Architecting Railway Systems for Resilience

Luke Church, MEng (Hons) MIET, Thales Ground Transportation Systems UK

SUMMARY

The primary purpose of a railway is to safely and reliably transport people and goods; resilience to disruption is crucial to railways achieving this purpose. A railway is a highly complex system, consisting of multiple interconnected sub-systems; enhanced resilience can be achieved by analysing the architectures of railways systems, for example signalling systems, which are the subject of analysis in this paper.

Achieving resilience through the principle of capacity, flexibility and tolerance is explored using examples from across the rail industry, in addition to other industries with a particular focus on similar critical infrastructure systems. The lessons learnt from this analysis are used to develop a novel system architecture utilising cutting edge technologies to deliver a highly resilient, efficient signalling system.

1 INTRODUCTION

Resilience is defined as *the capacity to recover quickly from difficulties* (OED, 2019)*;* in the context of a railway this includes maintaining a normal service during minor disruptions and ensuring graceful degradation and swift recovery from major disruptions. Resilience has been studied extensively in academic literature; one model which has been developed is the multi-phase resilience trapezoid (Panteli, et al., 2017), as shown in Figure 1. This model breaks system response down into three phases: disturbance progress, post-disturbance degraded state and restorative state. From this model it is clear that resilience can be enhanced by graceful degradation and fast stabilisation during phase 1, minimising the duration of phase 2 and rapid system restoration during phase 3.

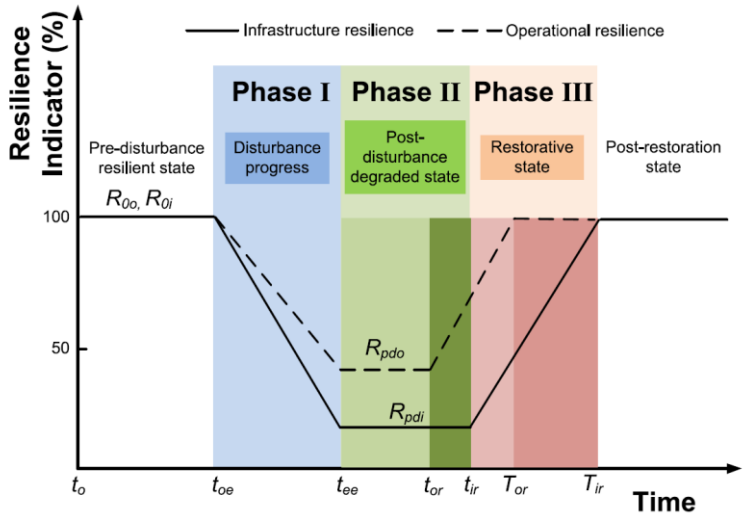

Figure 1 - Multi-phase resilience trapezoid (Panteli, et al., 2017)

Sillitto (2014) defines four basic principles of resilience: capacity, flexibility, tolerance and cohesion, the first three of which are explored in this paper. Railway systems are highly effective at achieving resilience through capacity, or more specifically redundancy, however railway systems have had less focus on achieving resilience through flexibility and tolerance.

2 CAPACITY

Capacity is the ability of a system which allows it to withstand disruption in its normal operating state; once the capacity of a system to withstand disruption is exceeded the system must rely on other resilience attributes to achieve graceful degradation and recovery. The capacity attribute can be further broken down into three further attributes (Sillitto, 2014):

- **Absorption**: Having margins within the system to withstand disturbances.
- **Physical redundancy**: Having back-ups for critical components.
- **Functional redundancy**: Having different means of performing critical functions.

Each of these attributes has been exploited to great effect to deliver resilience to railway systems, as is discussed in the following sections.

2.1 Absorption

2.1.1 Timetables

The ability of the railway to absorb disturbances is clearly demonstrated through recovery margins in timetables. Having recovery margins in timetables allows the timetabled service to be achieved without trains being driven at the maximum permitable speed, or with minimum dwell times in platforms. This means that should a minor disturbance occur the train can recover time through driving at the maximum permitable speed or minimising dwell time.

