Evaluating and Preventing Capacity Loss when Designing Train Control to Enforce NFPA 130 Compliance

Davis Dure
SYSTRA Engineering, Inc.
New York, NY

ABSTRACT

Many rail operators are wrestling with how to compensate for the capacity reduction resulting from having to enforce the NFPA 130 requirement that there be only one train occupying a tunnel ventilation zone at any one time.

Normally, capacity limits are experienced at the highest speeds, with lower speeds affording greater capacity by allowing the signal design to take advantage of the shorter safe braking distance using shorter block lengths. With the long blocks that result from NFPA 130 enforcement, this is no longer the case. Signal clearing times lengthen at the lower speeds, and capacity actually drops. Network simulation documents the capacity "tipping point" under these conditions: When trains enter the tunnel at very short headways, throughput actually drops significantly below what can be achieved when trains enter the tunnel on less aggressive headways.

This phenomenon hampers the ability of the system to recover from perturbations which deliver a "burst" of trains trying to enter the tunnel once the cause of the delay has been resolved. It also hampers recovery from any incident in which trains are slowed or stopped within the tunnel. Capacity is lost when it is needed most.

A major northeastern commuter railroad encountered this situation in its attempt to quantify recovery capacity in its new tunnels. The "tipping point" for throughput reduction occurred whenever a train encountered a "0" cab speed code, requiring it to reduce speed from either 40 or 30 mph to 15 mph, and then run its train length before accepting better code.

A solution was found by introducing an intermediate 20 mph speed code into the signal design, filling the gap between 30 and 15. This virtually eliminated the "tipping point" phenomenon and makes the full range of capacity available in recovery situations.

INTRODUCTION

A major northeastern commuter railroad recently confirmed its decision to comply with the recent NFPA Guideline 130 for designing a new line into the east side of Manhattan within a previously constructed but never used two track tunnel. The tunnel signal system will be designed to prevent more than one train from occupying a tunnel ventilation zone at any given time. This new constraint will result in the line capacity for the tunnels being reduced from its original capacity that was achievable under an optimized signal system.

The question has arisen: Can the peak scheduled volume of 24 trains per hour (TPH) originally planned for the terminal station and tunnels be sustained reliably in spite of the reduction in line capacity? If not, what measures must be taken and how many trains can be scheduled to ensure reliable operation?

The line capacity supported by a signal design which complies with NFPA 130, which effectively converts each ventilation zone into an absolute block, can only be partially documented using the generally accepted methods. The conventional assumption is that capacity increases at lower speeds, and thus it is normally assumed that capacity quantified at MAS, or at any given intermediate speed, is available at lower speeds as well.

This paper begins by describing the generally accepted methods by which capacity is calculated, in which the maximum possible throughput, sometimes called the theoretical capacity, is quantified either at MAS or a slightly lower speed corresponding to an intermediate speed code. This number is then reduced by 20-25% to yield what has been called "practical" capacity, which is generally considered to be 75-80% of theoretical.

The paper then describes the unique phenomenon under the one-train-per-vent-zone constraints in which capacity actually drops when speed drops below a certain point. This speed drop will frequently occur when trains enter the tunnels at too high a frequency, with entering headways that are two short. This capacity loss at lower speed is referred to in the paper as the "tipping point" phenomenon. This discovery has significant practical implications, namely that capacity which is documented using generally accepted methods of calculation is not available when it is needed most, during recovery from perturbations.

The following question remained: Taking this phenomenon into consideration, could the railroad schedule trains right up to the 80% practical limit and still have a reliable operation? If not, what is the practical limit for scheduling purposes to avoid minor operational variations cascading into significant delays? The paper concludes by describing the solution that was identified which eliminated the "tipping point" phenomenon and allowed the documented capacity to be fully available in recovery situations.

DEFINING RAIL LINE CAPACITY

For capacity measurement purposes, the theoretical headway at a specific signal location is the shortest time a following train can pass that signal location at the same speed as the first train. It is normally assumed that this is the maximum speed allowed by the best speed code or signal aspect available entering each block, subject to civil speed limits, although a fixed underspeed associated with each speed code can be assumed where it is known that the average speed maintained for a code is somewhat less than the permitted speed.

