TRAVEL TIME RELIABILITY IN A NETWORK WITH DEPENDENT LINK MODES AND PARTIAL DRIVER RESPONSE

Agachai Sumalee
Research Fellow
Institute for Transport Studies
University of Leeds
36 University road
Leeds, LS2 9JT, UK
Fax: +44-113-343-5334
Email: asumalee@its.leeds.ac.uk

David Watling
Senior Lecturer
Institute for Transport Studies
University of Leeds
36 University road
Leeds, LS2 9JT, UK
Fax: +44-113-343-5334
Email: dwatling@its.leeds.ac.uk

Abstract: Previous attempts to model the impact on travel time reliability of link failures in a road network have assumed that links fail independently, and that drivers are fully adaptive to the realised conditions. In practice, causal factors may lead to dependent link failures, and not all drivers will be aware of the degraded conditions. A framework and algorithm are presented here for estimating bounds on the probability of a path travel time exceeding some threshold, for the case of dependent multi-mode link failures, where only drivers on directly affected links are adaptable (partial equilibrium). A simple illustrative example is presented.

Key Words: Network reliability, Partial equilibrium, Dependent link failure

1. INTRODUCTION

In system engineering, *reliability* may be defined as the degree of stability of the quality of service that a system normally offers (Bell and Iida, 1997). In a transport system such as a road network, there are two key elements that contribute to the level of service: the travel demand flows and the physical network. The degree of stability of the transport network system can be referred to as the ability of the network to meet the expected goals under different circumstances (e.g. variability in flows and physical network capacities). For example, it might be expected that the network should be able to cope with the variation in demand over different days of a week by maintaining a constant average travel time between different origin-destination pairs (Ang and Tang, 1990). Alternatively, it might be required that the network is able to maintain at least one possible travel path between all O-D pairs under unexpected phenomena ranging from accidents to environmental factors, which cause some link failures (Iida and Wakabayashi, 1989). Clearly, many other definitions are possible.

Such unreliability has potentially major impacts on both individuals' lives and the efficiency of the economic system (SACTRA, 1999). From an individual's viewpoint, the feeling of travelling without control over one's travel time, and hence the schedule of one's activities, is very stressful. The effect is that higher redundancy is incorporated into planned travel times, to minimise the impact of unanticipated delays on scheduled activities. This may then lead to degradation of the welfare of society as a whole. Commercial organisations also often experience the problem of unpredictability in transport costs due to the uncertainty in the transport system; hence the risk may be transferred to the price of commodities. Apart from the effect on normal day-to-day life, transport systems are an important lifeline in the event of a natural disaster, e.g. snowfall, floods, land slide, earthquake or hurricane (Du and Nicholson, 1997). While other lifelines (electricity, water supply, communication networks) can be restored in a relatively short period of time, the restoration of these systems depends highly on the accessibility to the failure points using the transportation infrastructure.

The research area of 'transport network reliability' is a relatively new but wide one, encompassing: both major disasters and more minor, recurrent events; the impacts of demand-side variation, network link variation and behavioural response; and a range of alternative measures for quantifying the abstract notion of reliability. In the present paper, we shall begin (in §2) by briefly setting out past developments in the field in the context of our assumed framework for reliability analysis. In §3, an algorithm for bounding a measure of reliability is presented for the case of multi-mode, dependent link failures. In §4, it is explained how this approached may be combined with a model of partial user response, to provide bounds on network travel time reliability. A simple example is presented in §5.

2. FRAMEWORK FOR NETWORK RELIABILITY ANALYSIS

Figure 1 illustrates the general conceptual framework adopted here for the analysis of transport network reliability. The transport network can be seen as a system in which the interaction between demand and supply in the network is the main mechanism defining the state of the network, e.g. the link flow volumes in the network. The important characteristic of this system is its exposure to various *causal* sources of variation. This variation can be an element of both demand and supply sides of the network, it can be caused by an expected or unexpected incident, and the impact can be permanent or temporary. Usually, the network supply interacts with various external factors, such as weather conditions or natural/man-made disasters, which all can cause variation in the link capacities of the network. On the other hand, the demand also fluctuates, both in the recurrent and sporadic (e.g. special event) sense, with variation both within the day and between days. All of these causal factors in the system lead to a variable service state of the network.

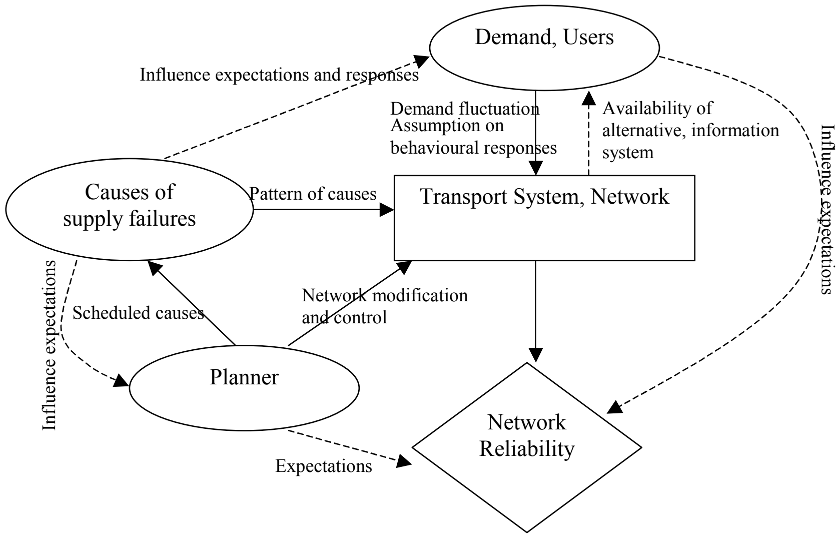

Figure 1: Conceptual framework for the analysis of transport network reliability

For any particular application of such a framework, modelling assumptions are therefore required regarding three key elements: (i) supply-side variations in terms of link failures; (ii) the response of network users to the unreliable network; and (iii) the criterion used to measure network reliability, relative to its expected function. We review the existing literature on transport network reliability below, with respect to each of these elements in turn.