A further example of timetables being used to makes the system more resilient are seasonal timetables, for example leaf fall timetables in the UK. In the autumn thousands of tonnes of leaves fall onto railway lines, sticking to damp rails with trains compressing them into a slippery layer which reduces adhesion between the wheel and rail (Network Rail, n.d.). To compensate for this, drivers must brake and accelerate more gently to avoid wheel slip/ slide, special timetables are therefore published to allow for the resulting journey time increase and maintain recovery margins.

There is, however, a direct trade-off between line capacity and recovery margin. The demand for faster, more frequent services can therefore come at the cost of the timetables' ability to absorb disruption, potentially leading to knock-on disruptions across the network as trains miss slots at key junctions. A solution which has been deployed to mitigate this effect is Traffic Management Systems (TMS), which are discussed later.

2.1.2 Energy distribution networks

A similar strategy is utilised in energy distribution networks (i.e. electricity and gas), which are also critical infrastructure systems. Rapid fluctuations in demand and supply of energy requires supply side infrastructure to maintain a capacity margin to meet these demands, an example of these margins in the UK energy networks is shown in Figure 2.

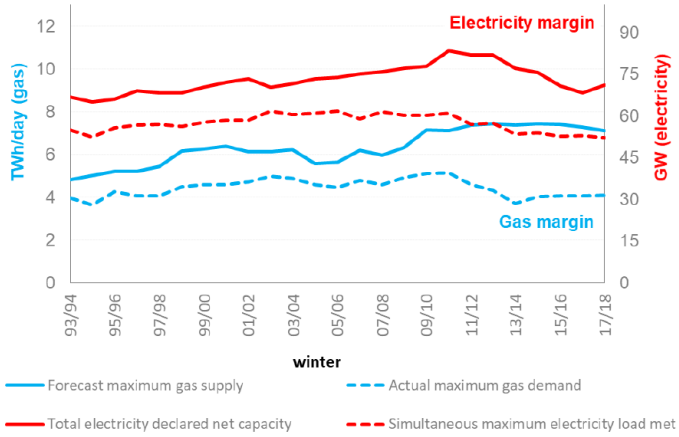

Figure 2 – UK gas and electricity capacity margins, maximum supply and maximum demand 1993/94 to 2017/18 (BEIS, 2018)

There is a balance to be struck between the cost of constructing and operating spare capacity and need to ensure constant power supply. Smart systems are being developed and deployed to enhance the system's flexibility, optimising the balance of supply and demand using techniques such as demand-side response. An example of this is with smart charging of electric vehicles, whereby charging is automatically managed to flex with local grid demands, this can have the added benefit of using lower tariffs to charge vehicles when demand is low. It also opens an opportunity to use stored energy in car batteries to supply energy to the grid when demand, and therefore energy tariffs, are high further adding to the networks resilience. Of course this technology needs to take into consideration primary purpose of the vehicle, ensuring the vehicle has full charge when it is needed, though smart configuration (OFGEM, 2017).

2.2 Physical Redundancy

Resilience is achieved though having a layered defence against failure, endeavouring to avoid any single points of failure. Conducting a Failure Mode Effect and Criticality Analysis (FMECA) is a key tool in the study of a system to assess redundancy and identify potential vulnerabilities, the result of which may be the introduction of backup systems to duplicate critical systems.

2.2.1 Failure Mode Effect and Criticality Analysis

A FMECA is performed to consider how the failure of a component or signal within a subsystem affects a system, and ultimately the railway. The FMECA process determines the possible failure modes of the components and then determines their effects on the complete system. Because it considers all component failures, this approach is particularly good at detecting conditions where a single failure can result in a hazardous situation. However, this technique does not consider multiple failure conditions, these fault trees are analysed using techniques such System Hazard Analysis, usually from a safety perspective to minimise the risk of a wrong side failure.

