



INDEPENDENT  
TRANSPORT  
SAFETY AND  
RELIABILITY  
REGULATOR

## **INFORMATION PAPER**

# **DRIVER SAFETY SYSTEMS AND AUTOMATIC TRAIN PROTECTION**

# Table of Contents

1. The purpose of this paper.....	3
2. Suggested terminology base.....	3
3. Driver safety systems.....	8
4. The border between driver safety systems and ATP system.....	8
5. A brief history of ATP.....	9
6. The current status of ATP development.....	12
7. The Perth system as an ATP example.....	15
8. Key major functional differences amongst some existent or planned systems.....	16
9. Potential negatives of full ATP systems.....	19
10. Safety integrity levels.....	21
11. Criteria for considering the installation of an ATP system.....	21
12. Appendix 1 – an example of a discussion paper about ATP Criteria.....	23

## **The purpose of this paper**

The Waterfall Inquiry, as well as broader safety management processes, have highlighted the need for the NSW rail system to consider its position on Driver Safety Systems and Automatic Train Protection (ATP). Moreover some work is being done – especially the “Advanced Train Management System” under development by ARTC and Lockheed Martin – which will potentially create new systems which may prevail only on particular segments of the Australian rail network. This paper is therefore intended to provide a perspective of the issues involved, without attempting to draw conclusions as to optimal paths to adopt.

The paper has been prepared by the NSW Independent Transport Safety and Reliability Regulator (ITSRR) for information purposes only. Inasmuch as it does include commentary or implied evaluations of any matters, it is not intended to represent a policy position by ITSRR, but is purely to provide a basis for discussion.

The aim has been to provide an outline in a style suitable for non-specialists in the area.

## **Suggested terminology base**

The whole area of driver safety systems and automatic train protection is greatly hindered by extreme variations in terminology, and by the same terms sometimes being used to mean different things in different jurisdictions. It should be noted, particularly, that the Waterfall Commission of Inquiry report uses substantially non-standard terminology which has led to some potential for misinterpretation.

It is therefore important to select a single terminology base at the outset of this paper. The definitions chosen herein would not necessarily be universally accepted (because there ARE no universally accepted definitions and are not likely to be) but they seek to establish internal consistency. They will be used as far as practicable by the NSW Independent Transport Safety and Reliability Regulator (ITSRR) in its communications with the industry.

### *Driver safety systems (DSS)*

Systems directed to confirming the capacity of the driver to continue to operate the train, and to stop the train if this capacity is not confirmed. These directly include deadman systems, vigilance control devices and physiological vigilance measurement devices; various other systems such as AWS may also be substantially regarded as driver safety systems.

There are also various systems which may contribute to driver safety without having the potential to intervene, but such systems are excluded from this definition.

### *Automatic train protection (ATP) systems*

Systems which provide security against inappropriate movements of a train if the driver is not managing the train appropriately. Notionally, this security may be anywhere from a very low level to a very high level, but the term “ATP” in isolation is normally applied to systems at the upper end of the range of control, in particular to what are described as “full” ATP systems below.

#### *“Full” ATP systems*

Systems which prevent a train from proceeding at greater than the authorised speed at any location (where “authorised speed” takes account of track limitations, the location of other trains, and any other constraints externally imposed).<sup>1</sup>

Such systems are predictive rather than reactive, in that they receive, transmit and analyse information in such a way as to act sufficiently early to prevent an unsatisfactory outcome from arising.

#### *“Partial” ATP systems*

Systems which comply with the basic ATP definition but do not meet all the criteria, e.g. those which will stop a train if it passes a signal at stop but will not control its speed under other circumstances.

### *Automatic train control*

This term nowadays refers to systems in which the complete operation of the train (except maybe doors) is controlled automatically, i.e. without driver involvement. (However in the past the term ATC has been used in other contexts, e.g. interchangeably with ATP, and is still so used in some places.)

ATC systems are not discussed in this paper, because there appears to be no suggestion that any railway in NSW (nor probably elsewhere in Australia) will head to ATC in the foreseeable future. ATC is also commonly referred to as ATO (Automatic Train Operation).

### *In-cab signalling systems*

[commonly referred to simply as “cab signalling”]

Systems which give an indication of movement authority to the driver within the cab, without relying on him/her observing signals fixed outside the train. Such systems are commonly but not inevitably associated with some level of ATP.

---

<sup>1</sup> This may be viewed in terms of controlling both overspeed and SPAD risk in cases where the concept of “SPAD” is meaningful; but the concept is more general, e.g. an ATP system will often have no signals.

### *Fixed block*

Systems where the track is divided into fixed segments called “blocks”, in which trains are separated on the principle that there cannot be more than one train in each such block.

### *Moving block*

Systems where trains are separated by a certain distance (depending on speed and geography) but where there is no fixed block system – for instance such a system might allow two trains to move continuously with a constant separation of (say) 3 km at 150 km/h.

### *Continuous-update ATP*

An ATP system which interacts continuously between the train and the external environment, so that at any given time, the train can become aware of any change in external circumstances. For instance, under a continuous ATP system, the clearance of a signal ahead will immediately be notified to the train.

### *Intermittent-update ATP*

ATP systems where the external information provided to the train is only provided at specific locations (typically by transponders or “balises” fixed on the track), so that if the train is between those locations, it cannot receive any update of changed circumstances. So, for instance, a signal changing from stop to a proceed indication may not be immediately notified to the train, and so until the train reaches the next balise, it must continue to move as though the signal were still at stop.

