A Two Component Railway Model Exhibiting Service Collapse*

R. C. Ball[1], D. Panja[2] and G. T. Barkema[2]

September 4, 2016

[1]Dept of Physics and Centre for Complexity Science, University of Warwick, Coventry CV4 7AL, UK. r.c.ball@warwick.ac.uk.

[2]Dept. of Information & Computing Sciences, Universiteit Utrecht, The Netherlands. d.panja@uu.nl, g.t.barkema@uu.nl.

Abstract

We model a railway network in terms of the flow of trains, and of memory of how long crews have been on duty. We impose limits at stations on the throughput of trains, the duty time of crew, and the availability of fresh crew. In the regime where the traffic is limited by crew renewal, we find the system is linearly unstable with respect to spatial fluctuations leading to lengthenning queues of trains. Numerical simulations show that this instability can lead to a collapse in service in which nearly all the trains are trapped queueing at a minority of stations.

1 Introduction

The general idea that a network service becomes surprisingly prone to failure [1] when it has multiple interdependent flows has received conisderable attention in recent years, under the general theme of a Net-of-nets. We model a railway in terms of three key attributes: the infrastucture of tracks and stations, the trains running over this, and the crew staffing the trains. Service to passengers could break down due to failure in any one of these, but what we find under further model assumptions is that it is most vulnerable to a catastrophic instability associated with coupling between crew availability and train flow. This instability is associated with an inverted relationship between mean field flow rate and train numbers, which occurs when crew availability becomes the limiting factor. There exist railway network models with much higher levels of detail [2], but our prupose lies in exposing the big picture such that links to other problems beomce apparent.

*University of Warwick Research Archive Working Paper no. XXXX

We consider an interconnected network of stations which for simplicity we take to form a square grid with nearest neighbour connections. Each station i is characterised by a limiting capacity A_i being the number of trains it can admit, platform and despatch per unit time. Arrivals in excess of capacity lengthen the queues of Q_i trains held (just upstream of) each station. Each station is also presumed to have a limited supply of fresh train crew, sufficient to re-crew C_i trains per unit of time. Each train requires a unit of crew to run, and there is a limit on the time for which crew can serve until they are replaced by fresh crew: specifically we require that the trains leaving each station at any given time have crew with average time of service less than s_c. We will generally assume that $s_c > 1$ so that crew can serve for multiple stops, and correspondingly that $C_i < A_i$.

We analyse this model in a hydrodynamic limit, meaning that we work in terms of numbers of trains and fluxes of trains instead of tracking individual units. The numbers and fluxes of crew simply follow those of the trains, but we need to separately model the local totals and fluxes of accumulated crew service time.

For simplicity we work with a discrete timestep and we assume this absorbs the transit time between adjacent stations. We exactly solve the uniform case of the model and find that the fundamental diagram relating train flow to train numbers has three regimes corresponding to limitation by infrastructure, trains and crew respectively. Spatially resolved stability analysis shows that only the latter is unstable, and that it is always so. Simulations of an elaboration of the model reflect the theory, and show that the instability of the crew limited regime leads to service collapse. Simulations also show that the collapse can be triggered from the linearly stable regime by non-zero noise amplitude in the station capacities.

2 Model in Detail

We first describe the model without detail of how trains are routed, as for our theoretical calculations this is sufficient.

At time t each station receives its inbound trains which added to its inbound queue, give a total *load*

$$L_j(t) = Q_j(t-1) + I_j(t) \tag{1}$$

where $I_j(t) = \sum_{i \text{ neighbouring } j} J_{ij}(t-1)$ and $J_{ij}(t-1)$ is the number of trains despatched from i towards j at preceding time $t-1$. These loads have the convenient property that they sum to all the trains in the system. We have analogously a loading of staff-service-time comprising a queued contribution and an inflow, plus an extra one unit of service time elapsed per train in the local load:

$$L_j^S(t) = Q_j^S(t-1) + I_j^S(t) + L_j(t). \tag{2}$$

The average service time of the crew at station j is now

$$s_j(t) = L_j^S(t)/L_j(t).$$

We now assume the station seeks to maximise its outflow of trains $J_j(t)$ subject to this not exceeding either the available load $L_j(t)$ or the station train capacity $A_i(t)$, or the limit of crew service time on the outgoing trains, $s_c \geq s_j^{out} = J_j^S(t)/J_j(t)$. Making maximal use of fresh traincrew leads to an outflow of crew servicetime given by $J_j^S(t) = s_j(t)\left(J_j(t) - C_j\right)$ and hence we obtain $J_j(t) \leq C_j/\left(1 - s_c/s_j(t)\right)$ which only applies when the denominator is negative. We can collect all of these resutls together as

$$\frac{1}{J_j(t)} = \max\left(\frac{1}{L_j(t)}, \frac{1}{A_j}, \frac{1 - s_c/s_j(t)}{C_j}\right) \tag{3}$$

which correspond to being locally limited by trains, infrastructure and crew respectively.

