

Predictive maintenance of railway subsystems using an Ontology based modelling approach

Authors: Paolo Umiliacchi, CNC s.r.l., Bologna, Italy
David Lane, Heriot-Watt Univ. Edinburgh, UK
Felice Romano, AnsaldoBreda S.p.A, Napoli, Italy

1. Introduction

The trend towards a more efficient, cost-effective and performance oriented railway system is strongly pushed by political, economic and environmental motivations, as it represents an important and strategic asset for guaranteeing citizens and goods mobility at reasonable cost, involves a 35 billion €/year still growing market and is by far the lower impact transportation mode. Despite some sporadic and non-objective scepticism [1], railways are the best (maybe the only) candidate to achieve sustainable transport capacity and growth, provided that they can keep pace with increasing customers expectations, evolution of market demand and high safety/security levels.

While the railways' impressive technological progress is evident to all, a lot can still be done in system optimisation and efficiency improvement, in order to increase substantially their overall competitiveness and market share, not only referring to state of the art high-speed trains, but also to high-capacity conventional trains.

Information and Communication Technologies (ICT) have already brought important innovations to railways in terms of improved safety (signalling systems, like ERTMS), support to operations (monitoring and control systems) and customer services (passenger information, electronic ticketing and so on). They have been and will be essential in turning railways into an Intelligent Transportation System [3], following basic guidelines stated by the new Directive on the interoperability of the rail system within the Community [5].

An area where improvements are needed and can be achieved using advanced ICT solutions is that related to rolling-stock maintenance and condition monitoring. The principal goal is to minimise the probability of a mission-critical fault occurring during train service, with strong consequences in terms of quality of service and extra costs, while at the same time keeping maintenance costs as low as possible, in order to reduce life-cycle spending and improve return of investment.

2. Maintenance of railway subsystems

Trains today are complex, real-time, distributed and reconfigurable systems, incorporating many embedded subsystems, which concur together in performing a high quality transportation service. Failure of any of such subsystems can have heavy impact on the service itself, with obvious deterioration of performance, reduction of perceived quality and increment of costs.

Sometimes a trivial but not negligible misbehaviour (e.g. a door not closing properly) can have negative consequences as delays, troubles to passengers, cascade impact on other transport services and costs to take countermeasures. Moreover, disruptions in the regularity of the service can produce irregular traffic conditions which, statistically, decrease the safety of the system.

For such reasons monitoring and diagnostic systems on board trains have become more and more important, and also more sophisticated, in an attempt to keep up with ever growing expectations coming from train operators, maintainers and final users.

It is important, in fact, to allow an evolution of maintenance strategy and processes towards more optimised and cost-effective solutions (see Figure 1).

Corrective maintenance, after the fault occurred, is to be avoided due to high associated costs. Not only does an additional transport service need to be organised and provided, but also a maintenance team must be sent and the train towed to the closest depot. Furthermore the length of tie the track is blocked and the fees for the incurred service delay as well as for delays caused to other trains can sum up to an unbearable total cost. This brings to a different approach, aimed at preventing the fault from occurring, activating suitable measures in advance.

Planned maintenance is currently the most common solution, but it reduces the useful life of components, due to early replacement, and often implies unnecessary maintenance activities, as they are scheduled in advance according to a priori criteria. Increasing replacement or maintenance time intervals will increase fault probability, falling back into the previous case.

Important improvements can be achieved considering the real condition of parts based on direct measurement or estimation from actual usage, which represents a best maintenance practice. Significant cost reductions can be achieved, partially counterbalanced by increased subsystem costs due to additional sensors, a local control unit and remote query capability. A more flexible maintenance planning capability is required as well. Therefore some limited intelligence is added to the system, but not exploited as much as could be possible.

The next step is to move to predictive maintenance, trying to estimate the time when a fault is likely to occur and adapt maintenance interventions accordingly. This strategy allows further maintenance cost reduction with limited additional implementation costs.