The FMECA uses a functional approach to identify and classify key system failure modes, in terms of both frequency and severity in order to determine the risk and impact to service operation, an example of which is shown in Figure 3. This allows design effort to be directed at the areas of greatest need and provides a framework for more precise quantitative risk assessment as system design progresses.

Severity / Frequency	**5 - Critical**	**4 - Serious**	**3 - Severe**	**2 - Major**	**1 - Minor**
5 - Frequent	A	A	A	A	B
4 - Probable	A	A	A	B	C
3 - Occasional	A	A	B	C	C
2 - Remote	A	B	C	C	D
1- Improbable	B	C	C	D	D
0 - Extremely Unlikely	C	C	D	D	D

Risk Class	**Explanation**
A	Intolerable (shall be eliminated)
B	Undesirable (shall only be accepted when risk reduction is impracticable)
C	Tolerable (acceptable with adequate control)
D	Negligible

Figure 3 - Example of FMECA risk criticality matrix with risk classification. Severity/ Frequency must be clearly defined in accordance with system requirements; for example, a critical failure could be defined as line-wide closure for a week and a frequent failure more than once per month.

2.2.2 Backup/ Standby Systems

Physical redundancy is achieved through duplication of critical systems, for example by having a backup as a "hot standby" to switchover to should the primary component fail. An example of this is having Vehicle Onboard Controllers (VOBCs) in the front and rear of Communications Based Train Control (CBTC) trains, as shown in Figure 4. Should an error occur in the VOBC or associated subsystems (e.g. communications antenna) then an automatic switchover will occur so that the train can continue to function normally. In this scenario the system is resilient to a single failure; however it has lost its physical redundancy.

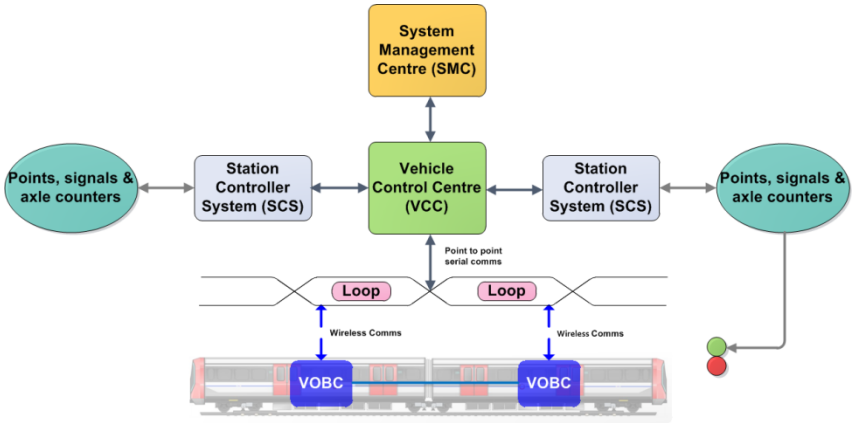

Figure 4 - High level architecture of Thales' Seltrac IS-Loop CBTC system, showing dual VOBCs for redundancy.

2.2.3 Virtual Machines

Critical computer systems with extremely high availability requirements take redundancy a step further than pure hardware redundancy through the use of Virtual Machines (VMs). A VM is a software computer which exhibits the behaviour of a physical computer; by utilising VMs multiple system environments can be run on a single physical computer.

One of the benefits of this architecture is the ability to switch to another virtual system should the primary system experience a software based fault. This may be particularly effective if the back-up VM is running diverse software, for example a previous version which is known to be stable. In this scenario automatic and immediate software reversion would occur a bug occur with a newer software version, which may not materialise for weeks or even months of operational use.

2.2.4 Cloud Computing

The use of VMs can be taken further with the advent of cloud computing. It opens the possibility for critical system switchover without the loss of redundancy as a new VM can be created in a cloud hosted environment, an example of such architecture is shown in Figure 5.

Major cloud providers offer solutions for guaranteed server availability, should railway systems be hosted on cloud platforms instead of on central hardware platforms the risk of hardware failure is effectively removed. This could be taken further through the use of diverse cloud service providers should one experience a catastrophic fault.