The minimal corresponding outflow of staff service time is given by

$$J_j^S(t) = \begin{cases} 0 & J_j(t) \leq C_j \\ s_j(t)\left(J_j(t) - C_j\right) & J_j(t) > C_j \end{cases} \tag{4}$$

and we note that in the crew limited regime the latter simplifies to $J_j^S(t) = s_c J_j(t)$.

Finally we retain the balance of the loads in queues, that is

$$Q_j(t) = L_j(t) - J_j(t), \quad . \tag{5}$$

and less trivially

$$Q_j^S(t) = L_j^S(t) - s_j(t)J_j(t) \tag{6}$$

where the decrement takes account of crews refreshed. Note also that in practical simulations the same variables can be used to hold both queues and loads.

2.1 Elaboration of train routing

The outbound direction of a train will depend on its route, which for simplicity in our simulations we take to be determined solely by the inbound direction. At each station we now need to keep separate queues and loads for each inbound route, but we pool the staff on the grounds that these could be swapped between trains. We then calculate the achievable pooled outflows of trains $J_j(t)$ and staff service $J_j^S(t)$, and divide these over the outbound routes in proportion to the distribution of train loads held.

3 Mean field results and Fundamental Diagram

We now consider the case where the system is and remains homogeneous. Starting from initial values of the loads L and L^S we have conservation of the train load L and hence the only dynamical variable is L^S. The update of the staff time load is given from Eq (2) using the queue from Eq (6) as $(L^S)' = L^S - sJ + J^S + L$, which in terms of $s = L^S/L$ becomes

$$s' = s(1 - J/L) + J^S/L + 1$$

where J is determined by eqn (3) and J^S by eqn (4). For $J > C$ we can substitute $J_S = s\,(J - C)$ from Eq (4) leading to

$$s' = s(1 - C/L) + 1,$$

where given that $C/L \leq C/J < 1$, this is a stable recurrence converging to

$$s = s^* = L/C > 1.$$

This fixed point has the simple interpretation that the decrease in aggregate servicetime due to refreshing at the crew capacity $s^*C = L$ balances the increase due to servicing the load for one timestep. The only alternative scenario for the recurrence of s is $J \leq C$ for which $J_S = 0$: then assuming $C < A$ the only reason to encounter this regime is that the flux is train limited with $J = L < C$, which leads to the immediate fixed value $s' = 1$.

Having found the fixed points for s we can now find the corresponding steady state fluxes from Eqn (3) using $s = L/C$ for the crew limited regime and hence

$$\frac{1}{J} = \max\left(\frac{1}{L}, \frac{1}{A}, \frac{1}{C} - \frac{s_c}{L}\right). \tag{7}$$

This is what is known in road traffic literature as the Fundamental Diagram[3, ?], in that it tells us the traffic flux as a function of traffic density reresented here by L. In the present case the middle infrastructure regime will only arise if $A < (1 + s_c)C$; otherwise there is a direct transition from the train limited regime to the crew limited regime at $L = (1 + s_c)C$. Note that for the steady state crew limited regime this implies $s = L/C \geq 1 + s_c$, where the offset by 1 arises because the constraint on crew service time is applied when trains are despatched whereas we measure the overall average s one time unit later when trains have arrived.

4 Spatial Instability

For the crew limited regime, our uniform flow Eq (7) shows decreasing flow for increasing density of trains. For one dimensional flows such as a single carriageway road traffic, it is very well known that where such a law applies locally the flow will exhibit spatial instabilities. Whilst our model is two dimensional with more complex local response due to coupling of trains and crew, we nevertheless find instability as follows.