### *Braking curve*

Full (and some partial) ATP systems need to know how fast a train can be travelling at a certain point if it is to conform with a lower speed at some subsequent point. For instance, if the train is required to be under 40 km/h in 500m, it obviously should not be doing 150 km/h now, because it won't be able to reduce speed sufficiently in that distance. A braking curve (so described because it is readily represented as a curve on a graph) gives the indication of the deceleration patterns which are practicable. A full ATP system will ensure that the train travels within the constraints of the relevant braking curves at all times; e.g. it will warn the driver if the train is travelling too fast for the desirable braking curve, and will actively intervene if the train is at threat of conflict with a limiting braking curve.

### *Intervention*

Driver safety systems and ATP systems are said to “intervene” if they take over control of the train from the driver, typically by de-applying power and applying braking.

### *Deadman systems*

Are typically a form of driver safety system which detects a continuous input from the driver, e.g. by application of force to a pedal or handle. Many such systems can only detect forms of incapacitation which cause that force to be released; so that sleep and inattention may well not be detected by deadman devices. The Waterfall accident demonstrated that even major incapacitation may still not result in the continuous input being released. Some deadman systems require periodic release and reapplication of the pressure, to avoid continuous pressure being accepted as legitimate input from an inert person or from a physical object inserted (deliberately or otherwise) to provide continuous pressure.

### *Vigilance control devices*

Typically require intermittent inputs from the driver, e.g. button pressing at fixed intervals or in response to flashing lights or sounds.<sup>2</sup> Modern vigilance systems are often task-linked, i.e. they accept direct evidence that the driver is actively controlling the train. To varying degrees, vigilance devices are little more than intermittent deadman devices, in that “automaton responses” can occur without conscious input by the driver – so someone who is asleep may still respond, someone who is dead will not. Various techniques (such as linking the frequency of required inputs to the speed of the train) may be applied to reduce the distance the train may travel before a non-response is detected and acted upon.

### *Physiological vigilance measurement devices*

Devices which will assess physiological measures of alertness, and will take action if that alertness is determined to be inadequate. For instance, they may act to warn the driver if alertness is detected to be inadequate, and may intervene to stop the train if that warning is not acted upon. Several such systems are currently under active review.

### *Train stop systems*

These may be mechanical or electronic, and serve to stop a train if it has passed a signal at stop. They may also be used in conjunction with timers, to stop a train if it is exceeding a designated speed. If such train stops are not adjacent to signals, they are *intermediate* train stops.

### *Conditional clearing*

A process wherein signals and/or train stops do not release to allow the passage of a train until that train has been verified to be moving sufficiently slowly.

---

<sup>2</sup> It is a matter of definition rather than substance as to whether a deadman device requiring intermittent release should be classified as a “vigilance control device”.

## *SPAD*

Signal passed at danger, i.e. when a train proceeds past a signal even though the signal instructed the train to stop.

## *AWS-type systems*

These require a response from a driver if a warning is given, e.g. at a signal showing a restrictive indication or on the approach to a speed restriction. If the driver does not acknowledge the warning in a few seconds, the brakes are applied.

## *TPWS*

Is a British system, combining AWS with what is effectively an electronic train stop system if a train passes a signal at stop, and with an intermediate train stop facility which will stop the train if it is going too fast at a designated point approaching a signal at stop or approaching some other hazard.<sup>3</sup>

Enhancements such as TPWS+ (for higher speeds) also exist but follow the same essential principles.

## *ERTMS/ETCS*

These acronyms refer to the European standard ATP systems under development.<sup>4</sup> While the lower-level versions of these may not have significantly greater functionality than ATP systems in use around the world for some time, they are being designed to have compatibility features which should establish interchangeability of different manufacturers' products and enable cross-borders utilisation (historically, ATP systems have been strongly proprietary). Higher level ERTMS/ETCS systems will involve use of modern remote communication technologies and will provide moving block capability.

## *Balises*

Objects placed between or adjacent to tracks, to convey information between external sources and a train. They may be passive (simply used to provide constant information, e.g. that there is a curve ahead with a fixed speed restriction) or active (providing or receiving information dynamically). Also known in some contexts as transponders.

---

<sup>3</sup> Systems such as TPWS and conditionally-cleared train stops may be regarded as examples of “speed traps”, which respond to excessive speed of a train at specific locations as distinct from monitoring train speed continuously.

<sup>4</sup> ERTMS = European Rail Traffic Management System. ETCS = European Train Control System, which constitutes the signalling element (including ATP) of ERTMS.

## **Driver safety systems**

As defined above, driver safety systems such as deadman devices and vigilance control devices have largely reached the limits of their potential, until such time as they can be designed to provide genuine assessments of *alertness/awareness* as distinct from mere robotic compliance.

The task-linked vigilance systems installed on RailCorp trains are reasonably state-of-the-art within their frames of reference, but should not be thought to have qualities which they do not have (in particular, the use of the term “vigilance device” should not be interpreted to indicate that the driver is “vigilant” in the common usage of that word; the driver could be effectively asleep or otherwise unaware of his environment, and yet still be accepted as responding to the device).

It would be possible to design devices which require less robotic responses than button pressing – for instance the driver could be made to select the correct response from some set of responses, either by mechanical actions such as choosing the correct button or by vocal response (uttering the correct word). However such techniques run risks of distracting the driver.