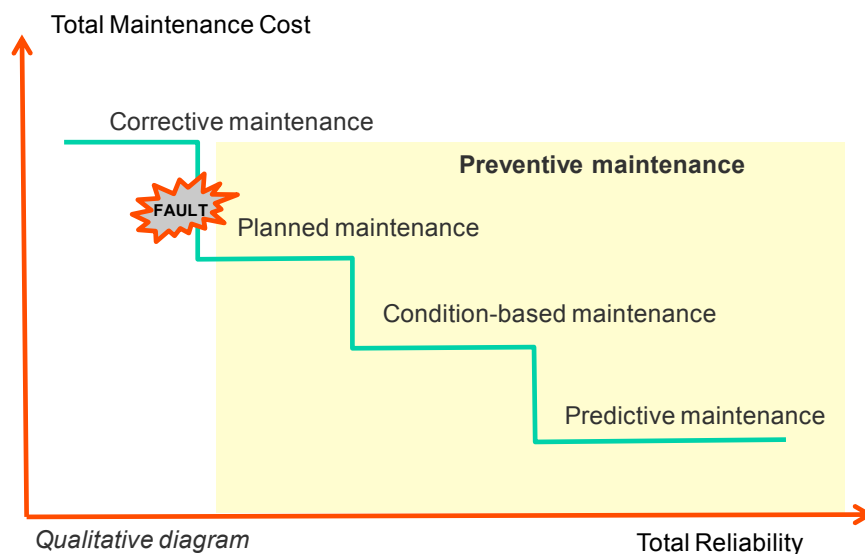


Figure 1 – Possible maintenance strategies

The problem is that this can only be achieved following a new approach in the conception, design and development of subsystems, so that appropriate methodologies for incipient fault detection and predictive maintenance can be really implemented. Technologies which allow a suitable level of intelligence to be embedded in the subsystems are required as well.

3. Previous projects

The need to improve diagnostic systems in order to better support maintenance processes was addressed by the EuRoMain FP5-IST project ([8], [9]). EuRoMain showed the possibility to implement a remote maintenance concept by means of on-board and on-board to ground communication (resulting from another FP5-IST project: TrainCom) and demonstrate the importance of a standard format for diagnostic data, so as to simplify their understanding and interfacing with ground tools.

This was the basis for Remote Condition Monitoring, allowing real-time checks of the actual situation of subsystems, in terms of a number of parameters and their values.

The EURoMain diagnostic data format was based on XML and Schemas, which were able to perfectly define the data syntax. This allowed maintainers to access data coming from different subsystems, trains and fleets based on a common language, so that the same diagnostic and maintenance tools could be used, e.g. to display information in a human understandable way.

This represented a good progress, compared to the situation when maintainers needed to use (and learn) different tools, in order to be able to access and work with data coming from different fleets.

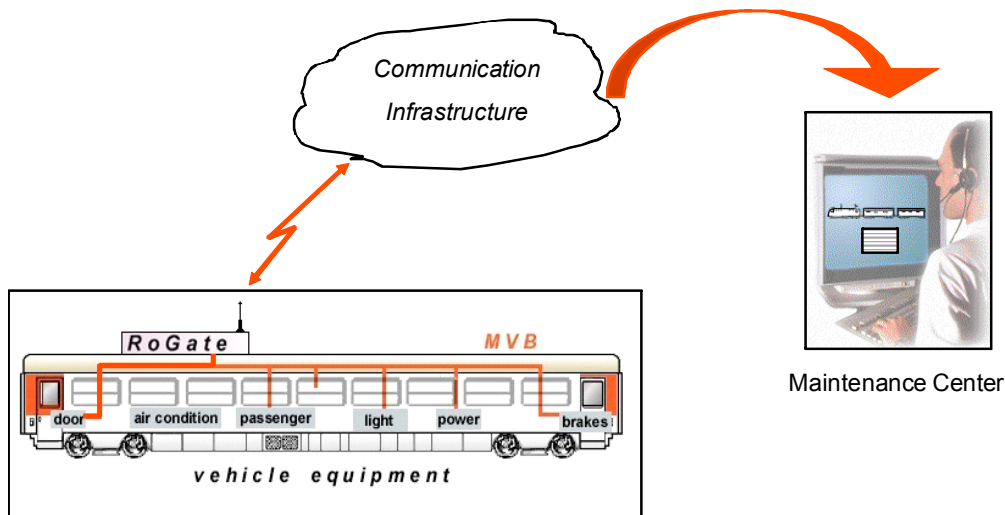


Figure 2 - TrainCom and EuRoMain maintenance concept

However, such results are not enough to achieve a full understanding about diagnostic data, as there are limitations concerning the semantic aspects. Adding a semantic context by means of an Ontology allows data to be turned into information which can be unambiguously understood and automatically elaborated by computers. This missing step had to wait a few years more, until it could be identified and developed in another research project.

Ontology is not a new idea. It is a formal language which allows any concept to be described within a domain, using logic rules to express it in terms of relations to other (previously defined) concepts. Starting from basic concepts (e.g. time and space) and moving towards more detailed concepts the goal is to fully describe the knowledge we have of such a domain.

ICT developments based on Ontology started more than 20 years ago, but became more popular with the wider availability of powerful computers. The first important applications have been implemented in the medical field, where it is important to analyse complex networks of symptoms in order to identify the real health problem.

Further advances in computer technology allowed this technology to be applied on board vehicles, where autonomous diagnostic and recovery capabilities required a more intelligent approach, for example in autonomous underwater vehicles, or remote off-shore plants ([10],[11]). The good success of such first implementations have pushed towards widening the application field.

