

Highly resilient FRMCS/5G design for future rail operation

White paper

Contents

Executive summary	3
Future rail operation	4
Use cases, scenarios and requirements	6
Use cases and requirements	6
Rail scenarios	7
Resilient FRMCS design	9
Reference architecture	9
Principle solution design aspects	10
Assessment of deployment options	15
E2E/overall security	17
E2E management and orchestration	20
Conclusion	21
Abbreviations	22

Executive summary

The rail sector in Europe is on the verge of the largest technology leap in its history, with the German sector initiative [Digitale Schiene Deutschland \(DSD\)](#) striving toward a high degree of automation in rail operation. As a prerequisite to this, the European railways are currently working on the introduction of the 5G-based Future Railway Mobile Communication System (FRMCS), which will provide the connectivity backbone for future rail operation.

Due to the criticality of this infrastructure, [Nokia](#) and DSD have invested in a collaborative project to make the FRMCS system more resilient. In this work, various architectural options and features available in 5G have been analyzed for their suitability in the resilience needs of future rail operation use cases. Redundancy concepts for the fiber-optic trackside transport network, as well as active 5G components and the compute platform have been analyzed. The applicability of various options to allocate FRMCS/5G functions to infrastructure elements has been explored together with their impact on costs. Additionally, orchestration and security have been considered to achieve a resilient FRMCS network design.

Future rail operation

Less traffic, less congestion, less particulate matter—and more people and more goods on the rails. The rail sector in Europe is on the verge of a technological leap into the digital future. The sector initiative Digital Schiene Deutschland (DSD) is taking advantage of this opportunity and bringing future technologies into the rail system. This benefits not only passengers, but also the climate and Germany as a business location—and all this without having to construct a single new track.

The foundation for this is being laid with the fundamental modernization and digitization of the infrastructure through the consistent introduction of digital control and safety technology. In addition, DSD is working on a far-reaching digitization of the railway system. For this, a system architecture will detail the tasks of individual components of the railway system, and how they should work together.

On this basis, numerous digital technologies will then be tested and further developed for use in the system; for example, an AI-based traffic and incident management system will provide intelligent and automated control of trains in the future. These will then run fully automatically and at an optimal distance from each other. The latest sensor technology for environment perception coupled with high-precision train location and an automated interruption detection are further technologies that will play an important role in the digitalization of the railway system. Overall, a significant improvement in capacity, punctuality and efficiency of the railway system will be achieved, all of which are requirements for more traffic on the railway and a strengthening of the railway as the climate-friendly mode of transport of the future.

The rail system of the future will be characterized by data-intensive and partially latency-critical applications, which is one of the reasons why the European railways are currently striving to soon introduce the 5G-based Future Railway Mobile Communication System (FRMCS). In this context, Deutsche Bahn AG and Nokia have just concluded an R&D collaboration on how to design a highly resilient FRMCS/5G system for future rail operation.

Standardization and introduction of mission-critical applications started in 3GPP with LTE. Now 5G goes much further in meeting the demands of the communications service provider market, and fully supports the mission-critical needs of industries and enterprises, including rail operators. As well as providing efficient broadband capabilities, 5G networks will offer measures to build ultra-high reliability networks with ultra-low latency. 5G can also provide the massive Machine Type Communication for sensors and predictive maintenance that train operators will increasingly require in the future to improve and optimize their services. With huge performance improvements over previous generations of mobile technology, 5G delivers high speeds of up to 10 Gbps and very low latency—the time for the network to respond to requests from the mobile device. Furthermore, 5G also achieves such performance at much lower cost than other technologies.

Even more important, 5G will bring additional communications flexibility for railway operators to use, besides their dedicated network for mission-critical applications—a commercial 5G network slice for additional non-critical services.

This will bring rail operators new opportunities and applications based on mobile broadband capabilities.

Applications can be categorized into three segments¹:

- **Critical:** applications essential for train movements and safety or a legal obligation, such as emergency communications, shunting, presence, trackside maintenance, Automatic Train Operation (ATO), Automatic Train Protection (ATP)
- **Performance:** applications that help to improve the performance of the railway operation, such as train departure procedures and telemetry
- **Business:** applications that support the railway business operation in general, such as passenger information or ticketing support.

Several reasons form the basis for choosing 5G technology for the future of rail communications. FRMCS based on 5G is expected to be introduced in Europe around 2025, by which time it is likely that the 3GPP will have ceased standardization work on 4G LTE technology. Additionally, 5G can be expected to serve railways for much longer, extending over the next decades.

Many of the future uses for railway communications will demand minimal latency between the radio and the network, as well as the ability to work in private cloud environments. Examples include ATO and broadband machine-to-machine communication. Today LTE already provides a high quality of service (QoS) and broadband capabilities. 5G is even more specifically designed for these kinds of ultra-reliable, low-latency mission-critical communications, which also include video applications. This is supported by 5G flexible deployment schemes using cloud and mobile edge computing.

There is also a huge effort to use 5G technology for the forthcoming Industry 4.0 transformation, and it is expected to become the dominant technology for vertical markets. Using 5G as a common technology will also make it easier for intermodal communication concepts, for cars and other vehicles to interwork with FRMCS to communicate with trains.

In this whitepaper, we describe how an FRMCS/5G system can be designed and deployed in order to fulfill rail-specific requirements for reliability, availability and IT security. We take the following into consideration:

- Functional split options, i.e., considerations on where which function of the 5G system should be physically placed (including impact on redundancy)
- Redundancy concepts for the track side transport network
- Edge and cloud system
- Orchestration
- Security.

The result is an optimized FRMCS/5G deployment and configuration, which leverages standardized technologies to provide the required QoS, reliability and security for various applications envisioned for future rail operations.

¹ https://uic.org/IMG/pdf/frmcs_user-requirements.pdf

Use cases, scenarios and requirements

In this section, we list some envisioned exemplary railway-specific use cases and their requirements. Additionally, we differentiate between various rail scenarios in terms of expected train density as well as radio coverage.

Use cases and requirements

Voice: Train drivers, maintenance staff at the tracks and dispatchers all communicate with voice. Railway voice services comprise features such as group call services and functional and context-dependent addressing of communication endpoints.