2.1 Modelling of supply deterioration

The initial impetus for research on transport network reliability appears to have arisen from the study of major natural events – such as earthquakes (Bell & Iida, 1997) – affecting the 'connectivity' of the network. Each link of a network is assumed to have an independent, probabilistic, binary mode of operation (Wakabayashi & Iida, 1992; Asakura, 1996; Bell & Iida, 1997; Asakura *et al*, 2001). At one extreme, the link states might represent whether the link is 'open' or 'closed'; more generally, Bell & Iida suggest it may represent subjective definitions by the planner of the successful function of a link, such as the flow to capacity ratio being less than some given value. There are, however, a range of less severe but more recurrent causes of supply deterioration, such as minor accidents, on-street parking violations, variations in weather or lighting conditions, road maintenance or traffic signal failures, all of which would lead to effects such as reduced capacity or reduced free-speeds. More generally, then, we may view such impacts as the change of the 'mode' of each link, where a link is assumed to have multiple modes of operation (Du & Nicholson, 1997). For example, the modes of each link may represent alternative, discrete reductions in link capacity, which can be caused by some natural or man-made incidents. In the worst case, the link capacity may be reduced by 100%, which means the link has failed in a similar way to the binary mode case. However, if the link capacity has been reduced but not necessarily failed, it is commonly referred to as a *degraded* link. A limiting case of the multi-mode model is where a continuous probability distribution of link operation is assumed, as for example might be the case if degradation is explicitly linked to a continuous probability density function of link capacity (Yang *et al*, 2000; Chen *et al*, 2002).

Aside from whether the mode of the link is assumed to be binary, multi-mode or continuous, a common theme to the approaches cited above is the assumption of *statistical independence between links* in the degradation model (an exception may be found in an extension noted by Chen *et al* (2002) to their basic model, with capacities assumed to be correlated across links, whereby an alternative, Monte Carlo simulation approach is proposed). In the binary mode case, such an assumption was apparently motivated by the general treatments of reliability under independent failures in order to ease the combinatorial overhead that naturally arises with such problems in large-scale applications. While the link independence assumption may be justified in some application domains, it is unattractive in the study of transport networks since the degradation of different links may often have common underlying causes. In reality, when a link fails or is degraded, the adjacent links or the links in the same area are also often threatened by the same cause of degradation (e.g. flood, snowfall or earthquake). On the other hand, while it may increase the realism of analysis, the modelling of dependent link failure probabilities has the potential to lead to much greater computational complexity, as well as the practical problem of specifying the full correlation structure.

2.2 **Modelling of user response**

In modelling the transport network users' responses in a variable environment, a range of alternative, context-specific assumptions are possible. In severe events such as a natural disaster, the drivers' main objective is likely to be one of safety, rather than executing their planned journey. In such a case, it would be anticipated that the demand response is likely to arise from adaptive travel decisions, made as the users encounter conditions *en route*, but greatly influenced by some 'network regulator', perhaps acting according to a predefined strategy (e.g. evacuation plan and deployment of first-aid emergency vehicles). For planning in such cases specialist tools will be needed (e.g. Sheffi *et al*, 1982; Kurauchi *et al*, 2001), the discussion of which is outside the scope of the present paper. In less severe cases, it might be expected that drivers will adapt their travel choices (such as choice of route) in response to the degraded conditions, yet this clearly depends on the level of predictive information they may reasonably be expected to possess. Two main approaches may be found in the literature. In the first case, any realised degradation (whether or not its occurrence is modelled deterministically or probabilistically) is assumed to be perfectly predictable by drivers, and they thus exhibit *fully adaptive* behaviour. In such situations, a conventional Wardrop User Equilibrium (UE) or Stochastic User Equilibrium (SUE) model may be justified in representing the user response, in the sense that any such realised (degraded) state of the network is assumed to satisfy UE/SUE conditions. For the purposes of efficient implementation, the equilibrium response may be approximated via sensitivity analysis, yet the underlying philosophy remains the same. The extensive use of such an approach may be found in the transport network reliability literature (e.g. Asakura, 1996; Du and Nicholson, 1997; Bell *et al*, 1999; Berdica, 2002; Kurauchi *et al*, 2001; D'Este & Taylor, 2001; Chen *et al*, 2002).

An alternative, rather different approach is to assume that drivers exhibit *unadaptive, risk-based* behaviour. In such a case, realised degradations are not predictable by drivers, and so they cannot adapt their decisions to specific network failures. On the other hand, from their experience over time with recurrent degradations, they are assumed able to form perceptions of the risks such degradations impart (in terms of delayed travel if risk-averse; or in terms of the positive opportunity to have a chance of an unexpectedly fast journey, if risk-prone). Thus, drivers form a single long-run travel decision in the face of such degradations, trading off the risks for their alternative choice options (e.g. alternative routes). Examples of approaches following this general philosophy include those of Lo & Tung (2000), Yin & Ieda (2001), and Liu *et al* (2002).

2.3 Expected network function and reliability indices

The analysis of the reliability of a network involves measuring the ability of the network to 'function', in meeting some expected criteria. The measure of reliability is thus related to this expected function, which in turn varies according to the severity and frequency of the underlying cause. A classification into 'sporadic' and 'recurrent' causes is used in the discussion below.