In this category, the future may lie in physiological vigilance devices, which can in principle detect states of sleep and physiological inattention. An example is the device colloquially known as a “Russian wrist watch” which is a Russian-invented device which can be worn on the wrist and purports to detect such matters. Although previous attempts to design such devices for the trucking industry have been less than successful, it is not yet clear how successful the Russian device will be. A number of alternatives are being developed.

It should be mentioned for completeness that some reminder devices are sometimes included under the category of “driver safety systems”, although they do not fall within our definition. A prime example is the British “Driver Reminder Appliance” (DRA) which is manually set by a driver when standing at a station with a signal at stop ahead. Setting the DRA prevents the train from powering until the appliance is reset. While this may or may not be a worthy device (some significantly adverse implications exist) it should be regarded as a SPAD-prevention initiative rather than a driver safety device in the context of this paper.

## **The border between driver safety systems and ATP systems**

Labels are often attached to types of systems as though these are inherently separate. This is not the case; indeed there is close to a continuum of system types. For instance the distinction between deadman devices and vigilance control systems is very tenuous, the border between DSS and ATP is also tenuous, and ATP systems can be anywhere between very primitive and very comprehensive.

The DSS/ATP border is worth considering specifically by reference to the British AWS system (also applied in Adelaide and Brisbane, and existing in a variety of similar-functionality types under various names around the world).

AWS is based on devices in the track, typically a short distance before a signal or major speed restriction. If the signal is showing a full clear indication, a bell sounds in the driver's cab and free running is permitted. If however there is a restrictive status ahead, a buzzer or horn sounds, and the driver has a short period of time to acknowledge that sound by physical action. If he/she does not do so, the brakes will apply.<sup>5</sup>

This device is clearly a driver safety system in a general sense, in that if the driver is dead or incapacitated, the brakes will ultimately be applied – but perhaps not for a very long time, in that there may be a long distance between AWS locations. It is equally clearly not an effective vigilance system in many respects, in that (even apart from the problem of the distance between activations) drivers hear the buzzer/horn so often that they tend to react subliminally. However it does have one element of an ATP system, in that it is related to the state of the track ahead, and will not seek to intervene unless there is some restricted situation.

An important point is that it is not a true ATP system, in that once the driver responds to the warning, the system will then not further intervene. A true ATP system will provide some ongoing monitoring to ensure that the driver does not initially respond favourably but then ignore the situation thereafter, and will also protect against the possibility that the driver's initial response was without adequate awareness of the intent of the warning.<sup>6</sup> In other words, a true ATP system will seek to protect the train irrespective of what the driver does or does not do.

### **A brief history of ATP**

Notwithstanding the previous paragraph, the history of ATP cannot be viewed in isolation from that of Driver Safety Systems, and indeed in the first half of the 20<sup>th</sup> century, systems generally analogous to AWS were often thought of as the best that could be achieved. Some ATP systems indeed evolved by progressive development from AWS-style origins.

American accident reports from the first decades of the 20th century featured many incidents of both SPADs and excessive speeds on curves. In-cab devices were therefore developed early in the USA, initially based on AWS-

---

<sup>5</sup> British AWS is a two-state system, i.e. warning or no warning. This means that the driver will get the same warning for any restrictive signal indication, regardless of whether it is a stop indication or something only mildly restrictive. Some non-British AWS systems have a higher level of differentiation between levels of warning, for instance the Irish CAWS system.

<sup>6</sup> There have been numerous accidents, some very serious, in which the driver correctly received an AWS warning and made the required physical response, but failed to act on the warning – either because the warnings occurred so frequently that responses became automatic without conscious thought, or because the driver interpreted the warnings incorrectly.

type philosophies where the driver had to acknowledge an in-cab warning if encountering a restrictive signal indication. Nevertheless, accident investigation reports from this era frequently suggested that something approaching modern ATP concepts should be developed, and this gradually occurred.

It is difficult to identify a “first” true ATP system in the US, but progress was well being made by the 1930s. Particularly, systems were available which prevented a train from exceeding “restricted speed” after passing a warning, requiring such restricted speeds to be sustained until a release from that state was provided at a subsequent location. This at least ensured that if an accident occurred, it would be at relatively low speeds.

It is somewhat easier to identify the history in Europe, by reference (for instance) to the Indusi system first developed in Germany in the early 1930s, and surviving in progressively refined forms to this day. "Indusi" is an acronym derived from "**I**nduktive **S**ignalsicherung", or Inductive Signal Protection. For a generally readable account of this system, see <http://www.sh1.org/eisenbahn/rindusi.htm> .

Essentially, Indusi worked by having a succession of track mounted electromagnetic devices, which would not only require an acknowledgment by the driver after passing a first restrictive indication, but would enforce progressive reduction in speed thereafter if such reductions were required.

Meanwhile (or indeed earlier – pre 1900), a different approach was adopted for some rapid transit (suburban) systems, in which the idea was not to prevent a train passing a signal at stop, but to make sure that it immediately stopped safely if it did do so. This was done by electro-mechanical (or electro-pneumatic) “train stop” or “trip” devices, where a metal arm was raised if the signal was at stop, and when another arm on the train struck this arm, the train brakes would apply. If a “neutral zone” (known as an “overlap”) was provided between the stop signal and the first obstacle, the train would be stopped without any harm being caused. This system was installed when the Sydney and Melbourne suburban systems were electrified in the 1920s.

