserted between WIUs and points machines;
- Motor and hand-operated points are the only wayside devices currently within the scope of AUSTRAC;
- Software emulation of track circuits, using automatic train location technology, in such a way that real track circuits can be added at a later stage without altering hardware or software;
- Normal operation at Level 30, with partial failure resulting in a degradation to Level 20 or Level 10 as appropriate; and
- The addition of an end-of-train subsystem to determine accurately when the train clears critical points.

CENTRAL OFFICE

AUSTRAC's central office will be located next to AN's headquarters at Mile End in suburban Adelaide. It consists of a dispatch system, safety interlocking system, and a data communications controller.

Dispatch System

The dispatch system is hosted on six operational, networked Sun Sparcstation (Sparc is a trademark of Sun Microsystems, Inc.) work stations and two spare Sparcstations (see Figure 2). Two of the operational Sparcstations are called train control work stations (TCWs) and will be actually located in Port Augusta where AN operationally controls TAR and CAR lines. The TCWs enable the train controllers to enter various authorities, block track sections, receive alarms or other information from the network, and perform other train management functions (see Figure 3). Two other work stations are designated as control computers (CCs). One CC is operational while the other is on standby. The CC controls the release of authorities and other commands from the central office to the field and permits history logging. Another Sparcstation is used as the system manager's work station (SMW). The SMW permits AN's AUSTRAC system management personnel to interrogate the network for fault diagnosis or other information. The SMW can also be used to remove AUSTRAC equipment from service or put it back into service.

AN's traffic information management system (TIMS) is accessed via the TIMS interface work station (TIW), which is the sixth operational Sparcstation. The TIW allows train in-

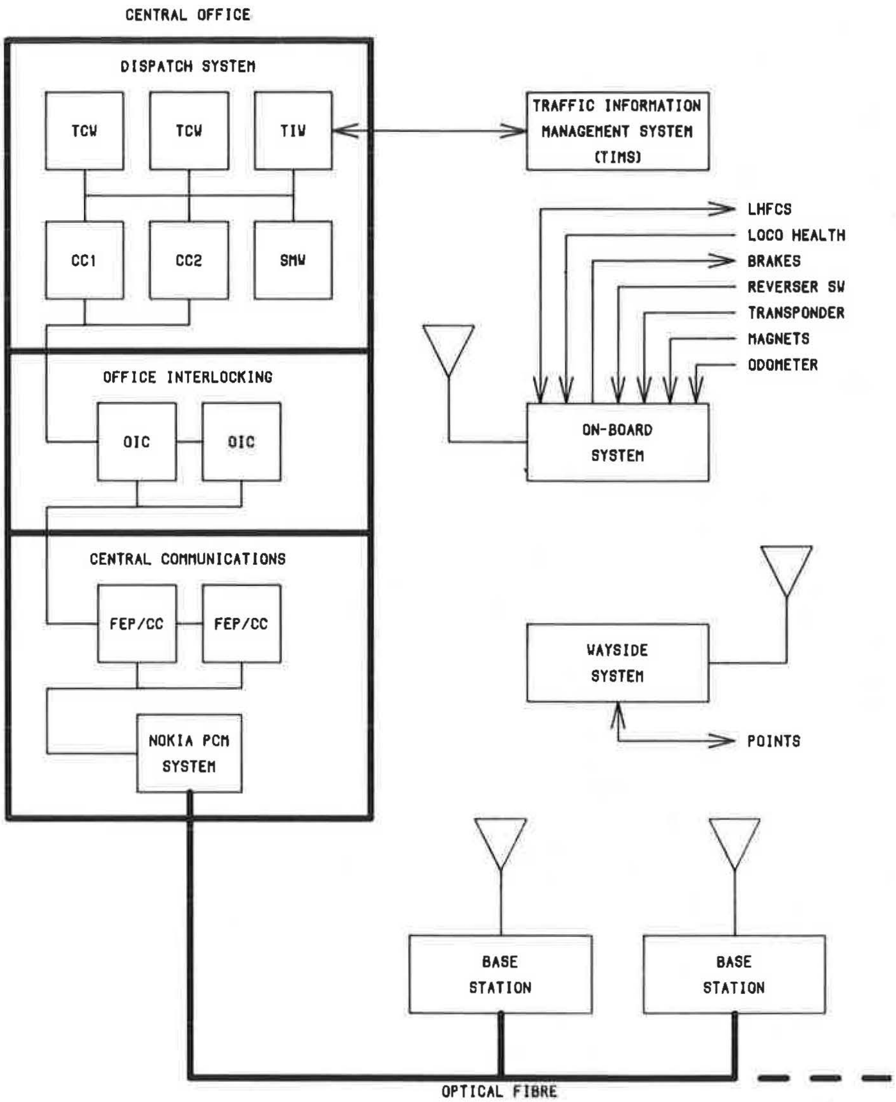


FIGURE 1 AUSTRAC architecture.

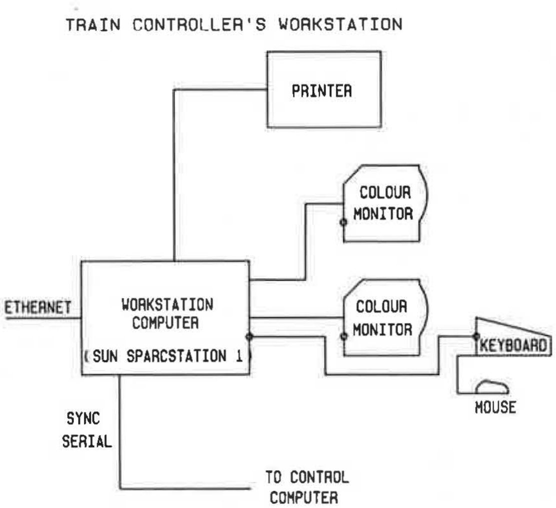


FIGURE 2 Central office.

formation, such as consist, to be downloaded to AUSTRAC. The TIW also provides train running data, such as train location updates, to be returned to TIMS.

Office Interlocking System

The office interlocking system is based on a pair of safety computers known as office interlocking computers (OICs). One OIC is active while the other is on standby. Each OIC has a communications processor associated with it called a quad-serial input/output (QSIO). The QSIO connects the OIC to the dispatch and communications systems. Each QSIO is controlled by an Intel 80C186 microprocessor. The role of the OICs is to ensure that no commands sent to the field could lead to a hazardous situation. To facilitate this, the OICs maintain data bases of train locations, points status, authority information, and other necessary data. The OICs also respond to emergency conditions, such as the overrun of authority

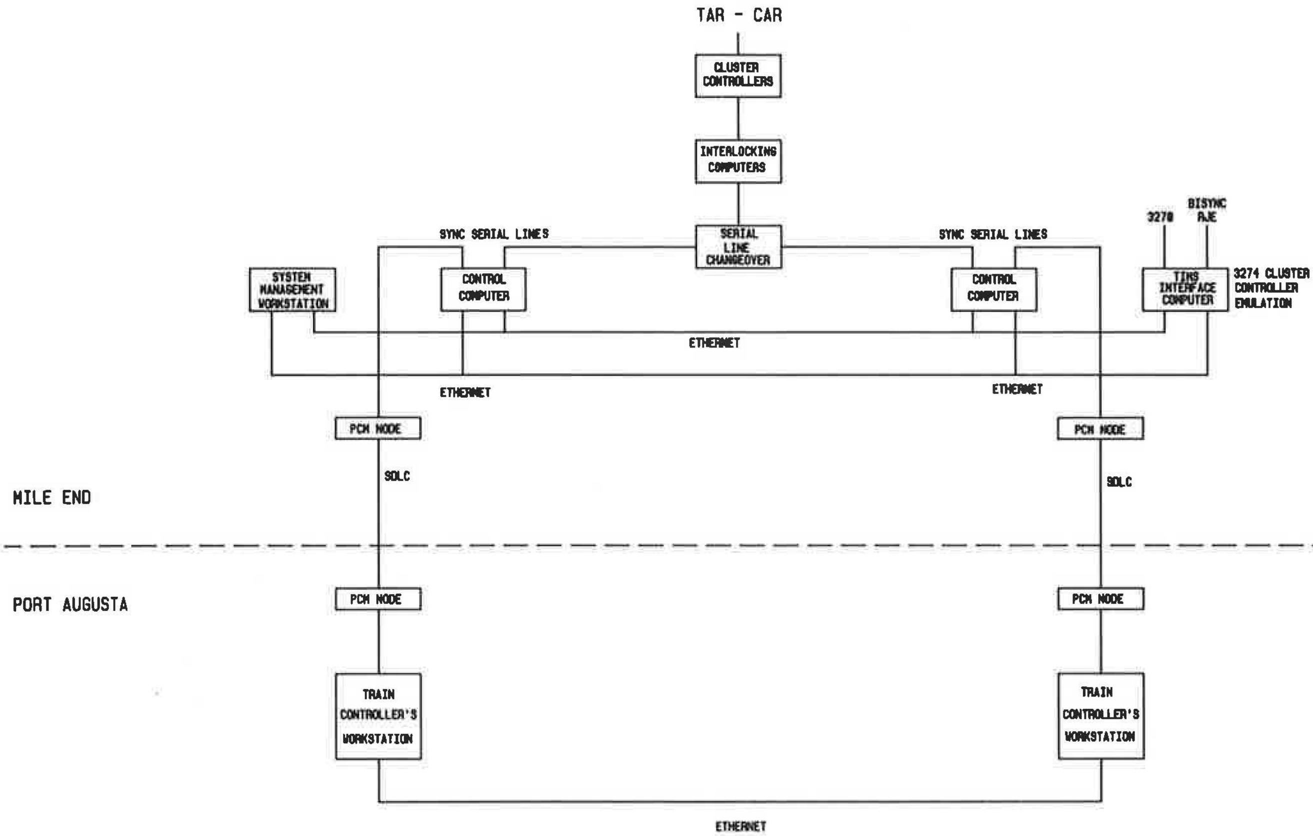


FIGURE 3 Train controller's work station.

limits, by issuing appropriate commands such as restricting the authority limits of opposing trains.

The OICs, as are all AUSTRAC safety computers (i.e., OICs, WIUs, and OBCs), are each based on a pair of VIPER microprocessors. The VIPER chip has undergone a high degree of formal specifications and analysis in order to prove that it can move between its defined states only in a predictable manner.

To enhance safety, the VIPER has no interrupts, no stack, a limited range of addressing modes, and a reduced instruction set. Each VIPER operates in a closely coupled pair. If any discrepancy is detected, such as illegal bus transactions, out-of-range addresses, or parity errors, the VIPER hardware will halt. Note that all VIPER memory contains 8 parity bits for each 32-bit word.

Each safety computer includes watchdog, fault detection, and fault injection circuitry on all vital hardware. This ensures that not only can they detect faults, but they can also determine whether the fault detection circuitry is itself faulty. In addition, software is able to detect other fault conditions such as incorrect I/O module addressing and I/O connectors that are plugged into the wrong I/O module. Any detected failures cause an automatic disconnection of the computer from its vital functions.

Central Communications System

The central communications node is a combined front-end processor and cluster controller (FEP/CC). The FEP/CC interfaces with the optical-fibre–based communications backbone to access the ATCS base stations. Refer to (*2*) for a description of the AUSTRAC communications backbone.

All AUSTRAC data communications conform with the applicable ATCS specifications (200 series). These in turn generally follow the Open Systems Interconnect (OSI) seven-layer model (e.g., ISO-3309, ISO-4335, and ISO-7809). This includes the X.25 error-controlled link-layer protocol (*3*) and a transport layer that features a 31-bit vital cyclic redundancy check (CRC) code. Only the safety computers know the algorithm to encode and decode the vital CRC. This ensures that only vital software can initiate vital actions. The FEP/CC is duplicated for reliability reasons. The FEP/CC can control up to 64 base stations. Provision has been made for an expansion FEP/CC to be added that would enable connection of a maximum of 128 base stations. The initial AUSTRAC configuration has 48 base stations. The FEP/CC comprise's one QSIO module for the FEP and one QSIO module for each CC. Each CC controls four base stations and up to 16 CCs can be installed in a single FEP/CC.

WAYSIDE SYSTEM

AUSTRAC's wayside system consists of a wayside interface unit (WIU), solar power supply, wayside communications package (WCP), and a set of vital interlocking relays.

Wayside Interface Unit

The WIU is comprised of a safety computer (similar to the OICs), a QSIO communications processor, an environmental

input module, and between one and three points control (PC) modules. The QSIO provides the communications interface with the WCP. The WIU's environmental input module permits the WIU to read the wayside identity and to detect conditions such as hut light on and hut door open. Each PC module can control one points machine (or detect one set of hand-operated points) and interface to two track circuits. The WIU ensures that the points are kept in the state previously set by the OIC while a train is in the vicinity of the wayside. The WIU continually monitors the health of all wayside subsystems. It will shut itself down safely if it detects that it no longer has control of its vital functions.

Solar Power Supply

The wayside system is powered by a solar power supply that delivers 24 Vdc to the points machine and 12 Vdc to everything else. The batteries have sufficient capacity to sustain operation of the wayside system for up to a week without sunlight, assuming normal usage rates are maintained. The power supply provides outputs to the WIU to indicate low battery voltage so that the WIU can shutdown the wayside installation at an appropriate time (e.g., not while a train is moving over the points).

Wayside Communications Package

The WCP consists of an ATCS-compatible modem supplied by Harris Corporation, a data radio supplied by UniLab, and an antenna. As the modem and radio are not fully compatible with each other, the WIU's QSIO must translate commands and responses between the two devices. The radio system has been designed to be fully ATCS-compatible, but operates in the 400 MHz band as required by the Department of Transport and Communications. It features forward error correction (FEC) and a channel access procedure that employs busy-bit insertion. The WCP communicates by radio with the FEP/CC via its nearest functional base station.

Vital Interlocking Relays

Unlike ATCS, AUSTRAC specifies the use of vital interlocking relays between the electronic circuitry and the points machines. Their usage makes it virtually impossible for the points to be moved unintentionally even in the highly unlikely event that an undetected software or hardware fault occurs. For each PC connected to a points machine, three vital relays are provided. These are the switch locking relay (SLR), normal control relay (NCR), and reverse control relay (RCR). The points can only be moved if the software operates the relays in the correct sequence. The WIU checks the state of the vital outputs every 50 ms to determine whether any fault conditions are about to attempt an unwanted relay operation. As each relay takes more than 50 ms to operate, AUSTRAC ensures that only intended actions can cause movement of the points.

ON-BOARD SYSTEM

For fully equipped trains (i.e., Level 30), AUSTRAC's on-board system comprises an on-board computer (OBC), mobile communications package (MCP), driver's display and control unit (DCU), train location subsystem, end-of-train (EoT) subsystem, enforcement subsystem and interfaces to the locomotive health system (LHS), and proposed long-haul fuel conservation system (LHFCS). A voice radio system is also supplied so that drivers can keep in contact with train controllers throughout AUSTRAC territory. Note that unequipped and partially equipped trains can also safely traverse AUSTRAC territory, although a higher level of driver and train controller interaction is required.

On-Board Computer

The OBC has a safety computer similar to that of the WIU, except that instead of PC modules it has interfaces to the EoT equipment, odometer, reverser switch, and enforcement subsystem. It also has an extra QSIO to interface it with the DCU, train location system, LHS and LHFCS. The role of the OBC is the safe operation of all locomotive functions. In particular, its role is to accept and validate authorities issued to it by the central office, to keep track of the train's location, communicate with the driver through the DCU, and apply enforcement where safety is threatened.

Mobile Communications Package

The MCP is very similar to the WCP. The main difference is that the EoT receiver (see below) is also incorporated within the data radio enclosure, which is then known as a lococom radio.

Display and Control Unit

The DCU permits text and graphics to be displayed to the driver. It has a touch-sensitive screen so that the driver can acknowledge various actions and provide certain information to AUSTRAC. In normal (Level 30) operation, the driver needs only to provide minimal information, such as the EoT serial number. However if, for example, the location subsystem should fail, the train reverts to Level 20 operation and the driver must enter the train location information manually. Because a touch-sensitive display is used, no valuable cab area is required for a bulky keyboard. The DCU has a 640 $\times$ 200 pixel display arrangement on a 190 mm-wide $\times$ 95 mm-high monochrome screen. Bit-image graphics permit animation (such as the approach to crossing loops) and the display of any international language (Danish, Thai, and Mandarin versions have already been demonstrated).

Train Location Subsystem

The train location system comprises track-mounted transponders, locomotive-mounted transponder interrogators, and in-

terfaces to the locomotive's own odometer and reverser switch. Together, these devices permit the OBC to determine its location to within the limits of the odometer, which is calibrated every time that a transponder is passed over. The odometer is calibrated by comparing its pulse rate with a known distance (i.e., the distance between two calibration transponders). Transponders are located at strategic locations, such as crossing loops, so that the odometer is not required to be especially accurate (AUSTRAC allows for up to 3 percent error caused by slippage and wheel wear without any danger of a train overruning its authority limits). The transponder equipment is sourced from a third party and interfaces directly to the OBC.

End-of-Train Subsystem

The EoT subsystem consists of track-mounted magnets located at important control points (such as the entry point to a crossing loop), a magnet detector and radio transmitter mounted on the coupling of the last vehicle of the train, and an EoT receiver in the locomotive that passes the magnet detection information to the OBC.

The EoT subsystem is used by the OBC to determine when the train has cleared critical areas, such as the points-fouling area. This is achieved by a data radio transmission from the rear of the train to the OBC whenever a magnet is detected. In the same way, the EoT subsystem also permits the OBC to determine that the train is intact. Because the transponder interrogator is located on the lead locomotive and the magnet detector is mounted on the rear of the last vehicle, the OBC can determine the length of the train by calculating the distance traveled between a transponder and a magnet.

The EoT subsystem is an AUSTRAC addition to ATCS. ATCS specifies the use of track circuits and duplicated odometers. The track circuits are used to determine when a train has completely cleared a critical track section. Duplication of odometers enables the OBC to reliably determine the location of the end of the train.

