The Institution of Railway Signal Engineers Inc

Australasian Section Incorporated

MOVEMENT AUTHORITIES – A SYSTEMS FRAMEWORK

Peter Burns
MBA, BAppSci (Elect), MIRSE, CPEng, MIEAust
PYB Consulting

SUMMARY

This paper on Movement Authorities is one of a series on the various elements of the Generic Systems Framework (see figure 1). The issuing of Movement Authorities is distinguished from the setting of a route and the general pre-conditions for the issuing of a Movement Authority stated.

Movement Authorities are shown to be found in all safeworking systems and having characteristics which are common to all of them. The process for issuing a Movement Authority may be characterised as the formation of a contract between the train and the interlocking.

Looking at fixed signal systems, the signal is found to fill three distinct functions, one of which is the communicating of movement authorities.

Turning to ERTMS and CBTC systems, it is shown that their central functionality is of a nature that does not require treatment as a movement authority. Benefits can be obtained by recognising the different natures of the three distinct functions which are replaced when ERTMS and CBTC systems replace fixed signals, and drawing the signalling requirements around those distinct functions appropriately.

1 INTRODUCTION

As signalling moves away from lineside signals as the primary conduit for providing movement authorities, it is important to understand the fundamentals of these authorities. In his president's address to the IRSE in 2004, J D Corrie [1] lamented the communications cost associated with the need to continuously communicate movement authorities to trains, what he termed in his address the "reversionary principle".

He expressed a desire for signalling to review its use of this principle and pointed (in the spoken version of the address) to a simpler world of "intermittent authorities" as seen with, for instance, train order working.

The commonly held view that fixed signal based movement authorities are continuous whilst some other types are intermittent represents a misunderstanding about how movement authorities work.

This paper places the concept of "movement authorities" into a systems framework and addresses:

- What are the fundamental requirements which surround a movement authority?
- How do the functions associated with fixed signals meet these requirements?
- What are the fundamental requirements for removing an authority once issued?
- What other functions do fixed signals fulfil?

This is a theoretical paper, not focussed on any specific jurisdiction or signalling method, and it is an overview. The intent is to present a common language which will support future generations of signalling as well as those we are familiar with. This language can assist in evaluating features in proposed systems in terms of what requirements those features set out to meet. It can also assist in seeing the gaps in proposed new systems without the need to refer to some existing signalling system the reviewer may be familiar with.

The reference for the discussion will be a user view deconstruction of basic signalling functionality which I term the "Generic Systems Framework" (Refer Figure 1).

2 NOTATION & ABBREVIATIONS

Abbreviations are as follows:

ATP:	"Automatic Train Protection" in its most general sense. Used here to represent any vital function included in the ATC (refer IEEE 1474.1 for discussion of ATC);
CBTC:	Communications Based Train Control;
ERTMS:	European Rail Traffic Management System;
ETCS:	European Train Control System – a component of ERTMS;
EVC:	Electronic Vital Computer;
RATP:	Regional ATP; the portion of ATP function implemented in infrastructure. This may be contrasted with VATP, which is the portion implemented on the vehicle;
TBTC:	Transmission based Train Control; The immediate predecessor to, or earlier generation of CBTC;

In figure 1: the fill colours used have the following meanings:

Green: Function carried out in infrastructure;

Yellow: Function carried out on the vehicle; where both colours are used, the function may be partly carried out on the vehicle and partly in infrastructure;

Blue: Function not allocated and may include functionality outside traditional signalling.

3 GENERIC SYSTEMS FRAMEWORK

The generic systems framework is a user requirements view of the processes needed to move vehicles through a controlled environment. Whilst the current context is railway signalling, it can equally be applied to road and air traffic movements.

The framework presented here has nine sections as illustrated in figure 1.

This top level view in the requirements analysis (URS – User Requirements Specification) is abstract and not intended to map 1:1 onto any physical systems and subsystems. The mapping to physical system and technology occurs through the lower level SRS (System Requirement Specification) and ADD (Architecture Design Definition). The requirements are also nest-able, so that for different modes of working within a single physical system the mapping of the requirements to function can differ.

4 MOVEMENT AUTHORITIES

4.1 When can you give a movement authority?

Much signal engineering time and expertise is put into ensuring that the route setting and point movement processes occur safely as they should. But these are not the topic for this paper. In considering an authority we can consider a train. The route ahead is set and all the points are in position. The train would like to enter the section, but to do this requires an authority.

Before a movement authority can be issued, the following need to be in place:

- The route (if any) must be set;
- Points and other apparatus called to position or state in the route setting process must be detected in the field to be in the correct position or state to the extent necessary to allow safe passage;
- Where the authority is applicable only for a particular class of train, the train must be confirmed to be of that class;
- The train may need to understand the geography and any important features of the authority section (route knowledge);
- The track ahead must be proved clear (optional – refer below); and
- Required “insurance” provisions must be in place according to the local practice (see below).

With all the pre-conditions in place we are ready to issue an authority. But to do this requires a process.

4.2 What is an authority?

We all understand that a green signal gives authority, as does a train order issued over the radio, but how do they work to give authority at a fundamental level? We notice for instance that a green light is not always enough. A staff or some other item might also be needed (eg for a single line section).

An authority is a contract.