Hosting railways systems in the cloud also removes the issue of hardware obsolescence, a major issue facing the industry due to the rate of development in computer systems rapidly outstripping the expected life of railway systems.

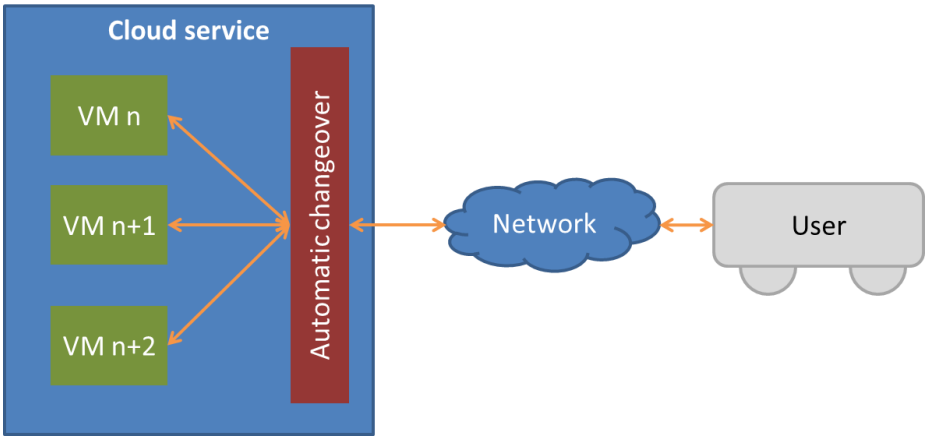

Figure 5 - Indicative architecture with multiple VMs hosted using a cloud service.

2.3 Functional Redundancy

Functional redundancy is achieved through multiple components of the system performing the same function independently and diversely. Should one component of the system fail then it can function with a secondary, fall-back, system. Although being effective at enhancing resilience, there is a high capital cost associated with the design and deployment as well as an additional operational cost of maintaining two discrete systems.

2.3.1 Secondary Signalling Systems

An example of where functional redundancy has been implemented is the use of secondary signalling systems, for example the Downtown Line in Singapore. The Downtown Line uses a CBTC system for Automatic Train Operation (ATO); however it also has a coded track circuit system for secondary train detection and fall back ATO train control.

The primary CBTC system communicates movement authority wirelessly based upon train positions determined by onboard sensors and trackside balises; the contrasts with the coded track circuit which determines train position through track circuits and transmits speed codes through the track circuit to give movement authorities for the train to move in ATO.

When utilising the secondary system there is a performance impact as a fixed block signalling system is used instead of moving block. However, this allows service to continue without significant human intervention which is especially important in a fully automated system (i.e. driverless) such as is deployed on the Downtown Line.

2.3.2 Power Generation

An example of where functional redundancy is utilised to good effect is in power generation, by using diverse energy sources. By having a diverse mix of energy sources a national power grid is resilient to disruptions to a specific source; for example geopolitical issues restricting the supply of oil, or unfavourable weather for renewable power generation as well as being able to meet peaks in demand. The diversity of energy sources for electricity generation in the UK is demonstrated in Figure 6; these are average figures over the course of a year but are highly variably on any given day due to factors such as fuel prices, weather and demand. The grid utilises the most efficient power sources to supply a base load, for example nuclear power, and meets demand fluctuations with responsive sources such as fossil fuel plants and energy storage. This is an efficient way of meeting demands for capacity with redundancy, whilst reducing operation costs; this is shown in Figure 7.

The nature of power infrastructure means that having a diverse mixture of energy sources also increases capacity, therefore further enhances resilience. This is not the case in the example of a secondary signalling system which is only used as a fall back; opportunities may exist for diverse signalling systems to combine to operate at peak capacity, with a fall back to using a single system with reduced capacity, or when demand is reduced. This could result in reduced capital and operational costs.