In the case of *sporadic causes*, there exist various further categories depending on the state and precise nature of the cause. Taking a well-studied example, Iida *et al* (2000) propose a useful classification of the various states of functional expectation following a major earthquake event (this classification is used here to introduce techniques that nevertheless clearly have a much wider applicability than earthquake events). According to this classification, then, the first period is the 'confusion state', where the main concern is the accessibility of rescue and emergency services. In this stage, the main reliability index to be considered should be *connectivity reliability* (Bell & Iida, 1997; Asakura *et al*, 2001). The objective of evaluating this reliability index is to compute the probability that a particular path or origin-destination movement will be 'connected' (or, more generally, will 'function' as desired). The second state, according to Iida *et al*, is the 'settlement state', namely the recovering stage after a major network disruption. In this stage, the demand pattern progressively adjusts back to 'normal' level, but the network is still seriously degraded. Reliability is then concerned with the ability of the degraded network to cope with the changing demand pattern. Nicholson and Du (1997) adopt the *flow decrement level* as a suitable measure, whereby the network is said to function if the degraded network can maintain the same acceptable level of activity (represented by the number of conducted trips under the economic concept of the demand-supply equilibrium). A different approach is to look at the effect of the reduction in capacity on the change in travel time, which is again directly considered by the user. From the planning point of view, the analysis of network reliability probably tends to be in a more active direction, i.e. the network planner may design the demand control regulation over the network to ensure a level that the degraded supply can accommodate, e.g. such that link flows are within the degraded capacity (Kurauchi *et al*, 2001). The third and final state of disruption according to Iida *et al*'s classification is the 'stability state', where demand patterns have settled back to their normal level, and the reliability analysis returns to that made for recurrent causes.

The second case, namely that of *recurrent causes*, distinguishes itself from the sporadic case in that activities, and therefore travel demand patterns, carry on as usual. The immediate impact of recurrent incidents on travellers is experienced in terms of the variability in travel times. If the users are the main focus in the analysis, the variance of the disaggregate travel times (say, by O-D pair or time-of-day) may then be analysed. On the other hand, from the network operator's point of view the total travel time as a whole may be a more suitable measure for examining the overall effect on the network. In the latter case, although some OD pairs may experience a high increase in travel time, if overall the travel time in the network is still acceptable, the network is considered to function. This latter concept is consistent with the original concept of reliability analysis in a system where although some of the degradable components fail, the overall system may be still functional. In this context, Cassir (2000) defines an acceptable level of travel time to be the travel time in normal conditions, plus a safety margin.

A range of alternative measures have also been proposed to evaluate reliability impacts. From the users' perspective, Bell and Schmöcker (2002) defined the '*encountered reliability*' to be the probability of not encountering a link degradation on the used path under the traditional UE condition. This measures the likelihood of users encountering a disruption on their preferred route. On the other hand, more from the network planner's viewpoint, Chen *et al* (1999) define the '*capacity reliability*' as a measure of the ability of the network to accommodate (in terms of link flows within capacity for all links) a specific demand level. A further possibility is the measurement of '*potential reliability*' or '*vulnerability*' of a network, where the aim is to identify potential weak points/problems and their effect (Berdica, 2002; D'Este & Taylor, 2001).

3. ALGORITHM FOR ESTIMATING BOUNDS ON RELIABILITY WITH DEPENDENT LINK MODES

Section 2 was concerned with describing a general framework for reliability analysis, within which many specific form of analysis could be cast. In the present section, a particular such form of analysis is proposed. It is supposed that the links of the network have multiple modes of operation, which occur with given, pre-defined probabilities. A natural interpretation of these modes are the possible capacity states of a link. We begin in §3.1 by making the common simplifying assumption that links fail *independently* of one another. While this in principle allows a network reliability measure to be exactly computed, the combinatorial overhead in large-scale networks is potentially prohibitive. Hence, an approximation method is presented which aims to find upper and lower bounds on reliability, by using a subset of the most probable network states. This type of 'bounding' approach is motivated by the works of Wakabayashi & Iida (1992), Asakura (1996) and Du & Nicholson (1997), for the case of independent link failures. §3.1 presents an algorithm for efficiently determining these most probable states in the simplest case of independent link failures. In §3.2, it is shown how the assumption of independent link failures may be relaxed, without the need to specify a full joint distribution of link failures. For this purpose, an adapted version of the algorithm presented in §3.1 is presented. Finally, §3.2 shows how information on the most probable states, from either the §3.1 or 3.2 algorithms, may subsequently be used to compute bounds on a reliability measure.

3.1 Algorithm for selecting the *m* most probable states: independent link failures

Following the algorithm due to Chiou and Li (1986), as noted by Du & Nicholson (1997), we assume there are M links in the network labelled $i \in \{1,2,...,M\}$, and N discrete modes of operation for each link labelled $j \in \{0,1,2,...N-1\}$. The modes of operation correspond to different levels of service for each link (nodes are assumed to be always operable). Let:

$$p_{ij} = \Pr(\text{link } i \text{ operates in mode } j) \quad (i = 1,2,...,M; \quad j = 0,1,2,....,N-1) \text{ where } \sum_{j=0}^{N-1} p_{ij} = 1 \quad \forall i \ .$$