As such it has:

- A meeting of minds:

  Both parties must have a common understanding of the matter being discussed (eg train and infrastructure must agree the version of the rules involved).

- Agreement between parties: An identified train and identified infrastructure.

  A number of accidents have occurred where a train has tried to accept an authority intended for a different train. Systems are provided to ensure that only the right train and the right infrastructure are involved.

  Communications are required and formal steps to evidence the agreement. This is a key challenge for any signalling system.

- A subject: Train movement over a section of infrastructure ahead of the train.

  The authority involves movement between a “start of authority” point and an “end of authority” point, but may include other actions to be taken in between which can be both arbitrary and complex. Alternately the authority may leave it to the train to do as it chooses within the defined section (at the other extreme).

  Insurance cover (overlaps, flank protection, etc. outside the limits of authority) are typically also included.

- Terms and conditions: The rules and procedures, the type of authority, etc.

  The authority in the form issued is typically the equivalent of populating a simple schedule (“tick colour of indication to be displayed”) to a quite complex contract (the rules and procedures) which are, as is the nature of contract, arbitrary for each jurisdiction.

  Sometimes the infrastructure owner tries to add special conditions by affixing a sign to a signal post, but the binoculars the train driver needs to read the sign are not always standard issue.

- Consideration: The train gets its path and the infrastructure gets a warm fuzzy feeling at minimum.

4.3 The process

Contract protocols are the original form of failsafe communications, predating even railway signalling.

An authority is put in place by a formalised, if arbitrary process of offer and acceptance.

Preliminary to the process is an “invitation to treat” step. The train initiates this by being in a particular area (eg sighting distance of a signal between the “first signal affected” and the commencement of authority (inclusive)) or by communicating with a control centre.

This is followed by the formal process which involves:

- An offer (by the infrastructure system to the train system)

  Offering the authority is done by displaying a proceed aspect on a signal, by transmitting an order to the train, or some other method (eg handing over the staff). This display or transmission continues until communication of acceptance occurs.

- An acceptance (by the train)

  The authority can be accepted by the train by taking possession of the offered object (real or virtual), or by commencing to perform the authorised activity.

- A communication of acceptance (received by the infrastructure)

  The train communicates acceptance by a formal process such as “reading it back” (for orders) or passing a signal (for fixed signals – the “first signal affected” or the first signal passed up to and including the one marking the commencement of authority – refer figure 2). It is important to note that there can be a gap between acceptance and communication of acceptance which can be invisible to the control centre. This gap is important, particularly if we need to renegotiate the contract later.

The gap in time between acceptance and communication of acceptance must be managed. From the point of view of the infrastructure, it is important (for safety) that the train is assumed to be in the process of accepting the authority in any case where this may be uncertain (a railway equivalent to the “postal acceptance” rule). This reflects the limited capability of the train to communicate with the infrastructure in some signalling technologies.

Once acceptance has been communicated the transmission of the offer can stop. From that point the train must remember the authority (vital persistence) by some method until it is completed.

In contract law, the issue of communication of offers and acceptance (and the possibility of miscommunications and delays) is an important topic which fills volumes of literature with cases and principles developed about how to treat each thing that can go wrong. What happens if the process is interrupted part way or a step improperly anticipated? These questions are important in contract, and no less important in railway signalling, particularly as we move towards communications based signalling. But there are more court cases than there are railway accidents, so the literature has more depth.

4.4 Cancelling an authority

The vast majority of authorities expire due to completion as they are performed. An authority can be recognised as completed either all at once or in defined stages. There is always a process involved in recognising the various stages of authority completion up to when the train reaches the next “start of authority” point.

In some other cases the infrastructure wishes to cancel the authority before it has been completed, usually before it has even commenced. This is a key process with risks which must be managed.

Removing an authority involves a renegotiation of the original contract which provided the authority. All the process steps involved in setting up the original contract need to be retraced for changing it.

In this case there is a revised offer initiated by the control centre putting a signal lever back to normal (or pulling the button) or transmitting a cancellation of authority order. The signal will revert to red, communicating the offer to cancel, but the authority is not yet cancelled.

The train accepts the changed authority by a formal response process or by observing the stop signal and stopping ahead of that signal. The authority remains in place in the meantime.

For fixed signals, the authority cannot be cancelled at all if the train has already passed the signal, since no method for communicating the revised offer or its acceptance is available. Being able to observe the signal ahead is of no help in this case either since that signal is not the one controlling the section to be cancelled.

The limitations of fixed signals (inability of the train to feed back to the control centre) come to the fore here. If the authority cancellation is offered with the train on approach to a fixed signal, the possible outcomes are:

- Train stops ahead of the signal: train has communicated acceptance of the removal of authority and the authority can be removed in the infrastructure;
- Train passes signal without stopping: the train has communicated acceptance of the original authority and communication of acceptance of the removal of authority is no longer possible. The authority must be retained in place by the infrastructure.

The above summarises the standard approach locking rules with which you will be familiar.

These rules have developed around the specific limitations associated with fixed signals. Communications Based signalling opens the way to performing the same task (cancellation of an authority) in a more comprehensive, safer and more elegant way.

The train could simply communicate acceptance ("I can stop in time") allowing the authority to be released immediately without waiting for the train to stop. Alternately the train could counter-offer with "brakes are on. I will definitely stop ahead of revised "Protection Point X", but not Point "Y". The infrastructure could then hold the authority in place up to position X and release beyond. This is flexibility beyond what our interlockings do today.