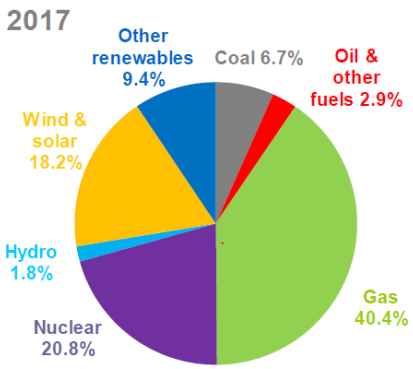

Figure 6 - UK electricity generated by energy source in 2017 (BEIS, 2018).

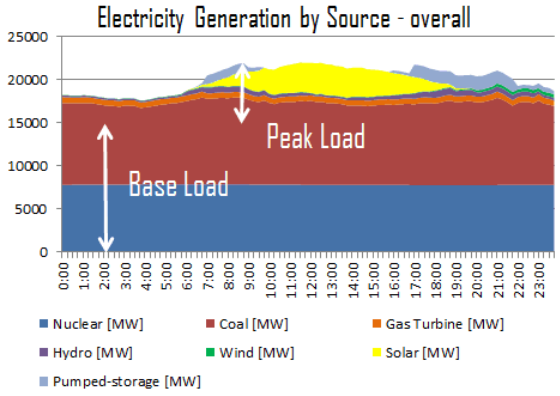

Figure 7 - Base and peak load power supply example (Nuclear-Power.net,

3 FLEXIBILITY

Flexibility allows a system to restructure itself when faced with disruption, for example by changing the system architecture or operating concepts. Railways exhibit flexibility when faced with major disruptions; however these often result in massive reductions in the quality of service, some examples are given in the following section.

3.1 Rail System Flexibility

CBTC systems, such as those implemented across London Underground, often utilise secondary train detection in the form of axle counters or track circuits. This allows the position of trains to be safely determined should all, or part, of the CBTC system fail. In this scenario the system is restructured to use fixed block signalling principles, such as discussed with Singapore's Downtown Line. However without full functional redundancy severe speed restrictions and restrictive operating procedures must be implemented to safely operate, resulting in major delays.

A further example of a change in operating concept is in the event of a lineside signal failure. In this scenario a driver would stop the train when no aspect is lit and require movement authority to be granted over radio by the signaller in order to proceed. This allows the service to proceed with a change in the primary operating architecture but again results in delays to service. An extreme example is the provision of rail replacement buses which deliver the same service of transporting trains but with a massively reduced quality of service.

3.2 Mission Command

In military doctrine "Mission Command" trains junior leaders to act on their own initiative in accordance with the commander's intent to achieve the goal of the mission when communications break down, regardless of unforeseen disruptions or obstructions not included in their orders. For this to be effective each component of the system needs to understand the overall purpose of the system, as well as having enough information and understanding to make appropriate local decisions (Sillitto, 2014).

3.2.1 Internet of Things (IoT)

Robust, high bandwidth communications networks along with the miniaturisation of technology have enabled the interconnection of physical devices, referred to as the Internet of Things (IoT).

Signalling systems tend to be highly centralised, and increasingly so in the UK with the roll out of Rail Operating Centres (ROCs) on mainline rail. This contrasts with IoT concepts which enable a system to be highly distributed; using this concept it is possible to architect a signalling system where the information and intelligence to control the local system is contained within wayside assets and on the train. Such system architecture would be highly flexible, utilising a kind of "Mission Command" structure, with local areas able to operate independently or perhaps take control of adjacent areas in the event of disruption.

4 TOLERANCE

Tolerance allows a system to degrade gracefully when faced with disruption, it is an area the rail industry where improvements can be made and important steps are being taken to achieve this, for example with the deployment of TMS.

4.1 Traffic Management Systems (TMS)

TMS primary function is to manipulate the timetable plan in near real-time to return trains to the correct timetable in the most efficient manner possible when there is a disruption, enhancing the networks tolerance to minor disruptions and speed of recovery from major disruptions. In addition, a TMS system can provide a wide variety of added value functions for operators including: platform management, passenger information, network availability, incident / delay management and service information (Stacy, 2014).