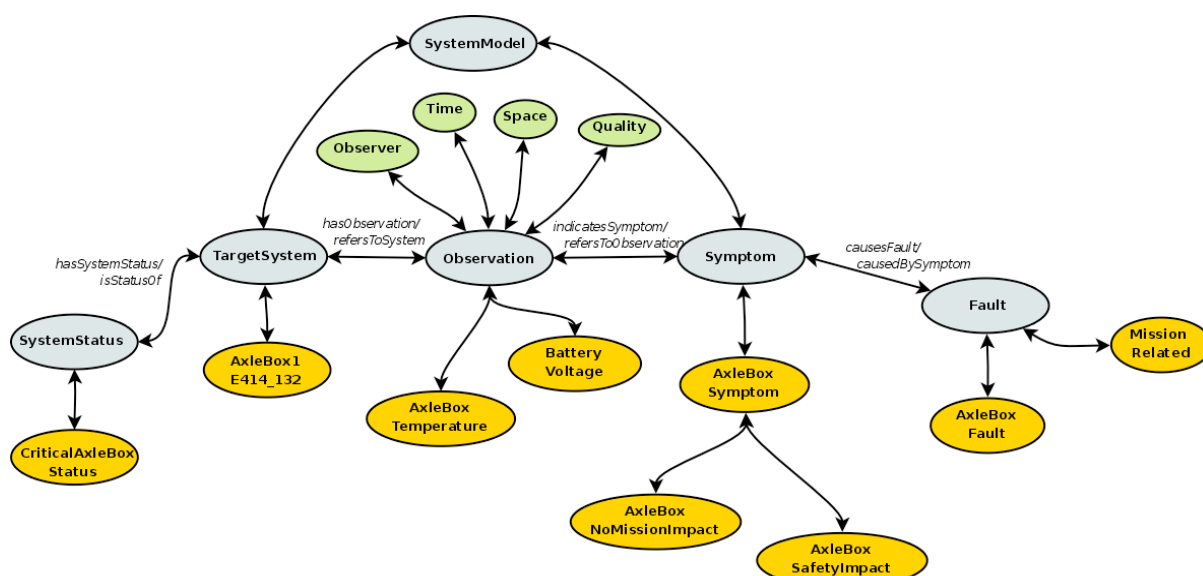


Figure 3 – A fragment of a possible Railway Ontology

Ontology based system approach in railways has few examples. The most important one is represented by the InteGRail, FP6 Integrated Project ([4], [7]), which defined an information sharing architecture based on Ontology and SOA (Service Oriented Architecture), showing how it could be effectively applied widely in the railway systems, for rolling stock, infrastructure, train operation and traffic management, bringing measurable improvement in the overall railway performance level.

In the specific area of diagnostics, InteGRail showed a prototype embedded system, working as a symptom and condition analyser on a locomotive and providing inputs for a more powerful ground system.