Voice services are already implemented in the current GSM-R network and are required to be used in the FRMCS system too. However, it is highly likely that voice services could be replaced in the long term to a large extent by data-based communication between automated units. Moreover, as we move on to Grade of Automation 4 (GoA4), the relevance of voice would be further decreased since the dispatching system would be automated. However, voice is still expected to play a part in emergency communication between the train and trackside.

European Train Control System (ETCS): Today's rail operation is based on block-centric train protection systems, where axle counters are used to ensure that at any point in time only one train can be in a certain segment of track. In the future, a so-called ETCS Level 3 Moving Block approach is envisioned, which operates in a train-centric manner and allows trains to travel a minimum distance through the usage of train-based integrity information and advanced train-based localization. From a connectivity perspective, ETCS involves the signaling of location and train integrity information from train to track, and the signaling of Movement Authorities (MAs) from track to train. For advanced train-based localization, digital map and GNSS correction data are required to be signaled from track to train.

A key requirement for ETCS is a highly reliable network since it relies on frequent updates of train information such as their position, speed and train-integrity; the network must also ensure that the trains receive the appropriate MA on time. Safety is guaranteed at the application level, and if the reliability is compromised due to disruptions in the network, the application would move toward taking corrective measures such as the reduction of speed and drive on sight. ETCS forms the basis of train movement and is complemented by other applications such as the ATO.

Automatic Train Operation (ATO): Future rail operations target the gradual change from a semi-automated train operation in GoA2 to a fully automated operation in GoA4. The application utilizes "journey profiles" that include information such as timetable, speed, acceleration and "segment profiles," which include infrastructure-relevant information such as the geographical description of the relevant rail segments and environmental elements. The ATO application works in tandem with the ETCS application. While the MA provided by ETCS is responsible for safely moving a train through the rail segments, journey and segment profiles address optimal acceleration and braking behavior. Except for a reliable network that supports frequent message exchange, ATO does not impose very stringent requirements on the network.

Incidence management: In the case of GoA4, when an incident occurs, emergency measures have to be taken by the remote staff. Such measures could include remotely driving the train to a safe place such as a nearby station and/or evacuating the train. In order to efficiently perform these actions, the remote team will need access to video/radar/lidar data from the train. Key requirements for this use case are the availability of higher uplink bandwidth (approx. 6-7 Mbps), low latency (≤ 10 ms when the speed of the train is ≤ 40 km/h) and high reliability, especially for remote driving.

Key requirements from the exemplary use case^{2,3}

The most relevant requirements for the solution design gleaned from the rail operation use cases are related to data rates, latency and reliability.

Data rates

The most interesting finding is that the envisaged use cases mainly require train-to-ground transmission, i.e. mainly uplink (UL), which is challenging for any cellular system. On the other hand, the needed data rates during normal operations are very small compared to the requirements of a typical mobile carrier network.

Latency

The latency requirements relevant for typical regular train operation is in the range of 100 ms, which is seen as possible to realize by 5G networks without specific ultra-low-latency designs. However, for GoA4 operation, and specifically in the case of remote driving, the requirement could go down to 10 ms or less, when the speed of the train is below 40 km/h. In this specific case, special considerations for low-latency designs need to be taken.

Reliability

Compared to typical mobile carrier network needs, ultra-reliability is clearly a key railway requirement, especially to avoid disruptions to the offered rail service.

Rail scenarios

The rail scenarios can be grouped into three main families: long-distance tracks, inner city and stations/shunting/depots. Figure 1 illustrates the different rail scenarios and Table 1 details the differences between the long-distance and inner city scenarios, mainly in terms of the density of trains and the expected radio coverage.

Along the long-distance tracks, the expected train density is lower and can therefore be served with a fewer number of sites. On the other hand, in the inner city as well as the station scenario, higher accessibility and throughput need to be ensured in order to handle the higher density of trains as well as support for high-bandwidth demands of use cases such as downloading of pre-journey maps and information and data off-load.

In this whitepaper, we focus on the long-distance tracks, but the lessons learnt can be applied to the inner city as well as the station scenario.

² TR 22.889: <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3162>

³ TS 22.289: <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3186>

Figure 1. Railway scenarios

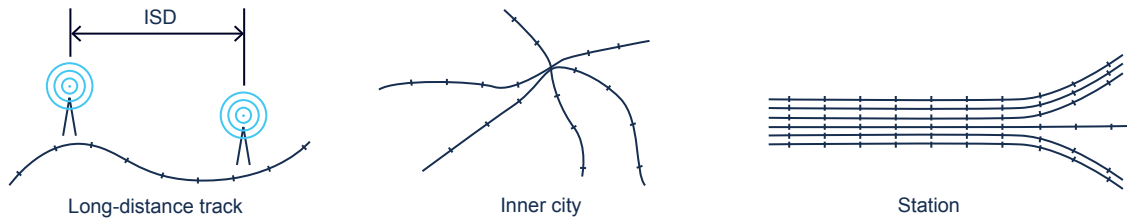


Table 1. Detailed comparison of various railway scenarios

Scenario		Long distance track	Inner city
Topology	Number of parallel tracks	2	2
	Max. train speed	350 km/h	80 km/h
	Average train density (all tracks)	0.5 train/km	1 train/km
Radio coverage	Average inter-site distance (ISD) (900 MHz)	8 km	4 km
	Average ISD (1900 MHz)	4 km	2 km