For simplicity we consider the highest wavevector in the first Brillouin zone for our square lattice of stations, corresponding to spatial antiphase between

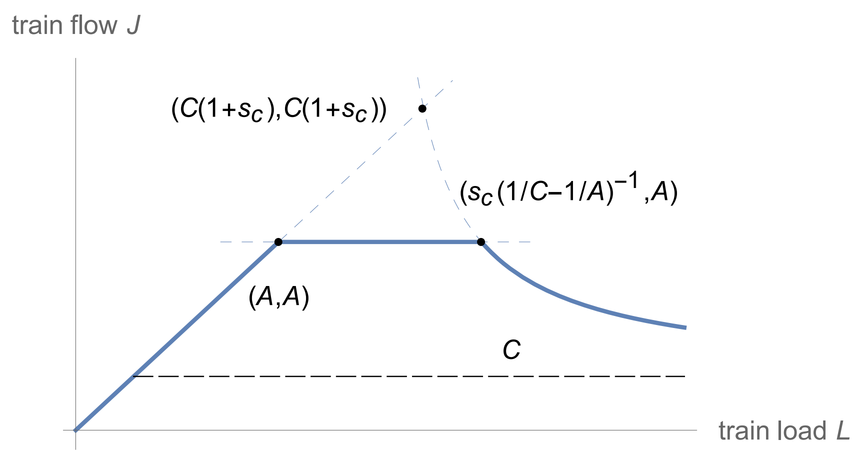


Figure 1: The Fundamental Diagram for flow vs load of trains, when all is strictly uniform. The three regimes are limited by available trains, station capacity and crew respectively from left to right, although the middle regime only occurs for station capacity $A < C(1+s_c)$. Note that the crew limited regime has negative slope of J vs L leading to the expectation that it could be unstable, as found in section 4. The dashed horinzontal line corresponds to the asymptote of the crew limited regime and is found to play the role of a tie line between the two loads with flux $J = C$.

nearest neighbour stations, a chequer board mode. We work to linear order in perturbations δL etc about the uniform solution for the crew limited regime. The analysis of this mode differs from the uniform case only in that the loads L can now have perturbation amplitudes,and that the outgoing and incoming fluxes no longer cancel as they carry different chequer board signs. The new equation for the train load pertrubation is then

$$\delta L' = \delta L - \delta J^{out} + \delta J^{in} = \delta L - \alpha \delta J$$

where δJ is the locally calculated amplitude of the outgoing flux and $\alpha = 2$ for the checker board mode with $\alpha = 0$ for a uniform perturbation as a check. The corresponding equation for the loading of crew service time is

$$\left(\delta L^S\right)' = \delta L^S - 2\alpha s_c \delta J - C\delta s + \delta L'$$

and the perturbation of the flux is given by

$$\delta J/J = -\frac{s_c}{s - s_c}\delta s/s.$$

In terms of new amplitude variables $\delta x_1 = \frac{1}{s_c}\delta L/L$ and $\delta x_2 = \delta s/s = \delta L^S/L^S - \delta L/L$ we then find $\delta x' = M.\delta x$ where the matrix of coefficients is given by

$$M_{\text{crew limited}} = \begin{pmatrix} 1 & \frac{\alpha J}{L}\frac{1}{s-s_c} \\ \frac{s_c}{s} & 1 - \frac{C}{L} + \frac{\alpha J}{L}\frac{s_c}{s}\left(\frac{1}{s-s_c} - 1\right) \end{pmatrix}.$$

It is easily checked that for $\alpha = 0$ we recover eigenvalues $1 - \frac{C}{L}$ for s and 1 corresponding to conservation of L. For the checkerboard mode we set $\alpha = 2$ and the matrix takes the form $M = I + \begin{pmatrix} 0 & p \\ q & X \end{pmatrix}$ where $pq = \frac{2J}{L}\frac{s_c/s}{s - s_c} > 0$ for $s > s_c$, and one its eigenvalues will always be bigger than unity, meaning that the mode is always positively unstable in the crew limited regime at finite load. The system approaches neutral stability as $L \to \infty$.

The above analysis is readily adapted to the train limited and station capacity limited regimes, and we find no linear instabillity in these regimes. Where the system is train limited we have

$$M_{\text{train limited}} = \begin{pmatrix} 1 - 2\alpha & 0 \\ \frac{1}{s} + \frac{2\alpha J}{L}\left(1 - \frac{1+s_c}{s}\right) & 1 - \frac{min(C,L)}{L} \end{pmatrix},$$

which is at worst neutrally stable. Where it is station limited the train fluxes are unperturbed and we have the same stability for checkerboard and uniform modes, with

$$M_{\text{station limited}} = \begin{pmatrix} 0 & 0 \\ 0 & 1 - \frac{C}{L} \end{pmatrix}.$$