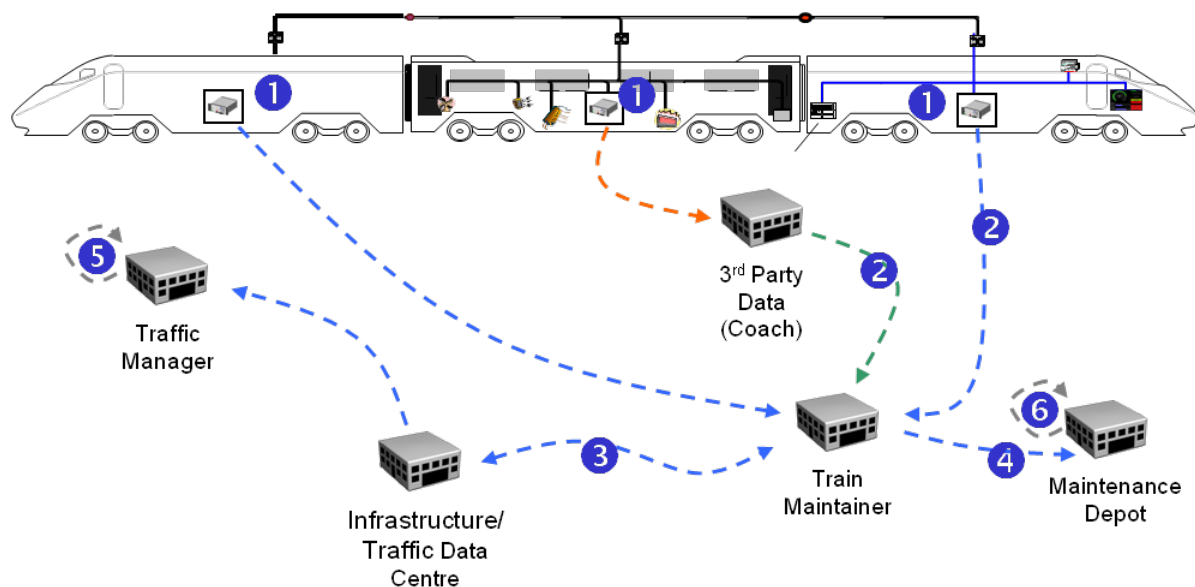


Figure 4 – InteGRail scenario related to diagnostics and maintenance

InteGRail was successful in demonstrating that cooperation between different applications was possible and the mechanism was working in a real environment ([2], [4]). It also identified the needed technologies and integrated them in order to achieve working prototypes.

The wide scope and size of the project allowed it to tackle all aspects of integration of railway information systems, defining an overall framework from which many specific, more focussed solutions can be derived. One of them is related to predictive maintenance applied to railway subsystems [4].

4. A new paradigm

Sometimes a higher level in the performance of a system cannot be achieved by simply improving current solutions. Instead, a completely different approach may be required. This explains why some early trials to achieve working predictive maintenance systems obtained quite poor results.

For example former tentative products based on “expert systems” did not come up to expectations because the concept was good enough, but the technology to implement it was not adequate.

Diagnostic tasks are today still largely under the supervision of human experts, who can take advantage of their wide and long-term experience in order establish the reason behind a fault and take appropriate actions.

This is not really satisfactory, in terms of reaction time, repeatability of results, possibility to record and transfer knowledge and the human errors that are always possible. Train operators are looking for new solutions which can better match the needs of current and future railway systems, improving their overall performance level and, in an indirect way, their safety.

InteGRail defined a new paradigm based on a combination of innovative technologies which together can bring a solution implementable in an embedded system [2].

A basic element of the new paradigm is that cooperation and information exchange between subsystems is needed. Some information is gathered by one subsystem which can have a wider

validity and usage (e.g. time, external temperature, train position and path). While it could be a wasteful cost to generate such information in all subsystems that need it, it is also important to ensure that consistent information are used throughout. Moreover, often the diagnostic process requires comparison between information coming from a number of subsystems: e.g. if all doors on a train take an average 4 seconds to close, while one of them takes 7 seconds, such door may be likely to break soon.

The key technology to enable wide peer-to-peer information exchange between subsystems is the SOA (Service Oriented Architecture) approach. According to SOA principles and the Reference Architecture provided by OASIS non-profit international consortium, the needed functionality is defined in terms of reusable services, which can cooperate in order to create flexible and agile applications. Mechanisms are provided in order to allow a device to discover other devices on the network and their associated services, to invoke the needed services and obtain the required outputs.

The current implementations of systems based on SOA use the technology of Web services, implemented in WSDL (Web Service Description Language), which provide a standard means of interoperating between different software applications, running on a variety of platforms. For its application in the field of industrial automation the profile DPWS (Devices Profile for Web Services) has been specified. It enables plug-and-play for networked devices. A device can detect DPWS-enabled devices on a network, then discover and invoke the Web service functionality each device provides. Abstraction from the heterogeneity of the underlying hardware device and operating systems services is thus achieved. The communication protocols used for message exchange are the standard protocols of the TCP/IP stack, which is in compliance with current trends in embedded systems.

A second essential element of the paradigm is to base the diagnostic process on an abstract behavioural model of the subsystem, which is able to capture the important concepts that form the basis of the subsystem functionality, removing the details of the individual product [6].

Traditional fault detection and diagnosis methods used in industrial systems rely on failure mode analysis and corresponding fault trees to look-up response options. These methods are not robust and do not work well where connected systems interact, or with complex systems where there are a combinatorial explosion of possibilities. They result in 'no fault found' messages, leaving human operators to use judgment and experience to choose safe actions and to identify repairs.

Recent developments in the semantic storage and use of information and its context have resulted in modelling techniques that make use of ontology to model engineering systems. The ontology stores the relationships between physical components in a system, as well as more abstract concepts about the components and their circumstances. The key benefit over simple databases is that reasoning can take place to infer the consequences of actions or changes in the ontology instances. In fault detection and diagnosis, this reasoning is the key ability that enables causes and effects of faults to be identified, even though they were not predicted a priori.