This Contract model for applying and varying movement authorities is in principle very flexible. Those familiar with train orders will know that the types of orders which, if enforced, can be safely issued, are many and varied and can allow huge flexibility of train movement. This contract negotiation model can open the way to that world using Communication Based signalling with all its safety layers. Splits, joins, reverses, stop shorts and temporary speed restrictions can in principle all be safely negotiated using the movement authority mechanism.

Clearly this is not a facility available to purchase today and exploring the detail further is beyond the scope of this overview paper.

The point is that an interlocking for a Communication Based system need not necessarily simply attempt to reproduce the rules associated with the existing fixed signal interlockings. To do that would be to forego the technology dividend associated with the new technologies.

I saw Communication Based Signalling described in a recent magazine article as Train Orders on steroids [2]; which can be true, but only if we develop the principles which will make it true.

4.5 Insurance

Any contract can include a component for insurance and so it is for railway signalling. The insurance component for signalling is found in such items as the overlap and flank conditions.

As with all insurance, policies can be complex, there is scope for endless discussions about what should be covered and what should not, and a variety of levels of excess can be considered. Policies are purchased at a price. In the case of signalling the price is denominated in real estate (physical space to fit the overlaps in), capacity (reduction of) and reliability (reduction of).

In signalling the design of insurance components can dominate the design effort. Overlap complexity can easily become the major contributor to interlocking complexity, occupies a major piece of the designer's thoughts, as well as the thoughts of those responsible for defining local standards and interlocking practices.

Interlocking designers tend to become insurance brokers.

But insurance and authority are separate things. If I take out fire insurance on my house, it does not give a person authority to light a fire underneath it or keep them from going to jail if they do. A train does not have authority to enter areas denominated as part of the insurance cover.

On the other hand, managing the insurance cover for the different levels of authority (particularly where authority change occurs for a particular change) is an important, if large, topic, beyond the scope of this overview paper.

The option opened by new generation signalling is that of self-insurance; ie the train can manage the insurance itself within the allocated authority area.

4.6 Memory and startup

I mentioned earlier that the vehicle must remember the authority once issued, a requirement sometimes found problematic with fixed signalling after the signal is passed.

A special problem for vital persistence in on board systems occurs for re-start events. The possibility of the vehicle forgetting its current authority during a vehicle re-start can be dealt with by allowing the vehicle to ask the infrastructure what its current authority is. If no response, it assumes none.

In many modern systems, this facility is used to request state information at a rate of once a second or more, regardless of a start-up event occurring.

It can be argued that the infrastructure systems can also re-start and the vehicle might not know. The vital persistence needs for the infrastructure system are in the route setting requirements. The appropriate authorities can be derived with the route states known.

The requirement to detect all trains at infrastructure re-start provides the basis for allowing existing authorities to safely complete. In modern systems where the infrastructure is required to communicate with the train at start-up, the management of existing authorities is enhanced as the train can acknowledge the start-up without any need for the continuous requesting mentioned above.

5 PROTECTION

Looking at the generic systems framework (figure 1), it can be seen that another function offered within the signalling system is that of protection (section 9). This comes in many guises. In the case of fixed signalling a signal can be but back to red as a "fail safe" in case of a fault or an obstruction ahead.

This functionality is quite different from the cancellation of authority functionality described earlier.

Although the signal may be red, the authority is retained in place by the infrastructure. The route will remain locked and, if the fault or the obstruction is removed the proceed aspect will again be displayed to the driver.

Protection systems can bypass the interlocking. For instance:

- The Earthquake Detection and response system for the Japanese Shinkansen uses the Overhead system, not the signal authority system, to protect its trains. If an earthquake above threshold level is detected in the substation the overhead is turned off. The train understands to interpret this event as an earthquake and applies its brakes. This well tested system operates on a weekly basis in Japan;
- ETCS can use the vital communication link between train and RBC in a similar way, although not specifically to detect earthquakes. In ERTMS level 2, if the communications link is dropped the train immediately stops. Whether this is a desirable use of the communications link will be discussed later in this paper.
- Aircraft collision avoidance famously operates independently of Air Traffic Control [Mid air collision near Swiss air space in 2002 [3] for example. Pilots needed to understand that the protection system was intended to take priority over an authority from air traffic control].

Putting the signal to stop ahead of the train is often just the “least worst” option available for providing protection. Fixed signalling protection systems are intermittent by their nature, not continuous. A train can only respond to a stop signal if that signal is visible. No protection is offered if an obstruction appears in a section the train has already entered, and the opportunity to stop safely can be lost in the time between the signal being put back to stop and the signal coming into view for the driver.

The red flag waved by the signaller or other person from the signal box is also available as supplementary protection, though this requires provision of a signaller and a signal box to be effective. With new generation signalling, we tend to forego this layer as we remove the local signal boxes.

6 ERTMS AND CONTEXT

6.1 Is ERTMS level 1 a signalling system?

ERTMS level 1 offers train authorities to trains as they travel over ballises. The authorities thus provided are intermittent in nature. This intermittency can be mitigated partly by providing additional infill ballises, but the problem remains that the infrastructure cannot communicate with the train when it is stopped.

