Smart Railway Maintenance – Challenges and Research Directions

João T. Fernandes[a], Marilia Curado[a], Howard Parkinson[b], Fernando Boavida[a]

[a] *Center for Informatics and Systems, Department of Informatics Engineering, University of Coimbra, Coimbra, Portugal*
[b] *Digital Rail Limited, Lancaster, United Kingdom*

Abstract

Railway maintenance is crucial for the operation of current railway systems, as these are subject to strict requirements in terms of efficiency, cost-effectiveness, quality of service and scalability. Existing solutions for railway maintenance are, traditionally, either reactive or periodical, which is far from optimal. Moreover, they tend to use expensive, specialized approaches, and do not take advantage of emerging technologic tools and techniques, such as 5G systems, cloud-based operation, and intelligent data processing and information extraction/inference for achieving predictive maintenance. In fact, predictive maintenance is the core paradigm of what is now called Smart Railway Maintenance. This paper provides an insight on Smart Railway Maintenance, by overviewing existing approaches to railway maintenance systems, identifying their limitations with respect to current requirements, presenting technologies with potential for supporting future Smart Railway Maintenance systems, and offering new research directions. Moreover, in line with the suggested research directions, the paper also proposes a novel architecture to address the encountered issues/limitations.

Keywords: Smart Railway Maintenance, Internet of Things Context-Aware, Software-Defined Railway Monitorization, Railway-Oriented Architecture

1. Introduction

Over the last century, rail transport has evolved from steam, low-velocity trains to electric, high-speed, and high-tech trains. In parallel with trains, which carry more people and freight, infrastructures also evolved in order to provide safety, comfort and robustness. Nowadays, trains can be used for travelling between cities and in urban areas (e.g., London Suburban Area), for delivering freights between two or more different points (e.g., between a seaport and an in-land industrial facility), or even for connecting different countries (e.g., the Thello train, which connects France to Italy). On November 2017, the Rail Statistics Compendium released a document with the numbers of passengers traveling in England. It stated that, from 1995 to 2017, that number increased 136%, with 5.1% growth between 2015 and 2017. On the other hand, freight rail usage only increased 2% since 1995, mainly due to the steep decline in the use of coal. Deutsche Bahn's (German Train's Company) Integrated Report for 2017[1], states an increase of 8.4% in voyages from 2016 to 2017, roughly corresponding to 200 million more passengers. In line with this, in their 2017 Annual Report, CP - Comboios de Portugal (Portugal's train company) reported a passenger increase of 6.3% in relation to the previous year, with expectations of continued traffic growth over the next coming years[2]. These numbers are typical of what is happening not only in Europe but also around the world, with general overall increase in passengers and freight, and, consequently, added pressure on both trains and infrastructures not to decrease the quality of service. Naturally, continuous operation with stringent quality requirements turns inspection and maintenance into critical activities, for which procedures, tools, and technologies must be developed and put in place. Moreover, constant monitorization of both vehicles and infrastructure is crucial.

Thus railway (meaning both vehicles and infrastructure) monitoring and maintenance activities are unavoidable. On the other hand, these must be done in a cost-effective way, taking advantage of all technological advances for simultaneously delivering the needed functionality and reducing costs. In this

[1] Available here: https://www.deutschebahn.com/resource/blob/1262924/d73cde1ca54b69f595e4bb5800ad011b/cms9-2016_duf_en-data.pdf

[2] Available here: http://web3.cmvm.pt/sdi/emitentes/docs/PC68257.pdf

respect, several emerging technologies are coming into play and must continue to be explored to their fullest potential, leading to what is now envisioned as 'Smart Railway Maintenance' (SRM). These are the Internet of Things (IoT) and Industry 4.0, Big Data, and, last but not least, 5G [1]. IoT can provide a large variety of sensors and tools for effective railway monitoring. Big Data techniques can and, in fact, must be used for processing the huge amounts of data that can and will be collected. 5G can provide the means for transferring massive amounts of data for subsequent processing, as well as for transferring real-time data with very low latency. In addition, software-defined approaches can bring much needed flexibility, interoperability, and ease the development of SRM systems and applications. Jointly, these technologies may lead to a reduction in repair costs and may increase the reliability of the overall railway system.

Some recent papers provided information on the potential and importance of IoT and Big Data for SRM. Bernal et al. [2] and Hodge et al. [3] proposed several sensor-based monitorization techniques to detect defects on both railway vehicles and infrastructure. In the same line, Fraga-Lamas et al. [4] presented a survey of communication technologies and IoT services that can be used for monitoring train components (i.e., coach temperature, door controls, brakes, etc). The current paper builds on the mentioned papers, as well on a variety of other relevant sources, in order to provide a clear, comprehensive view of current SRM, and to identify challenges and research directions for cost-effective, timely, functionally-rich, and robust monitoring and maintenance of railway systems. Given the above, the contributions of this paper are the following:

- Identification of the requirements of SRM;
- Overview of the state-of-the-art on train and infrastructure anomaly detection, together with an analysis on how the existing approaches address the identified requirements of SRM;
- Identification of tools and technologies with high potential for SRM;

- Identification of open issues and research directions for SRM, given the identified tools and technologies;
- Proposal of a high-level architecture for software-defined SRM.

In order to reach the objectives set out for this paper, Section II provides a detailed problem description, which includes an overview of approaches to SRM, a presentation of supporting technologies, and the identification of system requirements. The state-of-the-art is extensively presented in Section III from the perspectives of both infrastructure and vehicles. Section III includes an overview of existing, potential tools and technologies for future SRM. Section IV identifies open issues and proposes some research directions. Conclusions and guidelines for further work are presented in Section V.

2. Problem Description

Smart Railway Maintenance can be performed using one of several approaches, each of them having different consequences in terms of reliability, safety, downtime, human resources, and cost. Subsection 2.1 summarizes these approaches. In subsection 2.2, SRM supporting technologies are presented, comprising: (1) sensors, and the differences between them, (2) communication systems, with focus on 5G, and (3) data processing, more specifically, offline and online processing. Subsection 2.3 categorizes and explains the requirements of SRM systems.

2.1. Approaches to Smart Railway Maintenance

Smart Railway Maintenance (SRM) is an area inside the broader area of the Internet of Smart Trains [4], that poses considerable challenges due to the complexity and criticality of railway systems. When dealing with SRM, two complementary perspectives can be used: vehicle maintenance and/or infrastructure maintenance. Regardless of the concerned perspective, several approaches to SRM can be adopted, namely reactive maintenance, preventive maintenance, proactive maintenance, reliability-centered maintenance, and predictive maintenance. These will be explained in the current subsection. Complementary to the text, Table 1 summarizes the main features of each strategy and provides usage scenarios.

2.1.1. Reactive Maintenance

In this case, maintenance occurs only when a component fails completely. Despite being one the most used strategies (in combination with preventive maintenance [2] [5]), reactive maintenance has high repair costs. On the other hand, maintenance costs are low or even null because the vehicle simply runs-to-failure. This approach may raise safety concerns, and, in practice, is always combined with other types of approaches, like preventive maintenance. For example, a fractured wheel is a critical defect that needs to be taken care of as soon as it is detected. Continuing to operate until the fissure reaches a critical stage compromises safety. On the infrastructure side, similar problems can occur, leading to safety issues and accidents with potentially serious results. Moreover, this strategy requires more backup equipment or vehicles, because downtimes are typically higher in these cases.

2.1.2. Preventive Maintenance

Unlike reactive maintenance, this strategy makes use of scheduled maintenance, at specified intervals. It can be looked at as a compromise between reactive and predictive maintenance, with lower maintenance and repair costs when compared to reactive maintenance. Vehicles have less probability of having critical failures, and both maintenance and repair costs are stabilized in a lower landing. However, failures are only detected during the scheduled revisions and, although the risk of failure is low (when compared to reactive maintenance), there is a chance that this strategy still leads to critical problems.

Strategy	Main Features	Scenario
Reactive	Reduced maintenance costs. High repair costs. Safety issues. Longer downtimes.	The train operates until complete failure, even if anomalies occur during trips (as long as they do not prevent the train from operating).
Preventive	Lower maintenance and repair costs than reactive maintenance. Does not fully minimize risk of critical problems	Trains and/or infrastructure are maintained at fixed points in time or mileage, whether they have problems or not.
Proactive	Reduced failure probability. Higher cost of components and materials is compensated by higher reliability and longer times between maintenance.	Maintenance teams use the best possible tools, components and materials, in order to reduce the failure probability.
Reliability-Centered	Prevents unnecessary costs. Makes use of RCA methods to study the problem,such as Ishikawa diagram or Five Whys, to increase reliability and vehicle's uptimes.	A vehicle can run with a certain level of reliability. Repairs occur when this is compromised.
Predictive	Reduced repair costs. Higher upfront costs. Minimizes risks of critical problems.	Constant monitoring allows to predict/detect problems and timely schedule repairs.

Table 1: Maintenance strategies

2.1.3. Proactive Maintenance

Is this case, maintenance is performed using the best possible resources, tools, components and materials. By maximizing the quality of maintenance interventions, the probability of failures is drastically reduced, leading to longer operational periods. Naturally, this strategy can be – and normally is – combined with other strategies. Usage history and experience of repair teams dictate the time between proactive maintenance instances.