Enforcement Subsystem

The enforcement subsystem allows the OBC to apply the train's brakes if the authority or speed limits have been violated or are likely to be violated. Failure of the OBC will also cause enforcement to be activated. For AUSTRAC, enforcement would normally be enabled for the entire journey. Enforcement would be disabled while the train is outside AUSTRAC territory and when shunting inside a yard. The OBC continually calculates the train's speed, location, and stopping distance. The stopping distance is based on the track gradient, train speed, and train class. If this distance indicates that a certain limit may be exceeded, the driver is presented with a warning. If the warning is ignored (i.e., the driver fails to reduce the train's speed to within the appropriate limit), the OBC will apply enforcement. The train driver has access to an enforcement override switch. Such action will cause an alarm to be generated on the TCW.

Locomotive Health

The OBC has an interface to the existing locomotive health system (LHS) installed on most of AN's locomotives. The LHS simply permits locomotive running data to be transmitted back to the central office for examination by AN's maintenance staff. Consequently, up-to-date information will be available at all times instead of only when the locomotive returns to the maintenance yard. There are currently no plans for AUSTRAC to process these data itself.

LHFCS Interface

The OBC has an interface to the long-haul fuel conservation system (LHFCS) being developed by International Railroad Systems. This is a bidirectional interface to allow train running data from the OBC to be analyzed by the LHFCS and for driving advice (drive, coast, or brake) to be returned to the OBC for display to the driver on the DCU.

QUALITY ASSURANCE

Because partial or total failure of AUSTRAC has safety implications, a higher than normal level of quality assurance (QA) is being applied to the project. Some of the features of the QA used on the AUSTRAC project are as follows:

- Safety and hazard analysis by a third party;
- A company quality system based on AS3901 (ISO9001 equivalent);
- Application of standards such as software development to DOD-STD-2167A and specification practices generally conforming with MIL-STD-490A;
- A large QA team employed by both the contractor and AN to ensure compliance with all appropriate specifications, procedures, and standards;
- Formal specification of all safety-critical software;
- A team of verification and validation (V&V) experts employed by both the contractor and AN to analyze safety-critical software;
- Use of system specification methodologies and tools, such as Ward-Mellor, Statecharts, Excelerator, and Statemate (Excelerator is a trademark of Index Technology and Statemate is a trademark of i-Logix Inc.);
- Use of MALPAS (MALPAS is a trademark of Rex Thompson & Partners) (Malvern Programme Analysis Suite) static analysis on safety-critical code;
- Careful prototyping of hardware and software; and
- Rigorous formal testing of safety-critical hardware and software.

SAFETY CONSIDERATIONS

Unfortunately, safety can never be guaranteed 100 percent in any railway signaling system. As always, safety depends not only on the signaling system, but also on the associated operating rules and people who use the system. However,

AUSTRAC does have some important safety features, some of which have been previously described:

- Use of vital CRCs to ensure that only empowered software can generate or act on potentially hazardous commands;
- Design of fail-safe hardware that can only fail in predictable safe modes;
- An analysis of the hazards likely to be encountered by the system;
- A rigorous hardware and software development program;
- Use of the latest V&V techniques on safety-critical software;
- Diligent QA and safety teams; and
- A joint safety management group responsible for resolving hazards both during the development phase and over the life of the project.

CONCLUSION

AN's decision to initiate the development of AUSTRAC will put the railroad among the world leaders in railway control and information technology, and leave AN ideally placed to take full advantage of the Australian government's National Freight Initiative.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the assistance provided by colleagues at International Railroad Systems Pty. Ltd. in the preparation and review of this paper.

REFERENCES

1. A. Blakeley-Smith. *ATCS—The Universal Signaling System*. International Railroad Systems Pty, Ltd., Sydney, Australia, July 1989.
2. *System Architecture Overview*. ATCS Specification 100. ARINC Research Inc., Jan. 1989.
3. CCITT Recommendations X.1–X.32. *Data Communication Networks: Services and Facilities, Interfaces*. 9th Plenary Assembly, Melbourne, Australia, November 1988.

ASTREE: A Global Command, Control, and Communication System

PATRICE H. BERNARD

ASTREE is a research project on a global command, control, and communication system for train control and traffic management developed by the French National Railways (SNCF). It is based on maintaining a distributed data base of up-to-date, accurate, and comprehensive representation of route layout and train progress. Computer applications use the image of the actual situation gathered in the data base to make (or, in some cases, to help human operators make) all decisions regarding route setting and train control. The results are transmitted back to wayside equipment and to locomotives, where they are displayed and enforced, if need be, by a penalty brake application. Development began in early 1986 and has enjoyed important resources. After a few years of definition of requirements, technological component testing, and architecture design, the first integrated test is under way with extensions planned. A large-scale, yet not vital, test is slated for 1992. If the results from the tests meet the technical and economic expectations, a full-fledged implementation could be operational on one route by 1997.

In 1986 the French National Railways (SNCF) embarked on a large-scale research program to develop a global command, control, and communication system for train operation and railroad management called ASTREE. This paper highlights those points where the emphasis or the approach is felt to be different from those of Advanced Train Control Systems (ATCS) level 40 and reports ASTREE's current status and anticipated development.

ASTREE'S MAIN FEATURES

Integrated Approach

ASTREE is based on few simple ideas:

- Equip every train with location and communication capabilities.
- Connect a distributed computer system to switches—so as to monitor routes—and to trains to keep track of their positions.
- Keep in a distributed data base an up-to-date, accurate, and comprehensive image of the road and of the trains.
- Develop software tools that build on this representation to make (or help make) all decisions related to train control and network management.
- Transmit their output back to wayside equipment and to trains in the form of controls (such as lining switches), restrictions, authorities, or advice.

SNCF, Direction de la Recherche, 45, rue de Londres, F75379 Paris Cedex 08, France.

Sophisticated Train Location System

ASTREE's positioning philosophy relies on the following principles:

- The route set ahead of a train is monitored by ground-based equipment and the train progress on this route is determined by the train itself.
- The train location is defined by an interval for vital applications and by an average for nonvital applications.
- The location interval is determined by an odometric unit, which blends and processes information from a number of sources.

Ground-Based Route Monitoring and Train-Based Progress Measurement

Avoiding rear-end, head-on, or sideways collisions requires knowing not only on which track a train is running, which could be left to the train itself to determine, but also what route is set ahead of it. This is necessary to make sure that the authority to proceed granted to one train does not conflict with that granted to another train and not yet released by the latter train, or that access is not prohibited for some other reason such as maintenance of way.

Monitoring the route from the ground is also a natural consequence of the fact that in most cases routes are set by ASTREE.

The decision to let the train determine its own position on the known route is based on several considerations:

- A train must constantly know its own position and speed so that it may brake if, and only if, this is necessary to avoid exceeding its limit of authority.
- A train must be able to report its most recent position when the ground-based control center needs this information, so as to efficiently release and allocate track segments on a geographically continuous basis.
- This continuous on-board knowledge is best acquired from on-board sensors.

Definition of a Vital Location Interval

For vital applications, ASTREE defines two positions:

- A foremost position, such that the probability that the head end of the train be actually beyond this position is lower than a predetermined very demanding figure; and

- A rearmost position, such that the probability that the rear end has not cleared this position is lower than this predetermined figure.

The foremost position is used by the train itself for cab-signaling purposes or brake application. It is used by the ground control center to set switches early enough and issue warnings such as to a road crossing protection. The rearmost position is used by the control center to take track ownership back from one train and possibly grant it to another train, to unlock points, or to reopen grade crossing barriers.

The span between foremost and rearmost positions has two components:

- The train length, which therefore must be known in a vital manner (or, if not, an upper limit of which must be known) and which in most cases is derived from train consist (plus whatever slack there may be); and
- The uncertainty interval about the position of the head end.

Determining the head end vital interval is the task of the odometric unit.

Derivation of Vital Location Interval by Odometric Unit

The odometric unit is probably one of ASTREE's unique features when compared with ATCS. Because it represents a major development effort and will be a significant item of the on-board equipment cost, an explanation is offered on why SNCF puts such emphasis on keeping the interval small where a train location is known to be.

The volume of traffic on the French railroad is low by American standards; however, the traffic density expressed in number of trains/unit of track mileage is very high and generates severe capacity problems on at least some routes. This reflects not only the fact that the length of a French freight train is significantly shorter than that of its American counterpart, but also that passenger train traffic is very high. Keeping headways to a minimum is therefore a major concern, which translates into a desire to limit the uncertainty about a train's location.

It has been decided to base the train location not only on absolute fixes such as those derived from (satellite) radio determination, but on a measurement of the distance traveled. This integration approach entails a creeping error, which must be bounded by an adjustment process at some known places on the track.

The distance traveled is derived from a number of sensors, actually wheel rotation counters and Doppler radars, but other more exotic sensors are also being developed. Measurement based on wheel rotation is plagued by wheel slippage, especially where the tractive or braking effort is high. Doppler radars are not affected by these problems but may be subject to adverse atmospheric conditions that can, however, be overcome in most cases through efficient signal processing. Sensors measuring the pressure in the brake pipe or the tractive effort give indications about the confidence to be placed in one sensor or another. Through Kallmann filtering and the use of a model representation of the train behavior, an uncertainty interval can be derived, the limits of which can be relied upon as vital data. The computer that crunches the numbers must, of course, be vital. It can be seen that the odometric unit is a very significant component of ASTREE.

The location adjustment process is based on detecting microwave tags placed on ties through an on-board interrogator. It is another responsibility of the odometric unit to look for tags where they are expected to be detected and to adjust the uncertainty interval based on their detection. Failure to detect such a tag does not result in a degraded safety level but rather in an accuracy that further degrades until another tag may be correctly read.

The overall absolute accuracy is a function of the relative accuracy of distance measurement and of the spacing of tags. Current values are better than 1 percent for the relative accuracy, and 1 km for tag spacing, but the best return on investment may come from different values. It should be noted that a tag brings an improved accuracy downstream, which may lead to the installation of such tags at the limits of station areas.

Sophisticated Use of SNCF's Current Radio

The fixed-to-mobile transmissions rely on an advanced version of SNCF's current ground-to-train radio. This radio operates in the 450 MHz band. It is compatible with the International Union of Railways (UIC) standard.

Where U.S. radios operate in the VHF band for voice, and in the 900 MHz band for data, SNCF's UHF radio will transmit both voice and data. This decision is based on several reasons, an important one being that no frequency band in the 900 MHz range has yet been allocated to European railways. Other reasons for using this radio were that it already exists, which makes adapting it less costly than investing in additional facilities, and its good coverage quality (base stations are so closely spaced—4 mi odd—that when one is down, a train can still be reached through an adjacent station).

However, the most surprising reason for using this radio to carry ASTREE's data traffic is that is offers sufficient capacity. Estimates based on modeling ASTREE's behavior indicate that, for the most heavily loaded base station at the busiest hour, the highest priority traffic is on the order of 20 messages per minute, both in the train-to-ground direction, which consists largely of location reports, and in the ground-to-train direction which are mostly made up of authorities to proceed (consisting of a limit of authority along with a maximum speed indication and the braking parameters to be used by the train). The reason for this limited volume rests with ASTREE's synergy: the location reported by a train will be used by many applications and, conversely, an authority to proceed, although a simple message, results from a complex processing of a large amount of data.

The main technical differences between ASTREE's radio and that of ATCS will be discussed as follows.

Resource Sharing Between Voice and Data

UIC 450 MHz radio is basically analog, with a 12.5 kHz channel spacing. ASTREE data are staggered quarterly phase shift

keying (SQPSK) modulated at a rate of 2400 bits per second (bps). When no voice traffic is present, the full channel capacity may be used for data, at least in the ground-to-train direction. When voice traffic is present, it take precedence over data traffic but uses only three-fourths of the available capacity.

This is achieved through time compression-expansion. One second of speech, actually 1.040 sec, is digitized, stored and replayed in analog mode in 780 ms. At the receiving site, it is digitized and replayed in 1.040 sec. Voice quality is adequate. The remaining 260 ms every second still provide enough capacity to transmit two 255-bit frames in each direction.

When there is no voice traffic, the capacity available for data is eight 255-bit frames per second. Today, the actual use of voice radio is low (a few percent), which does not imply that it is unimportant. An intuitive feeling is that the use of voice radio will be even lower when most of the reasons for its use are taken care of by ASTREE.

It will be argued that using two radios rather than one offers a backup possibility. ASTREE will also be looking for some sort of redundancy, but preferably through the use of another medium rather than through a mere duplication of equipment. One such possibility would be using Locstar, the European cousin of Geostar, but it has not been possible yet to assess its ability to support ground-to-train traffic. Another possibility under consideration consists of using microwave tags that provide for two-way communication to add a limited data transmission capability to their location adjustment function.

Data Protection

A significant difference between ASTREE's radio and the one developed for ATCS is that, as far as we understand, the latter is considered as a vital component and therefore designed so as to allow an extremely low percentage of undetected errors.

ASTREE's approach, by contrast, is to put no strong demand on the degree of safety requested from the communications components. This by no means implies that ASTREE is not interested in a very low error rate, but rather that achieving an extremely low undetected error rate seems so important that the communications segment could not be trusted for fear that errors might occur in the terminal ends of the communication link.

Therefore, where vital transmission is required, it is the task of the application process that creates the data to supplement them with an extensive redundancy check sequence, and it is the responsibility of the application process that is about to make use of the received data to check their integrity against the received redundancy check sequence. Under these circumstances, it was not considered appropriate to spend any additional resources to improve the error rate on the communications segment itself.

In addition, failure of delivery of a message is not unsafe: a message will usually convey permission, not a restriction. In this manner, if for some reason the link is no longer available and no further message can get through, trains will come to a safe stop rather than blindly run farther. An exception is that if a dangerous event occurs at the last minute (e.g., an automobile falls from a bridge after an authority to proceed had checked that the track was clear from obstacles), an attempt will be made to transmit an alert to make the train stop.

It has been seen that safe transmission is obtained through an end-to-end error detection and retransmission mechanism. In addition to this, a forward error correcting (FEC) code is appended to the transmitted frames. This has nothing to do with safety because it is already provided by the error-detection scheme. FEC addresses two concerns, namely limiting the number of unwanted train brakings and reducing the cost for covering additional lines.

As will be indicated, the line access discipline is geared to waiting as long as possible before granting a train an extension of its authority to proceed. This attitude allows this extension to be made based on the latest possible known position of a train ahead, and therefore to grant the longest possible section of track, thereby minimizing the number of messages on the mileage covered. This attitude would be jeopardized if for some reason the new authority to proceed could not get through in time to avoid a train braking. One possible reason could be message collisions; the selected line discipline was chosen to minimize this possibility. Another possible reason rests with transmission errors. It is to decrease the occurrences of such errors that FEC is used.

Another potential advantage of FEC is to minimize the investment required to cover low-traffic tracks (with no voice capability on these lines). ASTREE does not demand a continuous coverage but only that the limit of an authority to proceed does not happen to be in a coverage gap, because a locomotive halted in such a gap would never receive the authority to proceed farther. It is felt that an appropriate attitude would be to install the base stations where they are easy to install and where most of the action takes place (i.e., in the stations, which are more closely spaced in France than in most of the U.S. or Canada). Although coverage gaps between adjacent stations would exist and would be accepted, it is felt that FEC may help reduce the number or size of such gaps.

A Line Discipline Based on "Intelligent" Polling

Line discipline is also an area in which ASTREE's options seem to differ from those of ATCS. ATCS radio has selected a contention scheme (free access) improved by the "busy-bit" technique, whereby a base station indicates to potential transmitters that it is already busy receiving a message. By contrast, ASTREE's radio makes use of an "intelligent" polling scheme.

This choice is based on three objectives: to maximize channel efficiency, to control transmission delays, and to cope with possible coverage gaps. It is also based on two features of ASTREE: the control center knows where the mobiles are located and is aware of their need to transmit. It is further influenced by a specific feature of UIC radio, whereby all the mobiles on a route transmit on one frequency, which may result in the transmission from one mobile at the fringe of a cell area being corrupted by the simultaneous transmission from a mobile at the fringe of an adjacent cell.

ASTREE's control center knows how far a train has been authorized to proceed, its current speed, and when it will have to begin braking if not granted a further authority. It knows what train has to report its position so that it may

release ownership of the track segment cleared by its rear end and so that part or all of this track may be allocated to the other train. The control center therefore knows when it needs the position of a train and can therefore keep the initiative to query this location.

This "intelligent" attitude was preferred to a systematic cyclic polling, which would translate into unnecessary messages or into using outdated locations, which would in turn result in smaller track allocations and in more frequent messages. It was also preferred to a basic contention method where it was feared that trains would unnecessarily report their position, or that message collisions would either entail a time-out and retransmit procedure or lead to using outdated train locations. Although a busy-bit technique does prevent most collisions, it was felt it could not avoid some undesirable transmission delays nor completely eliminate corruption from messages transmitted in adjacent cells.

The rigid discipline enforced by the ground control center does, however, leave room for free emergency reporting. To this end, some time slots are not assigned to a specific transmitter but rather left open for emergency messages.

The above line discipline is enforced by a communications controller. This controller is responsible for routing a message to the appropriate base station (using location from ASTREE). To better balance the radio traffic, the controller can force a train covered by several base stations to switch to the appropriate frequency. It manages ground-to-train transmission time slots, because it knows how urgent each message is and how to share a frame between packets or divide a long packet into multiple frames. It also manages train-to-ground time slots, avoiding collisions between messages transmitted by trains located in the same cell or in adjacent ones. It can also cope with coverage gaps.

Flexible Route Control

Route control is provided by ASTREE either directly—ASTREE then controls and monitors switches and takes care of interlocking—or indirectly, by connecting to existing interlocking plant. Each solution may be used where most appropriate. Over time, it is possible to migrate from one solution to the other.

Using Existing Interlockings

Using existing interlockings implies that routes will be controlled and monitored. This is an attractive solution in which technology is modern and allows for remote control. In cases in which technology is less modern, it may be necessary to control routes indirectly by displaying messages to a switchman; route monitoring is, however, direct.

To safely release routes, an existing interlocking must know the locations of the trains. This knowledge may be provided by ASTREE. It is, however, simpler to keep the existing island track circuits. Yet there is no need to keep wayside signaling, which is redundant with ASTREE's cab-signaling, or may even be more restrictive in some cases.

ASTREE-Provided Interlocking

If not provided by an existing interlocking plant, the interlocking process must be provided by ASTREE. The life-critical aspect of it—which demands that no section of track, switch, or apparatus be granted to more than one train at a time, and that no point be unlocked and operated if still allocated to a train—is part of the more general "free-track assurance" of ASTREE. This mechanism consists of allowing a train to proceed as far as a given point only if it is possible to grant it the (temporary) exclusive ownership of all the necessary resources. Compliance is enforced through brake application if a train's speed is such that the train is about to exceed its limit of authority or its allowed speed.