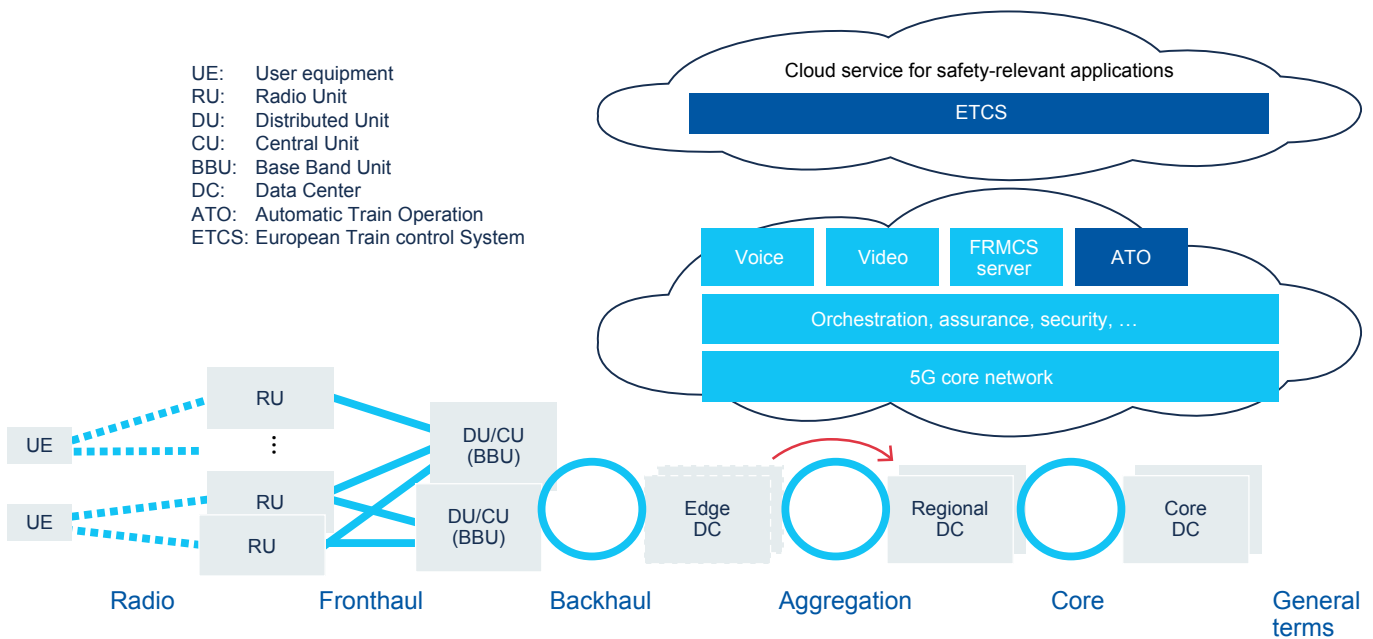
Resilient FRMCS design

In this chapter, we describe a generic system architecture that is based on the requirements described and then propose solutions to enhance network reliability.

Reference architecture

Figure 2 illustrates a reference model for the applicable FRMCS solution with its principal elements and structure. To cope with the reliability requirements, at least dual redundancy is considered on all levels.

Figure 2. Reference generic architecture



Starting from Radio Access, the reference model (especially for scenario one) is based on having the Radio Unit (RU) laid along the tracks and separated from the Base Band Unit (BBU), i.e., the combination of the Distributed Unit (DU) and the Central Unit (CU), which will “host” many RUs and will be deployed in fewer sites. So far the technology limitations foresee a maximum distance between the RU and BBU of 15 kilometers. The further aggregation of the CU function present in the BBU into regional centers can be evaluated as an alternative.

Edge data centers that are placed close to the end user are relevant in the case of applications with ultra-low-latency requirements. In the case of normal train operations, a three-layer separation (edge-regional-core) is not required because the QoS parameters for running a railway’s use cases can be easily fulfilled with only regional and core data centers.

Some of the 5G core network functions (i.e. UPF) will be distributed in regional data centers, where the most time-sensitive applications will run.

It is foreseen to have two core data centers where all the functions of a 5G core and the other FRMCS components (like MCX application servers or operations support system (OSS)) will reside in a geo-redundant architecture suitable for disaster recovery.

This architecture represents the best trade-off between fulfilment of a railway’s operator technical requirements and CAPEX/OPEX budget.

Principle solution design aspects

In November 2020, the Electronic Communications Committee (ECC) approved a draft for the official decision to allocate 2x 5.6 MHz in the 900 MHz band (FDD) and 10 MHz in the 1900 MHz (TDD) band to European railways under harmonized conditions for the availability and efficient use of radio spectrum for the Railway Mobile Radio (RMR).⁴ The European Commission published the implementing decision in September 2021.

A likely spectrum candidate for the first introduction of FRMCS will be the 1900 MHz band, which is therefore considered as a base layer for the solution design. The 900 MHz spectrum, currently already used for GSM-R services, would be fully available for FRMCS after the complete migration to an FRMCS based network. The option of using the 900 MHz partially for both GSM-R and FRMCS during migration is still under discussion and depends on improved co-existence solutions.

Obviously, the coverage in the 1900 MHz band will be smaller, and respectively will need a higher density of base stations. As such, the 900 MHz band can be considered as an overlay to increase reliability, but with less capacity and in hotspots where additional capacity is required.

The goal is to have “near zero service interruption.”

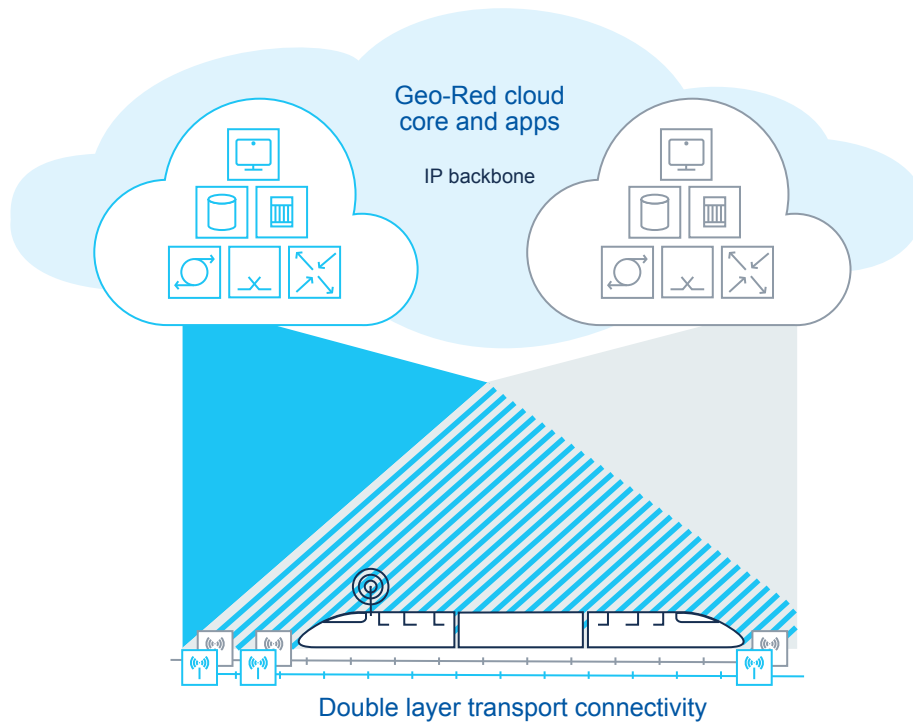
The proposed architecture is designed to minimize the impact of one failure (and in some cases multiple failures), tending to guarantee no service interruption.