However Sydney was a world leader in enhancing this system by using conditional clearing [see earlier definitions] on the most critical segments of line (initially the city underground). Under this enhanced system, a train would be prevented from approaching too fast. Thus closer headways could be obtained between trains, because the overlaps could be made progressively shorter as the speed of the train was confirmed to have been reduced.<sup>7</sup>

The development of ATP – in terms of functionality, rather than technology – can be seen largely as the development and combination of those initial

---

<sup>7</sup> This train stop system has stood the test of time very well, and is still functioning to good effect on the Sydney CityRail network. The conditional clearing option is generally only applied in the most critical areas, where trains need to be very close together and overlaps need to be minimised.

varieties from the early parts of the 20th century. Effectively, ATP systems have evolved by working on some mixtures of the philosophies that trains should be prevented from approaching obstructions too fast, and that they should either (a) be completely prevented from entering the actual obstructed area, or else (b) be prevented from entering it at a fast enough speed to do much or any damage.

Much of the early development was driven by a desire to control against passing signals at stop and/or against adverse consequences of SPADs. However in at least some jurisdictions, there was also belated recognition that the system could protect against excessive speed on curves and through turnouts. Modern ATP systems generally protect against excessive speed in any controllable form, whether it be related to conformity with signals (or equivalent on non-signalled lines) or to conformity with authorised speeds imposed because of track configurations.

Full ATP systems exist in numerous incarnations around the world. They are taken for granted in large parts of Europe; they are regarded as an essential basis for high-speed operations like the TGV or Shinkansen; and they are used even in various non-passenger applications. However regardless of the detailed implementation, full ATP systems can now generally be characterised by

- (a) their provision to a driver of information as to how fast the train is permitted to travel at any moment, including information about any further speed restraints for the immediate future
- (b) an enforcement process which will prevent anything but trivial exceedance of those authorised speeds.

A detailed historical review of all the different manifestations of these elements would be unproductive; suffice it to say that for “full” ATP systems, the nett effects are essentially the same irrespective of the detailed implementation, except in three key respects discussed below.

State-of-the-art ATP systems have commonly been installed onto new equipment or even new tracks; however it has been relatively rare to introduce new sophisticated systems (as distinct from lower level ones) onto an existing system of trains, except by fitting only a small number of units and then allocating them exclusively to the territory concerned. This issue may be specially relevant for the NSW context, since a large proportion of existing rolling stock was not designed with ATP fitment actively in mind.

## The current status of ATP development

Although the technology of the most modern systems may be very advanced – e.g. by using highly sophisticated computer-style control systems, high-level communications, or “artificial intelligence” – the actual train-movement intervention functions have not changed much, except in one respect, to wit “moving block” availability<sup>8</sup>. This is not because development has not advanced, but rather because there is not much more that can usefully be done in basic train management functionality, as distinct from ways in which this functionality can be delivered and ways in which that functionality can be combined with other useful services.<sup>9</sup>

It is instructive that a trial of a then up-to-date ATP system was conducted in NSW in the Liverpool – Campbelltown area in 1988. The trial was ineffectual and produced no demonstrated outcomes, but its existence demonstrated that the technology was acknowledged to exist by NSW rail administrators at that time.

Some recent “developments” have in fact not advanced the situation in terms of functionality. A key instance is British TPWS (soon to be applied on the higher-speed interurban passenger lines in Victoria, but otherwise not evidently having been taken up outside the UK). This system, which is essentially an extension from AWS origins, logically provides no inherently greater functionality than a system of conditionally cleared train stops, even though it does so at much higher speeds than have been applied with train stops; and in some respects it provides lower functionalities. It is an example of a system which has been developed to expand on existing systems in a supposedly cost-effective manner, rather than to advance towards full ATP which was perceived to be more expensive.<sup>10</sup>

The current development of ETCS/ERTMS as European standards appears to be driven not so much by gaining major new functionalities (at least at the lower levels) but more by the desire for interoperability, freedom from proprietary monopolies, and maintainability. However the highest level ERTMS systems (which don't yet exist in an ERTMS-compliant form) are intended to achieve full moving-block functionality by the highest levels of technological sophistication, and would presumably provide functionalities not otherwise available.

---

<sup>8</sup> The value of this will be discussed later herein.

<sup>9</sup> Examples of such “useful services” include granting authority for movements over single line sections, and providing better interactions between trains and level crossings.

<sup>10</sup> A good user-level account of AWS, TPWS and many of the associated issues is available at <http://www.railway-technical.com/sigtxt7.html>, although this article implies that TPWS is more widely installed across Britain than it actually is. Discussion of motivations for using TPWS rather than full ATP systems (primarily ERTMS) is provided at [http://www.rssb.co.uk/pdf/ertms\\_web.pdf](http://www.rssb.co.uk/pdf/ertms_web.pdf), emphasising a British philosophy that TPWS deals with “the majority of ATP-preventable risk”.

ERTMS quotes three “levels”. It should be noted that these classifications do not correspond well to those apparently adopted in the Waterfall Commission report:

Level 1: Balises as the basic control mechanism, with information being handled entirely locally and with the ATP system simply imposed onto a traditional signalling system. [It should be noted that the Perth ATP system, discussed further below, is effectively a level 1 system and not “level 2” as stated in the Waterfall Report.]

Level 2: Communication provided essentially via radio from a “radio block centre” with continuous updating of information, but still superimposed on some elements of a traditional signalling system (e.g. track circuits or axle counters retained, but lineside signals possibly abolished).