2.1.4. Reliability-Centered Maintenance

In this case, the focus is not only on the detection of problems, but also on determining if the problem is serious enough to compromise a given reliability target. Put in another way, if detected problems do not threaten the overall safety and are unlikely to cause failures, then repairing them is not needed and, in fact, would represent an unnecessary cost. This type of problems should be dealt with in preventive, scheduled maintenance operations. On the other hand, problems that may affect the system's reliability should be dealt with as soon as possible, in order to prevent failure costs that, in general, are always higher than maintenance costs. Reliability-centered maintenance uses Root Cause Analysis (RCA) to study the problems, using techniques such as Ishikawa diagrams or Five Whys [6].

2.1.5. Predictive Maintenance

With predictive maintenance, constant monitorization is performed in order to detect problems before they turn into failures. This maintenance strategy makes use of past and current data collected from the monitored system [5]. As it requires constant monitorization, there are higher upfront costs. These costs, although higher, are compensated whenever the system needs to be repaired because repair time and the risk of total failure are minimal [5]. Using a variety of sensors (e.g., accelerometers and sound sensors), it is possible to collect data that, for example, can then be used to assess the conditions of wheels, tracks, or track bed. In addition, by using historical data, the evolution of the target elements or system can be tracked and, consequently, this can be used for improving fault prediction.

Nowadays, companies focus on preventive and reactive maintenance, to minimize maintenance costs. Nevertheless, although these strategies can lead to controlled maintenance costs, they do not preclude high repair costs, as it is always costlier to repair vehicles or systems that have run to failure, not to mention the costs of dealing with potential disasters. This is the reason why interest in predictive maintenance is growing considerably, especially because this type of maintenance can explore the use of emerging technologies such as Internet of Things, Big Data analysis, and very low latency networks such as 5G. Moreover, being able to predict failures can help companies reduce both maintenance and repair costs. This, of course, does not preclude preventive maintenance, as vehicles and systems have a given lifespan and scheduled maintenance. Preventive maintenance is and will always be needed, but providing insights on the components can reduce critical failures and, at the same time, reduce repair costs and downtimes.

2.2. Supporting Technologies

Current technology allows collecting, transferring, and processing considerable amounts of data. This is so in smart cities, industry 4.0 systems, smart transportation, and connected vehicles. Thus, it is also important to understand how the technologies that make this possible can be used for SRM systems. This is, in fact, the purpose of the current subsection. We start by addressing the types of sensors that can be used for the problem at hand. Then, we proceed to analyze and discuss the role of communication technologies, of which the emerging 5G technology is the most promising one. Last but not least, we concentrate on the problem of processing the collected data, identifying alternative data processing strategies.

2.2.1. Sensors

Prior to developing any anomaly detection system, it is important to define which defects are going to be targeted [2] [3]. For example, if we are targeting wheel problems, several types of sensors can be used. On the other hand, there is no single sensor that can detect every possible defect that may exist in a wheel or track. For instance, in order to analyze fissures and cracks, the most common sensors are cameras (visual inspections). These are, normally, high-speed, high-cost cameras, able to work under adverse conditions (e.g., low light, unstable weather). On the other hand, low-cost sensors are increasingly being used, such as accelerometers, gyroscopes or ultrasonic sensors, with the objective of replacing or complementing other, more expensive devices. When compared to cameras, these sensors alone do not allow for 3D modeling of a wheel or track, making them more suitable for external defects detection (e.g., abnormal vibrations, sound variations, among others). However, Light Detection and Ranging (LIDAR) sensors are interesting for 3D scanning and, nowadays, they are affordable. In general, using this type of sensors is still novel and their potential has not been fully explored yet. To summarize:

- High-speed cameras are used to detect cracks and fissures in several elements of trains and/or infrastructure, and are capable of working under severe conditions. Due to their characteristics, they have high cost. High-speed cameras can be used for both train and infrastructure anomaly detection.

- Smaller, low-cost sensors are now increasingly being used to replace high-cost cameras. However, they are not capable of detecting the same types of problems as the cameras and, thus, should be looked at as complementary tools. These sensors collect data such as sound and vibrations in order to detect anomalies.

The use of inexpensive sensors opens up a whole range of possibilities, such as having a higher number of failure detection systems and installing them in all types of trains, not only in selected trains or in special, dedicated anomaly detection vehicles. Moreover, a large set of problems can be detected. For instance, using accelerometer sensors it is possible to gather data pertaining to the stability of the train and, consequently, use that data to detect problems in tracks. Similarly, the use of sound sensors can contribute to the analysis of the track-bed. To analyze the wheels, it is possible to use ultrasonic or LIDAR sensors (pointing to the wheel) to provide information in the external wheel layer. Despite the focus on infrastructure sensors only, Hodge et al. [3] show a variety of sensors that can be used for multiple purposes. This approach can help not only to reduce the amount of equipment needed to develop a reliable system, but also to ensure that the used sensors cover a wide variety of defects (for example, an accelerometer can detect vibrations on bridge and tracks, and detect dynamic acceleration and movement on the track bed. On the other hand, an acoustic sensor can detect cracks, fatigue, and stress).

2.2.2. Communication Systems

Currently, communication between trains and other systems, such as control and/or monitoring systems, resorts to Global System for Mobile - Railway (GSM-R) [4]. Nevertheless, as mentioned previously, SRM may generate considerable amounts of data and will require data rates and bandwidth that are much higher than the ones provided by GSM-R.

5G technology is currently being developed and it is expected that the first products will be available in the market by 2020. The main key features of these systems from the user perspective are very high bandwidth and extremely small latency. Nonetheless, current 4G technology can also provide sufficient features to be used within the railway area, with higher bandwidths and transfer speeds, when compared to the previous generation. These features can be crucial to improve SRM data transfer.

The opportunities offered by these new communication technologies can now be explored by the railway industry, leading to the replacement of older, bandwidth-limited GSM-R solutions and to added functionality, lower latency, and higher reliability of SRM systems.

2.2.3. Data Processing

As already mentioned, IoT, Industry 4.0, and 5G will allow collecting large amounts of monitoring data. This data will then be processed in order to support all kinds of maintenance activities, most notably for predictive maintenance. Regarding data processing, two options exist: online or offline. In the former case, collected data is immediately sent over the network to a processing site (typically in the Cloud) and it is processed as soon as it is received. In the latter case, processing is executed at a later point in time. This may resort to temporarily storing the data in a local device (such as an SD card), or to sending the collected data over the network to the cloud for subsequent processing, or both. Offline data processing typically occurs at regular intervals, such as every minute, every hour, or every day, for instance.

The computational power typically resides in the Cloud (possibly involving multiple servers, and exploring techniques such as neural networks). Processing results can then be sent to other systems, such as management workstations, smartphones, actuators, among other types of equipment. The choice of which processing approach to take – online or offline – depends on the scenarios and, in fact, a mixture of both approaches may be used at the same time in order to provide different views of the monitored systems. For instance:

- Offline processing can be used in situations where results are not needed in real time, e.g., when long-term behavior data of components is being collected. It may also be used whenever the circumstances require temporary storage of data, such as whenever there is some kind of congestion in the network or whenever the Cloud environment lacks resources, thus saving computational power for more important data.

- On the other hand, online processing can be used in critical situations. For example, if there is a known issue in a certain part of the railway track or if a vehicle has been down with recurring wheel problems, constant monitoring needs to be performed in order to understand changes over time. Naturally, this will require considerable bandwidth as well as low communication latency.

Data processing is a critical part of any SRM system, as it provides the information on which to decide what to do and how to act, in what maintenance is concerned. For this, all of the mentioned supporting technologies must be in place: sensors, communications, and processing power. For critical situations, online processing is desirable, as it is the basis for real-time monitoring and for efficient anomaly detection. Offline processing may be used in a complementary way, in order to support long to mid-term views of the system behavior. The adequate combination of all of these factors is crucial for constructing an effective SRM system.

2.3. Requirements

The previous subsections provided information on existing maintenance approaches and on the way new or emerging technologies can be used for the construction of future smart railway maintenance systems. It is now time to address the requirements of such systems, so that we can subsequently fully assess the existing approaches to SRM and the resulting challenges. Table 2 summarizes the list of SRM requirements, and can be used as guidance for the text in the body of this subsection.

SRM requirements can be organized into four areas – namely, data processing, anomaly detection, predictive maintenance, and scalability – the latter being orthogonal to the other three areas.

In what concerns data processing, an SRM system should be able to operate locally or remotely. In the case of remote data processing, one or more servers typically located in the Cloud process the data. The advantage of this approach is that storage and processing resources are elastic, i.e., they adapt to the needs of the system. Nevertheless, this approach requires adequate connectivity between the target system and the Cloud, in terms of bandwidth, reliability, and

Area	No.	Subarea	Requirement
Data Processing	1	Remote data processing	The system must send data to a remote server, in the Cloud, for processing.
	2	Local data processing	The controller must process data locally.
Anomaly Detection	3	Wheels anomaly detection (superficial)	The system must detect superficial defects, such as small holes, bumps, small deformations, etc.
	4	Wheels anomaly detection (internal)	The system must detect structural defects, such as fractures, serious deformations, etc.
	5	Tracks anomaly detection	The system must detect squats and/or fractures in the tracks, or major problems that can lead to vehicle derailment.
	6	Trackbed anomaly detection	The system must detect ballast bed problems, track bed structural problems, etc.
Predictive Maintenance	7	Predictive infrastructure maintenance	The system must continuously monitor the infrastructure for detecting anomalies before they evolve into failures or serious problems.
	8	Predictive vehicle maintenance	The system must continuously monitor vehicles for detecting anomalies before they evolve into failures or serious problems.
Scalability	9	Scalable System	The system must be able to support growth in terms of sensors, sensed data, communication needs, data processing, and also be able to deal with a higher number of vehicles.