While this formulation is general in allowing more than just two (operate/fail) modes of operation, a key restriction will be made that the modes of operation are *statistically independent* between links. For each link i, it is assumed that the modes are labelled in decreasing order of probability, such that:

$$p_{i0} \geq p_{i1} \geq ... \geq p_{i,N-1} \qquad (i = 1,2,...,M) . \tag{1}$$

A particular 'state' of the *network* (rather than of just an individual link) is fully described by the modes in which all its links are operating. The state variable of the network is thus represented by an M-vector $\mathbf{S} = (S_1, S_2,..., S_M)$, where $S_i \in \{0,1,..., N-1\}$ $(i = 1,2,..., M)$. The state space of possible such state variables $\mathbf{S}$ is denoted Ω. By assumption (1) regarding the ordering of the operation mode probabilities for each link, it follows that the most probable state is $\mathbf{S} = \mathbf{0} = \{0,0,...,0\}$, which occurs with probability $\prod_{i=1}^{M} p_{i0}$. The aim here is to generate the m most probable states. The algorithm to achieve this, to be described below, works by recursively generating new states from small changes to a 'current' set of states, starting from $\mathbf{S} = \mathbf{0}$. Given any single state $\mathbf{S} \in \Omega$, and given any link $i \in \{1,2,..., M\}$ and mode $j \in \{1,2,..., N-1\}$ (neglecting mode $j = 0$), then a new state $\varphi_{ij}(\mathbf{S}) \in \Omega$ is generated by:

$$\varphi_{ij}(\mathbf{S}) = \varphi_{ij}(S_1, S_2,..., S_M) = (S_1, S_2,..., S_{i-1}, j, S_{i+1},..., S_M) \ . \tag{2}$$

In fact, rather than operating on single states, the method works by operating on sets of states, in a special way. Now consider any non-empty set of states A: $A = \{\mathbf{S}^{(1)}, \mathbf{S}^{(2)}, \mathbf{S}^{(3)},...\}$ with $\mathbf{S}^{(r)} \in \Omega$. The recursion operates by starting from $A = \{\mathbf{0}\}$, and then generating a new, larger set $A \cup \Phi_{ij}(A)$ where

$$\Phi_{ij}(A) = \{\hat{\mathbf{S}} \in \Omega \text{ such that } \exists\, \mathbf{S} \in A \text{ with } \hat{\mathbf{S}} = \varphi_{ij}(\mathbf{S})\} . \tag{3}$$

In terms of implementation, $\Phi_{ij}(A)$ is generated from A by applying $\varphi_{ij}(.)$ to each element of A in turn, while taking care not to enter duplicate entries into $\Phi_{ij}(A)$. Such duplicates occur since if the two states $\mathbf{S}^{(x)}, \mathbf{S}^{(y)} \in A$ are equal except in their i^{th} element then $\varphi_{ij}(\mathbf{S}^{(x)}) = \varphi_{ij}(\mathbf{S}^{(y)})$, meaning that the cardinality of $\Phi_{ij}(A)$ may be less than that of A. In practice, duplicates are readily detected by keeping track during the recursion of which links i have been altered in forming new states on any previous pass. The only remaining question is how to select i and j for the purpose of the recursion (2)/(3)? This is done by ordering the set of state probabilities:

$$\{p_{ij} : i = 1,2,...,M;\ j = 1,2,...,N-1\} \tag{4}$$

in decreasing order; noting that in the set above, we have neglected p_{i0} $(i = 1,2,...,M)$. Thus, if in decreasing order these probabilities are (for example) $p_{23}, p_{41},...$ then in the first round of recursion one uses $i = 2$ and $j = 3$, in the second round $i = 4$ and $j = 1$, and so on.

Three further remarks may be made about the recursive generation of states by (2) and (3). Firstly, the method is exhaustive in the sense that it will ultimately generate (if not previously terminated) all possible states. Secondly, the probability of any new state generated by (2) is readily deduced from the probability of its generating state (since links independently operate):

$$\Pr(\varphi_{ij}(\mathbf{S})) = \Pr(\mathbf{S})\frac{p_{ij}}{p_{i,S_i}}\ . \tag{5}$$

Thirdly, the number of computational operations required is reduced if the recursively generated sets A are maintained ordered by probability. This is facilitated by the fact that the transformation $\varphi_{ij}(.)$ invokes a natural ordering; that is to say, for any $\mathbf{S}^{(x)}, \mathbf{S}^{(y)} \in A$ such that $\Pr(\mathbf{S}^{(x)}) \geq \Pr(\mathbf{S}^{(y)})$ where the transformation (2) generates two new distinct states $\varphi_{ij}(\mathbf{S}^{(x)})$ and $\varphi_{ij}(\mathbf{S}^{(y)})$, then $\Pr(\varphi_{ij}(\mathbf{S}^{(x)}) \geq \Pr(\varphi_{ij}(\mathbf{S}^{(y)})$. The recursion described yields a method for ultimately generating all possible states. In order to generate only the *m* most probable states, and the recursion is embedded in a two-phase algorithm, which in summary form has the following steps:

Cycling over each of the state probabilities p_{ij} *in* (4)*, in order of decreasing probability:*

Phase 1 *Continue cycling over the* p_{ij} *in this phase, recursively generating new sets of states via the recursion defined by* (2)/(3)*. In this phase, the set of states will progressively increase in cardinality. When the set first contains* $\geq m$ *states, terminate Phase 1 and move to Phase 2.*

Phase 2 *Starting from the last set of states generated in Phase 1, first 'prune' this set to contain only the m most probable of its elements. Then continue cycling over all the remaining* p_{ij} *(not considered in Phase 1) in this phase, again generating new sets of states via the recursion defined by* (2)/(3)*. In this phase, however, each time new states are added, the set is then immediately 'pruned' to its m most probable elements.*

Although this simple algorithm exhaustively considers all $M(N-1)$ link modes, for large networks this is many orders of magnitude smaller than exhaustively considering all network states (of which there are N^M).

3.2 Algorithm for selecting the *m* most probable states: dependent link failures

The algorithm presented in §3.1 was based on the assumption that links fail independently, thus allowing probabilities of network states to be defined simply as the product of the probabilities of the constituent arc states. In the present section we see how this assumption may be relaxed to accommodate a particular form of dependence structure, while using an adapted version of the independent-link algorithm.

