Availability target of the railway infrastructure: an analysis

Ambika P. Patra, MTech, Luleå University of Technology

Uday Kumar, PhD, Luleå University of Technology

Per-Olof Larsson Kråik, PhD, Swedish National Rail Administration (Banverket)

Key Words: Railway Infrastructure, Reliability, Maintainability, Availability, Capacity, Punctuality

SUMMARY & CONCLUSIONS

The railway has been accepted as one of the most environmentally friendly modes of transport for goods and passengers. However, the railway sector is striving to increase its capacity to meet the growing demand for the transport of goods and passengers with a high level of punctuality in its services. Higher availability requires the effective operation and maintenance of infrastructure, often necessitating the implementation of cost-effective preventive maintenance strategies. Therefore, a higher availability target means higher maintenance investment. However, the question of setting the availability target for the infrastructure is not easy, as it involves many influencing decision parameters, apart from a good understanding of the network configuration and traffic density. Railway networks that have a smaller number of trains and a low punctuality requirement do not require higher availability targets. The aim of this paper is to estimate the availability target for railway infrastructure based on the capacity and punctuality requirements of infrastructure managers and train operating companies. The objectives of the paper are to develop an approach to i) estimating the capacity of the infrastructure based on the design and operational characteristics and evaluating the influence of infrastructure availability on the required capacity and ii) estimating the volume of primary and secondary delay due to failures and maintenance of the infrastructure and establishing the relationship between availability and punctuality requirements. To achieve these objectives, an example is presented with parameters drawn from failure, maintenance and traffic data. Finally, a model has been developed in Petri-Nets to establish a relationship between availability, capacity and punctuality. Monte Carlo simulation is used to establish the relationship. The simulation results illustrate the effect of infrastructure availability on train delays and capacity.

1 INTRODUCTION

Rail traffic is the most important form of public traffic in Europe as the density of the railway network is very high compared to the other parts of the globe. To be in competition with other modes of transportation, railway traffic must be quick, comfortable, cheap and primarily safe. There have been contractual agreements concerning the targeted level of reliability and punctuality in the performance regime within the railway sector. The business needs of railway infrastructure can be defined as lower ownership costs, interoperability, enhanced safety, improved punctuality, increased capacity and reduced journey times. The availability of railway infrastructure plays a significant role in attaining a higher capacity and punctuality level of the infrastructure. The required level of availability determines the amount of maintenance investment in the infrastructure over a period of time. A higher availability target requires the effective operation and maintenance of infrastructure, often necessitating the implementation of cost-effective preventive maintenance strategies in combination with effective supply chain management. However, the question of setting the availability target for the infrastructure is not easy, as it involves many decision parameters, apart from a good understanding of the network configuration and traffic density. Railway networks that have a smaller number of trains and a low punctuality requirement do not require higher availability targets. In order to estimate the availability target, it is necessary to understand the capacity and punctuality requirements of the railway infrastructure. Section 2 illustrates the fundamentals of capacity and punctuality. A model has been developed using Petri-Nets to establish a three-way relationship between availability, capacity and punctuality. Section 3 describes the model with an example. Finally discussions and conclusions are described in Section 4.

2 RAILWAY CAPACITY AND PUNCTUALITY

An efficient utilization of the existing railway infrastructure is an essential component of a high-quality transportation system and has become a central task for railway infrastructure managers. Line capacity is, in essence, what the infrastructure managers have to sell as their final product. Although capacity seems to be a self-explanatory term in common language, its scientific use may lead to substantial difficulties when it is associated with objective and quantifiable measures. It is a complex term that has numerous meanings and for which numerous definitions have been given. In [1] it is stated that capacity as such does not exist. Railway infrastructure capacity depends on the way in which it is utilized. However, in [2] it is stated that capacity is a measure of the ability to move a specific amount of traffic over a defined rail line with a given set of resources under a specific service plan. As illustrated in Figure 1, capacity is a balanced mix of the number of trains, the stability of the timetable, the average speed achieved and the heterogeneity (mixed traffic with different train speeds) of the train system. It is, for instance, possible to achieve a high average speed on a railway network by having a high heterogeneity – a mix of fast and slower trains. However, the cost of maintaining a high average speed with a high heterogeneity makes it difficult to run a great number of trains with a high stability (punctuality) than if all the trains ran at the same speed. If one wants to run more trains, it is necessary to operate with less mixed traffic and thereby have a lower average speed, as in the case of metro systems.