Table 2: List of Requirements.

availability. Data sent to the Cloud is (mostly) raw, and filtering and noise reduction techniques are needed to 'clean' the data. As the available processing resources are virtually unlimited, processing time is usually low. Moreover, the distributed nature of the process enhances overall reliability. On the other hand, it may also be important to locally process the collected data, with the objective of reducing the amount of data that needs to be stored and/or sent over the network. Although local processing resources are, typically, limited, existing solutions such as specially built printed circuit boards (PCB) can combine effective processing and sensing hardware that can be used for collecting and immediately process sensed data.

Regarding anomaly detection, the focus is on two major components of the railway system: train (wheels) and infrastructure. In the former case, detection can address superficial problems, such as bumps or small deformations. Using specific, inexpensive sensors placed in the bogie structure, such as accelerometers and gyroscopes, these and other superficial problems can be efficiently detected. To detect internal wheel problems, expensive sensors (e.g., cameras) must be used. Although expensive, these sensors can provide information that can be useful to understand the behavior and structural condition of wheels. For infrastructure anomaly detection, the target systems are tracks and track bed. Monitorization of both tracks and track bed can be performed by sensors similar to the ones used for monitoring the wheels. However, to detect track bed problems, sound sensors can be used to listen to sound variations, instead of using accelerometers. Moreover, ultrasonic and LIDAR sensors can provide a 3D view of the head of the rail, to detect squats and small deformations.

Predictive maintenance strategies can play a crucial role in railway infrastructure and vehicles maintenance, leading to the optimisation of maintenance interventions, reduced costs, and higher reliability. The price to pay for this is that continuous monitoring must be performed in order to detect problems at an early stage and decide on appropriate actions. In combination with historical information, predictive maintenance strategies can be quite effective in infering the health of railways systems, assessing the criticality of defects, and planning maintenance interventions. With such strategies, breakdowns can be virtually eliminated, and downtimes can be minimised.

Last but not least, scalability is one of the most important features of an SRM system, as it provides system growth capability. As mentioned in the Introduction, railway traffic is growing at a fast pace, and it is expected to continue growing in the coming years. This means that more vehicles will have to be monitored, infrastructures will be subject to higher pressure and will eventually grow. Monitoring such systems will require more sensors, more sensed data, added communication needs, and increased data processing. SRM systems must be able to cope with new sensors and more data without modifications, i.e., they should be constructed in order to be scalable.

Meeting these requirements is essential for increased reliability of railway systems. Railway companies must rely on systems that are not only scalable, but also able to deliver precise and real-time information regarding their trains. Having described the main SRM system requirements, it is now important to look at the current state-of-the-art in order to identify specificities and limitations of existing systems.

3. State-of-the-Art

In the transportation area and, especially, in railway systems, considerable work has been done, mostly with the objective of equipping railway vehicles with a variety of sensors, in order to not only improve passengers' safety and comfort (Passenger Information Systems - PIS), but also to interconnect train vehicles with external operators [4], by using new communication technologies such as LoRa and LTE. Squats, rail deformation, wheel fractures, among others, can cause derailments, with significant consequences. It is necessary to monitor both tracks and wheels to detect problems and assess the degree of criticality of any fault. This section presents the state-of-art on infrastructure maintenance, railway vehicles maintenance, and the interaction between both infrastructure and railway vehicles. The work described in this section will be analyzed according to different characteristics: (1) sensors used to detect problems, (2) approach taken after a problem is detected, and (3) communication between sensors and processing equipment. Sections 3.1, 3.2, and 3.3 describe the state-of-the-art for infrastructure maintenance, railway vehicles maintenance, and railway systems (i.e., vehicles and/or infrastructure) maintenance, respectively. Section 3.4 identifies tools and techniques with potential for being used in the context of the problem at hand. Whenever appropriate, techniques presented in sections 3.1, 3.2 and 3.3 will be evaluated according to the requirements presented in section 2.3.

3.1. Infrastructure Maintenance

Some of the work found in the literature specifically addresses infrastructure anomaly detection. As shown in Figure 1, railway tracks are composed of several elements. Detecting anomalies on some of the individual components (e.g., rail or ballast) or combinations of components (e.g., track bed) can provide insights on the conditions of the track and, depending on the found anomalies, trigger notifications for subsequent maintenance/repair actions. There are two types of data that can be used for this purpose: (1) video data, collected from high-speed cameras capable of working under adverse conditions (e.g., rainy or sunny days, dusty tracks, low light), and (2) plain data (i.e., text), collected from a variety of sensors (e.g., accelerometers, gyroscopes, ultrasonic sensors).

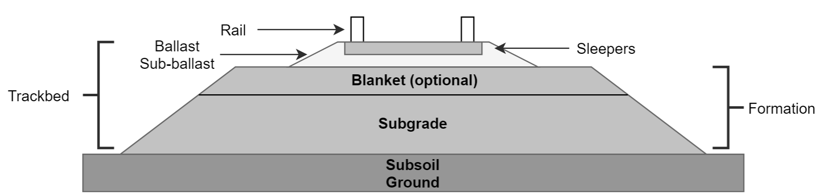

Figure 1: Characteristics of the track, from the rails to the subsoil. Image adapted from https://en.wikipedia.org/wiki/Track_bed

For instance, Rikhotso et al. [7] proposed the use of 3D image acquisition and modeling for assessing the condition of rails and detecting surface defects.

The authors used a structured light technique to analyze light patterns. Using triangulation between the camera, the projector/laser, and the object, the depth of the rail was calculated. They made use of MatLab to calibrate the camera and filter the image for the red light projected by the laser. The tests were performed in a laboratory.

Tastimur et al. [8] proposed a mathematical morphology-based method to detect deformities in rail track surfaces. The authors used two cameras (pointed to the left and right rails) to take images of the tracks. They used the Canny Edge algorithm that (1) converts the images to gray format, (2) smooths the image with a Gaussian filter, (3) extracts edges on the X and Y directions, (4) calculates the angle and gradient size, and (5) erases outliers. The authors used MatLab to process the data.

Liu et al. [9] proposed a method for automatically detecting faults in strands of the isoelectric line. Unlike the previous papers, they focused the detection on loose strands in the catenary. The authors divided the method into three stages, where (1) a convolutional neural network was used to extract the line features, (2) image segmentation was carried out based on Markov models, and (3) analysis of the results was performed by comparing the quantity of the independent connection regions and the pixel's standard deviation. They used a camera mounted on a catenary inspection vehicle (CIV) and the collected images were used for isoelectric line location, isoelectric line segmentation, and fault diagnosis. The used algorithm was an adaptation of VGG-16 image recognition algorithm [10] (Isoelectric Line Network - ILNET) for anomaly detection in catenary images.

Espinho et al. [11] proposed an algorithm for detecting tracks and turnouts, using edge detection, RANSAC algorithm, Histogram of Oriented Gradient, Template Matching and Support Vector Machines. The algorithm does not address anomalies detection, and uses a methodology that is different from the previously mentioned papers. Instead of relying on empirical thresholds and fine-tuned parameters, the authors developed a generic method to detect edges. First, they filter the images, then only parallel gradients are kept and, finally, the image is filtered with a mask and reduced.

Using a different technique, Vijaykumar et al. [12] developed a method to detect surface defects on railheads, using the Binary Image Based Rail Extraction algorithm. The authors used a 12Mp camera, from two different stations, and performed image enhancement to increase contrast.

Zeng et al. [13] developed a vehicle and associated sensors to detect height changes. This vehicle was able to detect failures and proceed with repair tasks. A Neural Network (NN) was used to process the collected data.
By also using an NN, Faghih et al. [14] were able to analyze images and detect surface defects. Using a Deep Convolutional NN, the raw image was used and there was no extra processing in the learning process. They used a mini-batch gradient descent method to optimize the network.

Feng et al. [15] proposed an automatic visual inspection system to detect missing fasteners, using a probabilistic model. The images were collected from two cameras (that were placed under the train) and sent to an onboard computer. The authors first detect the track and sleepers to, then, detect the fasteners. They used a variation of the Latent Dirichlet Allocation algorithm (LDA) [16], called Structure Topic Model (STM).

Parvathy et al. [17] developed a microcontroller-based system to replace manual fault detection. The sensing modules (temperature, accelerometer and ultrasonic sensors) are placed on the outer surfaces of the tracks, alternatingly and equidistantly. The modules can be powered by solar panels or piezoelectric energy methods.

Gan et al. [18] proposed a method to visually inspect railway surfaces, using background-oriented defect inspection. The inspection occurs in three phases: (1) pre-processing, by standardizing the values, regarding lighting and other variations, (2) background representation and defect determination, by modeling the background according to randomly selected samples, and (3) post-processing, by discarding false positive defects. The authors used a Dalsa Spyder camera, placed inside an inspection vehicle, with an onboard computer.

Ho et al. [19] evaluated track conditions by collecting the signature of the signals resulting from the interaction between the wheel and the track. The authors placed a Fibre Bragg Grating (FBG) sensor in the tracks and measured the same train over 2 months. The main idea was to study the track, every time a train passed by. The authors used one sensor and a database to detect outliers in the collected data.