Level 3: Entirely based on radio communication and on-train equipment, with moving block available and no need for most traditional signalling infrastructure (although hardware associated with points and other special functions is still required).

The ERTMS website at [www.ertms.com](http://www.ertms.com) provides much useful information, including reports on progress and a graphic illustrating the many different ATP systems around Europe (at least 15, from at least 8 different suppliers) which will supposedly be replaced by ERTMS.

It is understood, however, that there are currently some widespread concerns in Europe about the progress of ERTMS:

Firstly, there are major concerns about cost-effectiveness (both in absolute and relative terms), and some of the systems previously committed to ERTMS are said to have substantial concerns regarding the costs. For instance, the cost of retrofitting large numbers of motive power units is perceived to be seriously excessive.

Secondly, there are said to have been problems (in both ERTMS and comparable systems elsewhere) with data integrity and with the capacity of the radio networks to manage the data, and even with the information being supplied (e.g. the location of the ends of trains as distinct from the driving end). It has even been suggested that these difficulties may be great enough to lead to a “level 2.5” rather than the full level 3 as originally specified.

For instance, the SBB (Swiss Federal Railways, generally regarded as one of the most progressive) experienced problems with a pilot scheme between Olten and Luzern, and has accordingly reverted to some previous technologies and has reputedly suspended progress towards moving block systems.

A particular trend – as exemplified by the ARTC approach, as well as by higher levels of ERTMS and by some North American systems – is for ATP to be regarded in a broader context of communications.

Although this is multi-faceted, the most prominent element of the latest communications-based systems is that the whereabouts of trains, and instructions to them, are conveyed by a continuous communications link, which may be either hard-wired along the track or by radio.<sup>11</sup> Increasingly, much of the information may be derived from electronic communication sources (e.g. GPS). This reduces or eliminates the need for any on-the-ground hardware such as balises, track circuits, signals or indeed anything placed on or near the track.

Such systems inherently have major possible failure modes, the elimination of which may be difficult and costly. With radio and GPS systems, some potential failure modes to be addressed include loss of the radio or GPS signals, inaccurate location by GPS (e.g. if the GPS system believes that a train is on a different adjacent track), or communication with the wrong train. [See later herein for discussion of some other hazards.]

For a somewhat incomplete survey in relation to urban transit systems, see <http://www.tsd.org/cbtc/projects>, This document identifies systems available, in use and being installed, and also includes a good history of the chequered history of the New York Canarsie line project. It also highlights how the majority of such projects are related to completely new lines rather than retrofitting old ones.

In relation to non-urban railways, a description of some actually implemented American activities in these matters is available in a Federal Railroad Administration summary document dated December 15 2004, available at <http://www.fra.dot.gov/us/content/1265>. It shows how multiple parallel developments have been occurring, prompted particularly by the National Transportation Safety Board having pressed over many years for “positive train separation” as one of its key areas for safety improvement.<sup>12</sup> A key trigger in regard to longer-distance lines has been the high US incidence of fatigue-related freight-train SPADs with severe consequences.

However there are also problems being encountered in the US environment, some of them analogous to those described above re ERTMS, but also in regard to the proliferation of different designs. Unlike the recent unified approach in Europe, there have been multiple parallel developments in the US with little or no commonality, and yet there is actually wide through-running of locomotives in the US which may result in locomotives being far from their “home base”. According to recent reports, this situation has now been recognised as a problem by the Association of American Railroads, and

---

<sup>11</sup> Initial developments of this nature were implemented with continuous hard-wired equipment along the tracks, often described as “inductive loops”. Newer systems are radio based, although some railways have been reluctant to progress to the radio systems because the inductive-loop systems are regarded as proven and having lower safety and/or commercial risk.

<sup>12</sup> See also the presentation at <http://www.nts.gov/events/symp%5Fptc/presentations/03%5Fvanderclute.pdf> which seeks to put US developments into an overall industry context as at March 2005.

some steps are being taken to redress the situation. Lockheed Martin, the contractor for the ARTC developments, has reputedly been identified as a key contractor to rectify the American proliferation problem.

### **The Perth system as an ATP example**

Since the Waterfall Commissioner chose to quote the Perth system as a good example, it is opportune to enumerate some of its powers and constraints.

The Perth system is not by any means a “sophisticated” ATP system by current standards, and was not even state-of-the-art when chosen over 15 years ago. It was thought to provide optimal cost-efficiency for the likely traffic densities.

It is an intermittent system, where transponders (balises) are placed at selected locations (typically before signals), and each balise instructs the train about what speed (possibly zero, possibly not) it is currently required to achieve at some point further along the track. The distance to that further point is identified, and if a reduction from the current speed is required, the on-board equipment can assess what braking curve needs to be achieved to reduce to that lower speed.

If the driver appears not to be conforming to a desirable braking regime, the system will provide an audible warning. If the response is still not deemed adequate, the system will take over from the driver and impose a brake application to the required extent.

However, the system does not seek to control speed once it has got as low as 40 km/h (or in some special circumstances, 10 km/h). After those speeds have been reached, the system does not interfere unless the speed drifts above those thresholds again, or unless a signal is actually passed at stop (in which case emergency brakes are applied). Thus the system does not actually prevent SPADs; it merely ensures that they are not too severe when they happen (and overlaps are therefore still required, but only of a length to accommodate overruns at those controlled speeds). The question of why the system works in that way is (in substantial part) tied up with issues of continuous versus intermittent ATP, addressed later herein.