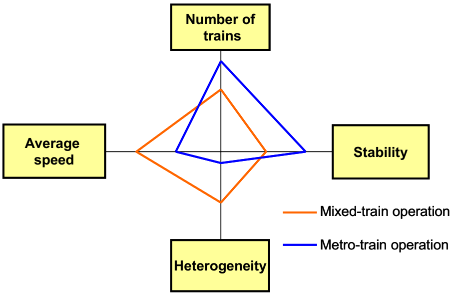

Figure 1 – Capacity balance [1]

Railway capacity has different values based on different criteria. In this paper the authors have defined railway capacity under three categories: inherent, achieved and operational capacity. The inherent capacity is the maximum capacity that a railway network can achieve. It is the number of trains that could run over a line or route, during a specific time interval, in a strictly perfect environment, with the trains running permanently and ideally at minimum headway. The inherent capacity is based on the infrastructure design. The achieved capacity is calculated under more realistic assumptions, which are related to the level of expected punctuality. It is the capacity that can permanently be provided under normal operating conditions. It is usually around 60–75% of the inherent capacity [1]. Banverket (Swedish National Rail Administration) [3] indicates a lack of capacity when the capacity utilization is above 80%, as higher capacity utilization leads to more delays of unexpected durations. The achieved capacity is the most significant measure of the track capacity, since it relates the ability of a specific combination of infrastructure, traffic, and operations to move the largest volume within an expected service level. The service level represents the punctuality level for the infrastructure based on a specified time table. If we change the time table, the achieved capacity also changes. The achieved capacity can be defined as the maximum capacity for a specified time table with defined operational headway where as the inherent capacity is the maximum capacity when there is no time table in place and the trains run at minimum (safety) headway. The operational capacity is less than the achieved capacity. This is the case if there is a prolonged shortage of facilities, e.g. due to accidents or weather conditions, but more generally due to failures in the infrastructure which disrupts the train operations.

Railway capacity very much depends on the headway time between the trains, i.e. both the safety headway and the operational headway time. Figure 2 illustrates the safety headway time between the trains. It is dependant on the distance that the trains maintain for safe operations on the track. As shown in the figure, the safety headway time is the summation of the travel time, braking time, release time and operating time. The travel time is the time required to cover the distance between two signals. The area between two signals is called a block and is controlled by a track circuit.

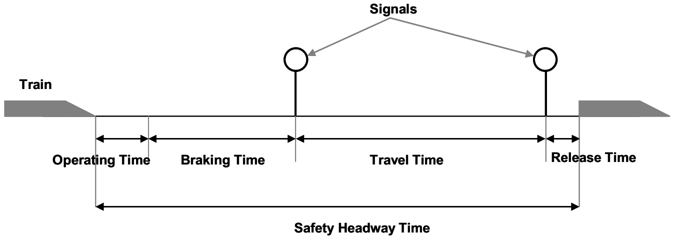

Figure 2 – Safety headway diagram

At any given time only one train can occupy a block section. The travel time depends upon the distance between the signals and the speed of the trains. The braking time depends on the braking distance i.e. the distance required to stop before a signal. It is calculated by considering the train speed and deceleration. The release time is the time required for the entire length of train to cross the signal. This depends on the length of the train and the speed of the train. The operating time is a safety time and is fixed by the infrastructure managers. The inherent capacity of a double track line depends on the safety headway between the trains. For example, if the safety headway is five minutes, the inherent capacity per track section will be 12 trains per hour. However, as discussed earlier infrastructure managers consider buffer time to accommodate delays. Moreover, the requirements of stakeholders set the operational headway time between the trains more than the safety headway. The operational headway is the actual time between two consecutive trains as per the train timetable.

Punctuality is defined differently by different infrastructure managers across the globe. A train in Sweden is considered punctual if it is less than five minutes off schedule at a station, otherwise it is delayed. Train delays may be classified into two major categories: primary delays and secondary delays. Primary delays are the delays undergone by the trains passing over a disturbed track section. Primary delays are the time differences between the normal and the disrupted journey. Secondary delays are delays of follower trains, which will not undergo the totality of primary delay, but which will undergo a delay because the previous train is delayed. This kind of delay happens when a failure is close to being restored. The principles of primary and secondary delay are illustrated in Section 3.