Most of the presented papers have some issues/limitations, such as being limited to laboratory tests, lacking important information (e.g., used processing techniques), or requiring the use of very expensive sensors (especially the cameras), among others. Table 3 shows how the presented papers address some of the requirements listed in Table 2, namely requirements 5, track anomaly detection, and 7, predictive infrastructure maintenance. Table 3 is limited to requirements 5 and 7 only, as the remaining requirements are not addressed by the papers under consideration. Although the papers propose and present several techniques for detecting defects, they do not use a fully predictive maintenance strategy, as they do not consider historical data over specific periods in time, simply relying on current data. For this reason, the check signs in the requirement 7 column are followed by an asterisk.

3.2. Railway Vehicles Maintenance

In this section, papers related to vehicle inspection (and, specifically, to bogie inspection) are going to be presented. Bogies are the only part of trains that are in contact with the rails. Thus, evaluating their health is critical. Any crack or fault in bogies can cause derailments or long downtimes for maintenance. There are two types of bogies:

- Free Bogie - this type of bogie does not contain a motor and the wheels spin as a result of the train's movement. Coaches and freight wagons have this king of bogie.

- Motor Bogie - this type of bogie contains the motors that cause the train to move. Motor bogies are heavier than free bogies because they contain

Paper	Requirements	
	5	7
Rikhotso et al. [7]	✓	✓*
Tastimur et al. [8]	✓	✓*
Liu et al. [9]	✓	✓*
Espinho et al. [11]	✗	✗
Vijaykumar et al. [12]	✓	✓*
Zeng et al. [13]	✓	✓*
Faghih et al. [14]	✓	✓*
Feng et al. [15]	✓	✓*
Parvathy et al. [17]	✓	✓*
Gan et al. [18]	✓	✓*
Ho et al. [19]	✓	✗

Table 3: Infrastructure maintenance requirements addressed in the papers.

at least one motor. Higher weight provides better adherence to the tracks and prevents wheels from slipping.

Bogies are complex systems, made of several components, including suspensions. Normally, bogies include two different types of suspensions. Primary suspensions take care of the main vibrations and oscillations resulting from the interaction between the wheels and track. They have a spring format, as can be seen in Figure 2. Secondary suspensions have the objective of increasing passenger comfort, and are similar to airbags. Usually, secondary suspensions exist in passenger cars only. In this subsection, the approaches presented in the various papers do not target any specific type of suspension, and so we will consider that they are applicable to both types.

Lu et al. [21] proposed an automatic fault detection system, for multiple components in the vehicle, using time-scale normalization. The authors deployed eleven cameras near the rails (to cover almost 270 degrees (bottom and sides). The anomalies were detected using image subtraction (target localiza-

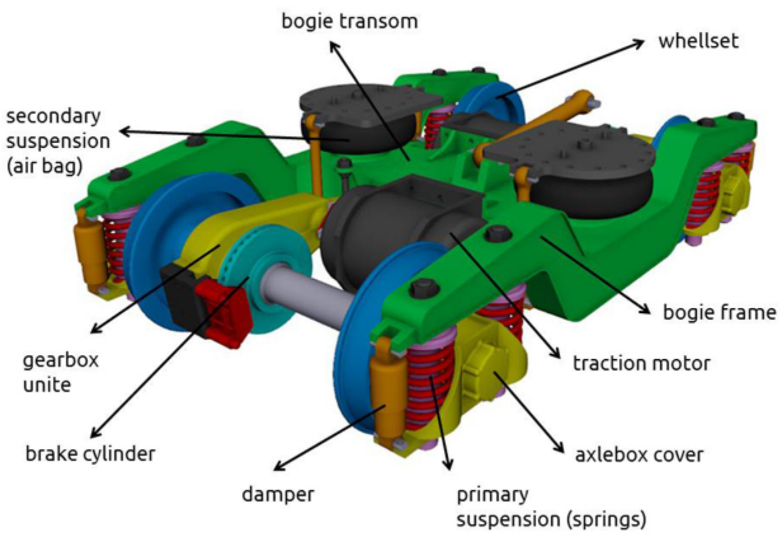

Figure 2: Bogie constitution [20].

tion).

Li et al. [22] investigated parameter estimation for railway vehicle suspensions, to provide support for condition-based maintenance. The main objective was to replace a calendar-based maintenance by a condition-based one (i.e., predictive). The authors ran simulations, with the objective of obtaining parameter values for different sensor configurations. Then, by using five sensors (four accelerometers and one gyroscope), simulations were matched against real data.

Ulianov et al. [23] developed a low-cost and portable system to predict how the track-wheel interaction would affect the running gear. The used freight vehicles (wagons) had a modified track-friendly suspension, and four sensors (accelerometers and gyroscopes) where attached to the vehicle in order to capture data from the vehicle dynamics. The authors used a Raspberry Pi to connect to the sensors and Wi-Fi hubs to start and stop the system and store the data into an SD card.

Ashwin et al. [24] developed a system, using MatLab, to detect and recognize patterns on captured images (e.g., primary suspension images). The authors analyzed cracks in the primary suspension because these are one of the major causes of derailment. The developed system targeted the prevention of over-excessive maintenance, thus reducing costs.

Liu et al. [25] developed a system, using Wheel Impact Load Detector (WILD), to detect wheel defects on high-speed trains. The main approach was to deploy twenty Fibre Bragg Grating (FBG) sensors along a 3-meter rail section, with 0.15 meters spacing between them, with the objective of collecting data whenever a train passed the section.

In a recent study, Bernal et al. [2] reviewed techniques for monitoring freight railway vehicles. Unlike the previously mentioned papers, the authors did not perform any experiment, but presented interesting conclusions regarding power consumption for several vehicle subsystems (i.e., carbody, coupler, wheelset). They concluded that, nowadays, it is a challenge to power an onboard system, for real-time monitoring. On the other hand, the authors also concluded that data processing algorithms are well established for detection of different vehicle components, and that current technology (i.e., microprocessors and sensors) is widely accessible. The work was based on the possibility of using real-time monitoring for predictive maintenance, along with IoT networks and solutions. In the final part of the paper, the authors present a table (Page 17, Table III) with a detailed review of different systems, including monitored anomalies, used sensors, sampling rate, data transmission method, power consumption, power source, and stage of development.

As it is possible to see in Table 4, the presented papers fulfill several but not all of the requirements identified in Table 2, especially 2, local data processing, 3, wheels anomaly detection (superficial), 4, wheels anomaly detection (internal), 8, predictive vehicle maintenance, and 9, scalable system. As with the previous section, the requirements that were not addressed by any of the papers were omitted. However, in several cases the papers lack specific information on several aspects, for instance on how data processing is performed. Moreover, although the authors claim they use predictive maintenance, by monitoring conditions of certain parts of the vehicles, only one paper explores the use of historical data. On other hand, the use of inexpensive sensors is explored in the work reported in some of the papers, in order to allow for low-cost solutions, which also opens the door to scalability and reliability. In addition, [2] also addresses the problem of energy expenditure for different types of sensors and components. It should be noted that, in Table IV, an asterisk indicates that the authors used predictive maintenance without considering historical data. The minus sign, in the last row, indicates that the authors address the topic by reviewing different solutions, but do not put them to practice.

Paper	**Requirements**				
	2	3	4	8	9
Lu et al. [21]	✓	✓	✗	✓*	✗
Li et al. [22]	✗	✓	✗	✓*	✗
Ulianov et al. [23]	✗	✓	✗	✓*	✓
Ashwin et al. [24]	✓	✓	✓	✓*	✗
Liu et al. [25]	✗	✓	✗	✓*	✗
Bernal et al. [2]	-	-	-	-	-

Table 4: Railway vehicle maintenance requirements addressed in the papers.

3.3. Railway Systems Maintenance

In the previous sections, the discussion was focused on analyzing the infrastructure (mainly by using sensors that were placed in dedicated maintenance vehicles) or vehicles. In this section, we will focus on solutions that target the infrastructure, the vehicles, or both. Typically, these solutions do not rely on special-purpose maintenance vehicles but, rather, on data acquisition equipment that is mounted on ordinary vehicles.

Yin et al. [26] proposed a method to detect squats using bogie acceleration (BA). The tests were performed using a SIMPACK simulation. The main objective was to understand and study how the bogie reacted to different squats' length and depth.

Lederman et al. [27] proposed a method for examining the tracks using an operational passenger train (light-rail vehicle). By using accelerometers placed on the train, they could gather data on the tracks' roughness. The sensors were deployed inside the cabin, central bogie, and roof. This low-cost approach also allowed to detect changes in the tracks.

Similarly to the previous paper, Bocciolone [28] proposed a technique for track condition inspection using accelerometers on operating bogies of Milan subway trains. With the data provided by the sensors, they were able to address three different rail structures: ballast, direct fastening, and rail track. The authors concluded that the structure over which the train travels has a strong influence on the vibration levels.

Chen et al. [29] proposed a technique for gathering information from multiple vehicles with multiple sensors. The idea was to combine data to have a more continuous and reliable track monitoring. This approach was validated using the data collected from the light-rail network in Pittsburgh. Their approach is computationally efficient and was able to detect track irregularities by comparing data being collected in real time with historical or baseline data.

Pau et al. [30] studied irregularities in the wheel-track contact. The authors used ultrasonic sensors to analyze the contact between wheels and track. This technique helped to detect grooves and imperfections (on the wheels), drills (on the track), and misaligned wheel-rail systems.