In particular the basic principles are:

- Geo-redundant core network and OSS functions, working in active-active configuration
- Dual radio coverage via two layers (by radio interleaved coverage, or a split of the 1900 band in two bands, or using the released GSM-R band as overlay) (see Figure 4) with dedicated equipment (radio access nodes, transport nodes) per layer. Each layer is linked to both core network sites. The choice of the dual radio coverage solution would have an impact on the deployment cost.
- Redundant IP backbone designed with multiple paths among the key sites.

⁴ [ECC Decision 20\(02\) https://docdb.cept.org/download/1446](https://docdb.cept.org/download/1446)

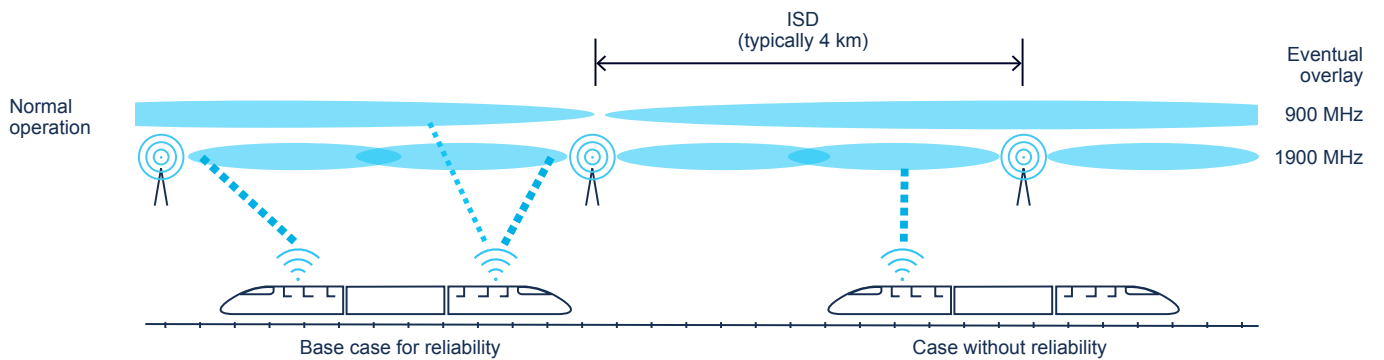
Figure 3. Overall redundant architecture



Multi-connectivity for high reliability

This section studies the network solutions to achieve highly reliable connectivity. It focuses on the long-distance tracks as it is the most demanding topology and specific to the rail environment.

Figure 4. Multi-connectivity for ultra-reliability



The proposed architecture guarantees a strong resiliency against various types of failure:

- Failure of one radio head

In case of failure of one of the radio heads that are covering the track, the design of the network, which foresees a redundant radio coverage, allows the user equipment (UE) to immediately attach to the next RU that guarantees the overlay coverage of the faulty unit. This is based on the assumption that the UE has dual connectivity to two RUs.

This can be achieved in two ways:

- i. The coverage derived by the overlapping neighbor node (having a dense remote radio head (RRH) layout implementing interleaved radio coverage)
- ii. The coverage derived by an “umbrella” 900 MHz radio coverage or two 1900 MHz bands.

Additionally, having at least two UEs (one in the head and one in the tail) in the train and with proper radio planning, the two UEs could be forced to stay as much as possible under two different RUs, so that the failure of one RRH can be seamless from the application point of view.⁵

- Failure of one train UE

As said before, this failure is handled by the availability of multiple UEs inside the train. A proper mechanism that manages the redundancy must be implemented on the application level in order to “mask” the failure to the upper levels and guarantee a seamless flow of data between the train and the ground (and vice versa).

Below, we differentiate between various levels of reliability in terms of the number of UEs and RUs with different frequency spectrum present:

- Single UE with single spectrum: In this scenario, the train is equipped with just one UE and only one frequency spectrum is available (see right-hand side of Figure 4). In this case, the failure of the UE or an RU that the UE is connected to results in the disruption of communication from the onboard to the track side.
- Single UE with double spectrum: This scenario is similar to the previous scenario with the difference being that both the spectrums are available. In case the RU for the 1900 MHz fails, the UE in the train is able to attach itself to the RU with the 900 MHz.
- Double UE with single spectrum: In this scenario, we can assume that one of the UEs is active and the other is passive and in case the active UE fails, the passive UE can be used. Alternatively, both UEs could be active at the same time and connected to different RUs in order to provide reliability in case of both a UE and a base station failure.
- Double UE with double spectrum: This is the best case scenario for reliability. As illustrated on the left-hand side of Figure 4, the train is equipped with at least two UEs, preferably one in the front and the other at the rear. These two UEs could be connected to different RUs of the same spectrum as well as to a RU with a different spectrum. In this case, the communication between the onboard and track side can survive the failure of an RU as well as a UE. In order for this solution to be used, the length of the train and the ISD play a part too.

⁵ This solution can be complemented by the support of onboard devices such as the MCX gateway and is still under discussion in the standardization bodies.

Fiber connectivity for high reliability

The proposed transport architecture is quite similar to the one implemented for GSM-R.