Other aspects of interlocking are dealt with in a nonvital way. One of them is route protection. Because the free-track assurance mechanism will, if necessary, safely halt a train before it crosses another train's route, only the best efforts will be made not to let it come close to this situation. Similarly, only in a nonvital manner will some routes be set in such a way as to minimize the consequences of a possible free-running of a train or a car, because a collision could only result from the simultaneous occurrence of a drift and of a faulty implementation of route protection.

ASTREE's route interlocking therefore allows any feasible route to be set, rather than only those routes built into the interlocking plant. This means that it may pay off to let ASTREE handle interlockings, even when an existing interlocking plant is of modern technology, when it is desired to modify the track diagram significantly. It may also prove worthwhile when bidirectional operation is desired.

If ASTREE provides for the interlocking function, it then must control and monitor switches directly. This control and monitoring may be centralized where the former interlocking plant used to be and then use the conventional point control and monitoring tools. It may also be distributed in the field. ASTREE allows for both types. In particular, a distributed system in which a central vital computer is connected through a multidrop link to wayside nonvital equipment is under development. In this system, the central vital computer transmits a random bit stream to wayside equipment that attempts to have this bit stream go through a contact to be monitored. If successful, the bit stream is transcoded in a device in the immediate vicinity of the contact in a way that is specific to this contact and then returned to the central computer. This computer checks the received bit stream against the anticipated result and decides that the contact was actually closed only if the redundant bit stream matches what it is supposed to be. A safe control is based on using the previously mentioned monitoring technique and relying on the fact that a transformer will let only alternative current through.

Simple Consist Acquisition and Train Integrity Checking

Knowing a train consist and the technical characteristics of the vehicles of which it is composed is essential for efficient traffic management and for safe train control. A continuous train integrity assurance is a necessity to ensure safety. Be-

cause continuous train consist checking is difficult with freight trains that do not have a train line, train consist acquisition and integrity checking are dealt with separately.

Train Consist Acquisition

Knowing a train consist is not only of commercial interest for work-order reporting. It is necessary to know how long a train is to safely control meets or passes, train separation, and point release. Knowing a train consist is one possible way to acquire this knowledge. Knowing the maximum permissible speed of each vehicle is necessary to determine that of the train. Knowing individual braking characteristics enable the aggregate braking characteristics of the train to be derived. Knowing the individual weight of each vehicle enables train tonnage to be taken into account when making dispatching or pacing decisions.

Automated acquisition of these data implies that each vehicle must be equipped with a device with some transmission capability. As far as passenger trains are concerned, it is possible to acquire this information through a dialogue between cars over the train line. Such a dialogue appears difficult to achieve with freight cars, except with some specialized fixed-consist trains. The difficulty stems from the absence of a train line that could serve both for communication and power supply. It was therefore decided to acquire the relevant information from wayside detectors.

Another difficulty comes from the volume of international traffic in Europe, which means that only internationally agreed solutions are acceptable. In this respect, it should be noted that automatic vehicle identification (AVI) has received new interest, which should lead to the installation of microwave tags on all cars involved in international traffic. These tags are readable when passing in front of ground readers. Their technology is similar to that used for location adjustment.

SNCF has tested a new generation of tags. These tags may be read at speeds in excess of 250 mph. They have a capacity of 2,048 bits, organized in 16 areas of 128 bits (but the new generation has twice this capacity). Some of these areas are read-only. They are used to carry the vehicle identity and its permanent technical data.These data are protected by a high redundancy. Other areas may be modified by a modulated microwave beam when passing in front of a reader. They carry information such as destination yard, payload, or commodity (the newer, higher-capacity tags might even carry some waybill information). A third type of area can be modified through an on-board series data link to report en route on internal status (temperature for reefers) or events (e.g., door tampering or time of a possible shock).

Readers are placed before and after those places in which train consist may be modified and establish this consist up and down the line. They are connected to an ASTREE control center. Some of them drive several antennas. This allows for a cost-effective way of detecting vehicles on multiple tracks at a given location and enables ASTREE to know which cars enter and leave specific tracks of a complex plant (one of the functions of work-order reporting). Train consist may also change in such places as a smaller industry track where it may not be justified to install and connect a reader. It is considered acceptable to enter the consist modifications manually and to take conservative values for some technical parameters until the exact consist may be checked when passing by the next interrogator.

Without waiting for ASTREE's deployment, SNCF has put this AVI subsystem into operational use on a limited scale. The yards of Lille, Avignon, and Marseille have been equipped with a total of 35 readers, as well as a specialized fleet of 400 freight cars that run between these yards at 100 mph.

Integrity Checking

For safety reasons, train integrity should be permanently monitored. This checking usually comes as an additional benefit when track circuits or axle counters are used for train location purposes. Because ASTREE does away with such trackside equipment, and because the spacing of AVI readers is too large, integrity monitoring must be performed on board.

One possible solution has been adopted by ATCS. End-of-train equipment is attached to the coupler of the last car. Over a radio link, it reports to the locomotive the pressure measured in the brake pipe. This solution has been tested by SNCF and it works successfully. It has not been retained, however, both because it was believed to place demands on the personnel (carrying a rather heavy piece of equipment) and because another solution (which may be more appropriate for Europe's relatively short trains than for America's very long trains) seemed to provide an even better answer.

This solution consists of continuously monitoring the pressure in the brake pipe, the air flow from the compressor, and the setting of the engineer's brake valve. The comparison of these data allows the detection of a train breakup. It also helps to check that the required brake tests were actually performed (but no indication is given as to the outcome). During these tests, it enables the volume of the brake pipe plus the auxiliary air reservoirs from the pressure variation that results from a known variation of the total air mass (itself, the product of the volume of air injected by the compressor times the pressure) to be derived. A change in this volume therefore indicates a consist modification, or the isolation of some brakes or of part of the brake pipe.

As for the AVI subsystem, it is possible that ASTREE's integrity monitoring subsystem could be put into operation before ASTREE's deployment.

ASTREE'S DEVELOPMENT STATUS

ASTREE's main features described above show that its ambitions are similar to those of a full-fledged level 40 ATCS. It has from the start been thought of as a global answer to the command and control problems of the entire French railroad.

Understandably, no management will want to commit itself to huge expenditure without some well-based assurance of the workability and performance of the system, as well as of its return on investment. The initial ideas gave enough confidence on both counts for the management to give a green light to the first steps in the development, but of course each

new important phase will have to be supported by convincing evidence.

Reported in this section will therefore be the earlier developments and the present status, followed by the anticipated next steps of development.

History and Current Status

ASTREE's research and development program was launched at the end of 1985. Interestingly enough, the emphasis was then laid on the improved level of safety that would accrue for the whole railroad system, whereas today much attention is also focused on avoiding or postponing investment in additional track capacity, better performance (adherence to schedule, energy saving, flexibility), and of course staff productivity.

From the start the effort was important and centralized. The assemble team included by the end of 1986 some 20 people working full time plus the part-time involvement of many others in the organization. The current level is 30 full-time SNCF employees, or a total of 50 if the staff from suppliers or software houses permanently assigned to the project are taken into account.

Although SNCF is fully aware that developing the future industrial products should be left to suppliers, its role in the development has been and still are not only those of a future operator and client but also those of a design consultant and integrator. The reason for this is that SNCF considers ASTREE as a potential major instrument in its future day-to-day operations and wants to make sure that all the major design choices are made in its own best interest (e.g., train integrity checking could be achieved by using end-of-train equipment or by in-train brake pipe pressure monitoring; this equipment is supplied by different vendors). This attitude is consistent with the approach adopted in the past for all major developments such as Train à Grande Vitesse (TGV). In the first stages of development, therefore, SNCF did the engineering, contracted to industry technological developments for individual components and progressively for subsystems, and acted as the integrator.

A first round with representatives from the various parts of the organization led to a tentative list of potential applications but not to a detailed functional specification. An initial system architecture was designed. Various component tests were performed extensively (e.g., Doppler radars or more exotic distance-traveled sensors, microwave transponders, point-monitoring devices, a new version of radio transmitters). After that, subsystems were contracted (the odometric unit first in a nonfail-safe version then in a fail-safe one, the integrity checking system). In most cases, SNCF kept to itself software development, even for modules as important as the communications handler or the control center main applications, in order to have the know-how and to be in a position to maintain and upgrade the system.

At first, tests involved cars in revenue service or test vehicles. For instance, a TGV set assigned to Switzerland traffic was used as a test bed for odometric sensors. Occasional special runs were also performed, however, for example to test performances at speeds in excess of 250 mph.

In 1990, the system level tests began on a very limited scale.The 5-mi line connecting Bondy, on the major trunk line from Paris to eastern France, to Aulnay, on the major artery from Paris to northern France, was selected, because the number of mobiles to be equipped was small—the passenger rolling stock is constant and few freight trains run during the day. Yet this line offers an interesting test bed: the route is partly double track, partly single track; one interlocking plant is a remote-controlled satellite of an important modern control center and another is an older lever-type one. There are grade crossings, industry tracks, and so on. A control center has been built in the station of Gargan (halfway between Bondy and Aulnay). Although it controls only a few trains and a few miles of track, it has been sized as though it were to control a very large area.

The purpose of this pretest, as it is called, is to demonstrate a preliminary, integrated version of all the functionalities of a full-fledged system. This gives users an opportunity to assess the services, the level of performance, and ergonomics. It also gives the users opportunities to see what changes are needed to this first version.

A major limitation of this test is that it is not implemented in a fail-safe way. No attempt has been made to make it fail-safe because the time needed to develop the system and subsequently certify it would have added undue delay before the first lessons could be drawn from this experiment. Accordingly, trackside signaling has been retained and takes precedence over cab signaling. In addition, most engineers' cabs are not even equipped with cab signaling (although the train sets are fully equipped with location, transmission, and integrity-checking subsystems).

In spite of this limitation, an extensive use is made of ghost (fictitious) trains to test the system's responses to all sorts of situations, including some potentially dangerous ones. These simulated trains appear to the system through real messages, which are actually transmitted over the radio to the appropriate base station (directive antennas are used to transmit to one base station or another). To add some complexity, not only ghost trains have been added, but also ghost switches and ghost tracks (but only ghost trains are allowed to run on ghost tracks).

The system now operates properly. The implemented trainborne functions include system setup (initializing a train when entering the ASTREE-controlled area), measurement of distance traveled, location adjustment, communications, cab signaling, integrity checking, and brake application. The implemented control center functions include train initialization, communications, display (train describer and time sheets), route monitoring, route control, consist acquisition, and authority to proceed. Dispatching makes use of a (modest) real-time expert system.

Next Steps

Some of the next steps have already received formal approval from SNCF's management as well as the requested appropriations. These next steps include extensions to the pretest and a regional test.

The major extension to the pretest will be a fail-safe version, which will have to be certified by the government. Only this will make it possible to operate real trains under ASTREE, and SNCF believes it is a necessary milestone to establish ASTREE's credibility among the potential users.

Other extensions to the pretest are not as well defined: it is anticipated that, on the basis of the lessons drawn from the initial tests, some modules will have to be modified, and some functions will have to be added. Which ones have not been determined yet, but provision has been made for a revised version with extensions.

One possibility under consideration would be to develop the fallback degraded mode of ASTREE into a low-end version, which could be implemented as an operational system on one selected branch line.

The regional test will offer the same functions as the first pretest, but will cover most of the trains running in the Paris-Est region, one of SNCF's 23 regions. About 400 track-mi will be covered, and 80 locomotives or train sets will be equipped. At the peak hour, 86 percent of the train-miles in the Paris-Est region will be covered (the remaining 14 percent will either be unequipped locomotives running through the region or unequipped branch lines).

The objective of the regional test is twofold: first, to demonstrate ASTREE's ability to cope with the processing workload and the communications volume; second, to demonstrate the ability to offer a continuous and dependable service. In particular, the regional test will demonstrate the possibility of reconfiguring the system when a control center fails, or when major parts of the communication network fail. In addition to these objectives, it is believed that much will also be learned from the variety of situations that will have to be faced during this full-size test. As indicated, this step has already been approved by the management of SNCF.

The next step will have to wait for the successful completion of the regional test and for a refined version of the economic evaluation before management approval is sought. This next step could be the implementation of ASTREE on a new high-speed line to be commissioned by the year 1997–1998. This could be the TGV-Est (Paris to Strasburg, and beyond to Mannheim–Stuttgart in Germany), the construction of which has been decided upon by the French government. Again, this next step in the development of ASTREE is purely speculative at this stage and will be subject to the successful completion of the regional test and to the further justification of its return on investment.

International Cooperation

The mere idea of a high-speed train running on a route connecting France and Germany shows that a new train control system can no longer be developed on a purely national basis. A TGV train set cannot be split up to get another locomotive when crossing a border. Were it possible, it would be unthinkable to waste so much time when so much effort is made to run at high speed. This proves that some form of international cooperation is a requirement.

The ultimate goal is the development of a unified European system. Short of this ambitious objective, a more modest one is to achieve enough compatibility for the rolling stock of one railroad to be able to run on the tracks of another—one of the goals of ATCS in America.

This compatibility demands that all interfaces between train and ground be standardized. Actually, there are only three such interfaces in ASTREE: the one between on-board microwave reader and transponders in the track to adjust the dead-reckoning process, the one between on-board radio (and on-board computer) and base stations (and various ground computers) for communication purposes, and the one between trackside microwave interrogator and on-board transponders for AVI.

Everything else may be specific to a railroad. The way one railroad controls and monitors a route, for instance, is of no concern to a foreign train, as long as the host railroad knows how to prevent that train from proceeding until the route has been set. Of course, it is desirable to go further in system harmonization, if only to benefit from a broader market, but each railroad may have its idiosyncrasies that result in its preferred solutions not being those of its neighbors.

Actually, a closer examination shows that a broader compatibility is desirable. A French engineer, for instance, will not drive the same way as a German colleague. The regulations are different today, and the information displayed is also different. For instance, one uses a target speed and the other uses a current maximum speed. Therefore, cooperation toward some harmonization goes beyond a mere technical compatibility.

The French and German railroads have acknowledged the necessity of a joint approach when defining their next-generation train control system. This is why a cooperative program called Artemis was set up between SNCF and Deutsche Bundesbahn (DB) at the end of 1989. In this program, SNCF builds on its ASTREE development. The unofficially stated objective is to install high-speed line equipment on the Paris-Mannheim-Stuttgart that hopefully would be also compatible with SNCF's current high-speed control system, TVM 430, and with DB's current system, LZB.

A broader European effort is sponsored by the European Community within the framework of its Euret (European Research on Transport) program. The objectives are the same, but the rail partners are the 12 railroads of Europe's 12 member states, plus the railroads of Austria and Switzerland. The first research contracts are expected to be awarded by the end of 1991.

What about ATCS? On the one hand, the cost of boring a tunnel between Europe and America does not make it strictly necessary to have a unified approach for some time yet. On the other hand, the concerns on one side of the Atlantic and the other are basically the same, although European railroads—with their more numerous shorter trains, their heavy passenger traffic, and their very high-speed trains—may place somewhat different emphasis on various aspects of train control. Certainly it is in the world railroads' interest to have as broad as possible a market to procure their equipment at the lowest possible cost.

Advanced ATP System for Improving Train Traffic Density and Control Efficiency

IKUO WATANABE AND TETSUO TAKASHIGE

A new automatic train protection (ATP) system allows railroads to use existing signaling systems effectively. This system can detect a train position continuously with an accuracy of less than 30 m by measuring the rail impedance shunted by train axles. Information such as the distance to the train ahead, speed restriction on switches, and so forth are transmitted to each train. These data, modulated by minimum shift keying (MSK) of the 3 kHz band, are also provided over the track circuit at a speed of 200 bits/sec. A train processor generates a brake pattern corresponding to the braking performance and actuates the brake automatically if the actual train speed is above the calculated value. As substantial information is given on a cab signal display, one train can run behind another at an extremely short interval, and an energy-saving operation can be attained. The function of automatic train operation (ATO) can be added easily. A level crossing is controlled safely on the basis of this train detection method. This system is applicable to any heavy-traffic density railway where various kinds of trains run: a high-speed train, a low-speed one, a freight train, and so forth. This system has much flexibility for its introduction and a train without a cab signal can be operated by using an existing wayside signal.

Because the crowding of commuter trains around the Tokyo area is worsening year after year, Japan Railway (JR) has had to increase the transportation capacity. Constructing a new line or increasing the number of tracks produces satisfactory results, but requires vast money and time. Increasing the train length increases transportation capacity, but it is difficult to extend the station platforms because of the lack of space and the high price of land. As a result, increasing the train traffic density is a more economical way to increase transportation capacity. Traffic density under the current signaling system allows 130 sec for a train to pass through a station. However it is not appropriate to reduce the interval between trains further because of the following:

- The system cannot identify train position continuously using a track circuit from 50 to several hundred meters in length.
- Train speed is controlled in a unit of blocks consisting of one or several track circuits, and the margin allowed for braking in blocks is stored.

By solving these problems, a new automatic train protection (ATP) system using the existing signaling systems can effectively reduce the train interval and improve the train traffic density. This system can locate train position continuously by measuring the rail impedance of track circuits shunted by the axles of a train. Train speed is controlled continuously by the digital data transmitted from the ground through track circuits.

Signaling Laboratory, Railway Technical Research Institute, 2-8-38, Hikari-cho, Kokubunji-city, Tokyo 185, Japan.

CURRENT SIGNALING SYSTEM

The train position is mainly detected with track circuits. The train speed is controlled in a unit of blocks consisting of one or several track circuits. The block length is generally several hundred meters. Only one train is permitted to proceed within the block to ensure its safety. There is a block signal around the entrance to each block that indicates the permitted speed to enter the block by the signal color. For example, as Figure 1 shows, red, yellow, yellow-green, and green respectively correspond to 0 km/hr, 45 km/hr, 65 km/hr, and 90 km/hr.

ATP systems with cab signaling have been introduced on the Shinkansen lines, Yamanote commuter line, and the others to ensure a high degree of safety. Only the speed signal is transmitted to the train processor through the track circuits in this system. The train processor receives the speed signal with antennas and displays it on the panel in the cab. If the actual train speed exceeds the transmitted value, the processor sends a command to apply the brake. If the actual train speed is below the value, the processor automatically stops the application of the brake (Figure 2).