In particular, a 'cause-based' model and algorithm is presented due to Le and Li (1989). The idea behind this approach is that of *conditional independence*. That is to say, there are assumed to be a number of potential discrete *causes* of link failures that occur with certain probabilities, but conditionally on any given causes having occurred, links fail independently. The combined effect is that the link failures are implicitly dependent, since the causes (when they occur) simultaneously affect a number of links. This framework appears not to have been well explored in the transport context, although in many cases it is natural to relate link degradation with common external factors. In this cause-based model, the causes are in general multivariate and dependent, a number of causes may simultaneously occur. For example, we might have a number of assumed binary causes such as 'snowfall' or 'flood', and then we would need to know the joint probabilities of the combinations of these causes, namely 'snowfall but no flood', 'snowfall and flood', 'flood but no snowfall', or 'no snowfall and no flood': each such combination across all causes is known as a *scenario*. Then, given each scenario, we must define the conditional probability that each link is in each of its possible modes given that this scenario has occurred, with these conditional probabilities assumed independent between links. Le and Li (1989) propose a simple two-stage approach for determining a set of most probable states (for given m_1 and m_2):

Stage 1: Considering only scenario probabilities, identify the m_1 most probable scenarios.

Stage 2: For each scenario identified in Stage 1, identify the m_2 most probable network states (conditional on the scenario).

In the case that there are a great many possible scenarios arising from independent causes, the problem in Stage 1 may be represented as a 'network' (decision tree), whereby the algorithm presented in §3.1 may be applied. Stage 2 is implemented by applying the algorithm of §3.1 to each scenario identified in Stage 1. The total number of selected states is therefore $m = m_1 m_2$. Using standard laws of conditional probabilities, state probabilities may be computed from a combination of scenario probabilities and the conditional state probabilities given a scenario.

In practice, an adaptation to this basic technique is applied, which specifically recognises that different number of states may achieve a different total probability coverage for different scenarios. Thus, m_2 is chosen as the largest number of states likely to be required to approximate any one scenario, and within each scenario the generation of new states terminated when the total probability coverage within the scenario achieves a given level (denoted $1-\varepsilon/2$, for $0 \le \varepsilon \le 1$). A second coverage check may optionally made in Stage 2; if the scenario are considered in decreasing probability order, then the consideration of new scenarios/states is terminated when the total probability coverage *across all chosen scenarios and chosen states* considered so far achieves a level of $1-\varepsilon$. It is consistent, then, to also apply some check based on ε to the Stage 1 selection process; Le and Li suggest a test based on whether the total scenario probability exceeds $1-\frac{\varepsilon}{2}$ (greater than the value of $1-\varepsilon$ used in Stage 2, to recognise the fact that not all states per scenario will be selected).

3.3 Forming bounds on a reliability measure

The reason to compute the *m* most probable states in §3.1/§3.2 is so that upper and lower bounds on a reliability performance measure may be efficiently estimated. Following on from the notation introduced above, let us suppose that for any given feasible state $\mathbf{S} \in \Omega$ in the state-space Ω, a known indicator function describes whether or not the network is considered to be satisfactorily operable or 'reliable', in the sense of delivering some subjectively defined level-of-service:

$$f(\mathbf{S}) = \begin{cases} 1 & \text{if network state } \mathbf{S} \text{ gives a reliable service} \\ 0 & \text{otherwise} \end{cases} . \tag{6}$$

For example, we may be purely interested in 'connectivity' of an origin-destination (O-D) movement, in which case $f(\mathbf{S}) = 1$ only if that O-D pair is connected. Each link only then needs two modes, a failure (disconnection) mode, and an aggregate mode representing all operable states. This case was considered by Wakabayashi & Iida (1992) for the special case of independent link failures. In the present paper, however, we shall be more interested in the notion of *travel time reliability*, in which case a network performance and user behaviour model is required to map **S** to some corresponding travel time state, with $f(.)$ then a measure of whether such a travel time state is subjectively acceptable. The precise mechanism for achieving this will be presented in §4. Based on the general definition above, the expected reliability is thus:

$$R = \sum_{\mathbf{S} \in \Omega} f(\mathbf{S}) \ \Pr(\mathbf{S}) \tag{7}$$

Having applied the algorithm in §3.1/§3.2, we then know the m most probable states (for some given m), the set of which we may denote $\hat{\Omega}_m \subseteq \Omega$. A lower bound on R is then obtained simply by evaluating those elements in the summation of R for the m most probable states, and making the pessimistic assumption that remaining states are not reliable (that is to say, assume $f(\mathbf{S}) = 0$ for $\mathbf{S} \notin \hat{\Omega}_m$). Likewise, an upper bound follows from the optimistic assumption that the remaining states are all reliable, i.e. supposing $f(\mathbf{S}) = 1$ for $\mathbf{S} \notin \hat{\Omega}_m$. Thus we can evaluate bounds as:

$$\sum_{\mathbf{S} \in \hat{\Omega}_m} f(\mathbf{S}) \Pr(\mathbf{S}) \ \le \ R \ \le \ \left(1 - \sum_{\mathbf{S} \in \hat{\Omega}_m} \Pr(\mathbf{S}) \right) + \sum_{\mathbf{S} \in \hat{\Omega}_m} f(\mathbf{S}) \Pr(\mathbf{S}) \ . \tag{8}$$

4. PARTIAL USERS' EQUILIBRIUM

Based on the general technique outlined in §3.3, a method for evaluating travel time reliability (on a path, O-D, or network level) may then be formulated. It is supposed that the network state vector **S** parameterises some network user behaviour model. This user behaviour model provides as output a travel time $t(\mathbf{S})$ (on an appropriate path, O-D, or network level) corresponding to the state **S**. Given some subjective definition of a maximum acceptable travel time τ, the binary reliability function (6) is thus defined as:

$$f(\mathbf{S}) = \begin{cases} 1 & \text{if } t(\mathbf{S}) \le \tau \\ 0 & \text{otherwise} \end{cases} .$$

In that case, interpreting **S** now as a random variable, the expected reliability (7) is then in fact the probability $\Pr(t(\mathbf{S}) \le \tau)$. The only remaining element to define is then the network user behaviour model. As noted in the review in §2, the great majority of approaches to transport network reliability assume *fully adaptive behaviour*, whereby $t(\mathbf{S})$ would effectively be evaluated by running a user equilibrium (UE) model based on the realised state vector **S**. While this would clearly be possible in the present context, we present below a model which assumes only *partially adaptive behaviour*. The precise form of this model is described below.