For fixed-block signal systems, both cab and wayside, the theoretical headway at each signal or cab-signal block is determined by calculating the "clearing time", which is the time it takes for the signal aspect or cab-signal code to restore to the highest speed code or best aspect available at that signal location after the passage of the first train. This includes the time it takes the signal apparatus to actually display the upgraded signal after the preceding train has cleared the control circuit, typically three seconds. An engineer reaction time is also identified, typically three to four seconds, allowing the engineer of the following train to know that the signal is clear and thus approach the signal with confidence at the same speed as the first train. This reaction time is added to the clearing time to yield the theoretical headway.

The clearing time at a given signal is determined by both constant (fixed) and variable factors. The cab-signal-speed-code or wayside-aspect control line of the signal, which spans the section of track beyond the signal that must be clear (unoccupied) before the signal can display that cab-signal-speed-code or wayside aspect, is fixed. Its length is determined during signal design based upon the worst-case safe-braking distance, which is the distance required for a worst-case train to safely come to a stop after it passes the signal at a given speed and begins braking.

The variable factors in determining signal clearing time include the speed and station stopping pattern of the first train as it passes through the control line of that signal. If there are multiple train types and station-stopping patterns, a clearing time can be calculated for each one. This can either be done in the field with a stopwatch or by using a Train Performance Calculator programmed with the civil characteristics of the line (grades, curves, civil speed restrictions, the maximum speed allowed by the signal code or aspect, and any underspeed assumptions), the unique acceleration and braking characteristics of each train type, and the dwell times at each scheduled station stop within the span of the control line.

Train capacity, or throughput expressed as Trains per Hour (TPH), at a given signal location is calculated by dividing one hour (3600 seconds) by the theoretical signal headway, which includes both signal clearing time and engineer reaction time as noted above. Where there are multiple train types and stopping patterns, a weighted average headway is calculated based on the proportion of each train type relative to the total number of trains.

It should be noted that the capacity calculated for each signal location assumes that a following train is arriving at that signal ready to accept the newly cleared code or aspect 3 seconds after it clears. Where a line segment has multiple signals, the longest clearing time defines the capacity of that segment. While some signals in a given segment may clear faster than others, trains operating on close headways cannot take advantage of these fast-clearing signals because their natural progression will reflect the slowest clearing signal.

The capacity of a given line segment, therefore, is based on the Theoretical Line Headway, which is the maximum of the headways calculated (or observed) at each signal on the line – sometimes referred to as the ruling or constraining headway. Throughput is calculated by dividing the Theoretical Line Headway into one hour in the same manner as for an individual signal block.

Normally the longest clearing time for a train making station stops is associated with a signal whose control line extends into a station platform. Station dwell time is usually the factor limiting capacity in such cases.

Where there are multiple train types and/or stopping patterns, an average Theoretical Line Headway is derived from the longest clearing time for each train type/stopping pattern, with a weighted average calculated based on the proportion of each type. In cases where the longest clearing times are at different signal locations for different stopping patterns, the weighting of clearing times is based on the number of times trains of one unique pattern follow trains of another unique pattern.[1]

Besides the two major variables of train type and stopping pattern, there are additional variable factors in determining the actual clearing time behind each train, namely the variations in engineer and equipment performance affecting rates of acceleration, deceleration and maximum speed. These are typically not included in clearing time calculations, but any measure of capacity must account for them in some way. Calculated capacity at a given signal that does not factor in these variable elements is often referred to as the "Calculated" or "Theoretical" capacity at that signal location.

Each of the variables not included in theoretical capacity can have a negative effect on capacity in one of two ways:

- By slowing a train as it approaches the ruling signal, making it unavailable to take immediate advantage of the clear signal when it becomes available. In the absence of buffer or schedule margin between trains this will cause every train that attempts to follow at the theoretical ruling headway to be late by the same amount of time that the first train was delayed.
- By slowing or delaying a train within the ruling control line such that the signal takes longer to restore to its best aspect or code.