3 AVAILABILITY TARGET ESTIMATION

Railway infrastructure consists of various sub-systems, such as the track system, the signalling and telecommunication system, and the power system. Each of these sub-systems contributes to the infrastructure availability. As discussed earlier, the railway infrastructure availability influences the capacity and punctuality of train operations. Therefore, when estimating the availability target, the capacity and punctuality requirements of the infrastructure must be considered. Different failure modes in the railway infrastructure induce different amounts of delay in railway network based on speed restrictions. The amount of delay depends on the occurrence rates and repair times of the failure modes. Figure 3 depicts the speed profile of trains due to a track circuit failure. A failure in a track circuit turns the signal red for that particular block and trains pass at a reduced speed over that block until the failure is rectified. This section develops a model which estimates the availability target. An example is provided to illustrate the model.

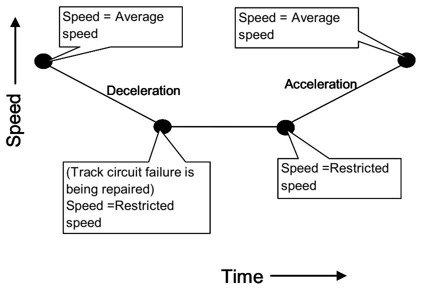

Figure 3 – Speed profile of trains due to failure of a track circuit

3.1 Example

Let us consider a double track railway line between two main stations. This line has multiple intermediate stations. Trains run with a uniform operational headway (OH) of 15 minutes and a safety headway (SH) of 5 minutes. All the trains that run on the track have the same speed pattern. As discussed earlier, the capacity and punctuality will be estimated for the line section between two adjacent stations. We consider the failures of three sub-systems that occur in this specific line section. The reliability and maintainability details of these sub-systems are given in Table 1. Occurrences of failures induce primary and secondary delay in the railway network as illustrated in Figure 4 & 5. When a failure occurs, trains reduce their speeds over the affected area (see Figure 3) and arrive late compared to their specified arrival time. The time difference determines the primary delay (PD). It is calculated by the kinematics equations of motion considering distance, acceleration, deceleration and speed. The number of trains that will be disrupted by primary delay is given by:

N_{PD} = Mean Down Time (MDT)/Operational headway (OH)

N_{PD} is an integer.

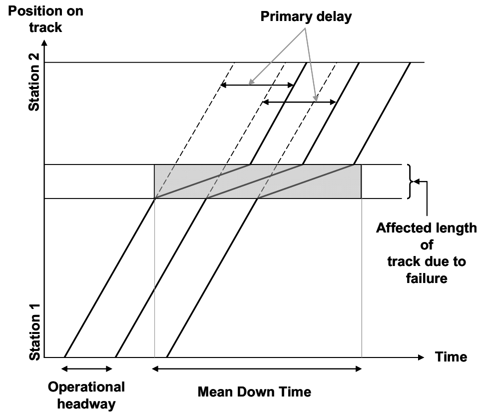

Figure 4 – Illustration of primary delay

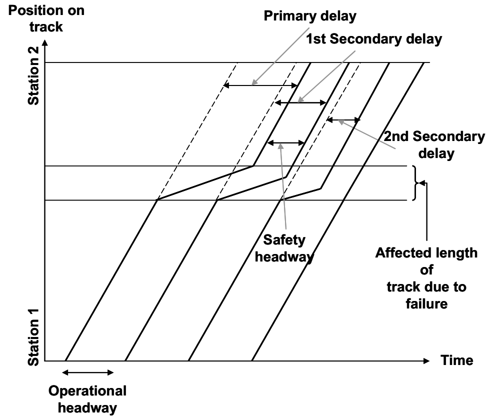

Figure 5 – Illustration of secondary delay

Secondary delay will occur if primary delay > (Operational headway – Safety headway). In this case the following train will have to slow down to keep the minimal distance (SH) from the last primarily delayed train. The secondary delay undergone by the 1st following train is given by:

$$SD_1 = PD - (OH - SH)$$

If $SD_1 > (OH - SH)$, the 2nd following train will be subjected to secondary delay. The secondary delay undergone by the 2nd following train is given by:

$$SD_2 = SD_1 - (OH - SH) = PD - 2*(OH - SH)$$

Secondary delays can be more generally expressed as:

$$SD_i = PD - i*(OH - SH)$$

The number of trains that will be disrupted by secondary delay is given by:

S_{PD} = Primary delay/(Operational headway – Safety headway)

S_{PD} is an integer.