This ATP system involves some problems. There are several speed signals, for example, 90 km/hr, 65 km/hr, 45 km/hr, and 0 km/hr. When a train is to be stopped, it must be slowed down below the speed signal indicated in each block. The driver must apply the brake off and on in each block, which gives passengers an uncomfortable ride. Each block length is appropriately selected with a margin and an idle running distance for braking. These distances are stored. Consequently commercial speeds become low and train traffic density cannot be reduced further.

The average track circuit length is about 200 m on the Yamanote commuter line, so the train position is identified in this unit. The track circuit must be cut to a shorter length to precisely detect the train position. However, a large number of track facilities, including the components for transmitting or receiving, is required to do this, and it also raises the construction cost and disturbs track maintenance work.

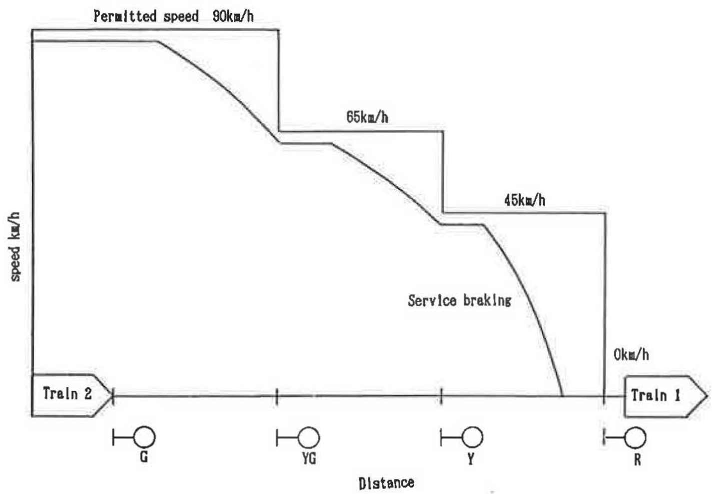


FIGURE 1 Blocking system with wayside signals.

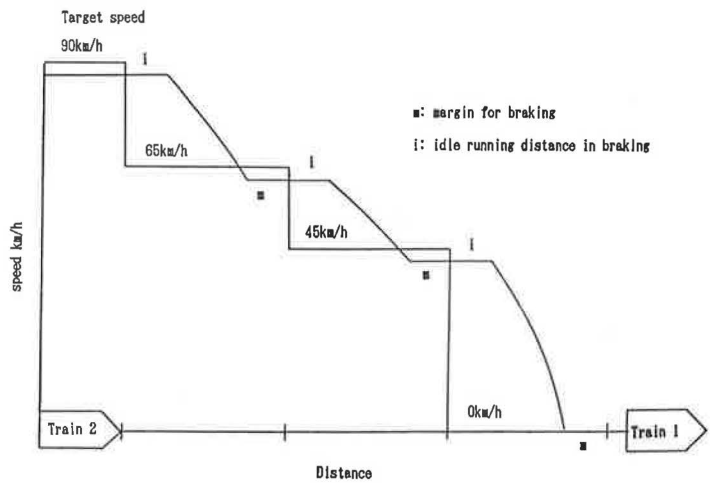


FIGURE 2 ATP system with cab signal.

ADVANCED ATP SYSTEM

Train Interval Control

An outline of the new ATP system is shown in Figure 3. This system consists of ground processors and train processors. The ground processor detects the train positions and transmits the data for controlling the train speed to a train processor. The train processor receives and decodes the data and controls the train's speed.

The train position is continuously detected as shown in Figure 4. The ground processor transmits signals to both the forward track circuit and the backward one at the junction point. The carrier frequencies of 1 kHz band are used (indicated by f1 through f4 in Figure 4). The processor inputs each voltage and current at the point, and the rail impedance shunted by axles of a train is measured by dividing the voltage by the current. The distance from the point to the train is calculated using this impedance. When there is no train in the track circuit, impedance larger than one corresponding to the track circuit length is registered, and the ground processor detects no train in the track circuit.

The ground processor generates such data as the distance to the train ahead or to a target stopping point to stop, or the speed restriction for passing a switch. It transmits the information to each train through the track circuit. These data are modulated by minimum shift keying (MSK). The carrier frequencies of 3 kHz band are used (indicated by F1 or F2 in Figure 4). These frequencies are not the same ones used for train position detection. The transmission speed is 200 bits/sec. The ground processor first calculates the distance from the junction point to the following train (L2) in the backward track circuit. When there is no train in the track circuit, the distance to the following train is equal to the track circuit length (L2′). Second, the processor calculates the distance from the junction point to the rear end position of the train ahead (L1). When there is no train in the forward track circuit, the distance to the train ahead is one transmitted from the forward ground processor. The system can continuously identify the length between the two trains by the sum of each distance (L1 + L2).

The train processor receives the data flowing on the rail through antennas consisting of two inductive coils installed at the front of the train. It generates a braking pattern corre-

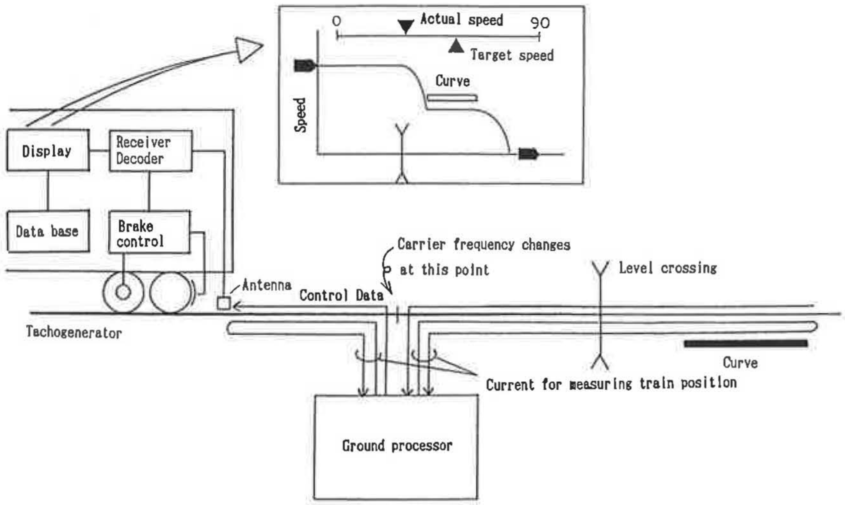


FIGURE 3 Principle of operation.

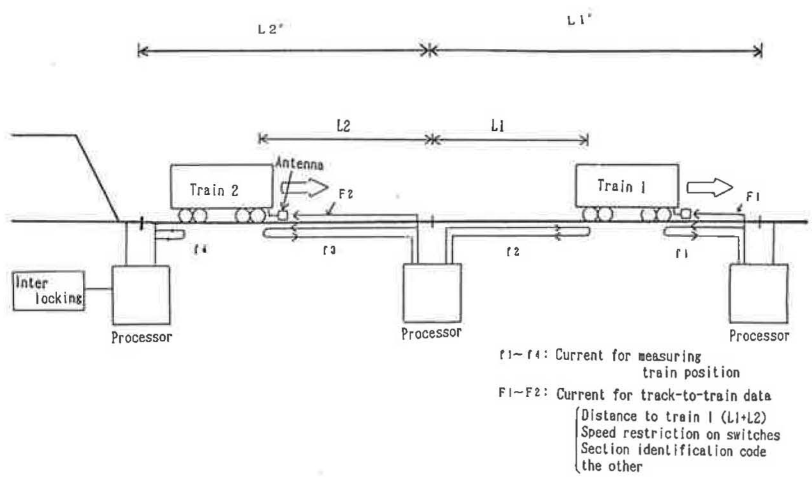


FIGURE 4 Principle of train position detection.

sponding to the braking performance and controls the train speed. Hence this system can be applied to any railway where various kinds of trains run: a high-speed train, a low-speed one, a freight train, and so forth. The train processor can detect continuously the rear end of the train ahead, so trains can run at an extremely short interval and an energy-saving operation can be attained by eliminating wasteful braking or acceleration. If the following train enters the same track circuit as the preceding one, it cannot receive the control data because the axles of the preceding train shunt the rail. Then the train processor holds the data just before it enters the circuit, and detects its own position and the distance to the train ahead by using both the data and the output from the tachogenerator. When the preceding train leaves the track circuit, the following one is controlled ordinarily to receive the information transmitted from the ground.

The interface between this ATP system and the interlocking system is realized as follows. The ATP system inputs the information about whether the route ahead is clear from the interlocking system. If the route is clear and the switches to be passed are locked, the train is permitted to proceed up to the rear of the train ahead. If the route is obstructed, the train is permitted to proceed up to the entrance of the route. Once the route is clear and the train proceeds, the ATP system outputs to the interlocking system information such as locked switches in the route until the train passes over them (Figure 5).

Expansion System Based on a New ATP

This ATP system can detect train position continuously in the whole area of the line, and it can transmit the controlling data of 40 bits to a train 3 times/sec. As a result, many expansion systems can be added.

There is a great deal of level crossing equipment on the commuter lines. This system can decide the accurate position at which to start or stop the warning signals, so that it transmits the timing of the warning to the level crossing equipment and controls it safely. Furthermore, because it can detect roughly the approaching train's speed by the change rate of the rail impedance, the constant warning can be given for trains at various speeds.

The train processors have the constant data on the line (for example, curves, grades, insulated rail joint positions, and

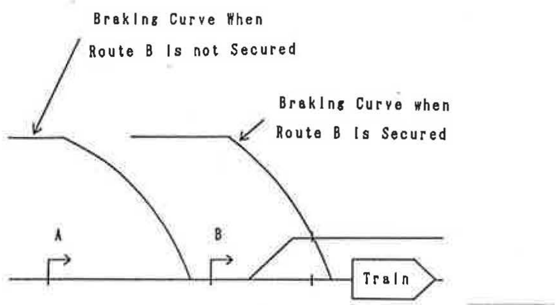


FIGURE 5 Interface with interlocking.

level crossing positions) and can detect the absolute position of the train by detecting both the change of the carrier frequency at the point of the insulated rail joint and the output from the tachogenerator. The position of level crossings, curves, and other information to assist the operator thus can be displayed in real-time on the cathode ray tube in the cab (Figure 3). Furthermore, automatic train operation function can be added to this system as follows:

- A few running patterns are stored in the train processor;
- The ground processor selects and issues the best running pattern to the train processor through the track circuit, depending on whether the train is behind schedule or on time; and
- The train processor controls the train speed by using both the pattern and the absolute position.

This system can be expanded to add the function of train-to-track data transmission. This function is realized when the transmitter is mounted on the train and transmits the data to the ground processor through the track circuit. The data are modulated by frequency shift keying (FSK) of approximately 20 kHz. The transmission speed is 1200 bits/sec. The ground processor can receive the data from the train processor within the 100 m area from the ground receiving point. It receives the train identification number, the condition of the train equipment, and so forth. It can receive much data at the home track circuit where the train stops.

If a train without cab signaling runs in this area, the current signaling system will be retained and its speed controlled with wayside signals. It happens that the wayside signal aspect indicates "stop" in spite of the cab signal aspect indicating "proceed." Therefore, when a train with cab signaling is approaching it, the ground processor knows that the train has the cab signal by using the train-to-track data transmission function and switches off the wayside signal (Figure 6).

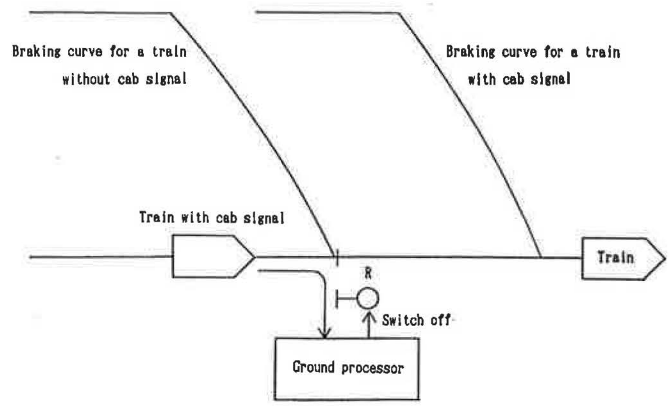


FIGURE 6 Control of wayside signal.

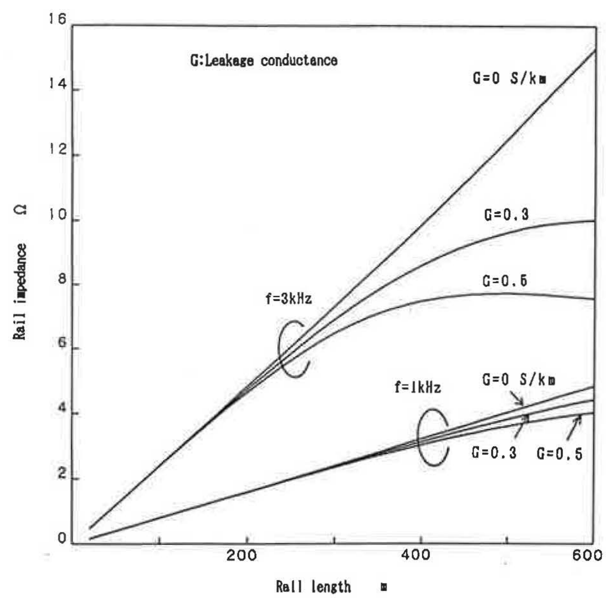


FIGURE 7 Relation between rail impedance and rail length.

DETECTION OF TRAIN POSITION

Method of Measuring Impedance

When a train enters a track circuit and its axle shunts the rails in ideal condition—the axle resistance is 0 ohm and the contact resistance between axles and a rail is 0 ohm—the relation between the shunted rail impedance and the signal transmission rail length is that shown in Figure 7. The frequencies are 1 kHz and 3 kHz. The leakage conductances are 0 S/km, 0.3 S/km and 0.5 S/km. The impedance is proportional to the transmission rail length, which is below 400 m for 1 kHz and 200 m for 3 kHz. So the 1 kHz signal frequency is better than 3 kHz in a long track circuit, 3 kHz is better in a short one.

The impedance is calculated by measuring the voltage between rails and by measuring the current at the point. The method of measuring the current is shown in Figure 8. The coils are fixed at the side of both rails and the voltage induced by the flux crossing the coil is measured. The polarity is set so that each voltage induced by the signal current for detecting the train position can be added together. Each voltage generated by the electric train current is then canceled out.

The voltage or the current is input to the ground processor through a matching transformer, an amplifier, a band pass filter, and an A/D converter. The ground processor calculates the rail impedance by using these and computes the distance to a train every 50 ms (Figure 9). When the change in the distance is much larger than expected, the processor makes the distance void and so can stay away from a momentary fault.

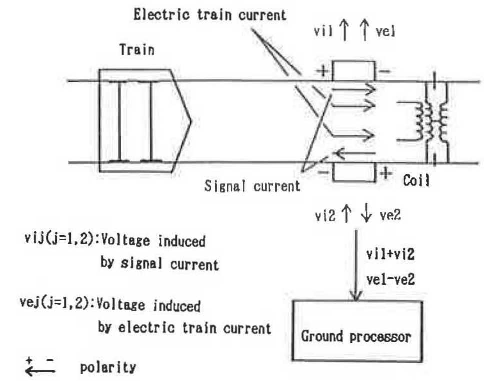


FIGURE 8 Measuring of current.

Error Factor in Measuring Train Position

The discrepancy between the measured and the actual distance is caused by various factors. When the impedance measured is smaller than the true one—for example, leakage conductance per unit track circuit length increases—the error is on the safe side. Because the system detects the train position as closer than the actual one, it actuates the following train's brake earlier than it normally would. However, when the impedance measured is larger than the true one, the error is not on the safe side. The following factors cause the unsafe error:

(1) Impedance between axles and rail increases when the rail surface is rusty;
(2) Rail is broken and does not permit the signal current to flow;
(3) Rail bond is worn out or disconnected;
(4) Sensor, filter, or amplifier is worn out; and
(5) Signal current is mixed with an unbalanced component of an electric train current.

The voltage between rails shunted by axles of a train is constant when a large voltage is given or an electric current flowing through rails is enlarged. The impedance between axles and rail then becomes small. Figure 10 indicates the relation between voltage given and maximum impedance between axles and rail. The impedance between axles and rail is reduced to less than 0.3 ohm by increasing the voltage given. The relation between the deviation in measuring train position and the impedance between axles and rail is shown in Figure 11. The deviation caused by (1) previously mentioned can be rendered smaller than +20 m by increasing the voltage given.

The deviation caused by (2), (3), or (4) previously mentioned is detected by monitoring the rail impedance of the track circuit when there is no train in the circuit because the impedance becomes larger than the normal one.

The railways in which this system is introduced use direct current. Analyzing the harmonics component of electric train current, the harmonics of 300 Hz—this is the sixth harmonics of the supplied power frequency in east Japan—is a large value. Therefore the error caused by (5) is avoided by using a single frequency different from the harmonics.

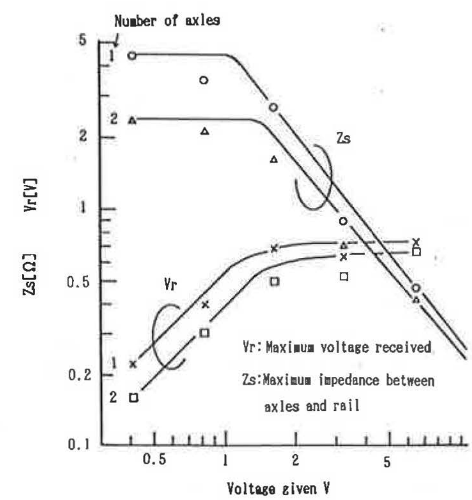


FIGURE 10 Relation between voltage given and maximum impedance between axle and rail.