TMS has been deployed in two main configurations: as a standalone system used as an operational decision support tool for signallers, as well as fully integrated with an interlocking to provide SIL4 movement authorities without intervention from a signaller. As a decision support tool the system if less complex with fewer challenging interfaces, however a fully integrated system delivers maximum system performance benefits as the railway automatically optimises train schedules.

TMS systems have been proven to enhance railway tolerance to disruption, as well as increasing system capacity and therefore recovery margins and further deployment will bring major benefits to the railway, an example of a control centre using TMS is shown in Figure 8.

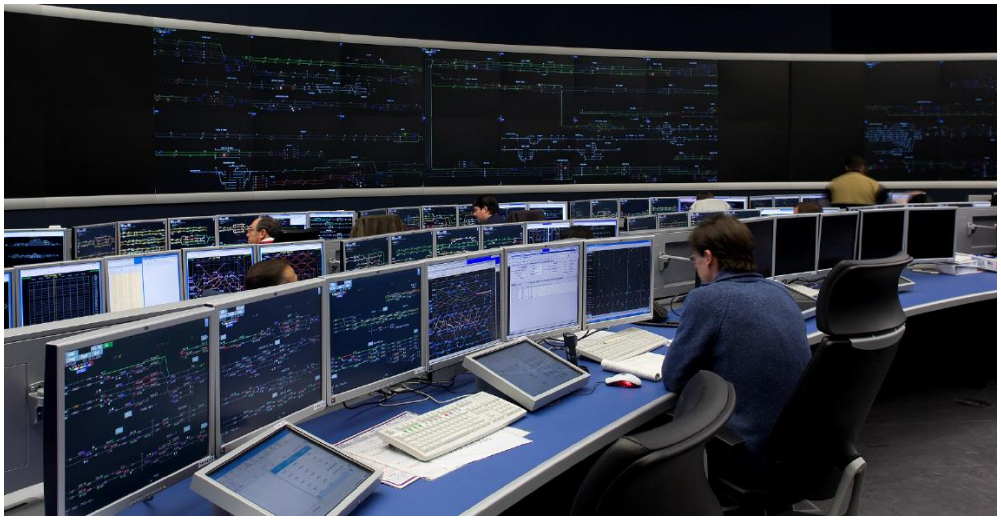

Figure 8 – Thales' ARAMIS TMS in operation in control centre in Portugal.

4.2 Maintainability

A further means of achieving system tolerance is by ensuring the system is maintainable. This means rapid and reliable system restoration can be carried out, using ordinary trained people and reasonable support facilities. Clearly the speed of which corrective maintenance of a fault can be completed so that the system can return to full operational capacity is crucial to minimising service disruption.

Railways have heavily restricted access trackside and are geographically distributed by their nature; therefore one means of enhancing maintainability in modern railway systems is through minimising trackside intervention needed to conduct both preventative and corrective maintenance activities.

Maintainability is further enhanced through designing the system to use Line Replaceable Units (LRUs). LRUs are components that can be replaced as a single complete unit by a single person, allowing for fast fault finding and corrective maintenance by replacing an entire LRU on site.

4.3 Graceful Degradation

An example of graceful degradation from computing is when a lower-resolution video is streamed in place of the high-resolution version when there is insufficient network bandwidth is available. A similar concept is used with

Smart Motorways, whereby variable speed limits are imposed or the hard shoulder is used to increase traffic capacity, when there are high traffic volumes vehicles. This helps to avoid cars bunching and improves average journey times, an example is shown in Figure 9.

Figure 9 - Smart motorway in operation (Boaden, 2017)

These are examples of where systems operate with reduced performance when presented with disruption instead of a major or sudden loss of service/ function. These solutions are comparable to approach taken with TMS; however further solutions to ensure the graceful degradation of railway control systems must be explored.