When the train is stopped within sighting of a signal, it cannot be offered a proceed aspect. This flaw, which is often overcome by leaving the fixed signals in place (restoring the ability to communicate with the driver) leads many to conclude that ERTMS level 1 is little more than a sophisticated ATP system (The “authority issue” function sits with the retained fixed signals).

In this view, ERTMS level 1 is not a signalling system, it may be characterised as part of the enforcement or protection systems instead.

But there is another way of looking at it. The requirements gap between ERTMS level 1 and a complete signalling system is small. It could be closed by any method of communication between the infrastructure and the train which is effective whilst the train is in sighting distance of the stop signal. For instance:

- Manual orders (delivered by hand)
- A 200m induction loop
- Limited coverage radio or wi-fi (again just 200m coverage needed)

With any of these measures in place the fixed signals can be removed leaving a fully functioning signalling system. This approach was that actually taken by some earlier generation TBTC systems to achieve the same end.

6.2 Fixed signals – what else do they do?

Before fixed signals can be replaced by another technology, we must also have an understanding of the other functions they perform apart from offering authorities. This section summarises those functions. Fixed signals perform three functions:

- Regulate train movements through interlocked areas

  Controlled signals are provided which require routes to be set before proceed aspects are given. This is the core functionality for the authority issuing processes discussed in this paper (refer figure 2 for diagram).

  The signal opens a gate which will allow a single train through the interlocking, generally closing it behind. The “stick” function ensures that the gate closes behind the train to keep the working of the interlocking flexible. This functionality is discussed in more depth later.

- Regulate spacing between trains

  Automatic signals are provided with no route associated with them.

  The operation of the signal is purely according to the location of the train ahead.

  These signals can be viewed as operating according to an “open road” concept (discussed in more depth below) for in that their primary reason for being is to prevent head to tail collisions between following trains (protection function).

Note that absolute signals at interlockings have this extra function bundled into their aspect sequence.

- Inform trains of failure

  Signals revert to red to inform and stop trains (though not with 100% reliability, as noted above) in case of certain failures in infrastructure.

  As noted above, the display of a stop aspect under this functionality is different from the removal of authority. Looked at from the infrastructure side, no process for route normalisation (the proper authority removal process) is initiated. If the fault is cleared the signal will return to displaying its proceed aspect (in general).

The fact that there is a single device for offering authority obscures the fact that there are three separable functions being performed.

6.3 In Cab Signalling functions

With modern signalling systems the functionality formerly provided by the fixed signal can be viewed as having been brought into the cab and made more versatile.

The three functions provided by the fixed signal can, in principle, be separated to the extent that they can be provided by three separate physical systems and discipline areas.

6.4 Commenting on each functional area:

6.4.1 Authorising movements in interlocked areas

This remains largely the domain of traditional interlockings. Within ERTMS this function should be the job of the Euro interlocking, a development not yet mature. Thus the principles for interlockings typically are the same ones found in 1960s and 1970s.

The capacity advances opened up by ETCS and CBTC are constrained by some of the characteristics of traditional interlockings, particularly their insurance components.

Currently the authority point protecting a junction is where the fixed signal would have been with the standard "one size fits all" insurance in place (ie one overlap for all classes of train) and routes which are relatively long.

A paper at Lyon in 2014 presented work Alstom [4] was doing at that time on next generation interlocking concepts, but this is beyond the scope of this overview paper.

6.4.2 Regulating spacing between trains

This is the task that both ETCS and CBTC have taken on at their core. It is the feature they are now selling to the world.

The schemes specified in Victoria tend to specify that all the points need to be removed. This removes the exposure to dot point 1 (above), allowing improved capacity without the need to face the challenges in that regard recognised in current interlocked areas.

Better options exist, which can allow interlocked areas to deliver the same capacity dividends as, say, CBTC. But these options are in the future, or in Japan.

Observation of Japanese infrastructure (Japan is the home of Group Running as well as running the highest capacity, safest and most reliable high speed network in the world) shows that points and operational flexibility do not need to be sacrificed to achieve high capacity or reliability.

Typical Shinkansen minor station

The question as to whether the protection against head to tail collisions a matter to be dealt with in the issuing of authorities, or another way, is discussed in the next section.

6.4.3 • Informing trains of failure

Fixed signals have been seen as convenient for informing trains of failures. As a system they have a number of deficiencies:

- It only works when the signal is visible.
  - Ineffective if the failure occurs when the train is in the section;
  - Action delayed ahead of signal till signal is sighted;
  - Taking account of the initial delay the train cannot always stop ahead of the signal;
  - Infrastructure faults and blocking trains can become confused in the driver's mind (vis Glenbrook accident [5])

Given these deficiencies, the failure cases are better treated on a case-by-case basis. Generally better alternatives can be found than simply stopping the train as a one size fits all approach.

As mentioned previously, more effective systems of fault detection and response are often available which bypass the interlocking entirely.

7 OPEN ROAD AND CLOSED ROAD CONCEPTS

In one paper written last century, the essential difference between US and UK train management was said to be that whilst UK signalling assumes a “closed road” concept, US assumes an “open road”.

An open road is similar to a highway (actually more like a turnpike, since the entry to the road is always gated and permission is required). The presumption is that, once entry is granted (at the gate) all comers can travel on the road freely unless told otherwise. Behaviour whilst on the road is governed by rules and instructions. More than one vehicle at a time may be on an open road, though separated to avoid head to tail collisions.