In the *Partial User Equilibrium* model, it is supposed that only a pre-defined subset of users in the network can realise the degraded condition of the network ("the affected users") and, thus, react to that change by diverting to alternative routes. In order to define the subset of affected users, a standard UE model is first run on the undegraded travel conditions (or some other nominal conditions). This yields a set of used paths, which when some state vector **S** is defined allows the identification of a subset of affected paths $\Psi(\mathbf{S}) \subseteq \Pi$ (i.e. a subset of used paths in the undegraded situation that use a degraded link under **S**, where Π denotes the full set of acyclic paths). It is noted that since UE path flows are in general non-unique, then $\Psi(\mathbf{S})$ is not a uniquely-defined set; while we do not consider it here, one way to circumvent this problem is to choose a set of the most likely paths from the original UE path flow solutions (see, for example, Larsson *et al*, 1998).

Once a subset of affected path flows has been determined, the Partial UE model may be formulated. The basic assumption is that the flows on unaffected paths (i.e. those not in $\Psi(\mathbf{S})$) are fixed at their undegraded level, while drivers on the affected paths aim to find a new UE based on the realised state **S**. Thus with Ψ denoting the subset of affected paths (the dependence on **S** will be dropped for simplicity of notation), the subset of unaffected paths is denoted $\Theta = \{q | q \in (\Pi - \Psi)\}$. The path flows from the undegraded UE on the unaffected path are denoted by $\tilde{F}$. Thus, by fixing the flows on unaffected paths and assigning the affected path flows to a new UE given these fixed flows, the condition for the Partial UE model is:

$$\begin{aligned} & F_p \cdot \left(\sum_{j=1}^{j} \left(\delta_{jp} \cdot c_j \left(\sum_k F_k \cdot \delta_{jk} + \sum_q \tilde{F}_q \cdot \delta_{jq} \right) \right) - \mu_i \right) = 0 \qquad \forall p,k \in \Psi; \forall q \in \Theta \\ & F_p \geq 0, \left(\sum_{j=1}^{j} \left(\delta_{jp} \cdot c_j \left(\sum_k F_k \cdot \delta_{jk} + \sum_q \tilde{F}_q \cdot \delta_{jq} \right) \right) - \mu_i \right) \geq 0 \end{aligned} \tag{9}$$

where F_p denotes the flow on path p, $c_j(.)$ is the travel cost-flow function for link j, μ_i is the minimum O-D travel cost for movement i, and δ_{jp} is a 0/1 indicator variable equal to 1 only if link j is part of path p. The existence and uniqueness of the such a Partial UE solution in terms of aggregate link flows follows the results in Smith (1979) for the standard UE model, since the only modification to the original UE problem is the addition of constants in terms of fixed link flows (from the paths in $\Theta = \{q | q \in (\Pi - \Psi)\}$) in the link cost functions, thus preserving the monotonicity assumption of the link cost function. In this paper, the assumption of separable link cost functions is used. Thus, the complementarity condition above can be reformulated as an optimisation program. The feasible space of the re-assigned path flow ($F_p; p \in \Psi$) and its related feasible link flow space for this optimisation program are closed, bound and convex. Thus, from the Weierstrass theorem, there exists a solution to the optimisation program (see Theorem 2.3 in Bazaraa *et al*, 1993). In addition, with the assumption of strict monotonicity of the link cost function, this optimal solution (in terms of link flows) is unique. The mathematical formulation of the optimisation program is:

$$\min_{\hat{\mathrm{v}}} \sum_j \int_0^{\hat{v}_j} c_j(x)\,dx$$

$$s.t.\;\; \hat{v}_j = \sum_p F_p \cdot \delta_{jp} + \sum_q \tilde{F}_q \cdot \delta_{jq} \qquad p \in \Psi; q \in \Theta; \forall j \in J$$

$$\sum_p F_p \cdot \Delta_{pi} = \sum_q \tilde{F}_q \cdot \Delta_{qi} \qquad \forall i \in I; p \in \Psi; q \in \Theta \tag{10}$$

$$F_p \geq 0 \qquad \forall p \in \Psi$$

The algorithm for solving the traditional traffic assignment can be used to solve this partial equilibrium problem, e.g. Frank-Wolfe algorithm. It should be noted that the traditional equilibrium condition in terms of generalised travel time amongst paths between an O-D pair might not be preserved anymore. Only the OD pair with at least one affected path will re-achieve the normal UE condition. Clearly, there is no condition to guarantee the UE condition for those unaffected O-D pairs (all paths for that O-D pair are not affected by the incident). This is the underlying intent for the construction of the partial UE. We argue that by reassigning the whole traffic to the new UE after the network is disturbed, one may run the risk of underestimating the effect of the incident across the network. In the partial UE case presented here, the information on the change of the network condition is assumed to disseminate only to travellers using the affected paths. This assumption replicates the fact that under such a abnormal conditions, most travellers on unaffected paths carry on their journeys based on their average experience of the network, whereas the indirect effect of changes in traffic conditions will not alter their behaviour or route choice. On the other hand, the travellers on the affected paths are clearly more aware of the incident and try to seek the next best route for their journeys. It should be noted that the improvement in disseminating the traffic information in the wider area of the network to wider groups of travellers (not only the information about the affected routes) may decrease the plausibility of this partial UE condition.