The contextual information provided by the ontology model allows a software package to infer information about the system without necessarily having been specifically coded for it. As an example, consider a train operator with two fleets of vehicles, one diesel powered and the other electric. In a system without an ontology, maintenance software would need to know about both types of fleet, and be specifically coded to establish the state of a diesel propulsion system, or an electric propulsion system from the condition monitoring data being returned. In an ontology driven system however, the ontology model shows that both diesel and electric motors are types of propulsion system; because of this the maintenance software can issue a generic request for the state of the propulsion system, and the appropriate data selected based on the propulsion system present on the vehicle in question. Additionally, rules encoded in the model can be used to infer the state of the motor dependent on its type, without the maintenance software needing to know those rules itself. As the operator expands its fleet, it only needs to add new instances to the ontology, requiring no change to be made to the maintenance software or to the existing fleets.

A final important aspect of Ontology-based modelling is its capability to embed in the model the existing diagnostic experience, coming from human experts and from the diagnostic system itself, thus enabling improvement in the diagnostic performance, and extending its benefits in a repeatable way to the complete fleet or to the whole system.

The third element of the paradigm is a direct consequence of the first two: distributed reasoning. The distributed nature of subsystems on board trains, of trains in the railway network and of

information repositories in the railway system, requires extensive and complete information processing to locate and access useful information wherever it is.

In fact, each reasoning module cannot rely on locally available information only in order to infer useful conclusions. In a distributed reasoning approach, all modules in the platform can decide to initiate a distributed reasoning process, and thus request reasoning processes on other nodes in the system to draw conclusions or supply information to provide input to the initiator at runtime.

As often huge repositories of historical data are maintained on ground sites, the distributed process must span on board and wayside subsystems. This implies that a reliable, seamless and hopefully continuous connection between the on board network and the ground network is in place.

In fact, some useful elaboration can better be performed on the ground, where more extensive data collection occurs (e.g. from the whole fleet), because some inference can only be possible comparing information from subsystems of the same kind, working on different trains.

The distributed, peer-to-peer approach enables the fault detection and diagnosis system to accommodate the dynamic nature of train configuration, where new subsystems can enter or old subsystems can leave at any time. An important point is also that the distributed reasoning approach must be able to operate within a constrained and embedded environment of subsystems.

A final consideration concerns the knowledge management aspect. During the lifetime of any diagnostic system, new knowledge can emerge from using the system itself. Re-occurring detection of certain fault situations in a distributed reasoning process can potentially be described in a more direct way into the ontology model, instead of having these situations repeatedly being detected by extensive distributed reasoning processes. Additional analysis by domain experts of the system's operation or detected faults can often result in enhanced constraints being represented in the ontology models describing the individual components and subsystems. Therefore, facilities must be provided for ontology version management and injection of new knowledge and models in the global fault detection and diagnosis platform.

The fourth and final element of the paradigm is related to the probabilistic aspects. Dealing with fault prediction means handling statistical data and this has several implications.

Dealing with probabilities is not a standard feature of an Ontology. Extensions are needed in order to embed and process the required information.

The ability of current-generation ontology technologies to handle uncertainty is rather limited. Different applications have different semantics, different knowledge and different data stores but want to use the same ontology. The data that answers to the ontology is often affected by uncertainty. This suggests that effective methods for representing and reasoning under uncertainty in complex, open-world environments is of vital importance for the success of an ontology-driven application.

For example, a train goes from being a good train to suddenly being a priority train. This is of course an evolving process. In a probabilistic ontology one could express that a train has already 75 percent chance of being a priority train. This percentage could fluctuate according to the characteristics of the train. These characteristics can be affected by uncertainty as well. For example, a door has a 15 percent chance that it will not close.

A reasoner could take advantage of this probabilistic information. For example, it could infer the probability that this train is a priority train using the probabilistic information about the characteristics of that train.

Such an approach is likely to lead to step changes in the prognostic capabilities of diagnostic systems, as prognostics is by its very nature uncertain. The use of belief based systems will increase the number of correct decisions leading to overall improvements in system metrics.

From the mathematical and statistical point of view, good results have been achieved using Bayesian Networks.

A Bayesian Network is a structure for performing reasoning using partial beliefs under conditions of uncertainty. These networks are often used in fields such as health care where decision-making occurs with uncertainty. Relationships are not directly explicit, but based on probability distributions, and therefore more knowledge can be gained of the confidence in any decision made. For each variable there exists an 'a priori' probability representing what is known about the variable without evidence. When there is evidence, for example clinical signs, then an 'a posteriori' probability can be calculated. This maps well to physical systems made up of many components each of which have particular curves of failure probability. Thus reasoning based on the differing probabilities can give good guidance on what requires repair. These can also be combined with

consequence of failure to support prognostics to allow maintenance to be performed before a failure occurs.

5. The door controller case

Door failure is a major cause of service disruption, repair costs and passenger complaints for railway transportation systems in urban, metropolitan and intercity services. A totally preventive maintenance approach is expensive to implement, and lowers the quality of service perceived by passengers when waiting until repairing is necessary. As a best compromise, predictive maintenance can schedule interventions as needed, ideally just before a fault will occur, optimising the cost/benefit ratio.