This approach may be contrasted with a closed road which is more like a gated path (or perhaps more like interlocked turnstiles). The presumption is that no-one can enter without permission, and then only one at a time. The conditions of entry cover the permitted behaviours whilst on the road.

Roads of either type can be provided utilising combinations of two authority types:

7.1 Single train authority

The most basic authority is the single train authority.

Under this system of authority the authority associated with a route is offered to a single train only. An authority is always provided only to a single train at a time (basic condition of contract – identified parties). Under this system, some infrastructure behind the train (the route) must be released and normalised after the train before an offer can be made (or route set) for a following train. A specialised function – the Stick – is provided to enforce this “one train at a time” separation.

7.2 Multiple train authority

This type of train authority differs from the single train authority in that no stick function operates.

A single authority is still offered to a single train at a time, but the gate is effectively tied open for the next train following. Since the stick does not operate, the route is not required to normalise between trains. The route can remain set and the trains can fleet through the section as though it were an automatic section.

Swapping between one authority type and the other in SSI is provided by the “set auto” functionality.

7.3 ABSOLUTE PERMISSIVE BLOCK

These concepts are elegantly combined in our method of Single Line working known as APB (Absolute Permissive Block – An American system as it turns out) where the authority is for one direction at a time (rather than one train).

In this system the single line section is gated at both ends (no entry without permission) until a train arrives to establish permission for one direction. Once that permission is given the single line section becomes more like an open road in that direction. Trains may be fleeted through with multiple trains permitted in the section (Permissive Block). Train separation is ensured by automatic signals through the section which are there just for train regulation.

Train movements in the opposing direction remain blocked (Absolute Block) till the section is completely clear and permission for the changed direction of movement obtained.

7.4 The train protection paradigm

As we look at the generic systems framework, we can ask which area is the one where we consider protecting against nose to tail collisions between vehicles. It can be done by issuing authorities (section 3) using closed road concepts, or by protection systems (section 9) using open road concepts.

For us, the traditional answer with fixed signals has been to place this functionality within the “Issue Authority” system functionality. This approach is carried forward into ERTMS level 2 where authorities are issued for line sections as short as 200m. You can imagine a long section of high speed line populated with a forest of gates opening at the rate of around one every 4 seconds ahead of the train and then closing behind (refer figure 3).

If you look at modern CBTC systems (and even ETCS), you will see that this split between traditional interlocking (issue authority) and collision protection (issue authority or protect train) has already occurred in the architecture and can be seen in systems being delivered today.

The shift from the “authorised speed” towards the “distance to go” approach opens the way to the new paradigm of long authority sections between junctions and managed by the interlockings. Within these long sections collision avoidance can in principle be treated as a train protection issue managed separately and with protocols different from those used for movement authorities.

A sniff of where this leads can be found in the February 2016 issue of IRSE News where we find a discussion of convoy mode in “ERTMS level 4” [6].

7.5 Why do we have warning signals (yellow ones)?

Looking at road vehicles operating on roads it can be seen that each vehicle maintains a safe distance between itself and the vehicle ahead without the use of signals (apart from tail lights, brake lights and turn indicators, but that is another story).

This is possible because a road vehicle can brake well enough to allow the driver to see an obstruction, react, apply the brakes and stop reliably within the distance reliably visible to the driver of the vehicle.

The reason trains are not able to do the same thing is that a train cannot in general stop from line speed within the distance reliably visible to the driver of the train. In other words, the train must commence to brake before the obstruction becomes visible.

Yellow (or distant) signals were developed to provide the facility to tell the driver to commence braking at a point where no obstruction is yet visible.

7.6 Fast forward to today.

With “distance to go” information brought into the cab and the on board EVC (Electronic Vital Computer), the problem of not being able to “see” the train or the obstruction ahead disappears. The train can reliably stop at the target stop point without the need for intermediate warning signals or (in principle) extra insurance.

Taking it a step further, if the train reliably knows [within its EVC] its own location, the geography it is in and its own braking characteristics (section 6 & 7 within the framework, outside scope for this paper), it can perform these functions quite efficiently with only basic (or no) information provided by any interlocking.

If the stopping point involved is just the train ahead, the interlocking may not need to be involved at all. If the train can obtain this information (the location of the rear of the train ahead in this case) directly (eg from the train itself or an RBC or RATP system), the whole collision avoidance functionality can be largely decentralised and not involve the interlocking (authority issue) functionality at all. The protocols for trains following trains start to look a bit more like those for cars following cars (something we are all familiar with). The need to issue or cancel movement authorities to support the functionality never arises.

The provision of the RBC in ETCS already sets us partway down that road, although you will notice it still issues movement authorities, some of which cannot be cancelled. So there is an element of the future about all this.

We are now ready to address the second question posed at the introduction. Is continuous communications required?

7.7 ERTMS level 2 continuous communications

It is a requirement of ERTMS level 2 that continuous radio communications be provided between the RBC and the train.

This decision to require continuous communication was controversial in the original design (it has even dropped out from some level 3 implementations, such as Bombardier’s ERTMS Regional). With the framework in mind, we can discuss the issues around it.