5. A NUMERICAL EXAMPLE

In this section, a simple network with five directed links and two O-D pairs is tested. The link travel time functions for each link, following the BPR standard function, are annotated in Figure 2. Two O-D pairs exist, from A-C and B-C. The traffic demands between these two O-D pairs are 200 pcu/hr and 100 pcu/hr respectively. Figure 2 illustrates the network structure where the degradable components are the directed links, labelled 1 to 5. In the normal condition (full UE with normal link capacities), there are five paths used between two O-D pairs in the network. The traffic condition on each path at the fully UE solution is shown in Table 1. To ease the explanation, we will refer to these paths as: Path 1: A-link1-link4-C, Path 2: A-link2-link5-C, Path 3: A-link3-link3-link4-C, Path 4: B-link5-C, and Path 5: B-link3-link4-C.

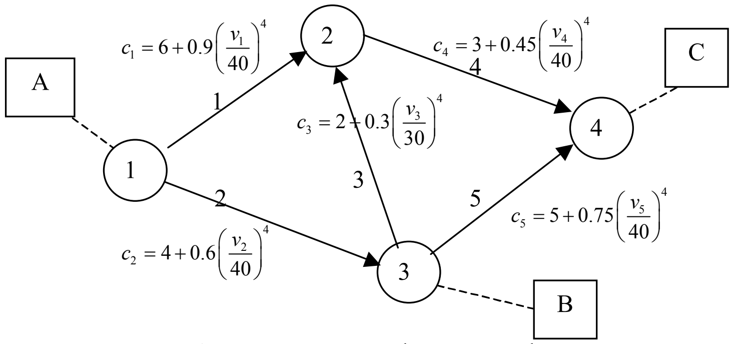

Figure 2: Example Network

Table 1: Traffic conditions in the undegraded state of the network

Path	Flow (pcu/hr)	Travel time
1	95.11	151.54
2	73.11	151.52
3	31.78	151.54
4	67.38	119.15
5	32.62	119.17

For simplicity, two possible causes of link degradation are considered, labelled causes 1 and 2. Table 2 shows the scenario probabilities. Each link can operate in three modes: fully operate (mode 1), degraded (mode 2) or fail (mode 3). The degraded mode is referred to as the condition in which the capacity of a link is reduced by 50% of its normal operational capacity. The fail mode is when the link is fully closed. The possibilities of the effects on the components from each cause are specified in Table 3.

Table 2: Scenario Probabilities

Scenario *X*	Probability	Coverage
(0 0)	0.855	0.855
(1 0)	0.095	0.950
(0 1)	0.045	0.995
(1 1)	0.005	1.000

(Element *i* of *X* = 1 if cause *i* occurs)

Table 3: Link Mode Probability Matrices

Cause	Link				
	1	2	3	4	5
1	(0.65,0.25,0.10)	(0.60,0.20,0.20)	(0.55,0.25,0.20)		
2			(0.65,0.20,0.15)	(0.65,0.15,0.20)	(0.65,0.20,0.15)

Entry located in row *i*, column *j* relates to p_{ij} (0.65,0.25,0.10) indicates that $p_{ij}(1) = 0.65$, $p_{ij}(2) = 0.25$, and $p_{ij}(3) = 0.10$. A blank entry means that cause *i* has no effect on component *j*.

The effect of a reduction in link capacity or link failure will be evaluated by comparing the modified travel time on each path under *Partial UE condition*, with different acceptable upper bounds on travel time relative to the undegraded condition as defined in Table 4. The network will be considered functional if the travel times on all used paths are lower than the acceptable travel time. In this test, we simply set three different criteria, corresponding to a 20%, 50%, and 100% increase in travel time on each path, compared to the undegraded condition shown in Table 1. The performance measures for these three cases are shown in Table 4. The interpretation of the different acceptable criteria of travel time shown is the different level of network performance expected. For example, if the travel time reliability index is 0.95 with 20% upper bound of the path travel time, it means under uncertain link capacity defined in Table 3, the possibility of all travellers in the network to carry out their trips with the travel time within 120% of the normal condition travel time will be about 95%.

Table 4: Acceptable travel time on each path

Path	Upper bound for the path travel time		
	Case		
	20%	50%	100%
1	181.85	227.31	303.08
2	181.82	227.28	303.04
3	181.85	227.32	303.09
4	142.98	178.72	238.30
5	143.01	178.76	238.35

In applying the method of §3.2, a value of m_1 = 3, m_2 = 10, and $\varepsilon = 0.05$ is used. For the scenario space (see Table 2), by inspection the first three scenarios achieve the desired probability coverage of $1-\frac{\varepsilon}{2}=0.975$. The number of states subsequently generated is 28 (out of 243), giving an actual accumulated coverage of 0.979 (see Table 5). The number in each column for each row defines the mode of the corresponding link in that state of the network. For example, with the state number 2 link 1, 2, 4, and 5 operates normally (mode 1) and link 3 is degraded by 50% reduction of the capacity (mode 2). From Table 5, we only have to test 28 states of the network to cover about 98% of the possibilities of the network states. After testing these set of states with the partial equilibrium, Table 6 gives a summary of the approximated upper and lower bounds for network reliability for the three cases. The interpretation of the results in this table is that if we allow at most 20%, 50%, and 100% increase of the normal travel time between each O-D pair, the probability that all travellers in the network with uncertain link performance (defined in Table 3) will carry out their trips under the specified acceptable travel times is between 90.1%-92.2%, 91.8%-93.9%, and 92.1-94.2% respectively. Figure 3 gives an example of the convergence of the approximation of the network reliability indices, for the case of 100% tolerated increase in path travel times.