The above results imply there is possible coexistence at flux $J = C$ between load $L = C$ in the train limited regime and limitingly large load $L \to \infty$ in the crew limited regime, both of which are at least neutrally stable in the analysis above. This is shown as a tie line in Fig (1)

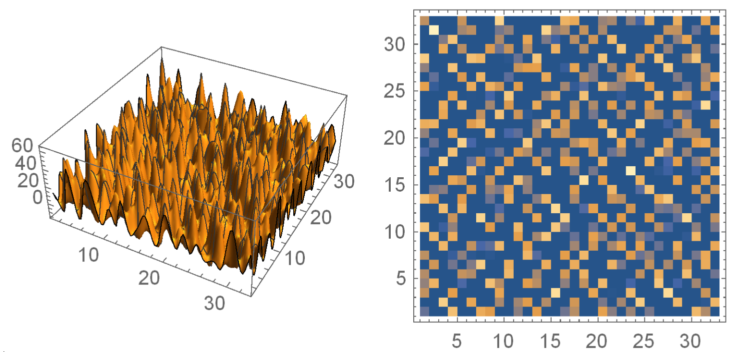
5 Simulations with Noise and Instability

To explore the full impact of instability we have explored simulations where the initial conidtion corresponds to a uniform steady state solution but the evolution included small independent random fluctuations in the staff capacity to feed instability. These were drawn independently from the uniform distribution at each time and station (white noise in time and space), with zero mean and amplitude characterised as a fraction of the mean staff capacity.

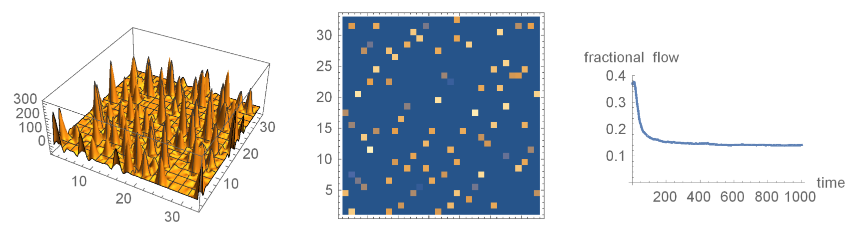
Figure 2 shows the evolution of a simulation expected to be unstable and duly proving so. The longer time behaviour shows coarsenning in which station queues become concentrated on a decreasign minority of sites. The total number of trains caught in these queues is non-decreasing but does saturate, as can be seen from the graph of the fraction of trains moving with time.

The limiting fraction of trains moving has a very simple explanation in terms of equilibrium between highly queued sites, which will emit a flux of trains $J_{limit} = C$ set by their capacity to send out trains with entirely fresh crew, and the remaining majority of stations which then operate at this same fresh crew only flow rate in a train limited manner without queues. Thus we expect the fraction flowing to be C/L and the observed asymptotes are consistent with this.

The above interpretation implies that isolated and persistently highly queued sites are not dependent on noise once established. As such we might expect that

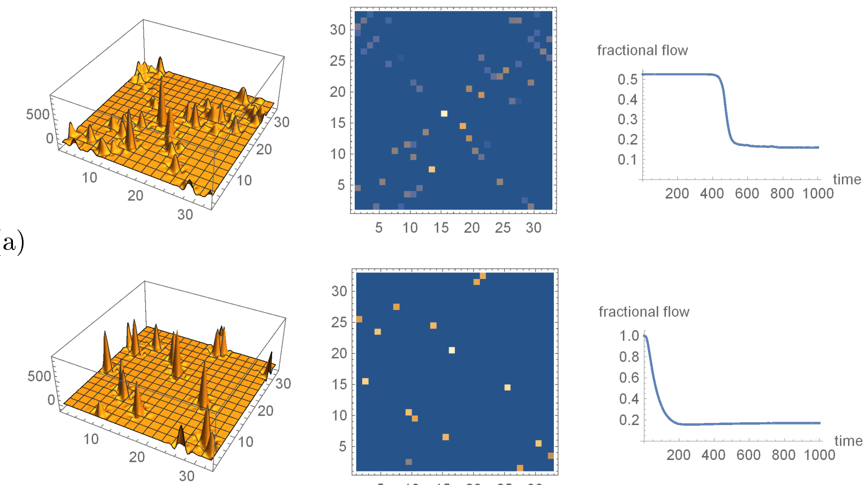


(a)



(b)