The train doors are operated by actuators mainly using air pressure power injected in appropriate cylinders associated with mechanical levers. The opening/closing action is electrically remote controlled by the vehicle on board computer of the Train Management System. Some sensors are provided on the actuator for movement detection.

Considering pneumatic doors for example, the main components can be identified and their relations mapped in a model, as here depicted:

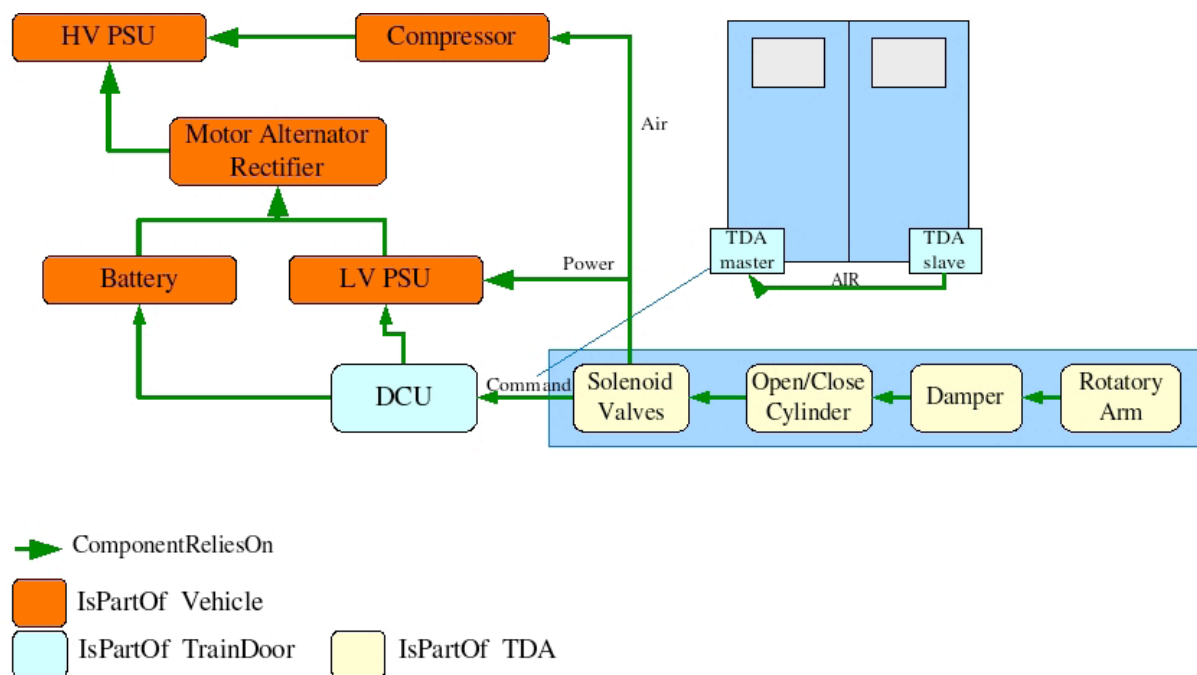


Figure 5 – Model of the pneumatic door subsystem

The model can be defined more or less in detail, as a compromise between complexity and accuracy. In Figure 5 some more detail has been given to the Train Door Actuator (TDA), compared to the Door Control Unit (DCU) and Power Supply Units (PSU).

The model can be coded using a suitable language, e.g. OWL (Ontology Web Language). This is normally not done directly, but using a specific tool (e.g. Protégé).

Of course, the diagnostic system must rely on a certain degree of observation capability about the subsystem status, which comes from sensors: typical measured parameters can be voltages, currents, air pressure, velocity, and so on.

The possible faults in the subsystems need to be analysed, considering each component (including sensors, of course) and functionality. For example:

- Door motor circuit failure. Motor current has not increased 10 sec after open/close control.
- Door failed to lock. Door is not locked after door close switch has been activated

The result of such an analysis is a fault classification which can be incorporated into the Ontology model, together with relations between faults and observed symptoms.

It becomes also possible to perform a root cause analysis, in order to understand the real problem to be fixed by means of maintenance activity. For example, in the case of a fault in the Low Voltage Power Supply Unit (LV PSU), the door system cannot work. Traditional diagnostic systems will react activating several alarms, e.g. fault in the Door Control Unit and fault in the Door Subsystem.

Following relations in the Ontology, the new diagnostic system can resolve the root cause and correctly issue only an LV PSU fault alarm.