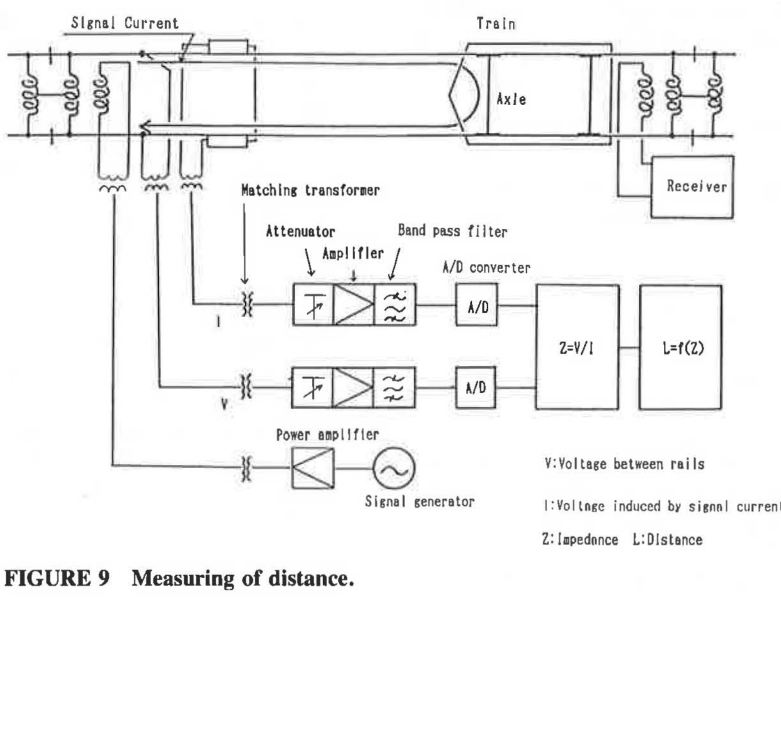


FIGURE 9 Measuring of distance.

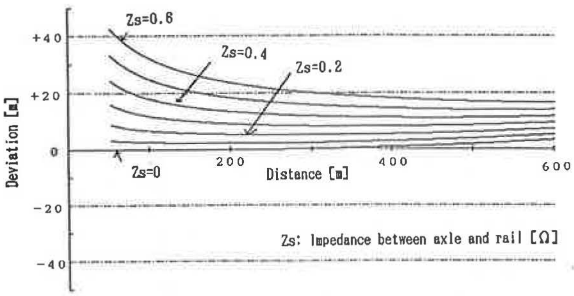


FIGURE 11 Relation between deviation and impedance between axle and rail.

Field Test for Measuring Train Position

The condition of the track circuit changes largely because of the differences in climate and ballast condition. The contacting condition between axles and rail changes is caused by the train weight and operational condition. For these reasons, the field test for measuring train position was repeated over a long time—for 5 years—and the track circuit to be measured was changed. The 1 kHz signal current is superimposed over the current track circuits. The voltage between rails and the current are applied to the equipment, and these are processed. The equipment produced experimentally for measuring the train position is shown in Figure 12. The block diagram is shown in Figure 9. The equipment shown in Figure 12 measures the train positions in three track circuits that are constructed using fail-safe processors. The duplicated microcomputers operate synchronized by a master clock, and the fail-safe comparator checks that each datum passing the bus is the same. The voltage and the current are applied to the microcomputer every 50 ms. The impedance is calculated by them, and the averaged impedance for 300 ms is accepted into the processor. This average period causes a train position detection error of only 13 m when the train runs at the speed of 90 km/hr. The processor estimates the position of a train axle by using the value. Figure 13 indicates the data measured.

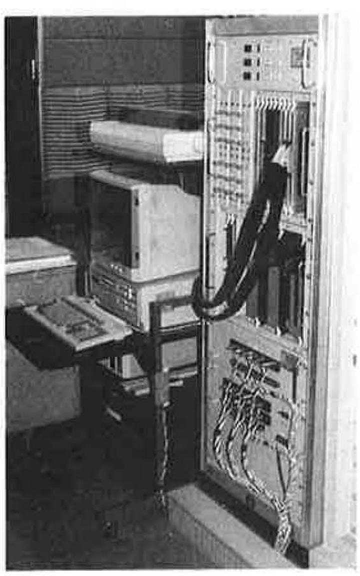


FIGURE 12 Measuring equipment in a signal house.

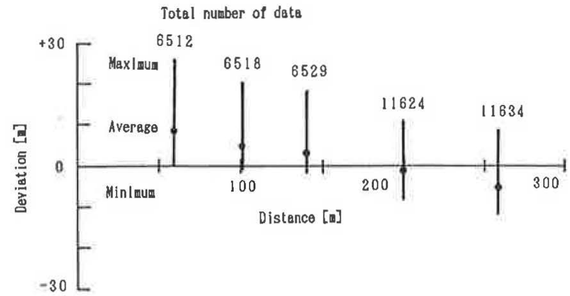


FIGURE 13 Results of a field test measuring train positions (f = 1 kHz).

The average, minimum, and maximum deviations of distance measured are indicated. The average deviation is a few meters larger in the area close to the transmitting point and is a few meters smaller over a long distance. The deviation exceeds the actual distance over a short distance because the resistance between axles and rail is about 0.2 ohms. The deviation is below the actual distance over a long distance because of the leakage conductance. The maximum deviation is 26 m.

DIGITAL CODED TRACK CIRCUIT

Information Transmitted from Ground to Train Processor

The data for controlling the train speed are transmitted to the train processor by an amplitude-modulated signal to which one frequency is assigned corresponding to the permitted speed in the current ATP system. A large volume of information cannot be transmitted using this method, so only a speed signal is sent. The driver cannot detect the time when the brake is applied.

In this system, the information is given to the train processor as follows:

(1) Distance to a train ahead or a target stopping point in a unit of 20 m;
(2) Speed restriction for passing over a switch; and
(3) Section identification code.

Eight check bits of cyclic redundancy check (CRC) code are added to this information. These are transmitted in a prefixed comma-free code. Consequently one frame is composed of 64 bits. The train processor is operated not synchronizing with the ground processor, so this code has the advantage of easily synchronizing the train processor with the data that are transmitted from the ground. This frame is received three times/sec. The transmission speed is 200 bits/sec to meet this requirement. It is advantageous to use as high a carrier frequency as possible to transmit a large volume of information. On the other hand, the information cannot be transmitted to a train over a long distance with a high frequency signal, so a carrier of 3 kHz—used in the current ATP system—is adopted. The system can transmit the information to a train at 600 m away under 0.5 S/km leakage conductance. The power is 40 dBm. The large harmonics of 300 Hz by electric train current appears, so the code is modulated by

MSK. This occupies a narrow band and the information can be transmitted without using these harmonics components.

Field Test of Digital Code Transmission

A field test transmitting digital code information to the train through track circuits was done. The equipment shown in Figure 12 transmits the digital coded data to the train. The data format is in a prefixed comma-free code. The transmission speed is 200 bits/sec. Two antennas, as shown in Figure 14, were installed at the front end of the train, and the equipment for decoding the received signal and for logging the data was housed in a box under the floor of the train, as shown in Figure 15. The train processor received a total of about 45,000 frames in this field test. It failed to receive the data three times. All of the failed data were detected with CRC. There were no successive mistakes in reception, so it was confirmed to transmit stably digital coded data to a train through the track circuit.

The train processor is operated not synchronizing with the ground processor. It can decode the transmitted data when it receives all bits of the frame, so generally it cannot decode any data of the first received frame when the train enters the track circuit. The average time to decode the received data after entering a track circuit was 0.6 sec; the maximum time was 0.9 sec. Therefore the train processor is to be designed to keep controlling for 1 sec by using the last received data without receiving the next data.

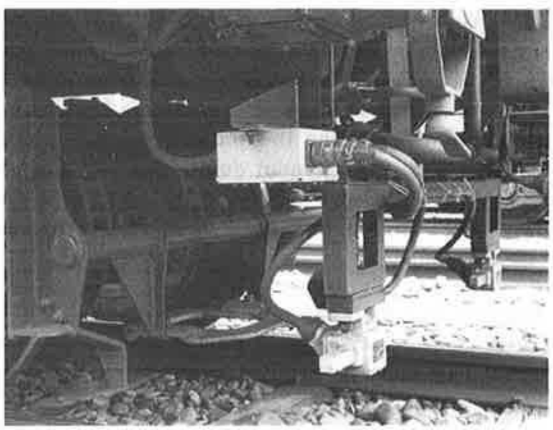


FIGURE 14 Antennas installed at the front end of a train.

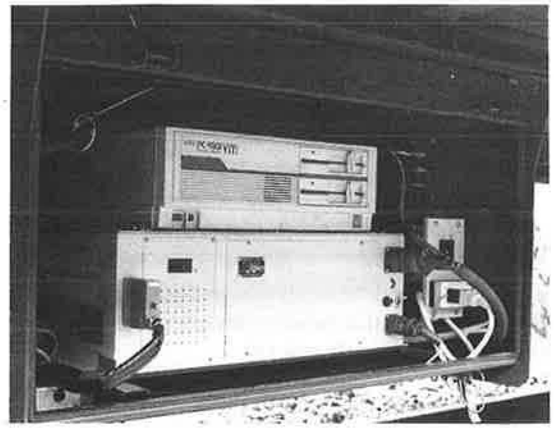


FIGURE 15 Decoder and data logger installed in a box under floor of a train.

TRAIN TRAFFIC DENSITY

The traffic density permitted under this system is estimated to be approximately 100 sec considering station stopping time (30 sec), train length (200 m), braking performance (2.5 km/hr/sec), accelerating performance (2.5 km/hr/sec), train speed at the station entrance (75 km/hr), accuracy of measuring the train position (30 m), and a margin for braking (50 m). As a result, this system can increase the traffic capacity by more than 20 percent.

CONCLUSION

This new ATP system reduces the interval between trains using the track circuits effectively. It is economical and easy to introduce because it uses much of the current signalling system. It can detect the train position continuously and can transmit the control data of 40 bits to a train 3 times/sec. The field test confirmed that these basic functions work stably.

This ATP system can be applied to any heavy traffic density railways where various kinds of trains run, and it has the advantage of not disturbing track maintenance works.

Microprocessor Diagnostic System for Heavy-Haul Trains in Actual Operation

JOÃO PAULO DO AMARAL BRAGA, NELYO CHOUCAIR DE OLIVEIRA, MARCOS BAETA MIRANDA, AND WELLINGTON SILVA

Operation with heavy-haul trains on Brazil's Ferrovia do Aço (Steel Railway) began in March 1989. Although the original project was for a totally electrified railroad using 25 kV, 60 Hz, the initial operation was with diesel-electric locomotives. Because this railroad has many tunnels, the longest one being 8.6 km, diesel-electric locomotive performance inside tunnels and questions such as type of consist, headway, and environmental conditions inside tunnels were intensively discussed. Because the dynamic operation of a heavy-haul train through long and difficult stretches is extremely complicated, and many random variables are involved, such an operation cannot be estimated in a simple mathematical simulation. Therefore, the necessity of preliminary tests in actual operation became a reality before the beginning of traffic activities. A system was developed to gather data from several points of the tunnels and another system to gather data from on board the locomotives. Both systems were computerized, coupled directly to sensors strategically located inside the locomotives and tunnels. The collecting of data both inside the tunnels and that concerned with the locomotive behavior was performed in real time, both systems being synchronized with the computer's built-in timer.

Brazil's Ferrovia do Aço (Steel Railway) is 300 km long and crosses the rugged Mantiqueira Mountains on a maximum gradient of 1 percent, with curves of 900-m minimum radius. The profile of Ferrovia do Aço ascends for about 200 km to a height of 1125 m, in the export direction; it then penetrates the Mantiqueira Mountains, descending on a consistent 1 percent grade for about 100 km. To accomplish this maximum grade and minimum radius curve, the construction required many viaducts and tunnels. There are 91 viaducts and 70 tunnels in all, the longest tunnel being 8.6 km long.

Because it was necessary to begin traffic operation with diesel-electric locomotives, operating conditions inside tunnels were discussed in depth, taking into account the type of train, train consist, headway, and environmental conditions inside tunnels. Theoretical studies were carried out by international consultants, but they were not definitive (*1*). Therefore, the necessity of preliminary studies in actual operation arose before the traffic activities with diesel-electric locomotives could commence.

Before heavy-haul train operation began, the main facts tested and discussed were the behavior of locomotives and the environmental conditions under which the train crew and the maintenance team would have to work when inside the tunnels.

Departamento de Planejamento e Controle da Manutenção, DEPCM-3, Superintendência Regional de Juiz de Fora, Rede Ferroviária Federal S/A, Av. Brasil, 2001, 5º andar, Juiz de Fora, Minas Gerais, Brasil, 36010.

After operation had begun, other problems appeared and so new tests had to be performed; some are still being carried out so that the failure in some locomotive systems can be solved and ideal conditions for handling trains in these long stretches can be found.

DYNAMIC BEHAVIOR OF HEAVY-HAUL TRAINS INSIDE TUNNELS

The idea of performing dynamic tests with a heavy-haul train inside tunnels arose from the need for studying more deeply the behavior of the train on this new railroad. There was a need to analyze the operational conditions and the demand on the equipment as a function of the general conditions of the train, as well as the traction theoretically dimensioned for the stretch.

The typical consists of iron-ore trains on the Ferrovia do Aço were initially four locomotives (3,000 HP) and 86 wagons (120 t). The first tests were performed with this consist. After 1 year of operation the consist changed to five locomotives and 108 wagons. Therefore, the necessity of other tests arose. The principal objectives were as follows:

- Dynamic behavior of locomotives inside tunnels.
- Data collecting while locomotives are inside tunnels and after they leave them,
- Analysis and diagnosis of the behavior of variables concerning tunnels and locomotives, and
- Operating feasibility of iron-ore trains hauled by diesel-electric locomotives inside tunnels of the Ferrovia do Aço (*2*).

Two systems, one capable of collecting data in several points inside the tunnels and the other capable of gathering data from on board the locomotives, were developed. The systems were computerized, directly coupled to sensors placed at several points in the locomotives and tunnels. Thus, it was possible to gather data both inside the tunnels and on the locomotives in real time, synchronized with the computer's built-in timer. The data sampling was about one datum per second. Thus, a more accurate analysis of trains' behavior inside the tunnels was possible.

Monitored Signals

The following signals were monitored inside the locomotive and in other parts of the train:

- Main generator voltage,
- Traction motor current,
- Train speed,
- Traveled distance,
- Brake cylinder pressure,
- External temperature,
- Air-intake temperature,
- Lubricant oil temperature,
- Cooling water temperature,
- Tractive effort,
- Acceleration notch, and
- Performance of the protection relays.

Description of Instruments

Temperature sensor—The thermocouple produces a small, variable voltage according to the temperature to which it is subjected. This voltage is amplified and read by the computer.

Speed sensor—The speed sensor consists of a pulse generator coupled to the wagon wheel. This gives a signal in a certain frequency, a signal that is converted into voltage and read by the computer.

Tractive effort sensor—This is based on a bridge of resistors that varies according to the effort to which it is subjected. This bridge nullifies the torque effort, registering only the movements of compression and expansion. These sensors are called strain gages and are connected to a coupler. The resulting signal passes through a conditioner equipped with a filter and an amplifier. This signal then enters a computer as voltage. Negative voltage means a compression; positive voltage, an expansion.

Brake cylinder pressure sensor—To measure the brake pressure, a sensor with a membrane connected to a strain gage is used. Pressure acts upon the membrane and a voltage proportional to its intensity is produced. The resultant signal is read directly by the computer.

Generator voltage sensor—This is based upon the Hall Effect. For the output voltage to be measured, a resistor was connected with a previously calculated value so that a current of a wanted value can be made to flow. The current flows through the shunt hall, and at its output, provides a proportional current with a reduced amperage. Finally, this output current flows through another resistor, which gives a voltage to be read by the computer. The scale of this signal is determined by the resistor's value.

Traction motor current sensor—To measure the traction motor current, a Hall Effect transducer was inserted at the armature terminal cables of the traction motor. The signal from this transducer passes through a condition-amplifier and then enters the computer as voltage.

Acceleration notch—The acceleration notches were monitored by installing a potentiometric transducer on the main lever of the leading locomotive. The signal from the sensor goes to the signal conditioner, which is an operational amplifier. From there, it can be read by the computer.

Data Acquisition System

A digital computer cannot store or process directly an analogical input signal. The computer does not understand something such as "5 V" as input data. It is necessary to transform this voltage into a digital representation that can be stored in a physical media, such as a disk. This transformation of an analogical input signal is performed by an analogical/digital (A/D) converter chip. As the input signals are all analogical, we use a board with an A/D converter that is able to read directly 16 input signals when connected to a microcomputer. To control this board, a program was developed in a scientific language. This program reads and stores information for subsequent use, and it also prints out reports and graphs.

The acquisition system of the locomotive has 32 analogical inputs in the range of −10 V to +10 V and a 17 ms-12 bit A/D converter. It also has 16 digital inputs for the locomotive protection signal indications, for relay devices, or for frequency signals. The signal multiplexing is performed by four 8-input multiplexers selected by a 6-bit word. These bits are directly transferred by the computer bus.

All the signals are isolated for 1,000 V, with a programmable gain via computer, of 1, 10, 100, and 1,000 times. This characteristic allows for a wide range of different indications, such as for thermocouples, strain-gage bridges, millivolts originated from resistor shunt, and so on.

A low-pass filter with a maximum range of 20 kHz and 3 dB is responsible for the high reading efficiency of the signals without noise interference due to the electric machines of the locomotive. The system can be said to be a board for a microcomputer slot that can be easily adapted and installed.

The software used has internal and external interrupt routines for acquisition data. During the test, the information appears on the screen, in windows, in the form of plotted color graph curves, during the test. Thus, a better analysis of the information is provided for immediate use.

The sampling rate can be programmed during the test. Tables and graphs can be consulted as well as the last 3 hr of the test without cutting off the signal acquirement. There are also alarm routines, such as loss of power in one of the locomotives, slipping, or operation of any locomotive protection system.

Some measures for protection were taken to nullify problems such as high vibration, temperature, and humidity. The worst of these problems is vibration. This was solved by creating a pseudo-driver for the operating system. At the end of each testing stage, the information is transferred from this pseudo-driver to the real driver where it is stored.

The system also has the track profile, which includes stretches of about 200 km mapped in memory in order to provide orientation during the tests. It shows the position of yards and important points, such as viaducts, bridges, and tunnels.

At the end of each test, the data are transferred to a work station where they will be analyzed, and graphs and reports will be produced.

Analysis of Test Results

The tests were performed using two types of locomotives, trying basically to keep all the operating characteristics, without involving any specific procedures. To accomplish so many specific studies, several tests were made inside tunnels to observe the behavior of trains while passing through them. These studies showed that the greatest problem was found in

the Cabritos Tunnel (T-0601). This tunnel is 908 m long with a grade of 1 percent for loaded trains and has the smallest cross section.