This architecture is based on two levels:

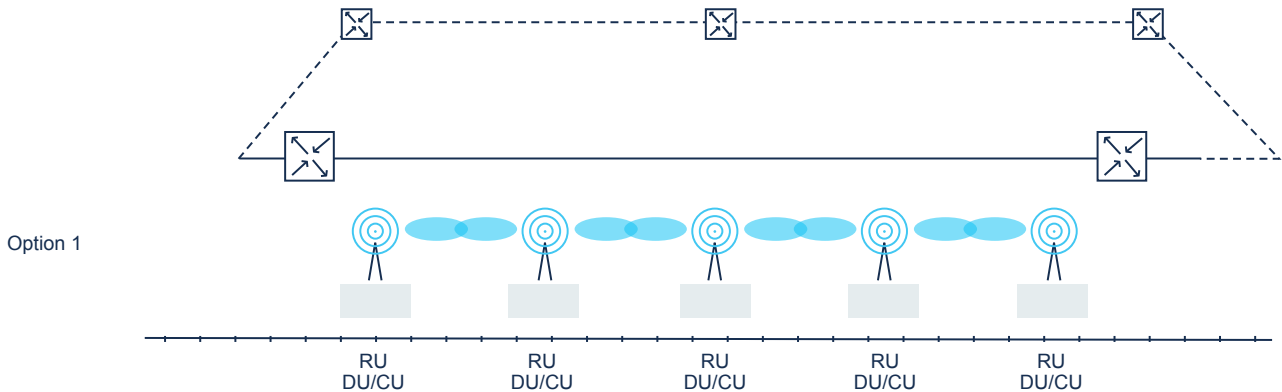
- National backbone, covering long-distance hops and connected to the core network sites
- Local/regional rings, dedicated to collect traffic coming from tracks and stations.

Regarding the second point, a couple of options (depending on the chosen radio access architecture) can be deployed:

Option 1: Ring structure with all-in-one base stations

As illustrated in Figure 5, co-location of DU/CU (Base Band) and RU (RRH: this is the classical model used already in GSM-R). In this case, the transport network will connect the full nodes to the backbone network in local rings. In case of site failure (or fiber cut), the connectivity is guaranteed by the other “leg” of the ring. This kind of solution has the drawback of higher CAPEX as every site is foreseen as a complete radio access node (BBU+RRH).

Figure 5. Ring structure with all-in-one base stations



Option 2: Distributed RU and “hoteling” of DU/CU

Distributing RUs and “hoteling” of DU/CU, as illustrated in Figure 6, is designed for long-distance tracks. This design helps to minimize the deployment of active network components. RUs are connected (in star or chain or loop) to the “centralized” DU/CU. According to the current latency requirements, a DU/CU hotel can support RUs located up to 15 km away. The xWDM technology depicted could be used in the case of a lack of sufficient fibers along the track and/or if IP connectivity is required for some of the base-stations, especially if they are all-in-one base-stations. The benefit of this approach is that the DU/CU hotels can be placed approximately 24-30 km apart. However, in case of a cut in the fiber, the connection between the RU(s) and the DU/CU hotel would be severed. IP connectivity can either be provided via a different fiber or via a wavelength in case xWDM technology is used.

Improvements on availability can be achieved by connecting multiple RUs in a loop (under study for FRMCS) or connecting a single RU to two DU/CU hotels, one acting as main and the other as backup (see Figure 7). In this case, the ISD between two DU/CU hotels would be approximately 15 kilometers, implying that the number of DU/CU hotels required would be a factor of 2 higher than in the previous case.

Figure 6. Distributed RUs and hoteling of DU/CU

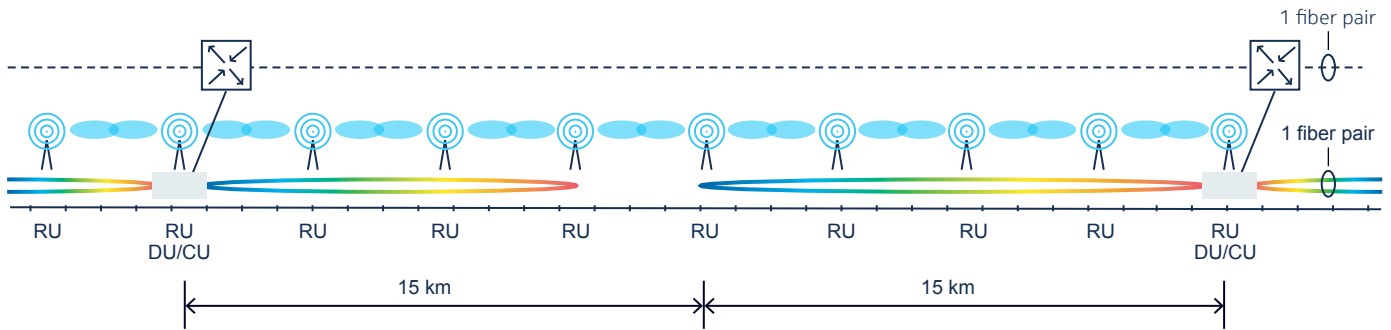
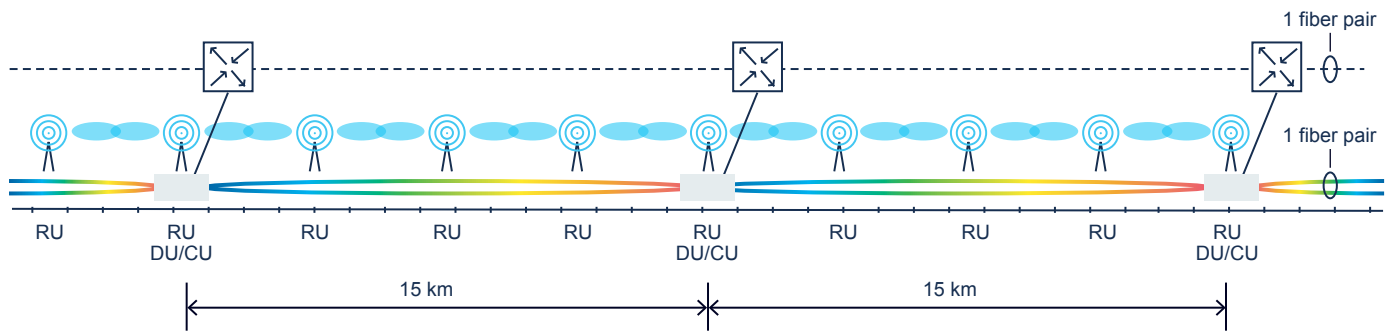