5 RESILIENT SYSTEM ARCHITECTURE PROPOSAL

The system architecture proposed in Figure 10 brings together a number of the resilience principles which are demonstrated in current rail industry applications, as well as those discussed from other industries. Key features are as follows:

Cloud based TMS: train scheduling and network management is carried out by a cloud based TMS system, utilising hot-standby VMs for switchover and a highly available cloud service provider. Signallers can access workstations remotely through the internet allowing total flexibility of control operations. The use of TMS enhances the systems tolerance to disruption as well as enabling more efficient use of recovery margins. This architecture also employs diverse TMS software, allowing for immediate reversion to a proven baseline should a fault occur.

Distributed interlocking: signalling assets each contain the interlocking capability to provide movement authorities in their local area, all networked together using IoT principles. Should one asset fail the functionality is duplicated in other local assets. Should it not be safe to issue movement authority with a failed asset then the disruption is highly localised and quick to diagnose and repair. It is also possible for the system to be flexible in restructuring its architecture, for example if smart devices are configured and capable of taking over adjacent signalling areas in the event of a major failure.

Communications network: the primary network for train and trackside asset communications can be a dedicated system, such as GSM-R. As a fall back, assets are equipped with communications equipment to utilise public cellular networks. Communications must be end-to-end encrypted and utilise leading cyber security techniques to ensure security of communications.

Train Control: in this architecture the primary means of train control is with ATO or in-cab signalling. Should this be disrupted functional redundancy can be implemented with back up conventional train detection systems and signals for fixed block signalling.

On-board controller: each vehicle will have two on-board controllers for physical redundancy. The on-board controller includes a positioning system which utilises on-board sensors and other inputs such as GPS in order to robustly determine its position and communicate it to local smart devices. If the train is operating in ATO, a train management system is control speed and braking when recieving movement authorities from the distributed interlocking, should the operator prefer in-cab signalling such an interface would be implemented to display movement authority to a driver.