Two alternatives were proposed for an 86-wagon train, when loaded: (a) four locomotives with no helper and (b) four locomotives with a helper at the end. Figure 1 shows the performance of the hot engine protection during the passage of an 86-loaded-wagon train hauled by four locomotives without a helper through the Cabritos Tunnel. Analyzing these graphs, one can evaluate the performance of the protection device, examining several parameters such as external temperature, traction motor current, and speed.

One can see that the protection device was operated when the water temperature reached 90°C. Due to this fact, both traction motor current and speed decreased. The problems were not greater because the protection device was operated just before the end of the tunnel.

Figure 2 shows the operation, at the same time, of the three locomotives' sprays and hot engine protection devices. The test was carried out inside T-0601 with four locomotives and no helper at the train end.

Because of the progressive heating of the locomotives from the first to the fourth one, the relays operation should have happened in the same order. This did not occur, however, probably owing to an error of adjustment, particularly for the spray of the second locomotive.

Figure 3 shows the test carried out inside Tunnel T-0601 and the comparative graph curves for the fourth locomotive external temperature and operation of several devices, but in this case the train had a helper at its end.

In this test, the sprays were operated for a shorter time, compared with the test without helper. Moreover, the engine protection did not actuate. Once again, the spray of the second locomotive showed it to be incorrectly adjusted, because the logical sequence would be to operate after the third locomotive spray operated, turning itself off before the third

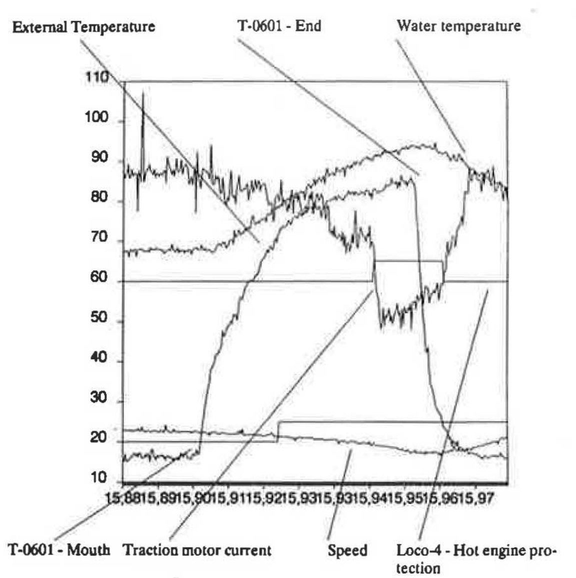


FIGURE 1 Hot engine protection (4 locomotives, 86 wagons, no helper) versus decimal time [external temperature (°C), water temperature (°C), traction motor current ($A \times 10$), speed (km/hr)].

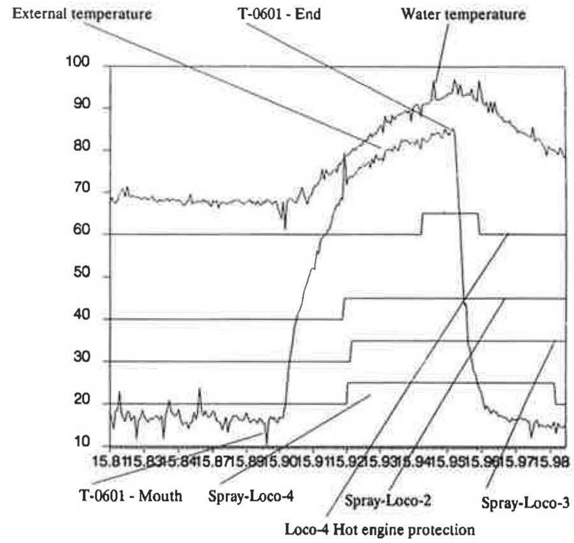


FIGURE 2 Hot engine protection (Locomotive 4) and spray (Locomotives 2, 3, and 4) (4 locomotives, 86 wagons, without helper) versus decimal time [external temperature (°C), water temperature (°C)].

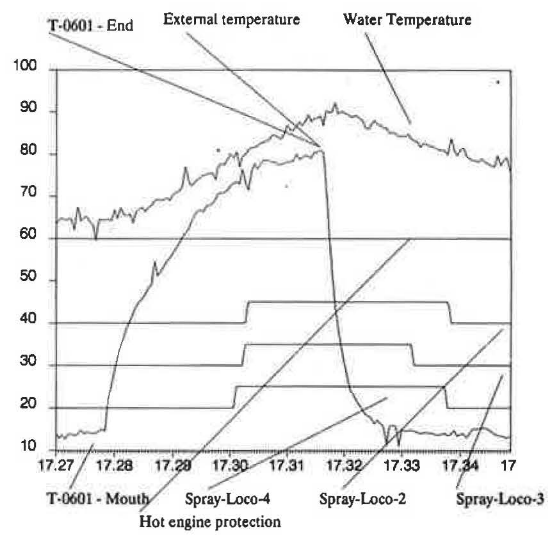


FIGURE 3 Hot engine protection (Locomotive 4) and spray (Locomotives 2, 3, and 4) (4 locomotives, 86 wagons, with helper) versus decimal time [external temperature (°C), water temperature (°C)].

locomotive did (because its temperature is lower as it receives hot air only from the first locomotive).

Figure 4 shows that speed was higher than the one for the experiment without a helper, and those problems that occurred with the previous experiment—when the hot engine protection almost shut down the fourth locomotive—did not happen this time.

Also, in all of the tests performed, it can be noted that, when the locomotive enters the tunnel, its power decreases and starts increasing only after leaving the tunnel. Such a fact explains the decrease in speed inside the tunnel.

Figure 5 was plotted from data collected during the passage of a train made up of four locomotives and 86 loaded wagons and no helper. A very fast increase in the external temper-

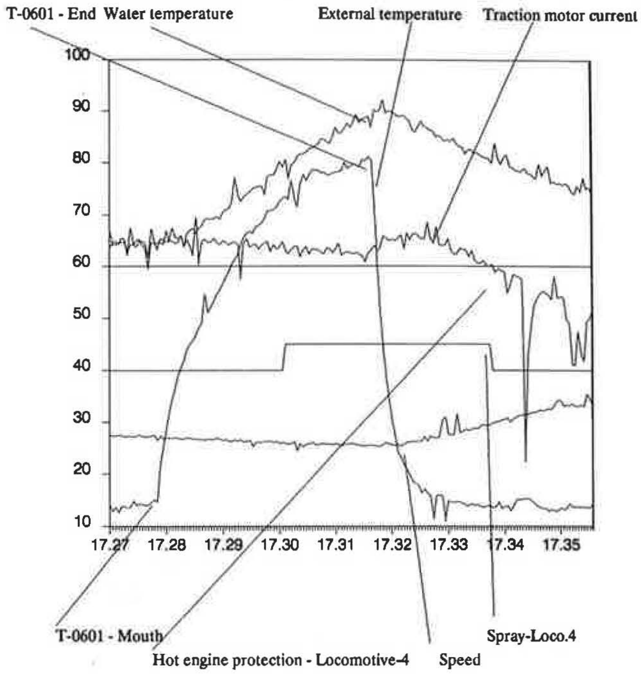


FIGURE 4 Hot engine protection (Locomotive 4) and spray (Locomotive 4) (4 locomotives, 86 wagons, with helper) versus decimal time [water temperature (°C), external temperature (°C), traction motor current ($A \times 10$), speed (km/hr)].

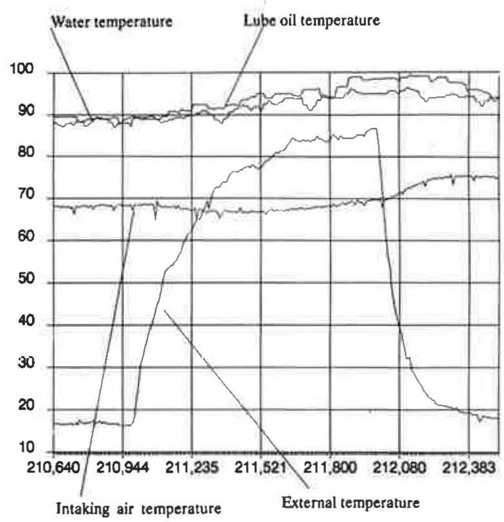


FIGURE 5 Data collected during passage of train (4 locomotives, 86 wagons, no helper): lubricant oil temperature (°C), water temperature (°C), external temperature (°C), and intake air temperature (°C) versus distance traveled (km).

ature inside the tunnel can be noted, reaching more than 80°C. A sudden temperature drop can also be seen as soon as the train leaves the tunnel. This graph also shows the temperature values for the air intake (monitored before it enters the combustion chamber), the lube oil, and the cooling water. The train speed when entering the tunnel is 16.5 km/hr, but inside the tunnel it drops to 14.5 km/hr.

ENVIRONMENTAL ASPECTS INSIDE TUNNELS

Vehicles fueled by diesel oil contribute to environmental pollution by burning this fuel owing to a set of chemical reactions that takes place simultaneously. These reactions happen because the diesel oil is composed of hydrocarbon (the boiling point of which varies from 170°C to 325°C) and gives out byproducts, that is, molecules with fewer carbon atoms, such as carbon dioxide, carbon monoxide, sulfur, and nitrogen compounds and water.

Carbon monoxide is extremely poisonous and can block the breathing process of living beings whose blood contains hemoglobin. Carbon monoxide competes with oxygen, and it is the first to reach the hemoglobin, forming a stronger bond, causing loss of vision and even death. As to the performance of vehicles, there is an alteration due to the decrease of oxygen percentage necessary for fuel combustion.

The principal objectives were as follows:

- Analysis of all components: SO_2, NO_2, NO, CO, CO_2, and fuliginosity in the air sampling inside the tunnels of the Ferrovia do Aço, and
- Analysis of the working conditions inside the tunnels for the locomotive crew and maintenance team.

Gases and particle samples were collected by selective monitors for specific analysis, according to the required technique, in the middle of the tunnel and at both ends, before, during, and after the passage of the train.

Then the following procedure assessed the concentrations of gases and particles found inside the tunnels and cabs:

Particles—The samples from which the particle concentration was to be determined were collected by passing a flow of air through a filter.

Gases—The measurement of gas concentrations was through direct reading using colorimetric reaction tubes equipped with electrochemical sensors.

This equipment was placed at several points along the tunnel and the sensors at different heights.

The study of the locomotive engineer's exposure to chemical agents inside the locomotive cabs was also carried out. The measurements were performed at the height of the breathing zone in normal conditions.

Monitored Parameters

- Direction and air speed,
- Air temperature,
- Carbon monoxide level,
- Oxygen level,
- Nitrous gases, and
- Other poisonous gases.

Data Acquisition System

The measurements of air speed, air direction, air temperature, and an analogical signal from 0 V to 5 V that comes from gas

sensors of CO_2, NO_x, CO, and O_2 were usually registered in EPROMs (erasable programmable read-only memory) from a data-acquisition system developed by the Universidade Federal de Juiz de Fora (UFJF) jointly with the Rede Ferroviária Federal S/A (RFFSA), according to the test operational conditions.

The portable device for collecting data inside the tunnels is a simple, small EPROM recorder for 32 kilobytes coupled to an 8-bit A/D converter multiplexed to three signals with sampling and recording performed by a real-time timer.

At the beginning of the tests, the timers were set in all the equipment in order to have the start-up synchronized. The boxes were distributed inside the tunnel, mainly inside the one that is 8640 m long. Normally, the sampling rate was of 1 sec. The storage capacity of the EPROM and the life span of the batteries permitted the system to record 6 hr of gas activity inside the tunnel, before, during, and after the passage of diesel-electric consist and wagon through it.

After each test, the EPROMs were taken from the boxes and input to a computer in which the information was analyzed. The previously mentioned information was also analyzed together with the data from the locomotive internal system. Because both signals were synchronized against the same timer, the results were easier to interpret.

Analysis of Test Results

Several tests were performed taking into account the results obtained for Tunnel T-0717, the longest one, which is 8.6 km long. Sample collection equipment was installed in three points along the tunnel (P1-17, P2-17, P3-17). The natural direction of air displacement inside the tunnel was the same as that of the train. The air speed at the checking points before the train passage was of about 0.5 m/s. Due to the movement of the train, it increased to 1.2 m/s and kept steady for 10 min at the P3-17 check point, 3 min at P2-17, and 4 min at P1-17 after the train had passed. Afterwards, the air speed kept constant around 1.0 m/s.

CO concentration can be seen in Table 1. The lowest dispersion time for the gas was reached at the P3-17 check point. This is the case because the air and train movements are in opposite directions.

The highest time for gas dispersion was at the P2-17 check point (20 min higher than the time found for P3-17). This was because there was less time for the "Piston Effect" to act on the air speed at this point and because a higher volume of contaminated air was displaced. The same factors were determined for the nondispersion of gas at check point P1-17.

TABLE 1 CO CONCENTRATIONS INSIDE TUNNEL

Check Points	Concentration of CO (ppm) by Train Passage: Before Max	Before Min	Inside Tunnel	After Max	After Min	Dispersion Time (min)
P3-17	1	1	15	24	0	40
P2-17	2	1	18	22	2	60
P1-17	3	0	19	23	15	—

At the P1-17 check point during 80 min after train passage, no significant dispersion of the gas was observed.

DOWNGRADE OPERATION

Requirements

In handling freight trains, downgrades of any significant length require observance of the following items:

- Balancing the grade, or holding speed steady at safe and practical values,
- Maintaining ample safety margin, or keeping speed within a value that will allow stopping the train anywhere on the grade within signal spacing or other prescribed limitations, and
- Balancing the use of dynamic brakes and air brakes, avoiding the overheating of the wagons' wheels.

Aiming at the improvement of this type of operation and the standardization of its procedure, the previously described data acquisition system was used to monitor the following parameters:

- Train speed,
- Traveled distance,
- Traction motor current,
- Position of the dynamic brake handle, and
- Brake cylinder air pressure.

Analysis of Test Results

Several tests were made in a 1 percent downgrade stretch of 100 km. The train under test had three locomotives and 86 wagons.

Figure 6 illustrates a type of operation in which the dynamic brake is kept in a fixed position and the speed control is performed by using the air brake several times. One can see the great speed variation over the distance traveled. Figure

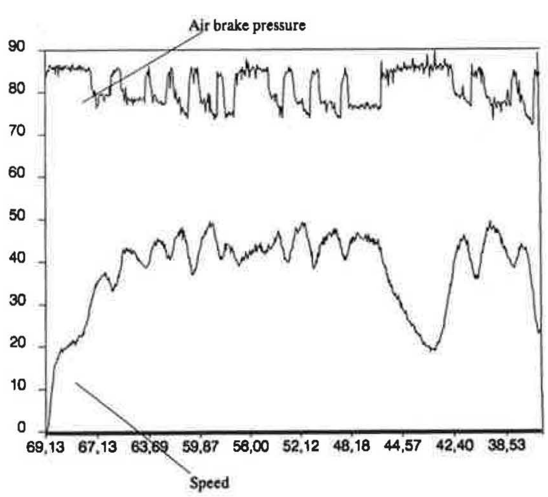


FIGURE 6 Air brake pressure (psi) and speed (km/hr) versus distance traveled in downgrade operation.

7 shows the graphic curve for traction motor current. The travel time was 50 min.

Another test was performed in this same stretch with the same type of train, but with a different train crew. The braking was carried out by using a minimum application of the air brake and using the dynamic brake to control the speed. It was noticed that the speed was kept within a smaller range variation, 45 km/hr on average.

The same distance was covered in 40 min, and the dynamic brake was used more than the air brake.

UPGRADE OPERATION

The data acquisition system was used to diagnose the problems in a 10-km stretch with an upgrade of 0.9 percent and in a 900-m tunnel at the top of the upgrade. The train type was four locomotives and 86 wagons.

Monitored Parameters

The following parameters were monitored in this test:

- Main generator voltage,
- Traction motor current,
- Acceleration notch,
- Speed,
- Brake cylinder air pressure, and
- Tractive effort.

Analysis of Test Results

The graphs of Figure 8 show not only the tractive effort curves, but also the speed and acceleration notch curves. In the tractive effort curve, two maximum values can be seen. The first occurred when the engine driver used notch 8 to accelerate just before entering the tunnel. The second occurred after the hot engine protection of the third locomotive was operated inside the tunnel.

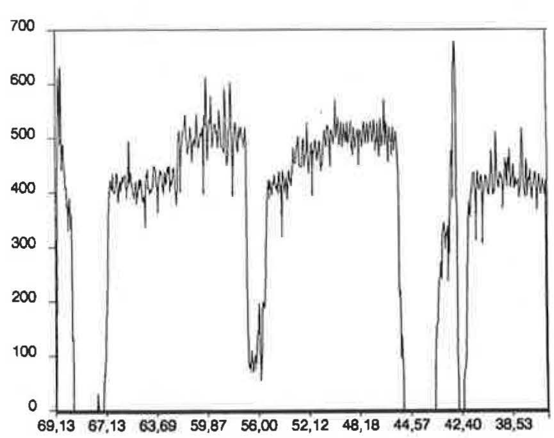


FIGURE 7 Traction motor current in dynamic braking (*A*) versus distance traveled in downgrade operation.

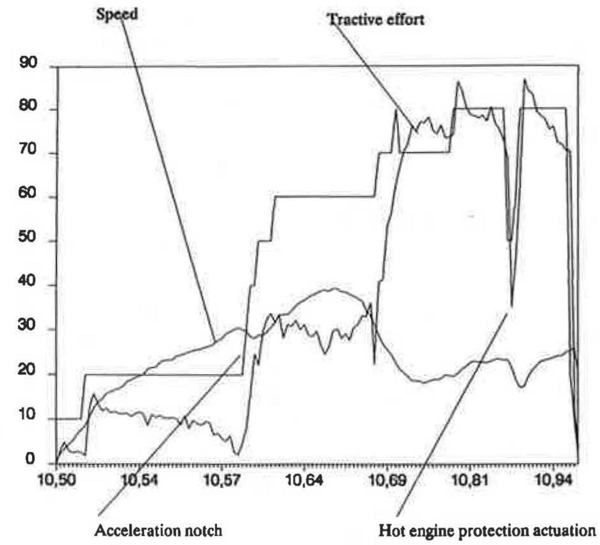


FIGURE 8 Tractive effort (kgf × 1,000), speed (km/hr), and acceleration notch versus decimal time.