Thus in respect of SPAD prevention and mitigation, it could be argued that the Perth system supplies little greater functionality than is supplied in those parts of the Sydney suburban system which are already equipped with conditional clearing.

In one respect, though, its functionality is significantly greater than is currently provided on the full Sydney system – in respect of providing control of speed approaching restricted curves and turnouts. Although the Sydney system of conditional clearing of train stops *could* in principle be used for such

purposes, albeit with limitations<sup>13</sup>, it is not thus used except in very limited contexts. However the Perth system readily provides such control, by the placement of permanent control balises approaching permanent speed restrictions, and by the provision of speed information from the normal signal-related balises on the basis of the relevant fixed signal indications.<sup>14</sup>

### **Key major functional differences amongst some existent or planned systems**

As noted earlier, there are high levels of commonality amongst the functionalities of higher-level modern systems. There are however three respects – two fundamental and one largely circumstantial – where very substantial differences exist, and these need to be considered in any decisions about where to head about the cost-effectiveness of installing new ATP systems, and then in selecting amongst competing offerings.

#### ***Continuous-update versus intermittent-update systems***

Many ATP systems – including some already in operation in Australia – communicate with the train by “balises”, i.e. objects (transponders) placed at fixed points and interacting with the train only at those points. Despite logic systems being provided to minimise the consequences of this non-continuity, there is the immediate consequence that for much of its journey, a train is under the influence of old information – for instance if a signal has cleared ahead, the train nevertheless cannot respond to that clearance until it passes the next balise.

On the Perth system (and various others), this problem was regarded as largely acceptable provided that trains were not constrained to excessively low speeds while waiting for the next balise. Thus a decision was taken not to govern the speed below a certain level, typically 40 km/h, so that once the driver saw that the signal ahead had cleared, he could at least sustain the speed at that intermediate level. The spacing of balises is chosen with due regard to providing information at sufficiently frequent intervals.

This threshold speed system provides a trade-off, in that ideally the train would be governed right down to a stop if that was necessary. Some people regard this trade-off as acceptable; others regard it as a “worst of both worlds” situation, in that the benefit of full control to a stop is not available but delays are still unnecessarily created because of the lack of instant response to signal clearing.

---

<sup>13</sup> A trainstop system for controlling turnout speed must act sufficiently early to reduce the speed of the train between the location of the speed check and the location where the turnout commences. This may be less effective and less responsive than a system which is able to monitor the braking curve on a more continuous basis.

<sup>14</sup> ATP systems can also provide control over temporary speed restrictions, e.g. by the placement of additional temporary balises and/or by modification to the information associated with existing balises. There may however be difficulties in ensuring the accuracy and effectiveness of such arrangements, especially in junction areas or in cases where the temporary speed information significantly conflicts with the information provided by the permanent system.

The effect is of course highly dependent on factors such as signal spacing, balise spacing, traffic density, implications of delays (e.g. blocking other trains), and the chosen threshold speed.

This dilemma is overcome in many ATP systems by the provision of continuous information. One of several common methods is “coded track circuits” (although these also have their limitations). In one version of this system, electric currents are carried through the rails at particular frequencies, each frequency representing a particular permitted speed, and the train detects which frequency (if any) and hence which speed is being conveyed. Failure of any part of the system will result in no frequency being present, which will be interpreted as a zero speed command.

However much of the development of the lower levels of ERTMS has used balises – thereby restricting its applicability to those systems where the consequences of non-continuous information are deemed acceptable, or where the expense and maintenance consequence of providing closely-spaced balises are deemed acceptable. So-called “infill loops” may also be provided which may provide information between balises.<sup>15</sup>

It is instructive to compare the balise situation with that provided with a mechanical train stop system. Both balises and train stops respond immediately to the relevant conditions, in the sense that at any particular moment, they accurately reflect the current status. However the balise system is associated with continuous monitoring of the train speed by the on-train equipment, so that the train speed control information cannot update until the next balise is reached. In contrast, in a train stop system, there is no enforced monitoring between successive train stops, so that the driver can react immediately when he sees the signal or train stop clear. There is thus a trade-off between the monitoring and information-flow aspects. In contrast to both these intermittent systems, a continuous system will provide both information and monitoring on the basis of the precise situation applicable at that moment.

### ***Moving versus fixed block***

Moving block systems have not been extensively installed as yet, although they do exist. Credit for the earliest functional moving block systems is commonly given to Alcatel, with a variety of installations typically on non-conventional and/or new rapid transit systems. The first commercial installation may have been the new self-contained Scarborough line in Toronto (Canada) in 1985. The version of the Alcatel system on the Docklands Light Rail in London is conveniently described at <http://www.xs4all.nl/~dodger/tech.htm>

Moving block is essentially a way of keeping *moving* trains separated by sufficient distance, particularly at higher speeds (but ironically, the initial

---

<sup>15</sup> A balise-driven system is adequate for many applications, and the cost of retrofitting existing lines with hardware such as coded track circuits may be considered prohibitive when a cheaper balise system is considered to have adequate functionality.

installations were not high speed). As a generalisation, it has greatest benefit (compared with an optimised fixed block system) when speeds are high, and the advantages diminish at lower speeds (because, inter alia, the separation distances which need to be attained are smaller, and the times to travel a given distance are greater).