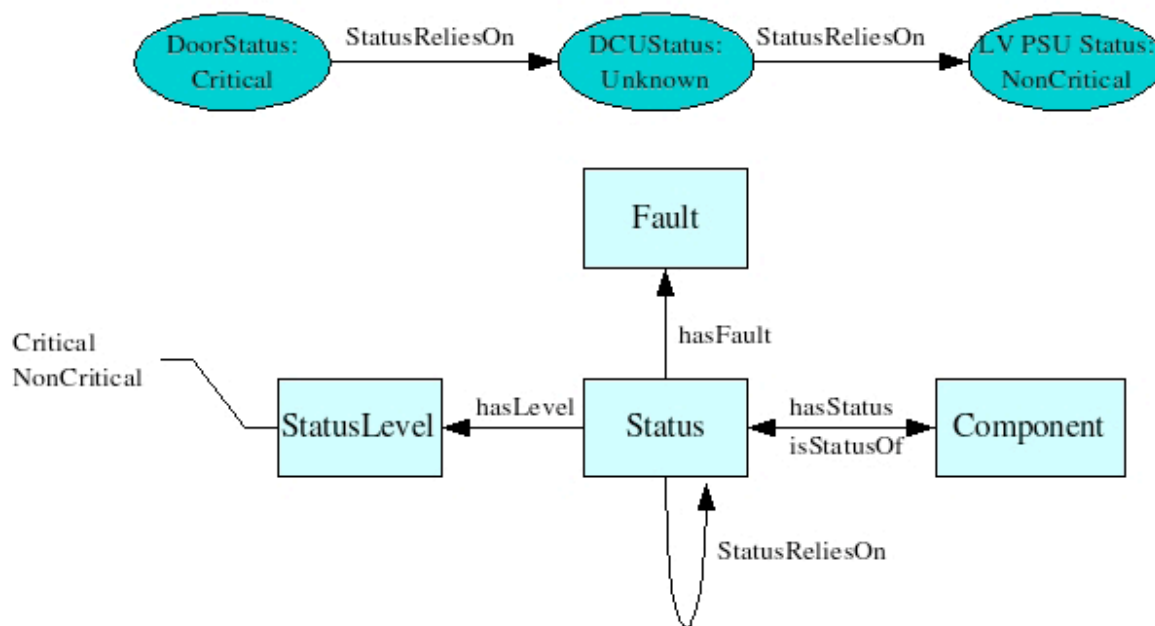


Figure 6 – Root cause analysis

From this, we have a basis upon which a predictive mechanism can be built, based on a probabilistic approach.

Observe the situation modelled in the Bayesian network in Figure 7. All three variables have two possible values T (for true) and F (for false). If we abbreviate the names of the variables to H = HandleFault, D = DoorFault and S = SignalFault than the joint probability function is given by $P(H,D,S) = P(D | H,S)P(H)P(S)$.

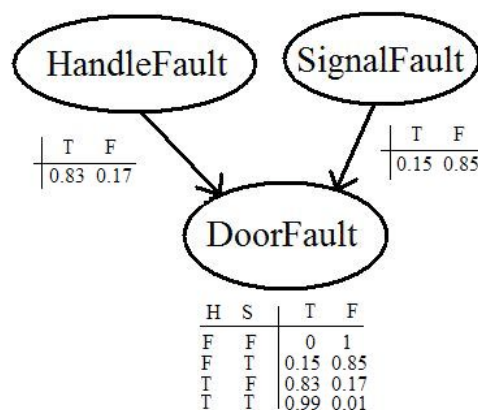


Figure 7 - An example of Bayesian network

The model can answer questions like "What is the likelihood that there is a handle fault, given the fact that there is a door fault?" by using the conditional probability formula and summing over all nuisance variables. This illustrates some of the basic capabilities offered by this approach.

Additional answers can become possible when extending elaboration (reasoning) to ground repositories of historical data, in order to identify in advance dangerous situations which can be hidden (or dormant) or incipient. Such analysis can identify also anomalous conditions, as recurrent faults at the same door and/or at a specific location. For example:

- Correlation between event and position. An alert is generated when a number of faults over a threshold occurs in the same location (environmental data reports for each event the Train number, Railway line, Position along the line, train mileage);

- Anomalous repetition of an event compared with the mean value, i.e. a particular door has a number of faults that greatly exceeds the mean value that would be expected considering the mileage of the train (thresholds could depend on the mileage itself: the oldest trains usually report a number of faults higher than the more recent ones);
- Anomalous repetition of temporarily fault events at the same door: it may be that the door is slower than expected in the closure or opening operation so that the coach reports that door n is not working (at speed equal to 0) for an unexpected time. This suggests the mechanical behavior of the door and the status of the door switches should be checked.

6. *Current and future activities*

The diagnostic system of the future can be seen as an ecosystem where many heterogeneous networked subsystems cooperate in order to achieve the common goal of efficient and reliable train operation.