As to the acceleration notch, it was observed that when the train started going up the grade the engine driver was using notch 6 and changed to notch 7; but notch 8 was only used when the train was approaching the tunnel. So the train speed, when entering the tunnel, was 22 km/hr. It dropped to 17 km/hr after the power decreased (hot engine protection operation).

To analyze the second test performed with the same stretch and locomotives, the same train type, one has to examine Figures 9 and 10.

The curves for tractive effort, speed, and acceleration notch are plotted in Figure 9. The tractive effort curve shows an irregular section due to slipping of the third locomotive (the test took place on a rainy night). As regards speed, it can be observed that the train entered the tunnel at 27 km/hr, owing to the different way of handling it by another engine driver. The train started up grade in notch 8.

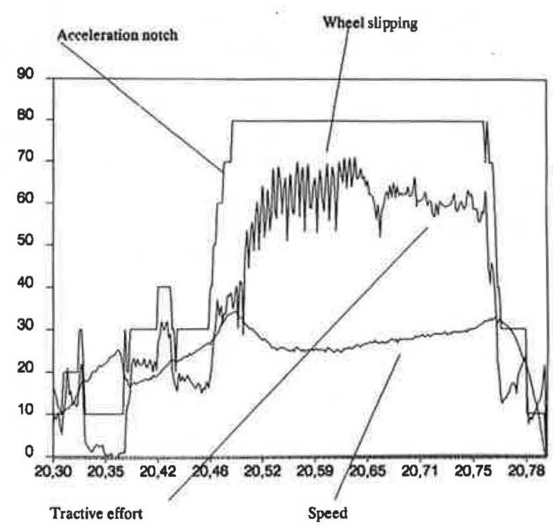


FIGURE 9 Tractive effort (kgf × 1,000), speed (km/hr), and acceleration notch versus decimal time: second test.

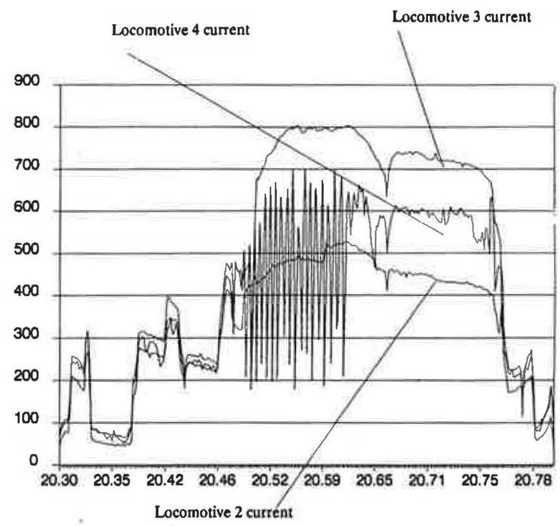


FIGURE 10 Traction motor current, Locomotives 2, 3, and 4, versus decimal time: second test.

Figure 10 shows the curve for the traction motor current. The slipping effect of the third locomotive can be seen.

TURBOCHARGERS OVER SPEED

In several tests performed to monitor the dynamic behavior of the locomotives when hauling heavy trains, the power decrease inside the tunnel was very evident. Some doubts about the locomotive turbochargers were raised. So, some tests with trains made up of 108 wagons and five locomotives were performed. The turbocharger of the fifth locomotive was instrumented, so that the actual operating conditions inside the tunnels could be registered.

Monitored Parameters

The following parameters were monitored:

- External temperature,
- Air-intake temperature,
- Turbocharger rotational speed,
- Turbocharger pressure,
- Traction motor current,
- Main generator voltage, and
- Speed.

Analysis of Test Results

Once again the previously mentioned data acquisition system was used. Figure 11 shows the following curves: external temperature, air-intake temperature, turbocharger rotational speed, and turbocharger pressure. This test was made in a 1 percent upgrade with a 900-m-long tunnel. The curve section relating to a sudden increase of the external temperature concerns the fifth locomotive when inside the tunnel.

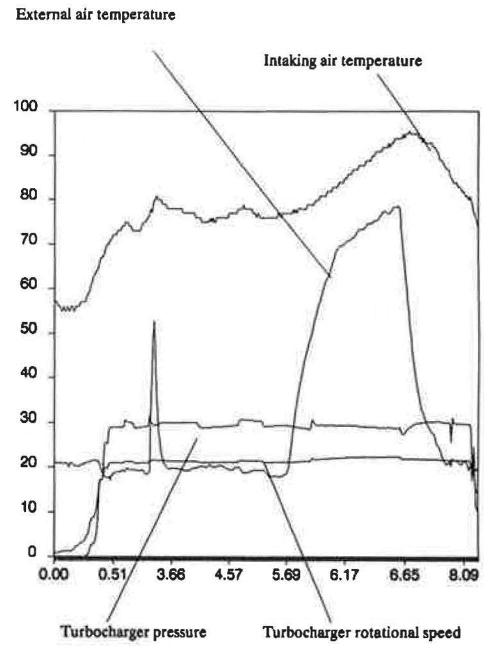


FIGURE 11 Turbocharger test on 1° upgrade and through 900-m tunnel external temperature (°C), intake air temperature (°C), and turbocharger rotational speed (rpm × 1,000)].

For a better analysis of the turbocharger behavior, Figure 12 shows a "zoom" of the curves of rotational speed and pressure when the locomotive is going through the tunnel. It is noted that there is a tendency for a pressure decrease, and the rotational speed increases by more than 1,000 rpm, compared with the reading registered before entering the tunnel.

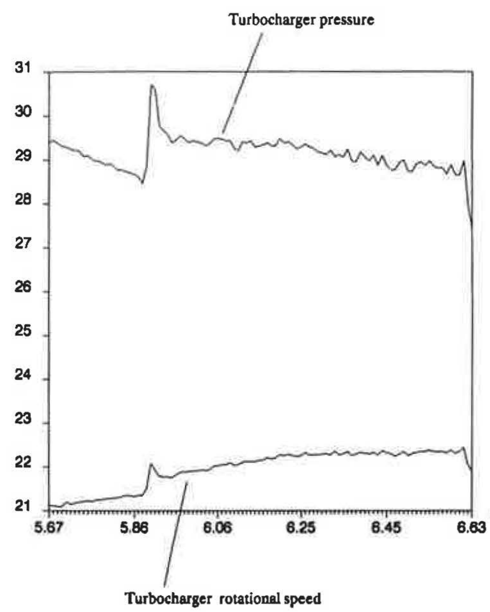


FIGURE 12 Turbocharger pressure (psi) and rotational speed (rpm × 1,000) versus traveled distance inside tunnel T-06-01.

CONCLUSION

The data acquisition system previously described in this paper has in its multiple applications the basis for a precise and appropriate diagnosis for the actual operating conditions of heavy trains. Some of the applications have already been mentioned.

The analysis of the data registered in each test provides a reliable diagnosis that make it easier to find a solution for problems relating to railroad transportation.

A deeper knowledge can be gained by passing from the academic research stage to actual dynamic operation tests; that is, we should not restrict our experience only to laboratory testing. Such knowledge is fundamental so that a definition of company investment priorities, based on the existing resources, can be achieved to increase transportation through lower costs and better quality.

ACKNOWLEDGMENTS

The authors are grateful for the support received from the Executive Board of their company, which believed in the serious development of this subject. Also, the continuous support and contributions from the Universidade Federal de Juiz de Fora (UFJF), the Centro Tecnologico de Minas Gerais (CETEC), and the Fundação Jorge Duprat Figueiredo de Segurança e Medicina do Trabalho (FUNDACENTRO) are gratefully acknowledged.

REFERENCES

1. J. P. A. Braga. Analysis of Dynamic Behavior of an Iron Ore-Train. In *Proc., 17th Panamerican Railways Congress*, Havana, Cuba, 1987.
2. J. P. A. Braga, N. C. Oliveira, M. B. Miranda, and W. Silva. Operational Feasibility of Iron-Ore Trains Hauled by Diesel-Electric Locomotives in Tunnels of Steel Railroad. In *Proc., 18th Panamerican Railways Congress*, Rio de Janeiro, Brazil, 1990.

Pilot System of CARAT on the Sanyo Shinkansen Line

HARUO YAMAMOTO, HIROTANE INAGE, AND YUTAKA HASEGAWA

The Computer- and Radio-Aided Train (CARAT) control system is a new train control system being developed by the Railway Technical Research Institute of Japan. Since 1987, the institute has examined individual, basic factors essential to this system, such as the quality of radio transmission, the accuracy of train positioning, and the control algorithm of trains. Next, the institute developed a pilot system for the purpose of confirming the synthesized performance under a more realistic environment. This system has been installed on the Sanyo Shinkansen between Shin-Iwakuni and Tokuyama. The pilot test began in March 1991. The test train uses an on-board system made up of three subsystems for (a) train position detection, (b) calculation of train target speed based on the control command received from the ground systems, and (c) radio transmission. On the ground, there are three sets of systems, each made up of three subsystems (a) the radio transmission subsystem, (b) a train tracking-controlling subsystem, and (c) a system simulating the movement of trains (except for the train equipped with the on-board system). The radio data are transmitted via leaky coaxial cable.

In Japan, train operation density of a commuter line in a big city must be increased and the interurban train speed raised. It is not economical to make these improvements by adjusting the current practice, because a large amount of work will be involved whenever the system is modified and, even then, traffic efficiency will not be improved that much. Besides, varying needs must be met flexibly and quickly. This has been a problem for train control systems that must be solved without increased cost or a deterioration of safety.

Against this background, the Railway Technical Research Institute (RTRI) of Japan has developed a new train control system, called the Computer- and Radio-Aided Train (CARAT) control system, under a subsidy from the Ministry of Transport. CARAT is equipped with a microcomputer, train position detection function, and driving-control function on-board, in which position information, interval control information, and so on are exchanged by radio between the on-board unit and the ground unit at a control center (*1*,*2*). Introduction of CARAT would realize a higher train operation density and an improvement in efficiency and economy at a stroke.

Figure 1 shows the development process. RTRI has worked on CARAT since 1987, and has carried out individual examinations of basic factors essential to the system, such as the quality of radio transmission, the accuracy of train positioning, and the control algorithm of trains. RTRI has achieved good results in putting theory into practical use. The next step was a pilot system that performs all these examinations at the same time. Installed on the Sanyo Shinkansen between Shin-Iwakuni and Tokuyama, the pilot system testing began in March 1991.

In this pilot system, one train is equipped with an on-board system made up of three subsystems that serve respectively for train position detection, train control, and radio transmission. On the ground, there are three sets of systems, each of them made up of three subsystems that serve for the radio transmission, the train tracking-controlling, and the train movement simulation. Radio transmission data are exchanged via leaky coaxial (LCX) cable. RTRI expects that this system will be able to confirm the performance of CARAT under a more realistic environment.

OUTLINE OF CARAT

System Concept

The following three points constitute the basic concept of CARAT:

1. Train position is usually detected on board; train interval and points are controlled on the ground according to position information transmitted from on board.
2. The on-board system determines the speed that guarantees safe train runs and controls the timing of the ground system with the train runs.
3. The information is transmitted between ground and on-board systems by a radio communication system covering one unit of the former to *N* units of the latter.

Figure 2 illustrates the system concept.

System Configuration

Figure 3 shows an example of the system configuration. The CARAT consists of ground systems, on-board systems, and radio transmission systems that exchange information between the train and the ground. The ground system is made up of a central system, control station systems, wayside systems, and a ground transmission system that connects the other systems. Table 1 illustrates the configuration of each system and their main functions.

Train and Traffic Control Laboratory, Railway Technical Research Institute, 8–38, Hikari-cho 2–chome, Kokubunji-city, Tokyo 185, Japan.

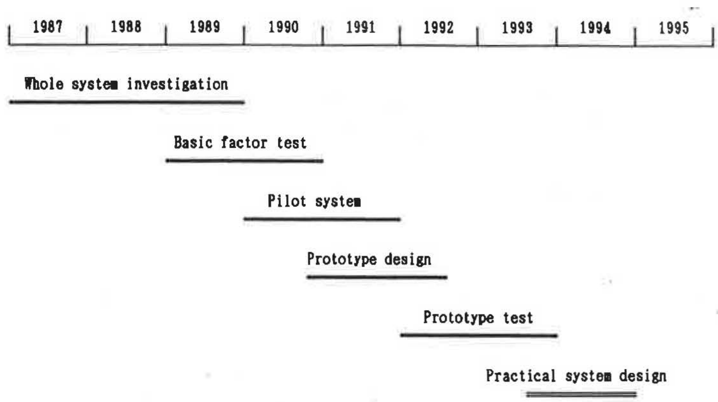


FIGURE 1 Development process.

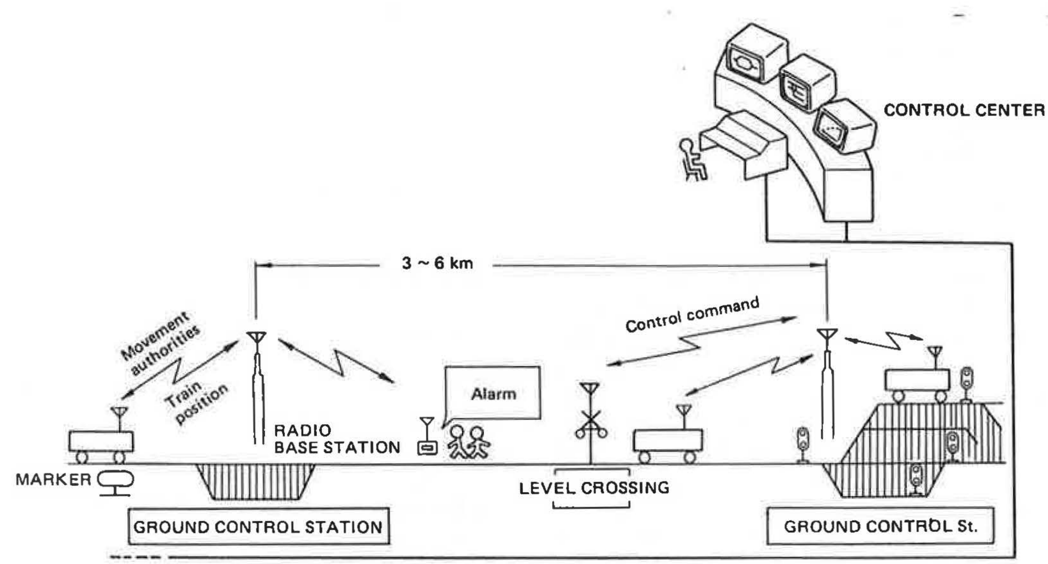


FIGURE 2 System concepts.

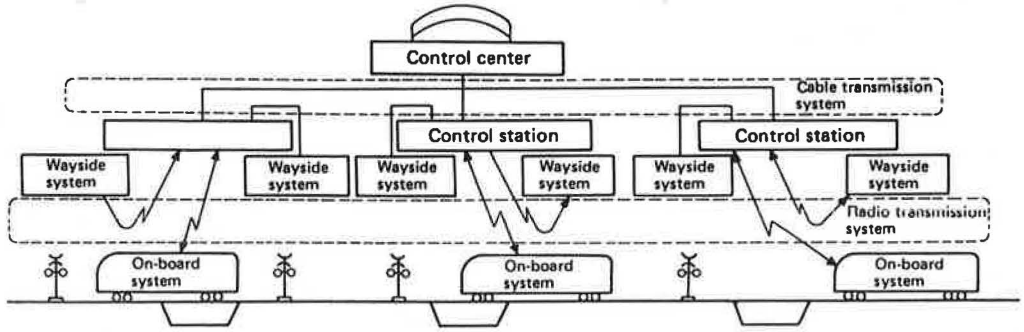


FIGURE 3 An example of system configuration.

TABLE 1 ARRANGEMENT AND FUNCTIONS OF EACH SYSTEM.

		Arrangement	Functions
Ground	Central system	Center	Information exchange medium between control stations Monitoring of whole system
	Control station system	Base station units	Ground functions of train interval control, station compound safety control and route control
	Wayside system	Level crossing units Work units	Level crossing control Train approach warning
	Ground transmission system		Transmission between ground systems (This system may be used in transmission between control station system and wayside one.)
Onboard		Train units	Onboard functions of position detection, train interval control, station compound safety, level crossing control and route control
Radio		Each system unit of control system, wayside and onboard	Information transmission between control station system and onboard one and between the former and wayside system

Examinations of Basic Factors

Radio Transmission

RTRI has measured electric field strength and bit error rate, and grasped characteristics of transmission quality and transmission distance about both LCX and space wave (quasi-microwave). Also, RTRI has established that in a tunnel it is possible to transmit—without taking special steps—over a distance of 4 km, because quasi-microwave propagates comparatively well in tunnels (*3,4*).

Position Detection

With two commuter trains, RTRI has tested the method in which running distance is calculated by rotation of axles. It has been found that it is necessary to sense rotation of two axles at least in order to limit errors under 20 m in slip and racing of wheels. For higher precision, rotation must be counted on three or more axles or software must be used to correct the count. But in this case, it is assumed that the spot coil of the ATS-S is located every 400 or 500 m (*5*).

Control Algorithm

RTRI has checked the basic algorithm using a laboratory prototype system. In this system, personal computers are allocated for processing units and a local area network (LAN) simulates the transmission between the ground and on-board units (*6*). RTRI has also finished the first examination of train interval control and level crossing control with this system in which some trains run on the imaginary railroad. Moreover, RTRI has compared the performance of a real-time operating system; the test results are reflected in this pilot system.

Calculation Unit for Safety Control

RTRI's purpose now is to check the algorithm using personal computers, but finally the software that deals with safety must be executed using a fail-safe processor. The processor for safety now used in Japan is 8 bits, which is insufficient for CARAT. Therefore, RTRI has been developing a 32-bit processor for safety (*7*). The basic hardware has been produced and its performance is being tested in the pilot project.

PILOT SYSTEM

System Configuration

Figure 4 shows the whole system configuration.

Test Purpose

The objective of the pilot system is to accumulate system performance estimation data and take error records in preparation for implementing a prototype. RTRI sees this pilot system as a means for an overall examination, including all basic factors except the calculation unit for safety control. Therefore, RTRI is relying on personal computers, except in the cases of the radio transmission system and the position detection system. RTRI operates the system (complete with the functions that have been developed in the laboratory prototype system over a long time in a more realistic environment), examining the communication of radio transmission data, train interval control, and position detection.