Goodmand et al. [31] developed a system to analyze track conditions using a wheelset of a boxcar. The system makes use of a MEMS device, with an accelerometer and wireless connectivity, to collect data from the wagon. The data collection lasted for 4 days and the test track contained 14 locations with several defects (e.g., lubricator, repair welds, bridges, turnouts, among others). The developed system detected ten out of fourteen possible features.

Table 5 presents a summary of the requirements addressed by the described papers, especially: 3, wheels anomaly detection (superficial), 4, wheels anomaly

Paper	Requirements						
	3	4	5	6	7	8	9
Yin et al. [26]	✗	✗	✓	✗	✗	✓	✗
Lederman et al. [27]	✗	✗	✓	✓	✗	✓	✓
Bocciolone et al. [28]	✗	✗	✓	✓	✗	✓	✓
Chen et al. [29]	✗	✗	✓	✓	✗	✓	✓
Pau et al. [30]	✓	✓*	✓	✗	✓*	✓*	✗
Goodmand et al. [31]	✗	✗	✓	✓	✗	✓	✓*

Table 5: Railway systems maintenance requirements addressed in the papers.

detection (internal), 5, tracks anomaly detection, 6, trackbed anomaly detection, 7, predictive infrastructure maintenance, 8, predictive vehicle maintenance, and 9, scalable system. The omission of the remaining requirements indicates that none of the papers address them. With the exception of paper [30] by Pau et al., all the others only used railway vehicles to detect problems inside the vehicle. The only paper that achieved anomaly detection on both track and wheels was paper [30], and it performed an analysis of the wheel-track contact. Most of the papers used different approaches, and considered the possibility of deploying monitoring equipment in some of the components of railway vehicles in order to monitor both tracks and wheels at the same time, so as to improve the efficiency of the approach and help reducing inherent costs.

3.4. Tools and Techniques

The previous sections provided information on existing solutions and approaches to infrastructure and railway maintenance. Nevertheless, one of the major problems of the overviewed papers is the lack of details on the algorithms and methods used for data processing. On the other hand, nowadays, the increase in sensing options, communication networks capability, and computational power, allows the use of powerful and efficient algorithms, tools, and techniques in various areas (e.g., agriculture, smart cities, among others), that can and should also be explored for the purpose of data processing and information extraction in Smart Railway Maintenance. As stated by Parkinson et al. [32], advanced algorithms can be used to reduce safety risks and also provide critical information on the day-to-day life of railway vehicles and infrastructure. This section provides a very brief overview of some of these tools/technologies, namely Neural Networks, Clustering and Pattern Recognition, and Adaptive and Multi-objective Algorithms.

3.4.1. Neural Networks

Neural Networks (NN) are inspired in the operation of the brain, namely on how neurons interconnect and interact in order to allow us to take decisions. Based on this type of operation, several machine learning algorithms were developed and, nowadays, they are used in a large variety of areas, such as social networks, video games, mobility, among others.

Sun et al. [33] developed a multi-task learning approach (MTL) to make accurate and context-aware delay estimations in bus transportation networks. By using MTL for transit short-term delay prediction, the authors reduced the possibility of having limited historical datasets. The authors consider static and real-time bus operations, weather conditions and scheduled events, to provide information regarding delays and service alerts. The data is divided into several JSON and GTFS files and processed using the developed multi-task neural network, with eight layers. The authors concluded that severe delays could be identified days ahead, providing critical information for commuters.

Ene et al. [34] demonstrated the possibility of using a feed-forward neural network (FFNN) for failure rate prediction. The authors used a Java application to simulate the FFNN network, and used real data stored in a text file. Using seven input neurons and seven hidden neurons, the authors achieved a 3% error in training, and a prediction error lower than 1.8%.

Smith et al, [35] developed and tested a prototype system for anticipating failures in airport ground transportation vehicle doors, that can support condition-based maintenance. The authors used a combination of statistical and neural networks (NN) approach. They compared three different NNs: backpropagation network (five hidden layers), cascade correlation network (sixteen hidden layers), and radial basis function network (fifteen hidden layers). Twenty-eight input variables were considered, such as closing energy, opening energy, closing time, opening time. The backpropagation network provided the best results (regarding training and testing root mean squared error), and after the normal parameters were known the algorithm was adjusted to be able to correctly predict vehicle door conditions. The authors were able to monitor doors operation, although online learning was not applied in this case (as the authors state, the normal behavior for a door depends on the system).

Zhi-Gang et al. [36] studied the use of backpropagation techniques (BPNN) to monitor the daily operation of a subway system, and a possible trend to failure, with the objective of applying them to predictive maintenance. This study was performed because subways normally adopt a preventive strategy and, most of the times, humans and resources are wasted. By using a BPNN to diagnose faulty pieces of equipment (e.g., brake shoes), they were able to provide useful information for maintenance, therefore reducing costs and saving resources.

3.4.2. Clustering and Pattern Recognition

Clustering and pattern recognition algorithms and techniques are very useful for data mining and data reduction, and are used in a large variety of fields, such as image analysis, market analysis, recommendation systems, among others. This subsection provides some use case examples, as a way to assist in assessing the potential of such tools for future SRM systems.

A. Kannan et al. [37] developed a gesture recognition system, for identifying static gestures, using a Microelectromechanical (MEMS) accelerometer (named ADXL335) and an ATMega 2560 microcontroller. The data was stored into several arrays: (1) feature set array, final set of features that define a gesture, (2) reference set, the average of the feature array, and (3) look-up table, that stores the reference set for all the gestures. In addition, gesture recognition used a lightweight approach to find the suitable reference in the look-up table, by using the Manhattan Distance algorithm. The higher the number of accelerometers, the higher the efficiency, with higher costs. The best trade-off for cost and efficiency was to use three accelerometers. This would allow having almost the maximum efficiency, with a smaller cost.

Mamun et al. [38] developed a k-means clustering method for detecting road anomalies or accidents in a certain area. The authors used an IoT-Fog server to process data in real-time, so end users could have the information in their smartphones. The authors clustered the collected data into two different categories: speed and accelerometer (z-axis). After defining thresholds, the authors evaluated the algorithm with respect to the speed over time, accelerometer over time, and a merge of the two. With this evaluation, it was possible to detect whether a certain variation was a pothole or a bump. On the other hand, if the speed over time was constant and suddenly the car slowed down, an accident had been detected. The authors developed a system that could fully work with a smartphone, without having to add any external sensor to the system. By using a k-means clustering algorithm, they were able to detect various problems that the driver faced (potholes, bumps or accidents).

3.4.3. Adaptive and Multi-objective Algorithms

In this section, two different algorithms are addressed. The first one is an adaptive algorithm that changes its behavior according to the retrieved data and the initial status. It is used in radar systems, to adjust the rate of false alarms. The second algorithm is a multi-objective optimization algorithm, applicable when it is necessary to achieve a trade-off between two or more objective functions, frequently used in the fields of engineering and logistics.

R. Kannan et al. [39] developed an algorithm that removes motion sensor bias. It can also successfully identify the duration of the motion. The authors modeled the noise in an accelerometer sensor in order to estimate linear motion coordinates. Using adaptive recalibration, the authors successfully detected and tracked all of the proposed scenarios, on multiple devices (Gear S3, Galaxy

A5, and Tizen Z3). They also performed rotational motion tracking, with a gyroscope and a magnetometer fused with an accelerometer, achieving good overall results.

A. Nunez et al. [40] proposed a system based on multi-objective optimization for condition-based maintenance in infrastructures. The authors used a Hilbert Spectrum approach to detect anomalies in the axle box acceleration measurement. The main focus of the authors was to optimise two distinct objective functions, in order to provide information on where the infrastructure company should focus in order to increase performance at controlled costs.

Summing up, there are several computational tools and techniques that have large potential for applicability in SRM systems. In section 3.4, examples of three such tools/techniques were presented, all of them exploring text-based data only. Image-based solutions exist as well, such as [15] and [41], but the cost to develop visual systems is high. Using accelerometers, microphones and/or gyroscopes, it is possible to provide text-based data that achieves similar results at a fraction of cost. This is the case of solutions [13], [14] and [27], that used a variety of techniques to find patterns or segments where the signals had abnormal behavior. The next steps are to fully explore the use of the mentioned tools and techniques in the Smart Railway Maintenance area.

4. Challenges

In light of the presented state-of-the-art, several open issues and research directions can be identified. These are addressed in the current section.

4.1. Open Issues

Despite the fact that railway maintenance is a universal need, the materialization of the Smart Railway Maintenance vision is still in its infancy, as many challenges must still be addressed before stable, widely-accepted, globally-applicable solutions for SRM are in place. These challenges mainly pertain to the requirements identified in Section 2.C, which fall into the following four categories: (1) data processing, (2) anomaly detection, (3) predictive maintenance, and (4) scalability, with the latter being orthogonal to the other three.

In what data processing is concerned, most of the papers that address it do not provide enough detail to reproduce the obtained results. In short, their focus is on the results and how they are useful for railway/infrastructure maintenance, and not on how the approach works and how it can be replicated in real, working systems. The tools and techniques presented in 3.4 can be of great help for data processing, as there is considerable knowledge on how they work, there are numerous, well-understood applications, and they provide the ability to extract useful information from large sets of heterogeneous data coming from a variety of sources.