Authorities provided by the RBC authorize the train to travel at a target speed over a section. The block lengths can vary (200m quoted for Swiss tunnel (high speed network) up to the same as a traditional fixed signal section) and the authority can include permanent and temporary speed restrictions (refer section 7 of the framework).

A train travelling at 180kph covers 200m every 4s. Higher speed trains may cover the distance faster so that the train may need to accept a new authority perhaps once every couple of seconds. The communications are continuous and there is no need for the train to acknowledge each authority.

But this is a different type of authority to the type discussed in the earlier sections – it does not require acknowledgement or even agreement between parties. Looked at in the framework of contract, the infrastructure has nothing to “offer” since the object of the agreement is the position of the train ahead which the infrastructure does not control (in most cases the section does not even contain points). The train is just progressing forwards under its own steam as it were. For the infrastructure to try to cancel such an authority is equally meaningless (auto signals do not have buttons to pull).

Looking at the safety consequences of lost communication through the process:

- If the train does not receive the offer, it will simply stop at the last received protection point. This is safe. If communications are subsequently restored and a new offer from further down the track is received, it is same to adjust the protection point accordingly (ignoring the non receipt of he “authorities” in between);
- As with fixed signals, there is no acknowledgement from the train. Acceptance is performance based as discussed;
- The infrastructure cannot cancel the authority, so communications are not required for this purpose;

In the event, if communications is lost the train regards this as a fault and must stop. But this behaviour is an artefact of the ETCS specification, it is not related to any fundamental need for the issuing of authorities. If any functionality is being supported by this continuous communications requirement, it is one of “fault detection” as discussed earlier. The question is then “what are the faults worth stopping the train for?”

It is also quite a communications intensive process. A current hot topic for ETCS is how to ensure sufficient bandwidth to support the specification’s requirement in heavily trafficked areas. Is it sensible to set up a communications based fault detection system where the most likely element to fail is the communications doing the detecting?

The alternative is intermittent communication between RBC and train. This approach is adopted in Bombardier’s ERTMS Regional which features intermittent communications. It sells itself as ERTMS level 3.

For ERTMS level 2, we can imagine the comms driving a process where virtual gates are opened and closed every couple of seconds for multiple high speed trains. This need is driven by the concept of authorities as closed road single train artifacts. As a process, it is hard to manage without high quality highly reliable communications infrastructure.

For ERTMS level 3 we can hope that "moving block" is really code for moving the whole function into the "protection" functionality where it arguably belongs. Here the communications requirement is less onerous because the quantity of communications needed is less, it is amenable to decentralisation and intermittency in communications can be tolerated.

8 CBTC

CBTC is proprietary (though generally self constrained to supersets of IEEE 1474 and/or IEC 62290), so I can be free to select a basket of typical characteristics between systems.

CBTC features use of a "distance to go" concept rather than a "target speed" concept. Some systems add an event horizon in case the actual "distance to go" is so far away as to be considered quite distant. "No need to stop within event horizon" is perhaps this concept.

Having said that, the "RBC" equivalent must generally manage the permanent and temporary speed restrictions, so target speeds for various sections must still be provided to the train. Intermittency in communications is generally tolerated in CBTC systems.

8.1 An alternate approach

The dominant paradigm for many current systems treats train "on board intelligence" as a scarce resource and "communications capacity" as a non-constrained resource.

This seems to work ok for practical CBTC systems (generally simple layouts) but maybe not so well for ETCS systems in higher traffic areas.

There is an alternative where "on board intelligence" (the on board EVC with the on board portion of the ATP) is treated as a less constrained resource and communications capacity treated as more constrained.

The Japanese Shinkansen provides a model for this. In the latest train versions, a configurable infrastructure model (including speed restrictions, topography and key infrastructure points) are stored in the on board EVC. Added to this, the EVC (as is common for modern on board ATP systems) understands the train's own braking curves. Thus the train is able to calculate its own braking curve for each location and calculate when to make a brake application to stop at a target point with extremely high reliability.

The infrastructure provides distance to go information to support the train as well as traditional speed steps (a hybrid for the brown fields situation), but temporary and permanent speed restrictions, station stops etc. can be managed by the train alone.

8.2 The future

Taking this concept a step further, the "distance to go" information for train separation can in principle be split from that required to protect junctions etc and be treated separately.

In this paradigm movement authorities can be issued for long sections between infrastructure protection points on open road concepts without reference to the position of any train ahead. Authority issue only needs to occur in the vicinity of interlockings.

In this paradigm, movement authorities are not used to separate following trains, train protection systems are. Such a protection system can operate "train to train" directly, or be mediated by infrastructure such as the RBC or RATP (or a combination of the two approaches) separate from the interlocking functionality. Train separation through interlocked areas can be managed in the same way.

In this world, the capacity benefits of the new communications based approaches move into the interlockings. "Signal to signal" concepts are replaced by "resource by resource" concepts and overlaps swapped out for self managed insurance provided by the train (refer figure 4).

Backward compatibility for non fitted trains can also be supported in the same way Windows will still allow old DOS programs to run (within limits).

The role of the signal engineer also evolves due to the need to take a much more active interest in the train-board equipment, particularly the potentially "smarter" on-board ATP systems (though not perhaps the multitude of other "smart" functions (eg those to do with doors) found in CBTC specs).

A time may come where the ATP component software on board the train is as important to signalling as the part which sits in the RBC, the RATP, or the interlocking on the ground.