In the limit, when trains are stopped, moving block has no benefit at all, except in relatively minor respects (e.g. one arriving train could follow a departing train slightly more quickly into a platform, but systems installed in the 1920s on the Sydney network to achieve effectively the same result were actually eliminated with more recent resignalling actions, in part because they rarely were utilised and had no practical benefit).

Similarly, moving block has no benefit if the trains normally run substantially further apart than the moving block separation distances. In country areas, for instance, there may be no point in spending a lot of money to keep trains running safely (say) 2 km apart if they do not need in practice to run less than 6 km apart and this can be comfortably achieved by a cheaper fixed block system.

Conversely, if moving block systems develop to the stage where there is no cost penalty involved relative to fixed block systems, one would ask "then why not use moving block"? This is one area where the economic factors are changing rapidly.

### ***Retention of fixed signals***

It is commonly believed that there is some necessary link between various levels of ATP and whether fixed lineside signals are provided. This is not true, except in respect of moving block.

Cab signalling systems without fixed signals have existed for very many decades without full ATP. Similarly many ATP systems co-exist with a full panoply of fixed signals.

With moving block, fixed signals could not provide the same information, since the block boundaries are constantly varying and yet it would be impossible to move the signals correspondingly. However even with moving block, it is sometimes thought desirable to retain some sort of lineside signal system (with fixed blocks) to cope with trains that are not fitted with the moving block equipment or on which it has failed.

With fixed block ATP, many systems have chosen to retain fixed signals, and many have chosen to eliminate or simplify them. In the former case, the decision is commonly based on coping with unfitted trains or equipment failure, and/or on providing information to the driver which may be different in substance or impact from what is provided by the ATP. (For instance, the ATP may only provide speed information, whereas lineside signals may provide route information, which may enable a driver to manage the train more effectively and may also enable him/her to challenge routes which have been

incorrectly set.) Even human-factors issues such as encouraging the driver to look at the track rather than at the in-cab equipment may be relevant.

On the other hand, some systems have chosen to eliminate or simplify the fixed signals because the cost savings are regarded as outweighing any advantages of retention.

A significant issue is that ATP systems can be configured on the basis of the braking capacity of an individual train – e.g. a train with greater braking capacity can approach a speed-reduction location less conservatively. Fixed signals, on the other hand, are typically configured to cope with the worst-braking trains – although use of multiple aspects sometimes provides information to enable drivers to match their responses to the capacity of their trains. An ATP system might therefore allow significant improvements in track capacity for some classes of trains relative to what is provided by a conservative fixed-signal system.

### **Potential negatives of full ATP systems**

Despite the obvious advantages of full ATP systems in controlling the train even if the driver does not, there are a variety of disadvantages which need to be factored into any overall risk or cost-benefit assessments. Some are listed below.

#### *Degraded working*

If the ATP system fails, what happens? Either trains stop altogether, or they are worked without the protections which the system normally provides. The risks created by such degraded working may considerably exceed the risks being protected against by the ATP system itself, especially where fixed signals are not provided as a back-up and trains are therefore operated entirely under human judgment.

#### *Human factors*

There has been surprisingly little detailed research on how specific ATP systems of different types affect the behaviour of drivers (or train operators, as they may be known in completely automated systems). However general human factors issues can readily be identified, of which the four most obvious are:

- (a) If the driver thinks the ATP system will protect him from any calamity, he may consciously or subconsciously lose concentration, so that when a hazard arises which the ATP cannot know about (e.g. a trespasser, a load shift on an adjacent train, or a vehicle on a level crossing) the driver may be insufficiently alert to see it and respond. There may even be temptations to do other things (e.g. read, use mobile phones), and someone may believe that fatigue is unimportant because if they do go to sleep, the system will still protect them.

- (b) Judgment may be impaired, e.g. the driver may become so accustomed to the system making decisions for him that he loses the ability to make the decisions himself (which then becomes a major risk when degraded working is needed).
- (c) Driver annoyance/frustration, where the automated system is preventing the driver from doing what he knows is reasonable (see next section), leading to other negative behaviours.
- (d) If the driver does not believe that the system is giving correct information, he may be tempted to override it (if such a facility is provided). This hazard also exists with systems such as AWS, TPWS and train stops, all of which have had major incidents with drivers simply rejecting correct information.

### *Train performance*

An ATP system typically has to work on a “worst case scenario” basis, assuming the most unfavourable state of factors such as train braking power, track adhesion characteristics, speedometer accuracy, and anything else that might determine the braking curve. Thus ATP may impose more conservative (and sometimes far more conservative) driving than would be applied by a competent driver, consequently extending running times (especially in times of disrupted services, when a competent driver can most exercise skill in recovering lost time while still acting safely).

This effect is readily observed when travelling on a suburban-style railway with ATP (e.g. Perth or the Sydney light rail) where it is frequently noted that the train is proceeding unduly conservatively.

In suburban systems (as distinct from high-speed systems like the TGV), theoretical studies typically show that at best small capacity benefits can be gained from ATP when compared with optimal conventional signalling, but in practice these benefits may be negative. Nevertheless, as noted earlier in connection with the issue of retention of fixed signals, there may be circumstances where there can be capacity improvements for some types of superior-braking trains which can perform more effectively than they can under a fixed-signal system configured for trains with lower braking characteristics.