Regarding anomaly detection, most of the work performed either light or deep inspections, mostly resorting to expensive sensors placed in the surroundings of the rails or in inspection vehicles. On one side, this approach can provide more detailed insights on the tracks and/or bogies, but, on the other side, inspection vehicles do not operate very often. In the future, the use of low-cost sensors (such as accelerometers) will help to improve monitorization and anomaly prediction. As stated in [3], several types of sensors are available for gathering data on a variety of parameters, with most of these sensors being low-cost. In light of this, it will be important to develop systems that will not only make use of such sensors, but also monitor infrastructure and vehicles (passenger or freight trains). Making use of regular services can help reducing costs, while constantly monitoring both trains (bogies) and infrastructure (rails or tracks). Moreover, this approach should be used for predictive maintenance also.

Last but not least, scalability is a crucial requirement and should always be a concern. With low-cost sensors, efficient data collection and processing algorithms, new IoT-related enabling technologies (e.g., cloud-based systems, custom PCB manufactures, among others) and large numbers of trains and infrastructure companies, building scalable systems is essential. For instance, deploying a new monitoring system or adding a new sensor to an existing system are two examples of frequently performed maintenance operations, with the purpose of providing more input data. The performance of the system should not be negatively affected by this and, in reality, it should become better, as more data should imply a decrease in misdetections and better insights on the behavior and status of overall railway system.

In addition to the referred requirement categories, other aspects equally pertaining to all of them should also be considered, as they can play an important role in future, reliable, globally-applicable SRM systems. These are: (1) standardization, (2) interoperability, (3) energy efficiency, and (4) security [4][42][43]. Standardization is essential for device compatibility and systems interoperability, thus fostering the development of widely used and scalable systems. On the other hand, emerging 5G solutions will not only provide resilient and efficient communications, but also secure channels to transmit and receive data [44]. According to the European Telecommunications Standards Institute (ETSI), the Railway Industry will have an active role in 5G systems, alongside with other application areas (Figure 3), due to the challenges that arise from communications, reliability, interoperability, and security in Passenger and Freight Information Systems (PIS and FIS), and Smart Railway Maintenance, especially when high-speed trains are concerned [45].

To summarize, emerging technologies and solutions such as 5G, Big Data analysis, and data processing approaches like neural networks, pattern recognition, and adaptive and multi-objective algorithms, have very high potential for Smart Railway Maintenance. Most of these solutions have not been explored in existing SRM systems and, thus, a whole range of opportunities open up for research and development.

One very important aspect that is lacking in all of the existing railway maintenance systems is flexibility. In general, the construction of railway maintenance systems uses a vertical approach, in which every piece of the system is specially built for the specific application at hand. This is a paradigm similar to the one existing in computer systems and networking systems several decades ago. It is apparent that we need a revolution in SRM systems, one that drives

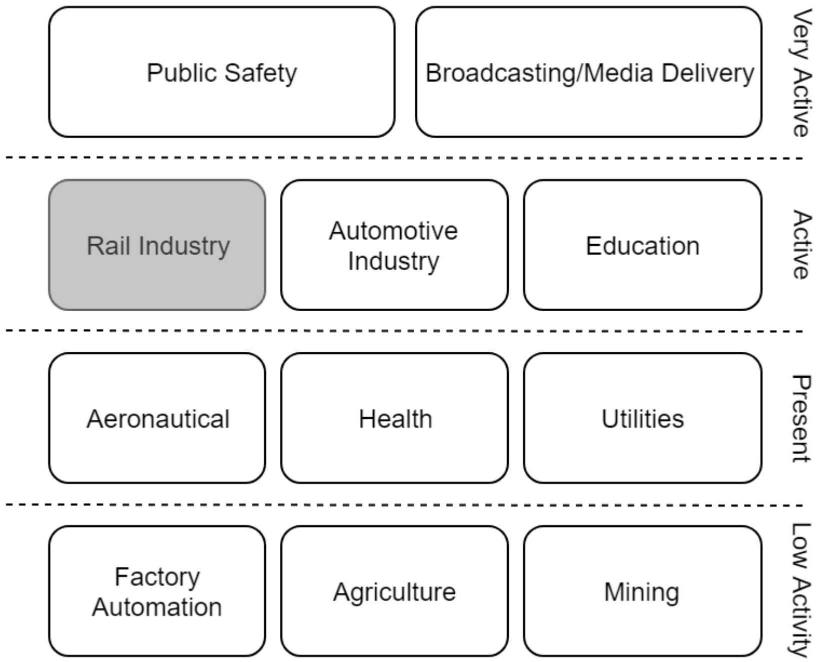

Figure 3: Areas where 5G communications are more important. Rail Industry is considered, by the ETSI, one of the most interested in this new paradigm. Image adapted from https://www.etsi.org/technologies-clusters/technologies/5g.

on the benefits that come from software-defined approaches, similarly to what is now being used in Software-Defined Networking (SDN). Such an approach to SRM system would bring several benefits, including:

- Virtualization – ability to deal with monitoring and maintenance resources independently of their physical details;
- Programmability – ability to change the monitoring behaviour of the system on the fly;
- Performance – ability to optimise the use of resources (physical resources can be shared by several monitoring applications);
- Service integration – modular approach allows code and service reuse, for ease of development;

- Openness – compatibility between different system modules, potentially from different vendors;
- Orchestration – ability to manage large numbers of devices, with full visibility over them;
- Dynamic scaling – ability to scale the system according to the application needs, through resource virtualization and cloud operation;
- Automation – ability to automate parts of the system monitoring application, leading to better performance and lower operation costs.

4.2. Research Directions

From the analysis carried out in the previous sections, several specific research directions can be identified. These include, but are not limited, to the following:

- Approaches to flexible, effective, efficient, and low-cost data collection for both railway vehicles and infrastructure monitoring, using regular trains;
- Data processing, reduction, and analysis in local controllers, and subsequent sending of that data to the cloud, for further processing;
- Online data processing systems, for real-time monitoring, using emerging communication technologies;
- Evaluation of 5G solutions for bulk, real-time, low-latency, reliable, and secure communications in SRM environments;
- Use of fog-based and/or cloud-based systems for data storing and data processing;
- Information extraction and inference from large data sets, for railway systems preventive maintenance;
- Integrated, interoperable, and scalable solutions for railway systems preventive maintenance.

With the above research directions/objectives in mind, we now propose a novel architecture that can be the basis for future SRM systems, which we name: Software-Defined Railway Monitorization (SDRm) architecture.

The proposed architecture is based on the paradigm of Software-Defined Networking (SDN) [46], and is depicted in Figure 4. According to it, SRM systems are organised into three planes, namely data plane, control plane, and application plane. The interface between the data plane and the control plane is called southbound interface, whereas the interface between the control plane and the application plane is called northbound interface. In the following paragraphs, a description of each of these architectural components is given.

The control plane deals with real, physical equipment, e.g., sensing equipment installed on railway vehicles and infrastructure. Moreover, this plane comprises local controllers, which are local processing units that interact with physical equipment for the purposes of configuration and data gathering. For instance, a local controller can request a given sensor to perform some specific readings over a given period of time at a given rate. Local controllers can also perform local data processing and/or temporarily hold some data until conditions for sending that data to a central controller are met. Typically, local controllers reside in the infrastructure or in vehicles.

One extremely important aspect of local controllers is that they (should) comply with a common, ideally standardised architecture, supporting a given and well-known set of operations for data gathering and equipment management. This is at the basis of features such as programmability, openness, virtualisation, and service integration, among other.

In the control plane, a central controller is in charge of performing high-level decisions with the objective of fulfilling application requirements. For instance, the central controller may decide on which data should be gathered from which railway vehicles or from which sections of the railway infrastructure, and instruct the appropriate local controllers to collect that data. A central controller may also want to know information on the features and sensing capabilities associated with a given local controller, so as to decide on subsequent sensing actions.

Communication between local controllers and the central controller is done via the southbound interface, using an appropriate protocol. The protocol must support actions such as feature collection, parameter configuration, read/write, and status information. It should be noted that, together with a standardised architecture for the local controllers, the southbound protocol will provide extremely important features such as the already mentioned features of programmability, openness, virtualisation, and service integration.

Last but not least, the application plane deals with the SRM application as a whole. In this plane, high level SRM decisions are made, such as guaranteeing that all trains are monitored periodically, critical infrastructure sections are subject to more specific attention, etc. At this level, resources are mostly high-level and/or virtualised, as there is no need to deal with specific physical details. Scaling and automation decisions may also be taken at this level.

The proposed SDRm architecture, presented in Figure 4, provides flexibility, code and service reuse capability, and ability to develop SRM platforms that can be used in a variety of contexts and by different railway companies, thus being able to cope with all of the important features that are missing from current railway maintenance systems.

5. Conclusions

Railway systems monitorization is essential for predictive maintenance. As seen in this paper, considerable work exists on railway maintenance, but the vast majority of it either addresses preventive maintenance or limited forms of predictive maintenance, often resorting to special-purpose, dedicated vehicles and/or specialized, high-cost devices, with little or no use of recent sensing, communication, and data processing technologies and techniques.