3.2 Petri-Net model for studying the relationship between availability, capacity and punctuality

The Petri-Net model (see Figure 6) has been developed for estimating the relationship between availability, capacity and punctuality on the line between two stations for the example described above. The reliability and maintainability parameters for the model are given in Table 1. As illustrated in Figure 6, places 1, 3 and 5 denote the working states of sub-systems 1, 2 and 3 respectively where as places 2, 4 and 6 denote the failed states. The transitions between these places operate according to the failure rates and mean down time of the sub-systems. The failure rates in this case are assumed to be following exponential distributions. Whenever any sub-system fails, the infrastructure goes to a failed state and it is restored to the working state depending on the mean down time of the sub-system. This is illustrated by places 12, 13 and 14. The firings of the transitions between these places occur at any time depending on the failures of the sub-systems. It needs to be mentioned that in this model we have assumed only three sub-system failures that affect the capacity and punctuality of the railway network. In other cases the number of sub-systems can be more dependent on the specific railway track design criteria.

Sub-system	Failure Rate-FR (per minute)	Mean Down Time-MDT (in minutes)	Primary Delay-PD (in minutes)
Sub-system 1	6.00E-05	75	12
Sub-system 2	1.00E-04	60	10
Sub-system 3	1.50E-04	45	8

Table 1 – Reliability and Maintainability data for sub-systems

Places 7 to 9 describe the movement of trains between two stations. The transitions between these places consider the fundamentals of primary and secondary delay illustrated in Section 3. Trains start at station 1 (place 7) and reach station 2 (place 9). Place 9 keeps account of all the primary and secondary delay that the trains undergo in the case of sub-systems failures. Places 10 and 11 calculate the capacity of the track section between station 1 and 2 depending on the arrival time of the trains at place 9. Place 10 calculates the number of trains that reach station 2 every hour. In this model we calculate the average capacity (trains/hour) of the track section over a period of time.

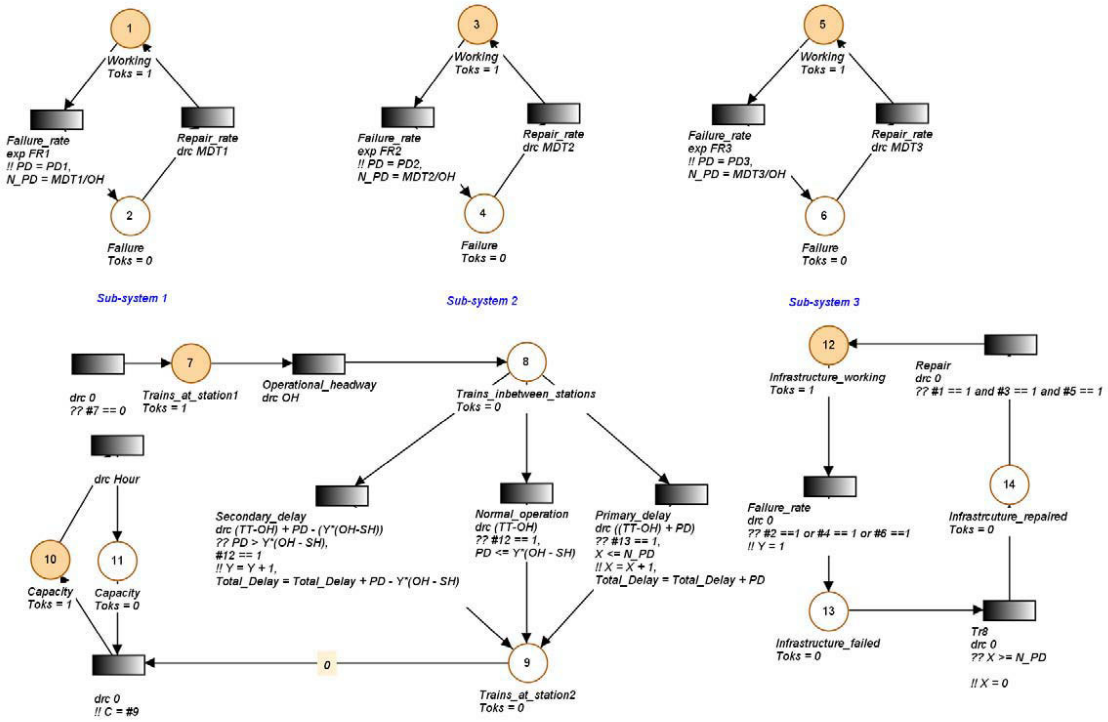

Figure 6 – Petri-Net model

The model also calculates the cumulative delay that the trains can undergo over a period of time on that track section. It is also considered that the trains take 25 minutes to travel between the stations i.e. the travel time (TT) is 25 minutes. The model will enable us to estimate the effects of the track system availability on the capacity and punctuality of that particular track section. The fundamentals of Petri-Net modelling can be found in [5].