The described approach also minimises the subsystem additional cost, as symptom analysis and predictive diagnostics are mainly implemented in software and its development cost is shared between all produced equipment.

Following prototyping and test results, the approach can well be extended to other subsystems, bringing even higher benefits to the railway business.

A number of initiatives are in progress which can bring important results in the near future. Several innovative aspects of the defined paradigm and related diagnostic platform need further investigation to be fully understood and developed. To this end, some European research project proposals have been submitted or are under preparation within the Seventh Framework Programme. Selected proposals can become a cluster of many small projects which can act as a bridge between the extensive results originated by the huge InteGRail project and the specific deployments focussed on individual achievements and upcoming products.

Investigations on the proposed approach are already in progress, carried out by the authors as a cooperation between their companies. Despite a slower pace due to reduced investment during the global crisis, some preliminary results have been achieved and more activities are currently in progress.

A number of component and subsystem manufacturers stated their interest in analysing the described approach as the basis for a new generation of products sharing a common interface at application level. This can bring about a higher level of subsystem interoperability and eventually achieve the final result of a real plug-and-play behaviour, with obvious benefits in terms of time and cost needed for integration and configuration of vehicles and trains.

The results will be considered for possible standardisation by the IEC-TC9-WG46 working group, dealing with multimedia applications on board trains. In reality, the WG46 scope of work is quite wide and covers all non-operational information, including driver, crew, operator and maintainer oriented services. The envisaged result will be the new standard series IEC 62580, currently at the level of Committee Draft, which will represent an important step towards a higher level of interoperability between train subsystems.

7. *Conclusions*

Starting from results achieved in other industrial sectors and in railway research projects, an innovative approach for a new generation of fault detection and diagnostic systems has been derived.

The described experience was very important in order to define a new platform and show that it can really work in the railway environment, bringing significant improvements. However, some more development work and field tests are required in order to achieve a fully engineered product which can be widely deployed in railways.

Integration with the on board communication network and train-to-ground radio system will enable remote monitoring and alerting in case of problems, which in turn will contribute to a reduction in any possible impact on normal operations. The use of existing and new international standards (as those coming from IEC) will also contribute to cost reduction and higher level of interoperability.

Bibliography

- [1] Kallistratos Dionelis - Traffic Technology International – January 2011
- [2] Michele Bernocchi, Emilio Miguelanez - “The contribution of European Research to more intelligent Railway Maintenance: the InteGRail project” - Euromaintenance 2010 International Congress - Verona Fiere - 14th May 2010
- [3] Paolo Umiliacchi, Didier van den Abeele, Pieter Dings, Valerio Recagno – “Turning Railways into an Intelligent Transportation System by better Integration, Management and Exchange of Information” - ITS World Congress – Stockholm – September 2009
- [4] R. Shingler, G. Fadin, P. Umiliacchi – “From RCM to predictive maintenance: the InteGRail approach” - 4th IET International Conference on Railway Condition Monitoring – Derby - 18th June 2008
- [5] DIRECTIVE 2008/57/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the interoperability of the rail system within the Community - 17 June 2008
- [6] Richard Lewis, Florian Fuchs, Michael Pirker, Clive Roberts, Gerhard Langer: “Using Ontology to Integrate Railway Condition Monitoring Data” – RCM-2006 – IET-International Conference on Railway Condition Monitoring - Birmingham – UK – 29-30 November 2006
- [7] P. Umiliacchi, R. Shingler, G. Langer, U. Henning: "A new approach to optimisation through intelligent integration of railway systems: the InteGRail project" - WCRR 2006 - Montreal - 7 June 2006
- [8] P. Umiliacchi, G. Cau, A. Gatti: “La ricerca europea per la manutenzione ferroviaria: il progetto EuRoMain”(European research for railway maintenance: the EuRoMain project) – XXI Congress A.I.MAN. – La manutenzione: processi e competenze – Milan – 16 September 2004
- [9] R. Shingler, P. Umiliacchi : “Advances in railways maintenance: the EuRoMain project” – WCRR 2003 – World Conference on Railway Research – Edinburgh – 2003
- [10] E. Miquelanez, P. Patron, K. Brown, Y. Petillot, D.M. Lane "Semantic knowledge-based framework to improve the situation awareness of autonomous underwater vehicles". IEEE Transactions on Knowledge and Data Engineering.
<http://www.computer.org/portal/web/csdl/doi/10.1109/TKDE.2010.46>
- [11] E. Miquelanez. D.M.Lane “Predictive Diagnosis for Offshore Wind Turbines using Holistic Condition Monitoring”. IEEE International Conference OCEANS 2010, Seattle, October 2010