Figure 7. Distributed RUs connected in a loop or to at least two DU/CU hotels

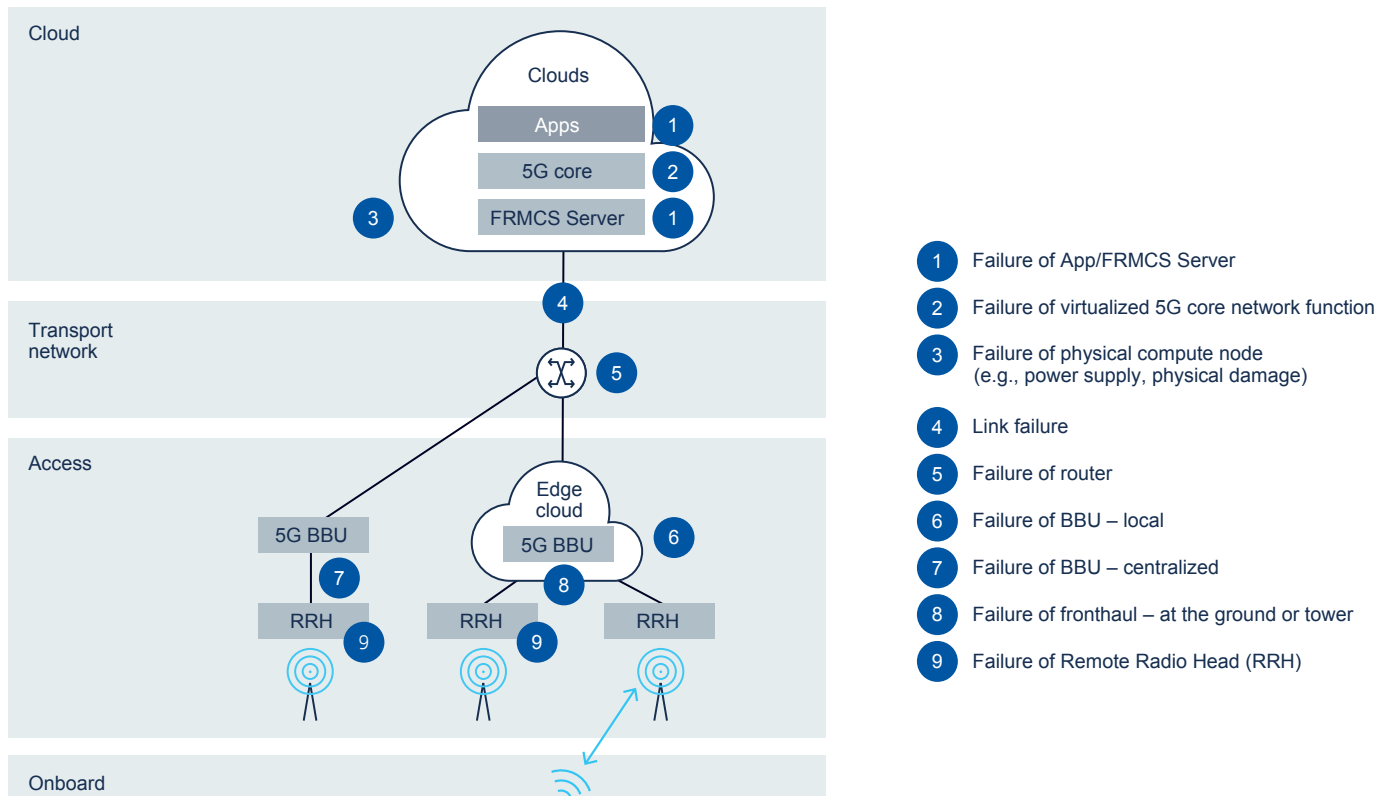


Failure summary

Figure 8 illustrates the various points of failure in the communication network. The failure of an application/FRMCS server (failure point 1), virtualized network functions (failure point 2), and physical compute (failure point 3) can be handled with data center redundancy as shown in Figure 2. Link failure (failure point 4) can be handled with a ring topology as shown in Figure 5 and Figure 6. Failure of routers (failure point 5) and local/centralized BBUs (failure points 6 and 7) can be handled with the help of redundant equipment and in some cases with geo-redundancy. In the case of split options, redundant CUs and DUs could be used as long as the distance limitations are satisfied.

Failure of fronthaul at the tower (failure point 8) and failure of fronthaul in the ground (failure point 9) can be mitigated with the help of dual connection as shown in Figure 4. Failure of an RRH (failure point 10) can be handled with the help of double coverage and/or the use of multiple spectrum (900 MHz and 1900 MHz in the case of FRMCS) as shown in Figure 4.

Figure 8. Various failure points in the trackside network



Assessment of deployment options

Four different deployment options are studied for the RAN (see Figure 9):

1. Classical approach with “all-in-one” base stations
2. Separated RU at the antenna locations, BBU (DU/CU) in a 12-15 km edge range (BBU hotel)
3. Separated RU/DU at the antenna locations, CU in regional center
4. Separated RU at the antenna locations, DU in a 12-15 km edge range, CU in regional centers.