Emerging IoT and Industry 4.0 solutions, 4G and soon-to-be-available 5G communication solutions, cloud-based storage and processing, and, last but not least, software-based techniques, open up the possibility for what is now called Smart Railway Maintenance systems, for which massive data collection and

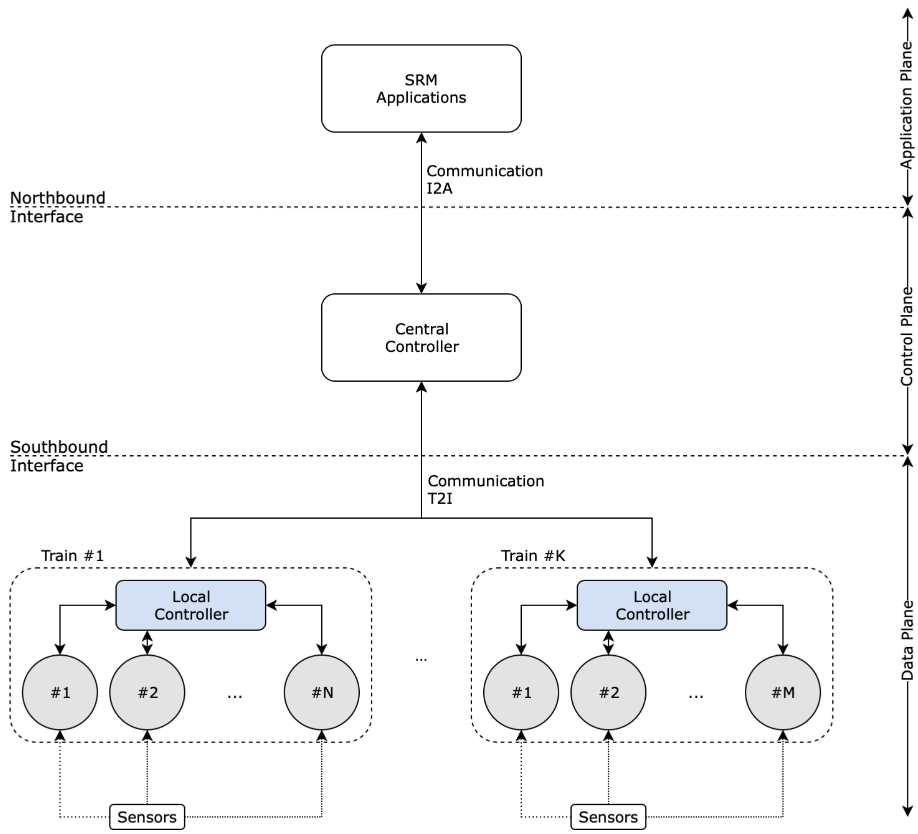

Figure 4: First design of the architecture, for a SDRm.

processing, high-bandwidth secure and reliable communications, intelligent data analysis and information extraction, and flexible development are crucial.

In this paper we have identified and described the key requirements and approaches to SRM, from the points of view of infrastructure maintenance, railway vehicles maintenance, and global system maintenance. The carried-out analysis has clearly shown that existing approaches have limitations, either by not addressing key requirements or by not exploring emerging technologies and/or computational tools and techniques. These have also been identified in this paper.

Given the above, we identified several open issues and research directions, spanning approaches to data collection, offline and online data processing, 5G communications, use of fog-based and/or cloud-based systems, information extraction and inference, and systems scalability and interoperability. Moreover, we proposed a novel architecture that explores the paradigm of software definition, with the aim of providing, flexibility, modularity, compatibility, reusability, scalability, and ease of development.

The identified research guidelines and proposal are the starting points for work that clearly needs to be done in order to develop efficient and effective Smart Railway Maintenance systems.

Acknowledgment

This work was carried out in the scope of the MobiWise project: From mobile sensing to mobility advising (P2020 SAICTPAC/0011/2015), co-financed by COMPETE 2020, Portugal 2020 - POCI, European Union's ERDF.

References

References

[1] A. Nordrum, K. Clark, Everything you need to know about 5g (Jan. 2017). URL https://spectrum.ieee.org/video/telecom/wireless/everything-you-need-to-know-about-5g

[2] E. Bernal, M. Spiryagin, C. Cole, Onboard Condition Monitoring Sensors, Systems and Techniques for Freight Railway Vehicles: A Review, IEEE Sensors Journal (2018) 1–1doi:10.1109/JSEN.2018.2875160.
URL https://ieeexplore.ieee.org/document/8488578/

[3] V. J. Hodge, S. O. Keefe, M. Weeks, A. Moulds, Wireless Sensor Networks for Condition Monitoring in the Railway Industry: A Survey, IEEE Transactions on Intelligent Transportation Systems 16 (3) (2015) 1088–1106. doi:10.1109/TITS.2014.2366512.

[4] P. Fraga-Lamas, T. M. Fernández-Caramés, L. Castedo, Towards the internet of smart trains: A review on industrial IoT-connected railways, Sensors (Switzerland) 17 (6) (2017) 1457. doi:10.3390/s17061457.
URL http://www.ncbi.nlm.nih.gov/pubmed/28635672http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5492363http://www.mdpi.com/1424-8220/17/6/1457

[5] R. B. Shetty, Predictive Maintenance in the IoT Era, Prognostics and Health Management of Electronics (2018) 589–612doi:10.1002/9781119515326.ch21.
URL http://doi.wiley.com/10.1002/9781119515326.ch21

[6] D. Li, J. Gao, Study and application of Reliability-centered Maintenance considering Radical Maintenance, Journal of Loss Prevention in the Process Industries 23 (5) (2010) 622–629. doi:10.1016/j.jlp.2010.06.008.
URL https://www.sciencedirect.com/science/article/abs/pii/S0950423010000756

[7] V. Rikhotso, N. Steyn, Y. Hamam, 3D rail modelling and measurement for rail profile condition assessment, in: 2017 IEEE AFRICON: Science, Technology and Innovation for Africa, AFRICON 2017, IEEE, 2017, pp. 1522–1527. doi:10.1109/AFRCON.2017.8095708.
URL http://ieeexplore.ieee.org/document/8095708/

[8] C. Tastimur, O. Yaman, M. Karakose, E. Akin, A real time interface for vision inspection of rail components and surface in railways, in: IDAP 2017 - International Artificial Intelligence and Data Processing Symposium, IEEE, 2017, pp. 1–6. doi:10.1109/IDAP.2017.8090267.
URL http://ieeexplore.ieee.org/document/8090267/

[9] Z. Liu, L. Wang, C. Li, Z. Han, A High-Precision Loose Strands Diagnosis Approach for Isoelectric Line in High-Speed Railway, IEEE Transactions on Industrial Informatics 14 (3) (2018) 1067–1077. doi:10.1109/TII.2017.

2774242.
URL http://ieeexplore.ieee.org/document/8114270/

[10] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, CoRR abs/1409.1556. arXiv:1409.1556.
URL http://arxiv.org/abs/1409.1556

[11] J. Corsino Espino, B. Stanciulescu, P. Forin, Rail and turnout detection using gradient information and template matching, in: IEEE ICIRT 2013 - Proceedings: IEEE International Conference on Intelligent Rail Transportation, IEEE, 2013, pp. 233–238. doi:10.1109/ICIRT.2013.6696299.
URL http://ieeexplore.ieee.org/document/6696299/

[12] V. R. Vijaykumar, S. Sangamithirai, Rail Defect Detection using Gabor filters with Texture Analysis, International Conference on Signal Processing, Communication and Networking 3 (2015) 1–6. doi:10.1109/ICSCN.2015.7219838.
URL http://ieeexplore.ieee.org/document/7219838/

[13] Zeng-Guang Hou, M. Gupta, A rail damage detection and measurement system using neural networks, in: 2004 IEEE International Conference onComputational Intelligence for Measurement Systems and Applications, 2004. CIMSA., IEEE, pp. 4–9. doi:10.1109/CIMSA.2004.1397218.
URL http://ieeexplore.ieee.org/document/1397218/

[14] S. Faghih-Roohi, S. Hajizadeh, A. Nunez, R. Babuska, B. De Schutter, Deep convolutional neural networks for detection of rail surface defects, in: 2016 International Joint Conference on Neural Networks (IJCNN), IEEE, 2016, pp. 2584–2589. doi:10.1109/IJCNN.2016.7727522.
URL http://ieeexplore.ieee.org/document/7727522/

[15] H. Feng, Z. Jiang, F. Xie, P. Yang, J. Shi, L. Chen, Automatic fastener classification and defect detection in vision-based railway inspection systems, IEEE Transactions on Instrumentation and Measurement 63 (4) (2014)

877–888. doi:10.1109/TIM.2013.2283741.
URL http://ieeexplore.ieee.org/document/6630109/

[16] D. M. Blei, A. Y. Ng, M. I. Jordan, Latent dirichlet allocation, J. Mach. Learn. Res. 3 (2003) 993–1022.
URL http://dl.acm.org/citation.cfm?id=944919.944937

[17] A. Parvathy, M. G. Mathew, S. Justus, A. Ajan, Automatic rail fault track detection for Indian railways, in: 2017 2nd International Conference on Communication and Electronics Systems (ICCES), IEEE, 2017, pp. 144–147. doi:10.1109/CESYS.2017.8321251.
URL http://ieeexplore.ieee.org/document/8321251/

[18] J. Gan, J. Wang, H. Yu, Q. Li, Z. Shi, Online Rail Surface Inspection Utilizing Spatial Consistency and Continuity (2018) 1–11doi:10.1109/TSMC.2018.2827937.
URL https://ieeexplore.ieee.org/document/8354938/

[19] T. K. Ho, S. Y. Liu, Y. T. Ho, K. H. Ho, K. K. Wong, K. Y. Lee, H. Y. Tarn, S. L. Ho, Signature analysis on wheel-rail interaction for rail defect detection, Railway Condition Monitoring, 2008 4th IET International Conference on (2008) 1–6doi:10.1049/ic:20080342.
URL http://digital-library.theiet.org/content/conferences/10.1049/ic{_}20080342