A capacity measure that includes all the factors not included in this measure can be considered "practical capacity".

Depending on the reaction times, vehicle acceleration rates, and the spread between cab-signal speeds, trains must be able to achieve and maintain the practical headway after routine and relatively minor disruptions. If there is sufficient recovery time built into the scheduled dwells at the terminal – in other words, if trains can be turned more quickly than the schedule requires – this lateness will not cascade to the opposite direction.

A "maximum throughput" network simulation is a second means of defining capacity. A simulation is run in which a throughput entering the line is deliberately greater than the theoretical line capacity, and the resulting throughput at the other end of the line is measured.[2]

The throughput achieved in the maximum throughput simulation approximates the throughput that could be achieved in recovery mode following an on-line or off-line failure. To the extent that this throughput is greater than the scheduled throughput, trains in the same direction will gradually become less late. Once trains are back on schedule, normal scheduled throughput will resume.

It should be noted that if trains become increasingly late during the course of the simulation, this level of throughput cannot be scheduled. It can (and should), however, be achieved in a recovery situation. Sensitivity tests should be performed in the simulation to identify the maximum throughput that can be achieved with a constant OTP at the leaving end.

Typically, plotting capacity with respect to speed results in a bell shaped curve in which capacity is low at low speeds due to the long time it takes to traverse fixed blocks, high at speeds in the 45 to 60 mph range when supported by appropriate intermediate-speed signal aspects or codes, and tapering down again at higher speeds which require longer safe braking distance. It should be noted that braking distances increase with the square of the speed, resulting in much longer signal blocks at higher speeds.

APPLICATION TO COMMUTER RAIL NFPA 130 CONSTRAINTS

Theoretical Line Headway results from the signal clear reports show that the NFPA-compliant design for a major northeastern commuter railroad comfortably supports their 24 TPH schedule requirement, with a capacity utilization at that TPH of 70-71%, well within industry practice. This was based on a theoretical line headway of 1:47 calculated at MAS using the methodology described in the previous section, which translates to a theoretical capacity of 33.6 TPH.

However, network simulation maximum-throughput tests paint a somewhat more complicated picture. Two levels of maximum throughput network simulations were run: One in which trains entered the tunnel at an average rate of every 2 minutes, and another in which trains entered the tunnel every 90 seconds. It was found that while 30 TPH could be sustained in the first test, throughput actually declined to 26 TPH in the second test, when trains enter the simulation at 90 second headways.

These simulations document the amount of capacity above the scheduled throughput that is available for recovery. Given the fact that the simulations do not account for engineer pre-action and variations in train performance or engineer conservatism and reaction time, documented simulation throughput rates should be derated by roughly two TPH to arrive at a recovery capacity that could be relied upon in practice.

[1] Pachl, Joern, Railway Operation and Control, p. 152
[2] Pachl, p. 157

The "Tipping Point" phenomenon

The following is a detailed description of what happens. The interlocking entering the tunnel, where trains merge from several different lines, is able to "serve up" trains at intervals as close as 1:28, with an average arrival headway of 1:46. At this average arrival rate, but with every fourth headway at 1:28, some trains are reaching signals faster than they can clear, and have to slow to the next lower code rate. This slowing is cumulative at this sustained high arrival rate, and the fourteenth train in the simulation encounters a section of 0 code approaching the longest vent zone, which is still occupied by the thirteenth train.

Even though the thirteenth train clears the vent zone within seconds and the code received by the fourteenth train upgrades, the damage is done, since the fourteenth train now must run its length at 12 mph (the speed allowed in the simulation under 0 code, reflecting a 3 mph underspeed) before accepting better code. Almost immediately the entire procession through the tunnel is alternating between 75 and Restricting code, and the average headway at the end of the tunnel increases from 1:50 for the first 13 trains to 2:17 for the remaining trains.