Figure 9. Different split options





	Fronthaul	Edge location	Midhaul	Regional DC	Backhaul	Core DC
RU+CU+DU (all-in-one)						5G core
RU		CU+DU				5G core
RU+DU				CU		5G core
RU		DU		CU		5G core

Table 2. Comparison among different split options

Different split options				
Features	Distributed all-in-one	Distributed RU	Distributed RU+DU	Distributed RU and partially distributed DU
CAPEX (equipment)	Most decentralized => most expensive	More centralized => less equipment required	More centralized => less equipment required	More centralized => less equipment required
CAPEX (infrastructure)	Less expensive (requires larger cabinets at base stations)	Expensive (requires separate locations to host RUs and to hotel CUs+DUs)	Expensive (requires separate locations to host RUs+DUs and CUs)	Most expensive (requires separate locations to host RUs, DUs and CUs)
OPEX ⁶	Most expensive, especially maintenance	Intermediate	Intermediate	Least expensive due to maximum centralization
Rollout by vendors	Most preferred by vendors and clients	Preferred among the various split options	Depends on demands	Depends on demands
Multivendor scenario	No issues	Effort required for interoperability tests	Effort required for interoperability tests	Effort required for interoperability tests
Coordinated Multipoint (CoMP)	More complex	Beneficial	More complex	Beneficial
Interference minimization	To some extent	Feasible	To some extent	Feasible
Overall reliability		Better and cheaper due to cloudified components	Better and cheaper due to cloudified components	Better and cheaper due to cloudified components
Reliability in terms of backup CU/DU	Easier	Depends on the availability of backup location with the distance constraints.	Depends on the availability of backup location with the distance constraints.	Depends on the availability of backup location with the distance constraints.

⁶ In principle, less active equipment in the field reduces the OPEX. Another factor is if the RU with integrated antennae is mounted on the top of the mast. This would require maintenance personnel to climb the mast, thereby resulting in higher OPEX. This applies to all variants.

Table 2 highlights how the all-in-one solution as well as the various split options being considered differ in the features they support. The different split options do not provide an advantage over the all-in-one scenario in the support of features such as carrier aggregation, dual connectivity, Dual Active Protocol Stack (DAPS), HARQ, Sub Carrier Spacing (SCS), Mobility, TDD UL-DL allocation and edge breakout and is therefore omitted from Table 2.

E2E/overall security

As the FRMCS network is a full data-driven architecture, by nature it is more vulnerable to security attacks. This chapter highlights some general aspects to be taken into account while realizing a new generation communication network.

From a security point of view, networks can be categorized as open or closed. Under this definition, an open network is one with little or no restriction on what devices connect to it, such as a mobile network operator or a public Wi-Fi system to which any paying customer can connect. A network might also be open to the public for specific tasks only, such as a passenger experience network providing interactive service information or ticket sales.

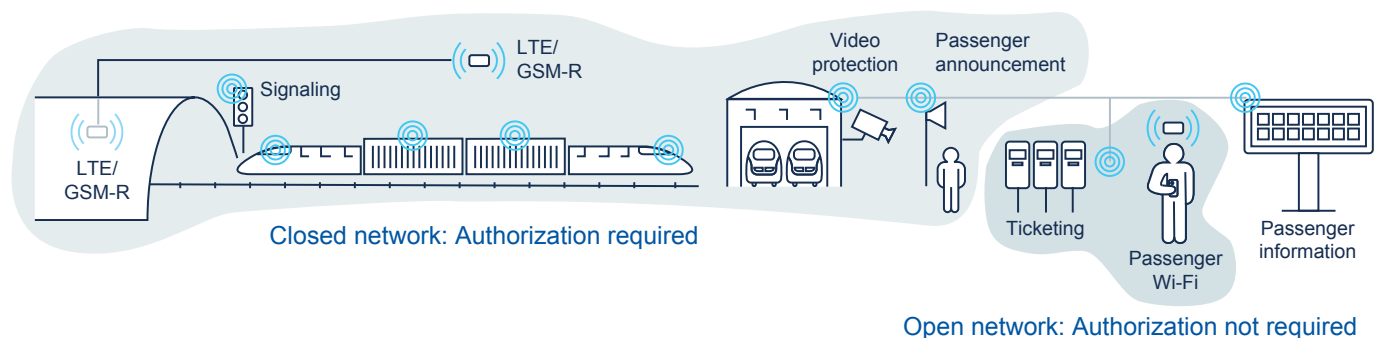
In theory, a closed network is one to which only authorized employees and authorized devices can connect. Examples of closed networks in railway systems include the operational networks (e.g. signaling, SCADA, controls, communications, supervision and OSS), safety system networks (e.g. security cameras, sensors and alarms, access control and emergency communications) and some passenger experience networks (e.g. public announcements and information boards). In practice, many supposedly closed networks are not completely closed.

Table 3. Definition and example of network types

Accessibility	Definition	Examples	Security
Open networks	Anyone/any customer can access	Telco network, public Wi-Fi	Low, anyone can access
Closed networks	Only authorized people/device may access	Corporate networks, industrial networks	Reasonable if properly deployed and secure
Air-gapped networks	No network access – may allow access via USB drives, for example	Nuclear control, secret/top secret government networks	High, but not perfect

As shown in Figure 10, railway networks are typically classified as a combination of open and closed networks.

Figure 10. Example of a railway network



Wireless networks are particularly difficult to secure because they are necessarily accessible by any device in radio range. Even physically secure networks with no external connections (“air-gapped networks”) can be breached by an insider attack or a supply chain attack.

In light of the continually evolving and unpredictable nature of cyber-criminal activity, successful defense of critical infrastructure requires the capacity to prevent, survive and recover from attacks. Moreover, the railway network, similar to other large-scale networks, consist of a large number of components and therefore security should be seen from a multi-dimensional point of view.

Prevention

Successful defense begins with denying unauthorized access to secure systems, and especially to critical systems, using mechanisms such as network segmentation, firewalls, identity management, antivirus software, and encryption.

Networks should be segmented so that critical systems, sensitive equipment, equipment holding critical data, safety-critical equipment, etc., are not directly exposed to untrusted networks, devices or individuals. Different layers of network segmentation would be required based on the security level.

Identity and Access Management (IAM): Steps must be taken to manage authorized access. Access must be permitted only where authorized and only by authenticated identities, and, as much as possible, all access must be audited. Authorized access should be minimized. The fewer the number of individuals, devices, software and traffic flows that are permitted in a secure area, the lower the risk of accidents or insider attack.

Insider attacks are a particular issue for critical infrastructure because nation state attackers are especially motivated and persistent, as opposed to other networks where attackers are more likely to be opportunistic. To reduce the risk of insider attack, role-based access control (RBAC) should be preferred over providing access to shared accounts. Both vendors and individuals should be vetted before being authorized to access any closed network.