Trains are operated for 18 hours a day and preventive maintenance is carried out during the rest six hours. Therefore, the availability of infrastructure considers only corrective maintenance. By performing Monte Carlo simulations on the Petri-Net models, we obtained the capacity variations over a period of one month (18x60 hours). Figure 7 illustrates the estimated operational capacity over a period of one month.

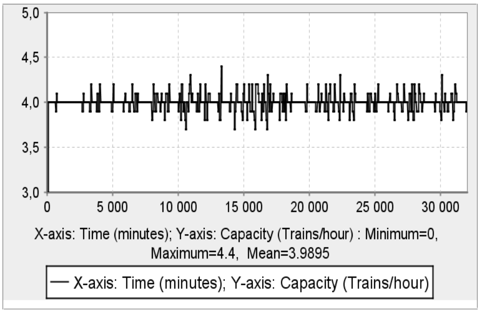

Figure 7 –Operational capacity over a period of one month

As shown in Figure 7, when a failure occurs, the train operation is disrupted until the failure is corrected and hence the capacity decreases. However, after the failure is corrected the operational capacity of the track increases because the delayed trains arrive in the same hour as the non-delayed trains. The achieved capacity in normal operation is expected to be 4 trains/hour as the operational headway is 15 minutes. However, due to failures, the mean operational capacity over a period of one month is 3.9895 trains/hour, which almost satisfies the achieved capacity requirement. Similarly, the cumulative delay that the trains will undergo over a period of 1 month is illustrated in Figure 8. The total delay occurring in one month is 355.6 minutes, which comprises of both primary delay and secondary delay. The average availability over that period is estimated to be 0.9826. In order to estimate the target availability, we need to perform sensitivity analysis on capacity and punctuality by changing the values of availability; i.e. changing the values of reliability and maintainability.

A relationship between the capacity, punctuality and availability is illustrated in Figure 9. As shown in the figure, the delay decreases with increases in availability value. However, in this particular case, the capacity remains constant with the change in availability values. This is due to fact that the timetable specified in the example absorbs all the delays because of the buffer embedded into it.

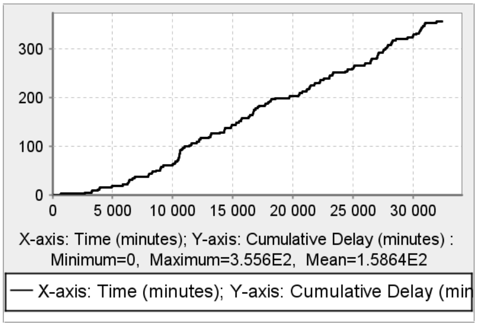

Figure 8 –Cumulative delay over a period of one month

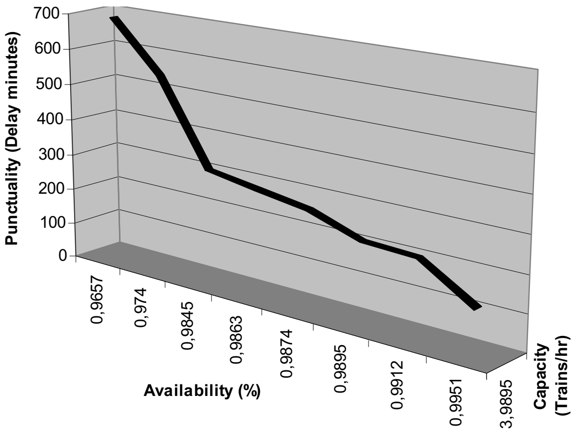

Figure 9 – Relation between capacity and punctuality with availability

In this particular example the operational capacity is almost equal to the achieved capacity for all values of availability. There might be a case when we reduce the availability to a much lower level then operational capacity will fall below than the required capacity. However, in this case we can estimate the availability target for the line section e.g. if the delay requirement should be less than 250 minutes, then our availability target should be 0.988. We estimated this availability target for a line section between two adjacent stations. In a railway network there can be many stations. The highest availability target that we estimate for any section will determine the availability target for the whole railway network. When we determine the availability target for a specific railway network, the infrastructure manager can estimate the maintenance investment in that particular network over a period of time to achieve that availability target.