### *Potential wrong-side failure modes*

There are many potential failure modes in an ATP system which are not present in a normal signalling system. Most obvious of these is the communication element, where information has to be transferred and interpreted through a variety of separate processes, each with its own possible failure modes. This is far more complex than a relatively simple system of illuminating signal lights and the driver viewing them, even though the latter system also has its failure modes.

Just to take one example, if an ATP system tells the driver that he can continue at normal speed when in fact he shouldn't, the driver may have no way of knowing that the instruction is incorrect (especially if there are no fixed signals) and he may even be less vigilant when an obstruction actually comes into sight. Moreover this false information may have been given through multiple possible causes – e.g. a transponder giving a wrong code, the train transponder misreading that code, the code being corrupted along the train's wiring, the on-train computer having some hardware or software bug, and so on.<sup>16</sup>

It follows that the certification of ATP systems is extremely complex, and moreover that they have to be maintained at a high level.<sup>17</sup>

It might be anticipated that a whole new host of potential failure modes will exist in, and will need be eliminated from, communications-based systems under development.

### **Safety integrity levels**

In greater generality than the above, it should be noted that any ATP systems must be designed with a clear understanding of the safety integrity levels which they are required to sustain. For instance, the ATMS system being developed for ARTC is stated to require safety critical software to at least SIL3, and wrong side failure integrity to SIL4.

This paper does not seek to canvass the criteria which might be applied to demonstrate that a particular ATP system is "adequate" in SIL terms. Any proposals for the installation of ATP systems should however explicitly consider SIL requirements in the overall context of risk assessment and prevention of adverse consequences.

### **Criteria for considering the installation of an ATP system**

This paper does not seek to direct the industry into particular lines of assessment or development of ATP. Nevertheless, as an indication of some of the sorts of considerations which *might* be appropriate, Appendix 1 herewith is an example of a discussion paper which was intended to trigger lines of debate.

---

<sup>16</sup> In practice, protections are provided against most of these contingencies, e.g. received codes need to satisfy various internal validity checks, and failure to do so will trigger the most conservative outcome. However this then creates more potential dangers from degraded working when the ATP functions are not fully available.

<sup>17</sup> In addition to matters directly relating to the physical maintainability of the equipment itself, there may be issues about the availability of qualified personnel, and about risks to personal safety if workers need to be on track to maintain objects such as balises. This should also be viewed in conjunction with an assessment of the reduction of risks which might result from removing existing equipment or changing existing operating patterns, for instance if the introduction of a need to maintain balises was accompanied by elimination of maintenance-intensive and failure-sensitive train stops.

It should be clearly understood that this Appendix is *not* a statement by the NSW Regulator of a preferred path in NSW, but is to be read purely as an *example* of matters which might be considered.

## APPENDIX 1

### AN EXAMPLE OF A DISCUSSION PAPER ABOUT ATP CRITERIA

**It should be clearly understood that this Appendix is *not* a statement by the NSW Regulator of a preferred path in NSW, but is to be read purely as an *example* of matters which might be considered.**

The existing electro-mechanical train stop system in NSW is actually higher in the ATP hierarchy than it is commonly perceived to be, because it is capable to some degree of monitoring the speed of trains and stopping them if going too fast. However the limitations are;

- Time/speed control is only applied on a small fraction of the system, and in a small fraction of circumstances (i.e. on most of the system, train stops are applied only to stop a train passing a signal at stop and not for speed control);
- Only trains fitted with train stops are influenced by them (which excludes all locomotives), and reliably fitting such equipment to locomotives is arguably not easy
- The system is only applied in the electrified area (and even then, not 100%);
- The system is very little applied for speed control other than in connection with signals, i.e. it does not (in current usage) protect against excessive speed on curves under clear signals;
- The facility for controlling speeds is typically only applied in lower speed ranges, and it is not clear how well it would function for higher speeds, especially inasmuch as it would need to be applied predictively rather than just at the location where a speed reduction is ultimately required;
- Except where conditional clearing is provided, overlaps have to be set to cope with “worst-case” scenarios, thereby potentially reducing track capacity.

Granted, however, that the train stop system is effective within its limitations, one would expect to look particularly at the following sorts of criteria for a possible new system:

- It provides significantly greater functionality than the train stop system
- It provides significantly greater compatibility (both across types of trains and across geographical locations, which may well mean across the nation since locomotives now wander far afield)

- It must be reliable and maintainable, and ideally should be less maintenance-intensive than existing equipment
- It is “future proof”, i.e. it will continue to be economically maintainable and have minimum obsolescence
- It will not degrade the capacity of the operational network, and ideally will increase it
- It must be no less safe.

Some other criteria for consideration, but not necessarily absolute, might be

- It has to provide continuous or near-continuous speed monitoring, not just monitoring at discrete locations
- It will physically prevent relevant types of incidents from occurring, rather than just reducing their effects when they do occur
- It has to be capable of intervening when a genuinely dangerous condition is appearing, but should not interfere unnecessarily in the normal operation of the train
- It has to be capable of progressive installation, granted that it would not be physically possible to convert everything “overnight”
- It has to have a reliability level sufficient to eliminate safety threats caused by trains having to work in degraded mode outside the protection of the system, and to eliminate service reliability threats
- It should not be limited by proprietary constraints of equipment supply
- It should be undertaken with due consideration of the operation of equipment outside the network which is under direct review.

Clearly the evaluation of such criteria, as well as matters of cost-efficiency, require a substantial ongoing level of investigation in any given context, and neither decisions nor installations can be done on a short time scale.