User credentials should not be shared: first, security levels are reduced, and second, the task of tracking actions back to individuals becomes much harder.

Two-factor authentication should be used, especially for applications classified with a high safety integrity level (SIL). A password, plus a one-time password generated by a trusted device, is an example of two-factor authentication. One-time-passwords delivered by SMS are relatively easily bypassed and other forms of two-factor authentication should be selected where possible.

All components should be kept up to date with patches, because a high percentage of patches address vulnerabilities that might allow a hostile actor to breach or compromise a system. It should be noted, however, that installing patches in an operational technology system is challenging since it requires homologation efforts.

Networks should be physically secure. Where networks cannot be physically secure, including all forms of wireless network, sensitive traffic must be protected using certified encryption, such as Network Group Encryption.

Survival

There is no absolute protection against volumetric denial of service attacks.

While having spare capacity can help in that it raises the volume of noise required to overwhelm a service, it is essential that safety-critical applications are safe if a service or network is temporarily unavailable.

Surviving a breach begins with detection of the attack. While attacks on the outside of the network perimeter are relentless and can, in sufficient volume, prevent legitimate customers from accessing internet-facing services, it is unexpected activity inside the perimeter that requires a response.

A defensible system must be able to detect unauthorized individuals, unauthorized traffic, unauthorized devices, and unauthorized activity, most especially unauthorized configuration changes. This requires the ability to reliably recognize what is legitimate, and to appropriately respond to everything else.

Network intrusion detection systems and host intrusion detection systems are necessary but are labor-intensive. Detection is too time-critical to be completely manual. Detection must be backed by automatic analysis and response. Many actions should only happen with approval, such as clearing unexpected equipment access alarms, admin login to radio equipment, or a password or configuration change.

Recovery

The final line of defense for critical infrastructure is the ability to recover from a security breach, a situation that even well-defended organizations have had to face many times. The following factors may heavily influence an organization's capability to recover:

- Effective disaster planning allows an organization to respond and recover more quickly, by exploring in advance decisions that may need to be made to minimize and rectify the impacts of the breach. Most critical of these is the ability to assess the risk to human life caused by the breach versus the risk caused by disabling or shutting down affected parts of the network.
- Complete and up-to-date backups, stored offline, allow an organization to rebuild damaged infrastructure, if necessary.
- High-availability configurations, such as redundant sites and redundant hardware, can help protect against some forms of cyber-attack, such as denial of service attacks, but do little to mitigate breaches—a cyber intruder who has been able to gain access to one site can generally use that access to attack the other site as well.

E2E management and orchestration

Finally, the management of a nation-wide railway network would benefit from a high degree of automation. As we move toward higher degrees of virtualization, a closed loop operation would provide the following benefits:

- **Deployment and activation:** Services can be automatically deployed, reconfigured and activated based on orchestration documents, desired network topology and the load in the network.
- **Monitoring:** The activated services can be automatically tracked. As new services are activated or existing services are deactivated, or in the case of failure, an up-to-date view of the network can be maintained.
- **Reaction:** Automatically react to changes in the network based on service policies. If required, automated recovery actions can be triggered toward orchestration.

Conclusion

The use of Future Railway Mobile Communication System (FRMCS) allows railway operators to increase efficiency, offer new services and decrease total cost of ownership. The underlying technology (5G) makes available multiple architectural choices.

In this whitepaper, a joint collaboration between [Nokia](#) and Deutsche Bahn within the sector initiative [Digitale Schiene Deutschland \(DSD\)](#), we explore design choices for a resilient FRMCS architecture. On one hand, DSD aims to modernize rail operations to support new use cases that have demanding data rates, low latency and high reliability. On the other hand, Nokia is obviously developing 5G products that cater to the needs of the various operators and industries.

Our joint analysis of the various architectural options and features available in 5G leads us to the conclusion that it is feasible to find a technical solution that is able to guarantee a highly reliable network for railways at a reasonable cost. As part of the implementation of a concrete FRMCS network, a detailed analysis is needed to find a solution with the right trade-off between costs and reliability.

Abbreviations

3GPP	3rd Generation Partnership Project	GSM-R	Global System for Mobile Communications - Railway
ATC	Automatic Train Control	ISD	inter-site distance
ATO	Automatic Train Operation	LTE	Long Term Evolution
ATP	Automatic Train Protection	MA	Movement Authority
BBU	Base Band Unit	MCX	mission-critical services
CAPEX	capital expenditure	OPEX	operational expenditure
CU	Central Unit	OSS	operations support system
DC	data center	QoS	quality of service
DU	Distributed Unit	RAN	Radio Access Network
ETCS	European Train Control System	RU	radio unit
FRMCS	Future Railway Mobile Communication System	RRH	remote radio head
GoA2	Grade of Automation 2	UE	user equipment
GoA4	Grade of Automation 4	UPF	User Plane Function

About DB Netz

With more than 33,000 km of track, DB Netze Track operates the largest rail network in Europe. More than one billion train-path kilometers are traveled each year on the tracks in Germany. DB Netze Track is responsible for managing infrastructure operations as well as for securing long-term infrastructure quality and availability, and non-discriminatory access to train-paths and service facilities. This includes preparing schedules in close cooperation with customers, operations management, construction management and maintenance.

Theodor-Heuss-Allee 7
60486 Frankfurt am Main
www.dbnetze.com

About Nokia

We create technology that helps the world act together.

As a trusted partner for critical networks, we are committed to innovation and technology leadership across mobile, fixed and cloud networks. We create value with intellectual property and long-term research, led by the award-winning Nokia Bell Labs.

Adhering to the highest standards of integrity and security, we help build the capabilities needed for a more productive, sustainable and inclusive world.

Nokia is a registered trademark of Nokia Corporation. Other product and company names mentioned herein may be trademarks or trade names of their respective owners.

© 2021 Nokia

Nokia OYJ
Karakaari 7
02610 Espoo
Finland
Tel. +358 (0) 10 44 88 000

Document code: 1392551524174482819 (October) CID210670