4 DISCUSSIONS

The increasing complexity of modern technical systems has resulted in high reliability, maintainability and availability requirements. These requirement need to be met by the system owner to remain competitive. Setting these requirements is difficult when the systems have many stakeholders. Railway infrastructure is one of these systems. The system availability of the railway infrastructure directly affects the punctuality and capacity of the railway network. Failure to meet the capacity and punctuality requirements of the railway network incurs a penalty for the infrastructure manager who owns the system. Therefore, the infrastructure manager needs to estimate the availability target that it intends to achieve so as to meet the capacity and punctuality requirements. The availability target also enables the infrastructure manger to estimate the maintenance investment over a period of time. In this paper we have developed a model for estimating the availability target of the railway infrastructure in Petri-Net. The model considers the design, operation and failures of the infrastructure to derive a relationship between availability, capacity and punctuality. The relation is derived by means of an example presented in this paper. When the capacity and punctuality requirements are known, the infrastructure manager can estimate the availability target for this particular scenario explained in the example. However, the results show that the operational capacity does not change with the change in the system availability. This is due to the fact that the operational headway is kept large enough in the train timetable to absorb the delays, to keep the operational capacity of the network close to the achieved capacity. In general, if the operational availability changes, the operational capacity has to change. It actually does change for the hour during which a failure occurs (see Fig. 7); but during the next hour, the operational capacity increases as the delayed trains arrive along with the scheduled trains and we estimate the operational capacity as the average of all the hours. However, if we lower the operational headway, we can observe that the operational capacity will change with the change in system availability. This model will help the infrastructure manager to estimate the availability target of the railway infrastructure based on the capacity and punctuality requirements. A proper setting of the availability target will allow the infrastructure manager to avoid penalties due to delay and lower capacity. This will also help to estimate the total maintenance investment that the infrastructure manager needs to make over a period of time.

REFERENCES

1. UIC “Capacity”, UIC Leaflet 406, International Union of Railways, Paris, 2004
2. M. Abril, F. Barber, L. Ingolotti, M. A. Salido, P. Tormos, and A. Lova, “An assessment of railway capacity”, *Transport Research Part E: Logistics an Transportation Review*, (Sept) 2008, pp. 774-806
3. Banverket, “Calculation manual: Aid for cost-benefit analysis in the railway sector”, Handbook BVH 706.00, Banverket, Borlänge, 2001 (in Swedish)
4. B. Nyström, “Use of Availability Concepts in the Railway System”, *International Journal of Performability Engineering*, Vol. 5(2), 2009, pp. 103-118
5. T. Murata, “Petri Nets: Properties, Analysis and Applications”, *Proceedings of the IEEE*, Vol. 77(4), 1989, pp. 541 – 580.

BIOGRAPHIES

Ambika Prasad Patra
Luleå Railway Research Center
Division of Operation and Maintenance Engineering
Luleå University of Technology
SE-97187 Luleå, Sweden

e-mail: ambika.patra@ltu.se

Ambika Prasad Patra is a PhD candidate at the Division of Operation and Maintenance Engineering, Luleå University of Technology. His research topics are RAMS and LCC analyses for railway systems. He received his MTech degree in 2002 in Reliability Engineering from the Indian Institute of Technology, Bombay, India.

Uday Kumar, PhD
Luleå Railway Research Center
Division of Operation and Maintenance Engineering
Luleå University of Technology
SE-97187 Luleå, Sweden

e-mail: uday.kumar@ltu.se

Dr Uday Kumar is Professor of Operation and Maintenance Engineering at Luleå University of Technology, Luleå, Sweden. He also is the Director of Luleå Railway Research Centre (www.jvtc.ltu.se) at the same university. His research interests are maintenance engineering and management, reliability and maintainability analysis, product support, life cycle costing, risk analysis, system analysis, etc. He is also a member of the editorial board and a reviewer for many international journals. He has published more than 150 papers in international journals and conference proceedings.

Per-Olof Larsson Kråik, PhD
Swedish National Rail Administration,
SE-97102, Luleå, Sweden

e-mail: per-olof.larsson-kraik@banverket.se

Dr Per-Olof Larsson-Kråik co-ordinates research and development activities for Banverket (the Swedish National Rail Administration). Dr Larsson-Kråik has written over 50 articles and case studies in the areas of tribology, reliability, maintainability, logistics, and supportability. He has carried out several research and consultancy projects for several organizations in Europe and Australia. Dr Larsson-Kråik’s research interests include systems engineering, rail wheel (w/r) interaction, w/r lubrication and strategic operation and maintenance decision making in an organisation.