With a sound understanding the fundamentals of movement authorities – what they are for, what needs to be provided by one and what does not need to be provided by one, we can move towards that future with confidence.

9 CONCLUSION

ETCS and CBTC have become popular vehicles over the past decade for implementation of new signalling schemes.

By understanding the fundamentals of Movement Authorities it can be seen that the strengths of these two technologies lie in functionality which, although it is a replacement for fixed signals, is better characterised as "system protection" than "movement authority".

Recognising the differences between these two types of functions can help ETCS and CBTC better reach their potential with more cost efficient and higher capacity implementations.

It can be seen on the other hand that the Interlocking, where Movement Authorities are bread and butter functionality, are still waiting to catch the wave of advancing technology.

Many schemes treat Interlockings as Brownfield components scheduled to retain the legacy standards and functions appropriate for an earlier generation of relay technology.

By understanding the fundamentals of Movement Authorities, Signal Engineers can develop and apply the interlocking rules appropriate for integrating with a modern CBTC or ETCS scheme in the new generation of interlockings.

10 REFERENCES

[1] Corrie, J.D. 2004, *Proceedings, Institution of Railway Signal Engineers*, Presidential Address: 'I Bear a Far Shining Sign", 2004-2005, pp 20 – 26;

[2] Lindsey, Ron 2012, 'Traffic Control: Science or Art?', *Railway Age*, Sept 2012, pp 42-49; The actual quotation is: "*moving-block is actually dark territory on virtual steroids*"

[3] German Federal Bureau of Aircraft Accidents 2004, Investigation Report 02 AX001-1-2/02 May 2004, Ueberlingen;

[4] Denis, Guillaume 2014, *Alstom presentation*, *IRSE Convention – Lyon – June 2014,* 'URBALIS Fluence, Train-centric CBTC solution';

[5] McInerney, Peter Aloysius 2000, *Special Commission of Inquiry into the Glenbrook Rail Accident*; *Interim Report*, June 2000;

[6] Mitchell, Ian 2016, 'ERTMS Level 4, Train Convoys or Virtual Coupling', *IRSE News*, Issue 219 Feb 2016, pp14-15.

AUTHOR

Peter Burns

Peter Graduated in Applied Science (Electronics) from Melbourne University in 1981.

He commenced as Professional Engineer with Victorian Railways that year in Signal Design., then progressed through various roles in Signal Design, Test and Development, Maintenance Management. He completed his MBA at Monash University in 1991.

Peter left PTC in 1994 to work for companies such as Alstom and Bombardier in places such as Sydney, Melbourne and Copenhagen.

In 2003, he joined with 2 colleagues to form *Rail Networks*, providing assistance over a number of years to Government as well as to Connex, the then franchisee for the Melbourne rail network.

Peter is currently director of the small consultancy firm "*PYB Consulting*" and as Chartered Engineer fills wide and various roles in rail organisations when they seek assistance. Recent projects have included RRLCMR and Signal Advisor to teams bidding various Melbourne Grade Separation projects including the successful Burke/North/McKinnon/Centre Rd bid.