Entering the tunnel the range in headway variation was between 1:28 and 2:19 – a spread of 50 seconds. By the time the trains reached the other end, the signal system had compressed this variation down to an 8 second range around 1:50 before the No Code "tipping point", and to a 10 second range around 2:17 after the tipping point. It should be noted that these results were obtained both with and without 10 minute "recovery" turns at the terminal – without the turns there were no eastbound train movements and all trains had a clear path to the terminal.

It should be noted that a simulated turn time of 10 minutes at the terminal at the end of the tunnel was assumed for all maximum throughput simulations. This is because the only time trains will arrive and depart the terminal at a rate of more than the scheduled TPH is when a perturbation has backed up the flow. It is therefore safe to say that when "the dam breaks", the terminal will be relatively empty, and that every effort will be made to turn trains quickly to speed the recovery. Therefore, a test of recovery capacity that attempts to simulate a higher throughput than 24 TPH can appropriately assume that all turns can be achieved in 10 minutes until the onslaught of backed-up trains has been processed.

As a sensitivity test to determine at what entering headway the "tipping point" phenomenon disappeared, and a single throughput could be sustained indefinitely, a 120 second (2:00) average put-in headway (30 TPH) simulation was run with a similar but slightly wider range of variation – between 1:28 and 2:50 for a spread of 1:22 entering the tunnel. This 2:00 average headway was sustained at the opposite end for the entire duration with no degradation, with a headway variation of between 1:51 and 2:26 for a spread of 35 seconds. Eastbound, a 2:00 average headway (30 TPH) was also sustained through the entire tunnel, reflecting a "steady state" operation.

The "tipping point" phenomenon results in the counter-intuitive situation where output actually drops, significantly, when input is increased beyond a certain volume – in this case, beyond a rate of 30 TPH. While the tunnels and terminal can process an average rate of 30 TPH indefinitely, the tunnel throughput drops to 26 TPH after the first 14 trains when input exceeds 30 TPH. In other words, recovery capacity drops when it is needed most.

This tipping point appears to occur due to the disproportionate impact of the long vent zones at low speeds. To test the relative impact of low speeds in other locations, headway reports were run for a sample train in "normally signaled" territory on a segment of railroad not subject to NFPA 130 constraints. In these tests, the train was limited to progressively lower speeds reflecting each speed code down to "0 code". The same was done for a sample train in the vent-restricted tunnel under study. At each speed, the clearing time to the code which supports that speed was documented for each location, and the ruling headway determined.

Table 1 shows that in territory with "normal" unconstrained signal design in which trains are allowed to enter short blocks under "0 code", the theoretical line headway actually decreases slightly at "Restricted Speed" (15 mph) supported by 0 code, yielding the highest capacity of any of the speeds. Of course, run time at 15 mph is severely lengthened. However, in recovery situations, throughput is what matters.

Code Rate	Maximum speed of train	Theoretical Headway at specified code rate: Unconstrained Design (Line segment A)	Theoretical Headway at specified code rate: NFPA 130 Compliant Design (Line segment B)
180	90 mph	2:28	N.A.
270	60 mph	2:02	2:07
120	45 mph	2:18	1:58
75	30 mph	1:59	2:12
0	15 mph	**1:56**	**3:01**

Table 1. Headways supported by representative sample signal designs at different speeds.

By contrast, in a different territory with an NFPA compliant design, the supported headway increases dramatically at "0 code". The three-minute theoretical line headway shown in the table at 15 mph (12 mph with underspeed) implies that with sustained operation at 0 code over the entire segment, theoretical throughput is reduced to less than 20 TPH, with practical throughput even less (by at least 20%).

Typically, however, when the "tipping point" is reached, operations do not immediately degrade to a sustained operation at 0 code. Rather, 0 code is encountered briefly, while the preceding train is still occupying the next vent zone, before upgrading when the train ahead clears the vent zone. Thus, while TPH is only 20 TPH calculated using theoretical clearing times at 15 mph in the NFPA compliant segment documented in Table 1 above, a 26 TPH steady-state recovery capacity was documented over the same segment after the first 14 trains in the maximum throughput network simulation described earlier.