[20] RailSystem, Bogie, [Accessed on March 08, 2019] (2015).
URL http://www.railsystem.net/bogie/

[21] S. Lu, Z. Liu, Y. Shen, Automatic Fault Detection of Multiple Targets in Railway Maintenance Based on Time-Scale Normalization, IEEE Transactions on Instrumentation and Measurement 67 (4) (2018) 849–865. doi:10.1109/TIM.2018.2790498.
URL http://ieeexplore.ieee.org/document/8268578/

[22] P. Li, R. Goodall, P. Weston, C. Seng Ling, C. Goodman, C. Roberts, Estimation of railway vehicle suspension parameters for condition monitoring, Control Engineering Practice 15 (1) (2007) 43–55. doi:10.1016/j.conengprac.2006.02.021.
URL https://www.sciencedirect.com/science/article/pii/S0967066106000463

[23] C. Ulianov, F. J. Franklin, P. Hyde, R. Shaltout, Experimental investigation of wheel-track interaction characteristics using test track bogie dynamics measurements, in: Proceedings of 2016 IEEE International Wheelset Congress, IWC 2016, IEEE, 2017, pp. 117–121. doi:10.1109/IWC.2016.8068379.
URL http://ieeexplore.ieee.org/document/8068379/

[24] Ashwin T, S. Ashok, Automation of rolling stock examination, in: 2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies, IEEE, 2014, pp. 260–263. doi:10.1109/ICACCCT.2014.7019442.
URL http://ieeexplore.ieee.org/document/7019442/

[25] X.-z. Liu, Y.-q. Ni, Wheel tread defect detection for high-speed trains using wheel impact load detector, The 2017 World Congress on Advances in Structural Engineering and Mechanics (ASEM17)doi:10.12989/sss.2018.21.5.687.
URL http://asem17.org/index.php?aside=aside{_}01{&}article=article01{_}01

[26] X. Yin, X. Wei, L. Jia, Detection of railway track squats by using bogie acceleration measurement, in: Chinese Control Conference, CCC, Vol. 2015-Septe, IEEE, 2015, pp. 6297–6302. doi:10.1109/ChiCC.2015.7260628.
URL http://ieeexplore.ieee.org/document/7260628/

[27] G. Lederman, S. Chen, J. Garrett, J. Kovačević, H. Y. Noh, J. Bielak, Track-monitoring from the dynamic response of an operational

train, Mechanical Systems and Signal Processing 87 (2017) 1–16. doi:10.1016/j.ymssp.2016.06.041.
URL https://www.sciencedirect.com/science/article/pii/S0888327016302230

[28] M. Bocciolone, A. Caprioli, A. Cigada, A. Collina, A measurement system for quick rail inspection and effective track maintenance strategy, Mechanical Systems and Signal Processing 21 (3) (2007) 1242–1254. doi:10.1016/j.ymssp.2006.02.007.
URL https://www.sciencedirect.com/science/article/pii/S0888327006000434

[29] G. Lederman, S. Chen, J. H. Garrett, J. Kovačević, H. Y. Noh, J. Bielak, A data fusion approach for track monitoring from multiple in-service trains, Mechanical Systems and Signal Processing 95 (2017) 363–379. doi:10.1016/j.ymssp.2017.03.023.
URL https://www.sciencedirect.com/science/article/pii/S0888327017301516

[30] M. Pau, B. Leban, Experimental detection of wheel-rail contact irregularities, Proceedings of the 7th World Congress on Railway Research (WCRR) (2006) 4–8doi:http://www.railway-research.org/IMG/pdf/133.pdf.
URL http://www.railway-research.org/IMG/pdf/133.pdf

[31] D. L. Goodman, J. Hofmeister, R. Wagoner, Advanced diagnostics and anomaly detection for railroad safety applications: Using a wireless, IoT-enabled measurement system, in: AUTOTESTCON (Proceedings), Vol. 2015-Decem, IEEE, 2015, pp. 273–279. doi:10.1109/AUTEST.2015.7356502.
URL http://ieeexplore.ieee.org/document/7356502/

[32] H. Parkinson, G. Bamford, B. Kandola, The Development of an Enhanced Bowtie Railway Safety Assessment Tool using a Big Data Analytics Approach, International Conference on Railway Engineering (ICRE 2016)

(2016) 1 (9 .)–1 (9 .)doi:10.1049/cp.2016.0510.
URL http://digital-library.theiet.org/content/conferences/10.1049/cp.2016.0510

[33] F. Sun, A. Dubey, C. Samal, H. Baroud, C. Kulkarni, Short-Term Transit Decision Support System Using Multi-task Deep Neural Networks, in: 2018 IEEE International Conference on Smart Computing (SMARTCOMP), IEEE, 2018, pp. 155–162. doi:10.1109/SMARTCOMP.2018.00086.
URL https://ieeexplore.ieee.org/document/8421344/

[34] A. Ene, C. Stirbu, A Java application for the failure rate prediction using feed forward neural networks, in: Electronics, Computers and Artificial Intelligence (ECAI), IEEE, 2016, pp. 1–4. doi:10.1109/ECAI.2016.7861197.
URL http://ieeexplore.ieee.org/document/7861197/

[35] A. E. Smith, D. W. Coit, Y. C. Liang, Neural network models to anticipate failures of airport ground transportation vehicle doors, IEEE Transactions on Automation Science and Engineering 7 (1) (2010) 183–188. doi:10.1109/TASE.2009.2020508.
URL http://ieeexplore.ieee.org/document/5159354/

[36] M. Zhi-Gang, L. Han-Bin, Study on the Maintenance of Subway Equipment Based on Data Mining Techniques, in: Proceedings - 2nd International Conference on Smart City and Systems Engineering, ICSCSE 2017, IEEE, 2017, pp. 209–215. doi:10.1109/ICSCSE.2017.59.
URL http://ieeexplore.ieee.org/document/8120546/

[37] A. Kannan, A. Ramesh, L. Srinivasan, V. Vijayaraghavan, Low-cost static gesture recognition system using MEMS accelerometers, in: GIoTS 2017 - Global Internet of Things Summit, Proceedings, IEEE, 2017, pp. 1–6. doi:10.1109/GIOTS.2017.8016217.
URL http://ieeexplore.ieee.org/document/8016217/

[38] M. A. A. Mamun, J. A. Puspo, A. K. Das, An intelligent smartphone based approach using IoT for ensuring safe driving, in: ICECOS 2017 - Proceeding of 2017 International Conference on Electrical Engineering and Computer Science: Sustaining the Cultural Heritage Toward the Smart Environment for Better Future, IEEE, 2017, pp. 217–223. doi:10.1109/ICECOS.2017.8167137.
URL http://ieeexplore.ieee.org/document/8167137/

[39] R. Kannan, S. Jain, Adaptive Recalibration Algorithm for Removing Sensor Errors and Its Applications in Motion Tracking, IEEE Sensors Journal 18 (7) (2018) 2916–2924. doi:10.1109/JSEN.2018.2804941.
URL http://ieeexplore.ieee.org/document/8289379/

[40] A. Nunez, A. Jamshidi, H. Wang, J. Hendriks, I. Ramirez, J. Moraal, R. Dollevoet, Z. Li, A Condition-Based Maintenance Methodology for Rails in Regional Railway Networks Using Evolutionary Multiobjective Optimization, in: 2018 IEEE Congress on Evolutionary Computation (CEC), IEEE, 2018, pp. 1–7. doi:10.1109/CEC.2018.8477842.
URL https://ieeexplore.ieee.org/document/8477842/

[41] P. D. Ruvo, G. D. Ruvo, A. Distante, M. Nitti, E. Stella, F. Marino, A Visual Inspection System for Rail Detection and Tracking in Real Time Railway Maintenance, The Open Cybernetics & Systemics Journal 2 (1) (2008) 57–67. doi:10.2174/1874110X00802010057.
URL http://benthamopen.com/ABSTRACT/TOCSJ-2-57

[42] L. Zhu, F. R. Yu, Y. Wang, B. Ning, T. Tang, Big Data Analytics in Intelligent Transportation Systems: A Survey (2018). doi:10.1109/TITS.2018.2815678.
URL http://ieeexplore.ieee.org/document/8344848/

[43] Q. Li, Z. Zhong, Z. Liang, Y. Liang, Rail inspection meets big data: Methods and trends, in: Proceedings - 2015 18th International Conference on Network-Based Information Systems, NBiS 2015, IEEE, 2015, pp. 302–308.

doi:10.1109/NBiS.2015.47.
URL http://ieeexplore.ieee.org/document/7350636/

[44] M. Liyanage, I. Ahmad, A. B. Abro, A. Gurtov, M. Ylianttila, Regulatory Impact on 5G Security and Privacy.

[45] M. P. Papaelias, C. Roberts, C. L. Davis, A review on non-destructive evaluation of rails: State-of-the-art and future development (Jul. 2008). doi:10.1243/09544097JRRT209.
URL http://journals.sagepub.com/doi/10.1243/09544097JRRT209

[46] D. Kreutz, F. M. Ramos, P. E. Verissimo, C. E. Rothenberg, S. Azodolmolky, S. Uhlig, Software-defined networking: A comprehensive survey, Proceedings of the IEEE 103 (1) (2015) 14–76. doi:10.1109/JPROC.2014.2371999.
URL http://ieeexplore.ieee.org/document/6994333/