Figure 1 - Generic Systems Framework

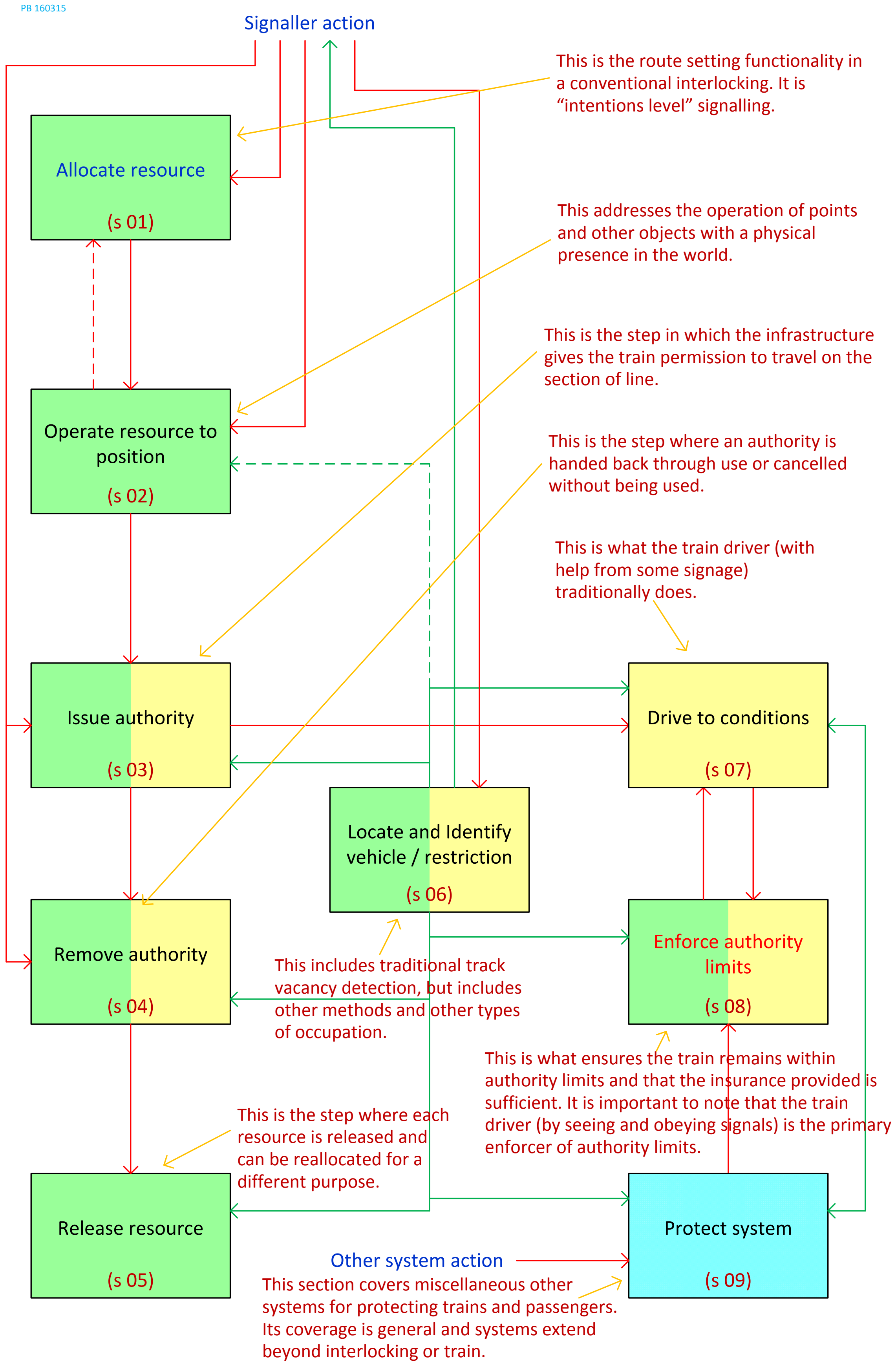

Figure 2: Fixed signal sections

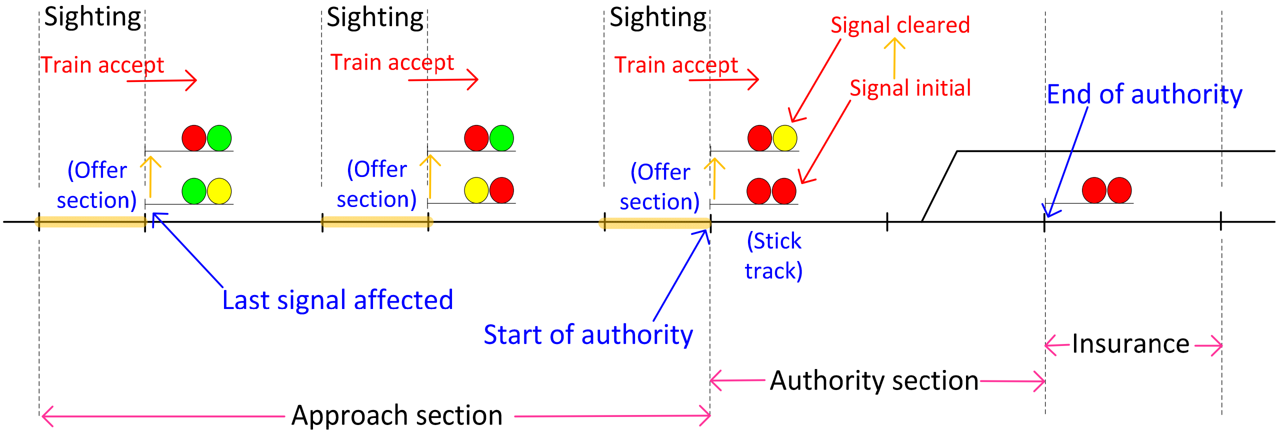

Figure 3: Closed road ETCS

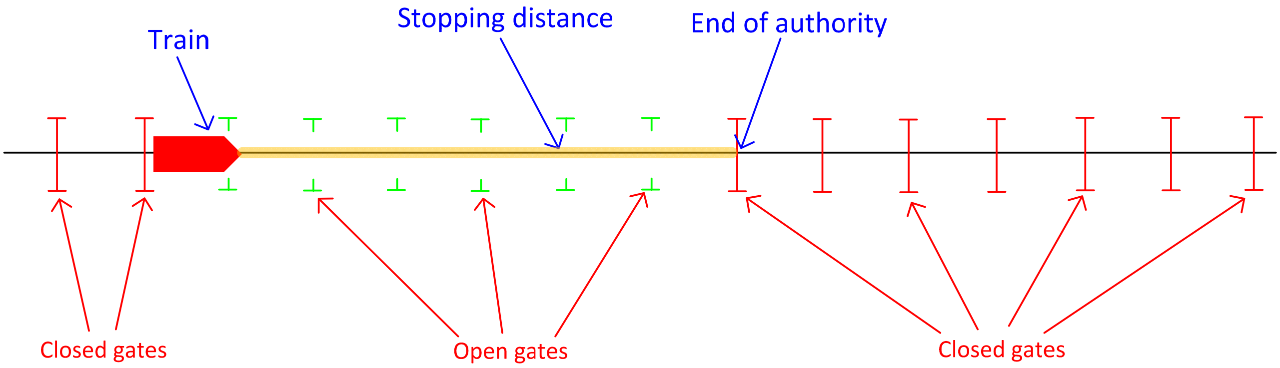

Figure 4: Open road concept