Figure 2: Unstable evolution for a simulation where the load L is 10% into the crew limited regime from a station limited regime, and with 10% noise in the station capacites. (a) shows the station queues after 30 timesteps, where in the projected view blue is low (no queue) and orange high; the influence of checkerboard instability can be seen. (b) shows such a simulation out to 1000 timesteps, together with a graph of the fraction of trains moving vs time. The queues become more concentrated on fewer stations and the fraction of trains moving levels out at a value much lower than the inital value from the uniform case. These simulations used a 33x33 grid of stations with periodic boundary conditions slewed by one unit so that trains were not trapped in a single row or column. The parameters were $s_c = 5$, $C = 2$, $A = 7$ and $L = 15.4$ where the latter is 10% into the crew limited regime.



(a)

(b)

Figure 3: (a) Late nucleation of service collapse for a simulation with load 5% below transition from station limited to crew limited. The initial fractional flow matches the uniform solution, and the later value the fresh crew only limit C/L. Parameters $s_c = 5$, $C = 2$, $A = 7$ and $L =$13.3. (b) Service collapse for a simulation with load 1.5% below transition from train limited to crew limited. Parameters $s_c = 5$, $C = 2$, $A = 70$ (effectively infinite) and $L =$11.8.

noise can nucleate the system into such a state even from a stable regime and instances of this are shown in figure 3. Nucleation appears to be harder from below the train limited to crew limited transition than for the transion from station limited, but both are observed.

Discussion

Our key result is that as the availability of train crew becomes rate limiting on how many trains can be despatched, the resulting delays to trains feed back on this unstably due to expiry of limited crew service time. This occurs despite our model fixing the number of trains, which already implies that as trains run slower due to delays the service frequency is being reduced by cancellations: this is the reason we do not see instability within the homogeneous solutions themselves.

Our model does not include some features which are important for regular service running but we would argue become unimportant as the system comes under stress and trains are running late and/or short of staff. The first is timetable keeping and any slack in train running capability which this has built in, which has effectively all gone by the stage when station capacity or crew availability limits service. A second is the availability of accumulated unused crew at stations: drawing this down could serve as a short time buffer but only the supply rate of new crew can sustain a steady state.

Some layers of management are implictly included in our model but others not at all. Our station capacities do provide a reflection of what local train management at the station level delivers in terms of track switching and platform allocation, whereas higher level interventions such as rerouting trains are not modelled. Similarly on the crew side, we have assumed a fixed distribution of the supply of fresh crew but not taken account that management might direct crew to travel to different stations where they were more needed.

The Fundamental Diagram of train flow vs train number density plays the same role in our problem as does an isotherm giving pressure vs density for particles in a fluid. In both cases negative slope is the hallmark of instability, and two phase regions emerge where states at same respectively flow and pressure can coexist, whihc is all well appreciated in the traffic literature. What is different about the train problem is that we have coexistence beween a low density bulk and islated very high density sites rather than between two bulk regions. This is because the high density phase only approaches stability as its density tends to infinity, and any extended region of it with finite density is unstable. The result might more accurately be described as a dilute microphase than a two phase region.

We have seen that modest fluctuations in the supply of crew can nucleate the service collapse regime even where the system remains on average stable, and it remains to be explored just how far the possible analogy with the role of temperature might be taken. We would further expect that the full rangle of possible collapse regime, which means all of the tie line between $L = C$ and $L = \infty$, can also be reached by system histories other than ambient fluctuations. Once reached these states are expected to be stable and hence persistent just like any thermodynamically stable two phase region, although they should coarsen over time with the dominant queues focussing on fewer sites and ulimately just one.

Acknowledgement

RCB acknowledges visitor funding from Utrecht University Complex Systems Programme which assisted the present research. The authors would also like to acknlwedge stimulating discussions with D Huisman and collaborators associated with NS Rail.

References

[1] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley and S. Havlin, Catastrophic cascade of failures in interdependent networks, Nature 464, 08932 (2010).

[2] N. Bešinović, V. Cacchiani, T. Dollevoet, R. M. Goverde, D. Huisman, M. P. Kidd, L. G. Kroon, E. Quaglietta, J. Rodriguez, P. Toth and L. Veelenturf. Integrated Decision Support Tools for Disruption Management. In RailTokyo2015: 6th International Conference on Railway Operations Modelling and Analysis, Narashino, Japan, 23-26 March 2015.

[3] Bruce D. Greenshields. A study of traffic capacity. In Highway Research Board Proceedings, volume 14, pages 448–477, 1935.