The simulations discussed above focus on recovery capacity from an "upstream" failure outside the tunnels – the ability to sustain recovery headway through the tunnels once such a failure is over, as opposed to recovery from failure inside the tunnels or a problem at the terminal.

Recovery from Downstream Perturbations

Two types of simulations have been conducted which address failure scenarios inside the tunnels. One set of maximum throughput simulations documented a recovery capacity of only 25 TPH following a switch failure lasting for 20 minutes at the interlocking at the end of the tunnel. This represents a recovery capacity no greater than the 24 TPH scheduled throughput when the variability of engineer and equipment performance is factored in. Contributing to the low recovery throughput was the fact that trains were spaced one per vent zone at the start of recovery, and proceeded at restricted speed until they ran their train length.

The second set of simulations was performed to evaluate the impact of a circuit failure in the tunnel at the beginning of the AM peak period and lasting for the duration. It was assumed for worst case purposes that the circuit at the downstream end of the vent zone was the one with the failure, meaning not only that a train had to run the entire vent zone at restricted speed, but that a determination could not be made that the train had cleared the vent zone until it had also cleared the first circuit in the second vent zone. This had the effect of preventing trains from entering the affected vent zone until the preceding train had cleared the first circuit in the next zone, with a devastating impact on throughput. These simulations resulted in a catastrophic back-up, with throughput of only 15 TPH.

It should be noted that in a long block between vents, there may be several track circuits, and perhaps an intermediate signal. In the case of a track-circuit false occupancy, it may not always be necessary to run the entire vent zone plus one train length at Restricted Speed, and it will be possible to allow the following train to enter the zone immediately after the preceding train has cleared it.

Introducing a new lower speed code

In order to reduce the likelihood of trains encountering 0 code during recovery from a perturbation, a new 20 mph code rate was introduced into the NFPA compliant design as a test. With this new code, each 120-0 or 75-0 step-down location in the design became a 120-20 or 75-20 step-down instead, with a further step-down to 0 code occurring closer to each occupied block at a location determined by the SBD from 20 mph.

This change almost completely eliminated the "tipping point" phenomenon in the 90 second headway put-in simulations, and resulted in a sustained 30 TPH throughput in both the 90 second headway simulations and the switch failure simulations. This means that the system with the added 20 mph code rate can support 30 TPH in recovery mode.

This change will offer next to no improvement in the event of a circuit failure, however, since by definition a circuit failure results in operation for long periods over a significant stretch of track under 0 code.

SUMMARY

A major issue of concern is the impact of the NFPA 130 one-train-per-vent-zone restriction on recovery from major perturbations, including those occurring en route to the tunnel, within the tunnel, and at the terminal. This impact, as documented in the previous sections, can be severe in certain situations, although its impact on steady-state operations with no code drops is extremely minor at simulated volumes up to 30 TPH.

While two different NFPA-compliant signal designs were modeled – a 2004 design and a 2009 design – both assumed that 30 mph was the lowest non-zero code available. With both designs, major reduction in throughput capacity occurs whenever trains operate at Restricted Speed, due to the time required to clear the long NFPA 130 compliant vent zones When this happens – typically during recovery from perturbations – simulated capacity immediately drops from 30 TPH to 26

TPH or less. Sustained operation at Restricted Speed results in throughput as low as 20 TPH, and when a circuit failure effectively lengthens the one-train-per-vent-zone restricted area to include the first circuit in the next vent zone, throughput can be as low as 15 TPH.

However, when a 20 mph speed code is added to the 2009 signal design, throughput remains at 30 TPH during recovery from perturbations. With this added assumption, all capacity measures discussed in this paper indicate that 24 scheduled TPH can easily be accommodated under the 2009 signal design, and that NFPA 130 compliance can be achieved without degrading practical capacity below the 24 TPH benchmark.

References:

International Union of Railways (UIC), "Method to be Used for the Determination of the Capacity of Lines", 1979, revised 1983.

Pachl, Joern, Railway Operation and Control, VTD Rail Publishing, 2002.

Parkinson, Tom, TRCP Report 13, Rail Transit Capacity, Transportation Research Board, 1996.