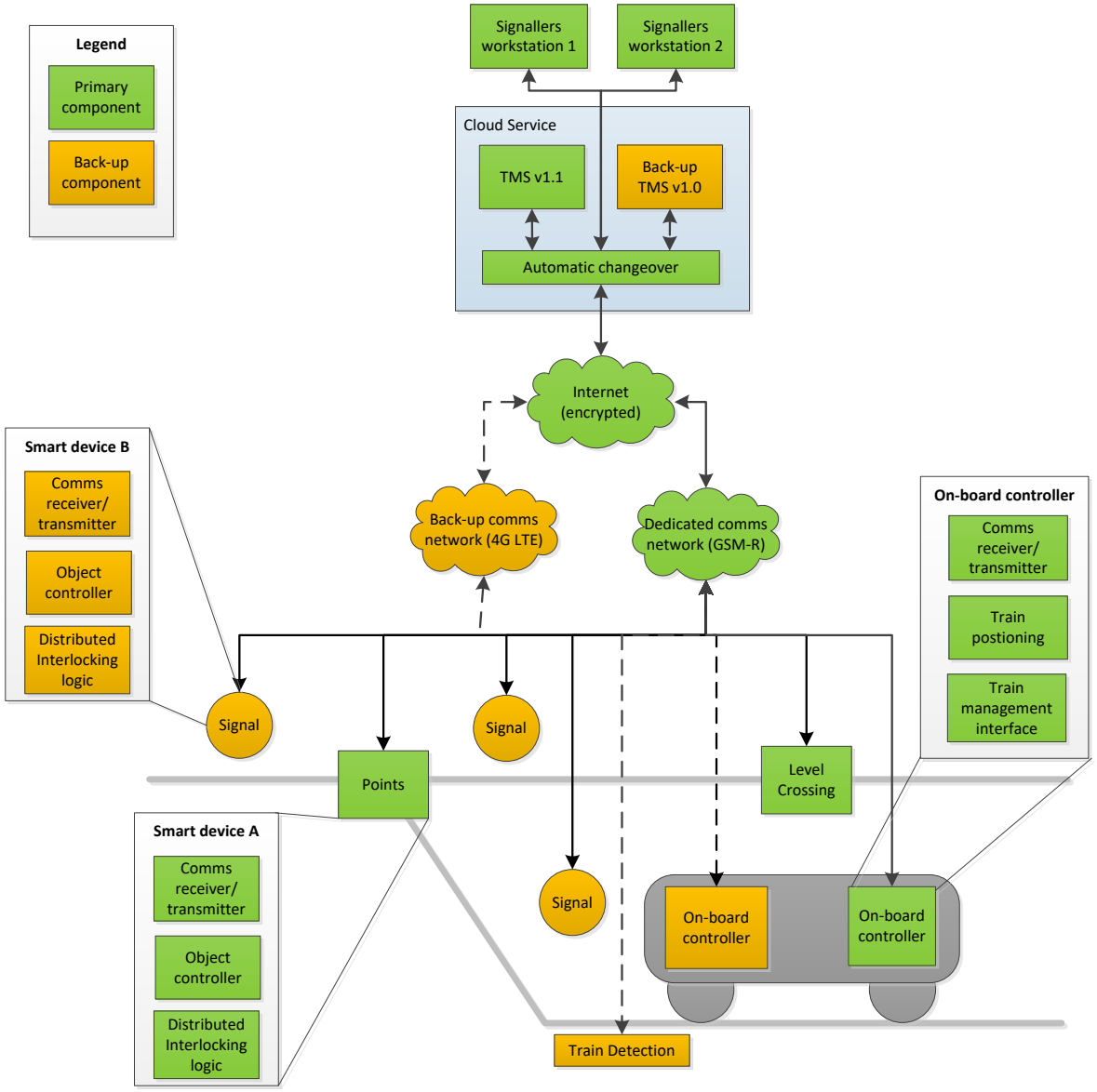

Figure 10 - Proposal for resilient system architecture

6 CONCLUSION

Threats to the railway from sources such as cyber-attacks and extreme weather are increasing. At the same time more train services are being introduced through modernisation of infrastructure, which reduces the margin to withstand disruption through absorption, and limits are being reached in what can economically be achieved through redundancy. Railway performance can be improved by architecting new and existing railway systems for resilience using the principles of capacity, flexibility and tolerance at the design stage.

These principles can be implemented and railway resilience can be enhanced by learning lessons from other industries and embracing new technologies, such as IoT, cloud computing, smart devices, TMS and mobile connectivity.

7 REFERENCES

1) BEIS, 2018. *UK Energy In Brief,* London: National Statistics.
2) Boaden, B., 2017. *Smart motorway in operation.* [Online]
   Available at: https://www.geograph.org.uk/photo/5251763
   [Accessed 30 May 2019].
3) Network Rail, n.d. *Leaves.* [Online]
   Available at: https://www.networkrail.co.uk/running-the-railway/looking-after-the-railway/delays-explained/leaves/
   [Accessed 28 May 2019].
4) Nuclear-Power.net, n.d. [Online]
   Available at: https://www.nuclear-power.net/wp-content/uploads/2017/08/Base-Load-Load-Follow-Peak-Load.png
   [Accessed 28 May 2019].
5) OED, 2019. *Oxford English Dictionary.* s.l.:Oxford University Press.
6) OFGEM, 2017. *Upgrading Our Energy System: Smart Systemsand Flexibility Plan,* London: HM Government.
7) Panteli, M. et al., 2017. Metrics and quantification of operational and infrastructure resilience in power systems.. *IEEE Transactions on Power Systems,* Issue 32, p. 4732.
8) Sillitto, H., 2014. *Architecting systems.* London: College Publications.
9) Stacy, M., 2014. *Traffic Management Systems – Train Regulation Made Easy?.* [Online]
   Available at: https://www.railengineer.co.uk/2014/05/30/traffic-management-regulation/
   [Accessed 30 May 2019].