Examination Place

The system is installed on the Sanyo Shinkansen about 900 km from the starting point between Shin-Iwakuni and Tokuyama. LCX is laid in three zones over a distance of 5.7 km, and a shed for testing is built 897 km from the starting point to house ground system equipment. The train carrying the on-board system runs twice or three times a day between Shin-Osaka and Hakata, and communicates traffic control information via LCX with the ground system on the testing section. Position detection is tested over the whole section.

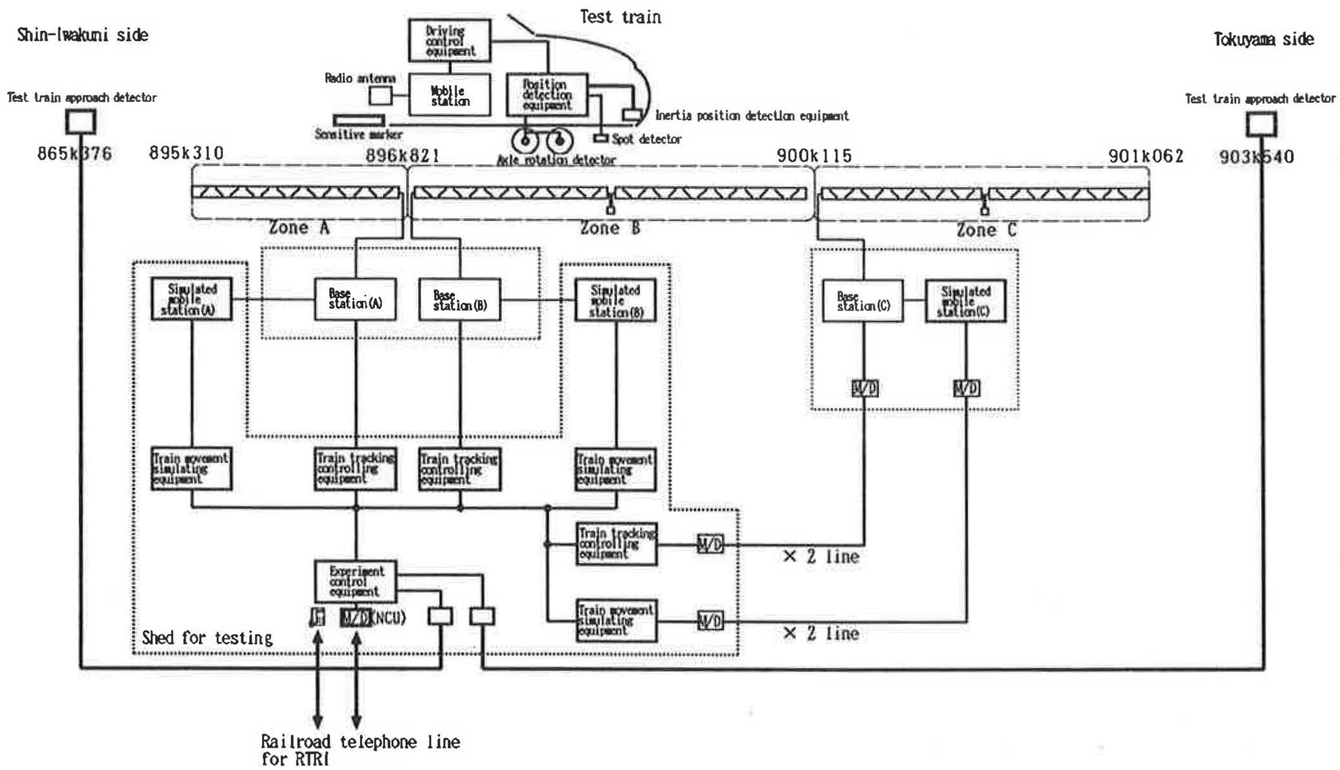


FIGURE 4 Whole equipment configuration.

Equipment Configuration

On the ground there are three sets of systems, each of them made up of a base station connected to LCX, train tracking-controlling equipment, a simulated mobile station, and train movement-simulating equipment. Experiment control equipment connects the train tracking-controlling equipment and the train movement-simulating equipment by LAN. On-board are position detection equipment, driving-control equipment, and an on-board station with a radio antenna. The pilot system uses personal computers except in the equipment related to radio and the driving-control equipment, and uses a real-time multitasking operating system. Application programs are described in C-language.

Testing Method

The pilot system is basically designed so that the ground system makes an experiment in interval control of a test train when the train runs on the test section installed with LCX. First, on the ground, each subsystem finishes its initial processing, because an approaching test train detector installed at both stations of Shin-Iwakuni and Tokuyama senses the arrival of the test train. And, as the test train comes into the test section, a train position message is transmitted from on-board to the ground and an interval control message is transmitted from ground to the train by radio.

Because the test section is only 5.7 km long and there is only one test train, the train movement-simulating equipment simulates trains ahead of and behind the test train. On the ground, RTRI confirms the functions of position tracking and interval controlling for each train. In on-board drive control, RTRI confirms the function of drive controlling and on-board signaling indication while running on the test section.

In this pilot system, radio communication data transmitted in the testing section are recorded by the train tracking-controlling equipment on the ground and by the driving-control equipment on board. The former also records the checking results of received train position messages. The detection error data between Shin-Osaka and Hakata are recorded by the position detection equipment on board. The data obtained on the ground are transferred from the train tracking-controlling equipment to the experiment control equipment, and transmitted to RTRI via telephone line. The data obtained on board, however, must be retrieved by a worker.

Subsystem Functions

Figure 5 shows the subsystem configuration and main exchange data. The following describes details of each subsystem.

Train Tracking-Controlling Subsystem

This subsystem does train tracking, route control simulation, interval control, train position indication, and logging. Each train tracking-controlling equipment covers the test section divided as shown in Figure 4.

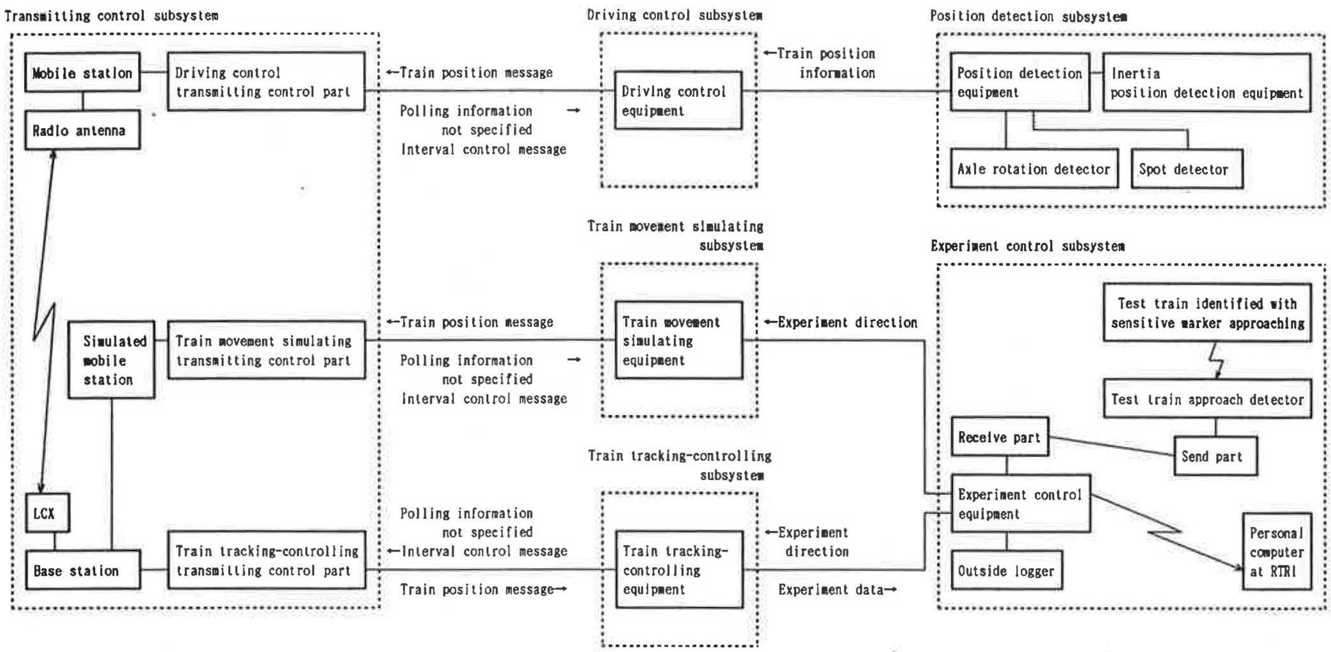


FIGURE 5 Subsystem configuration and main exchange data.

Train Position Tracking Train tracking-controlling equipment receives some train position data periodically from the on-board system and train movement-simulating subsystem, and then updates the train position data. Position data obtained here are provided to the functions of interval control, route control simulation, and so on. Position on the track is determined by train tracking and controlling section number, tracking block number, and distance from the starting point in the block, because test sections are divided by the so-called tracking block in the data. This equipment tracks trains by referring to a train table that grasps positions of each train and to a tracking block table that grasps trains that exist in each tracking block. Received train position data are checked for their contents.

Route Control Simulation In this system, a route is set up at the boundary of every tracking block, and the entrance into the tracking block is permitted via this route. The trains request route control in case the route within a certain distance is not being controlled. On a submain line, when trains start the route, they request route control at a fixed time after arriving at a given platform. Alternatively, route control simulation can be done by hand, and if all of the following conditions are satisfied, the route is controlled: the relevant route or a competitive one is not yet controlled; unlocking is not prohibited; there is no train in the clear section of the competitive route; and control request timing of the relevant route comes earlier than that of the competitive route. The route is restored when the train head comes into that route.

Interval Control To prevent a collision of trains or trains coming into an uncontrolled route, conditions ahead of each train are checked, and stopping target points are transmitted to each train. There are four kinds of stopping targets: the tail position of a train running ahead, the origin of a route not yet controlled, a railroad end position, and the position ahead at a certain distance set when the former three stopping targets do not exist.

Train Position Indication Position, number, velocity, route condition, and so on of each train running in the tracking-control section are indicated on the railroad diagram format.

Logging Process The radio communication data and checking results of received train position messages are logged.

Train Movement-Simulating Subsystem

Train movement-simulating equipment is installed at each radio station. It simulates train movement within the communication section.

The simulated trains are controlled with a deceleration control, which holds the train under a safe speed calculated by the driving-control function inside the equipment, and with an acceleration control, which acts when the difference between the current speed and the safe speed exceeds a certain value.

Safe speed is calculated from conditions such as current position, stopping position, restricted speed between those two positions, gradient, and so on. The train position update is calculated by averaging the speed from the last retrieved position to the current one. Stopping point is set at 10 m ahead of the stopping target.

Because more than one train is simulated, the current speed, stopping target, restricted speed, permissible speed curve, and

so on of a designated train can be indicated. Train movement conditions in the simulated section can be indicated also on the train diagram format.

Position Detection Subsystem

The position detection subsystem is basically designed so that it can determine the train position from rotation counts of axles and by detection of the spot coil. Apart from this, there is position detection equipment that uses the inertial method, and RTRI is investigating its precision and performance for making revisions when wheels slip or race.

The rotation sensor is an eddy current system that can count rotations exactly from low speed to stopping. As a countermeasure for slipping and racing of wheels, RTRI has adopted a method that uses software to process inputs from two detectors. In this processing, the slipping and racing are detected every second and the distance run in the meantime is revised using the limit value of acceleration and deceleration, and the operation control signal of the vehicles.

Four kinds of spot coils are used for special speed signal, train radio frequency switching control, open-close control of the train ventilation damper at the gateway to the tunnel, and so on. One spot coil is located every 700 m on the average. The computer for position detection is equipped with a table of kinds and intervals of all spot coils between Shin-Osaka and Hakata. It determines the train position automatically by consulting this table from detected data on kind and interval from whatever station the train starts. With the position thus determined, the train is revised referring to the position table every time a spot coil is detected.

The position detection equipment that uses the inertial method has three sets each of fiber-optic gyros and acceleration sensors. Because an error in this method tends to grow with the passage of time, this equipment—receiving speed data and position from the position detection equipment by wheel rotation—cancels error accumulations. The two position detectors log the distance-run into each logger every time the spot coil is detected. RTRI analyzes the characteristics by comparing both sets of logged data.

Driving Control Subsystem

The driving control subsystem consists of driving control equipment and an indicator that uses a liquid crystal display. The indicator is connected to the position detection subsystem in GPIB interface. While running in the test section, this subsystem exchanges control data between the train and the ground, handles safety control, cab signal indication, and the logging of data communicated by radio.

The permissible speed is calculated based on the interval control information received from train tracking-controlling equipment on the ground, train position input from the position detection subsystem, and wayside stored data. The results of this calculation guarantee safe driving up to a target point, but are not reflected in the actual train driving. Cab signaling indicates train position, wayside data, permissible speed curve, and so on. Radio communication data for the ground are logged into the equipment's logger.

Experiment Control Subsystem

The experiment control subsystem consists of the experiment control equipment connected to each subsystem on the ground, test-train-approaching detectors and their send-receive parts, an outside logger, and a personal computer connected to the experiment control equipment via telephone line at RTRI for data collection. The operator inputs experimental conditions here. At the start of the experiment (timed to the arrival of the test train), the conditions are transmitted to each subsystem. When the experiment is over, data logged in the train tracking-controlling subsystem are transferred to its outside logger; if a personal computer at RTRI requests the data, this subsystem transmits them. This subsystem monitors all the other subsystems on the ground throughout the experiment.

Transmitting Control Subsystem

The size of one input-output message from each control equipment is 160 bits, and a message converted to HDLC format of about 200 bits is transmitted to the radio system. The data-transmit speed is 4.8 kbps. An encoder of the radio system divides the data into 132-bit segments, appends a 40-bit flag sequence and 48-bit BHC marks to each data, and then transmits them to the radio station. At the radio station, data-transmit speed is 8 kbps by digital modulation system. A polling method with a period of 200 ms for the transmitting control uses one 12.5 kHz channel each for sending and receiving between ground and train systems.

Some polling information specifies a train and some does not. A train entering a radio zone first answers the polling information that does not specify the answering train, and requests a registration of the train for the train tracking-controlling subsystem. The train tracking-controlling equipment is already in receipt of a notice on the incoming train from the neighboring equipment and begins to send polling information that specifies the train after checking the train number and a train number received by radio.

Outline of System Processing on Test Train Runs

Having detected the approaching test train at the station, the train movement-simulating subsystem produces a simulated test train. The simulated test train answers the polling information that does not specify the train, and sends a train position message. After this, the subsystem simulates a train run exchanging the train position message and the interval control message. Upon coming into the test section, the test train receives polling information that does not specify the train, and answers with a train position message.

Then the train tracking-controlling subsystem switches train tracking from the simulated test train to the test train. After this, the subsystem sends an interval control message and receives a train position message for the test train. When the test train comes into the next section, the same communication sequence is repeated in the new zone. In the former zone, the tracking ends upon receiving the incoming information on the test train.

When the test train leaves the test section, the train movement-simulating subsystem recreates the simulated test train, and the train tracking-controlling subsystem switches the tracking train. Finally, when the simulated test train leaves the tracking section, both the tracking and the train movement simulation end.

The simulated trains that run ahead or behind of the test train are produced together with generation of a simulated test train. There is no switching of tracking for these simulated trains. At the boundary of train tracking-controlling sections, these trains are switched between neighbor train movement-simulating subsystems via LAN.

FUTURE CHALLENGES

Functions Buildup

Inputting the point detection information installed on the ground of Shinkansen, RTRI can grasp the detection position error of the test train real-time. RTRI also intends to implement an on-board point control function.

Addition of Test Train

By adding test trains equipped with on-board systems, RTRI intends to confirm the tracking performance of any test trains in the test section.

Development of Exclusive Equipment

RTRI would like to make a trial product that is closer to the real equipment than the personal computer. Then RTRI intends to introduce the calculation unit for safe control now being testing for performance, at the point handling the vital data.

ACKNOWLEDGMENT

This pilot CARAT system lacks some of the functions designed during the first phase of its development, because the development period was too short. In the second phase, RTRI would like to add the reserved functions and improve the system.

The development of CARAT began in 1987 with a subsidy received from the Ministry of Transport. In the examinations of basic factors, RTRI was able perform the radio transmission test and the position detection test thanks to the cooperation of East Japan Railway Company and West Japan Railway Company. The latest version of the pilot system has come into being thanks to cooperation of the West Japan Railway. We express sincere thanks to all the people concerned.

REFERENCES

1. Y. Hasegawa, H. Hayashi, M. Ikeda, and H. Inage. A New Train Control System by Radio. *Proceedings of IEEE VTS, National Conference,* April 1989.
2. Y. Hasegawa. Innovative Train Control System by Radio. *QR of RTRI,* Vol. 30, No. 4, November 1989, pp. 181–189.
3. H. Hayashi. 1.5 GHz Band Propagation Characteristics from Train to Wayside. *1990 Autumn National Convention Record* (in Japanese): Part 2 (Communications, Electronics), The Institute of Electronics, Information, and Communication Engineers, p. 9.
4. H. Hayashi. 1.5 GHz Band Data Transmission Quality from Train to Wayside. *1991 Spring National Convention Record* (in Japanese): Part 2 (Communications, Electronics), The Institute of Electronics, Information, and Communication Engineers, p. 409.
5. M. Ikeda and Y. Hasegawa. Characteristic of Position Detection and Method of Position Correction by Rotating Axle. *RTRI Report* (in Japanese), Vol. 5, No. 4, April 1991, pp. 43–51.
6. H. Inage, H. Yamamoto, and Y. Hasegawa. The Laboratory Prototype Phase 1 of the New Train Control System by Radio. *Proceedings of 26th Symposium on Cybernetics for Railway Applications* (in Japanese), Japanese Railway Cybernetics Association, No. 417, February 1990, pp. 226–230.
7. H. Nakamura, H. Yamamoto, T. Sasaki, and K. Akita. Fail-Safe Comparator for 32–Bit Strongly Fault-Secure Microprocessor System (in Japanese). *IEICE Technical Report: Fault Tolerant Systems.* The Institute of Electronics, Information, and Communication Engineers, Vol. 90, No. 76, June 1990, pp. 55–60.