Table 5: Most probable states generated and the accumulated coverage probability

Number of states	**Mode of the link**					**Cumulative probability**
	Link 1	Link 2	Link 3	Link 4	Link 5	
1	1	1	1	1	1	0.888
2	1	1	2	1	1	0.901
3	2	1	1	1	1	0.909
4	1	1	3	1	1	0.916
5	1	2	1	1	1	0.923
6	1	3	1	1	1	0.930
7	1	1	1	1	2	0.933
8	1	1	1	3	1	0.937
9	3	1	1	1	1	0.940
10	1	2	2	1	1	0.943
11	1	3	2	1	1	0.947
12	1	1	1	2	1	0.949
13	1	1	1	1	3	0.952
14	1	1	3	1	1	0.955
15	2	1	2	1	1	0.958
16	2	1	3	1	1	0.961
17	2	2	1	1	1	0.963
18	2	3	1	1	1	0.966
19	1	2	3	1	1	0.968
20	1	3	3	1	1	0.971
21	1	1	2	3	1	0.972
22	1	1	1	3	2	0.973
23	1	1	2	1	2	0.974
24	1	1	1	2	2	0.975
25	1	1	3	1	2	0.976
26	1	1	1	3	3	0.977
27	1	1	2	2	1	0.978
28	1	1	2	1	3	0.979

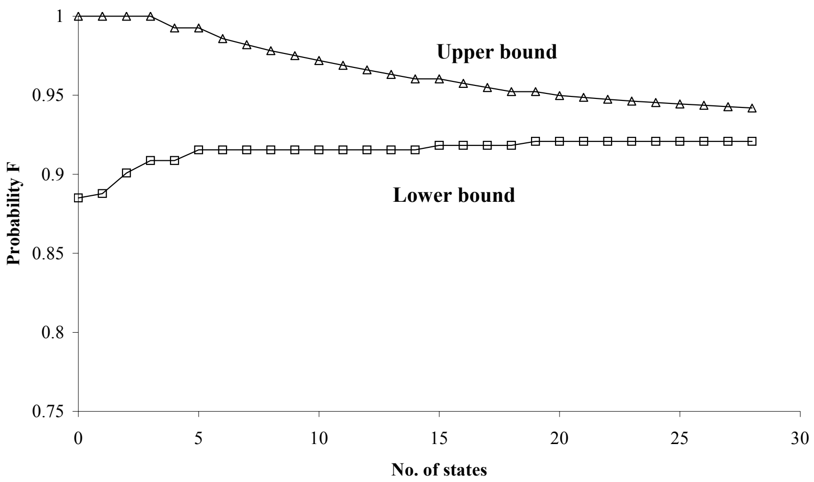

Figure 3: Convergence of approximation (100% tolerance on travel time increases)

Table 6: Summary of upper and lower bounds and error approximation

Number of states	Travel time increase 20%			Travel time increase 50%			Travel time increase 100%		
	Upper bound	Lower bound	Relative error (%)	Upper bound	Lower bound	Relative error (%)	Upper bound	Lower bound	Relative error (%)
1	1	0.888	*11.2*	1	0.888	*11.2*	1	0.888	*11.2*
2	1	0.901	*9.9*	1	0.901	*9.9*	1	0.901	*9.9*
3	0.992	0.901	*9.2*	1	0.909	*9.1*	1	0.909	*9.1*
4	0.985	0.901	*8.5*	0.993	0.909	*8.5*	0.993	0.909	*8.5*
5	0.978	0.901	*7.9*	0.993	0.915	*7.9*	0.993	0.915	*7.9*
6	0.971	0.901	*7.2*	0.986	0.915	*7.2*	0.986	0.915	*7.2*
7	0.967	0.901	*6.8*	0.982	0.915	*6.8*	0.982	0.915	*6.8*
8	0.964	0.901	*6.5*	0.978	0.915	*6.4*	0.978	0.915	*6.4*
9	0.96	0.901	*6.1*	0.975	0.915	*6.2*	0.975	0.915	*6.2*
10	0.957	0.901	*5.9*	0.972	0.915	*5.9*	0.972	0.915	*5.9*
11	0.954	0.901	*5.6*	0.969	0.915	*5.6*	0.969	0.915	*5.6*
12	0.951	0.901	*5.3*	0.966	0.915	*5.3*	0.966	0.915	*5.3*
13	0.949	0.901	*5.1*	0.963	0.915	*5*	0.963	0.915	*5.0*
14	0.946	0.901	*4.8*	0.96	0.915	*4.7*	0.96	0.915	*4.7*
15	0.943	0.901	*4.5*	0.96	0.918	*4.4*	0.96	0.918	*4.4*
16	0.94	0.901	*4.1*	0.957	0.918	*4.1*	0.957	0.918	*4.1*
17	0.937	0.901	*3.8*	0.955	0.918	*3.9*	0.955	0.918	*3.9*
18	0.935	0.901	*3.6*	0.952	0.918	*3.6*	0.952	0.918	*3.6*
19	0.932	0.901	*3.3*	0.95	0.918	*3.4*	0.952	0.921	*3.3*
20	0.93	0.901	*3.1*	0.947	0.918	*3.1*	0.95	0.921	*3.1*
21	0.929	0.901	*3.0*	0.946	0.918	*3.0*	0.949	0.921	*3.0*
22	0.927	0.901	*2.8*	0.945	0.918	*2.9*	0.947	0.921	*2.7*
23	0.926	0.901	*2.7*	0.944	0.918	*2.8*	0.946	0.921	*2.6*
24	0.925	0.901	*2.6*	0.943	0.918	*2.7*	0.945	0.921	*2.5*
25	0.925	0.901	*2.6*	0.942	0.918	*2.5*	0.945	0.921	*2.5*
26	0.924	0.901	*2.5*	0.941	0.918	*2.4*	0.944	0.921	*2.4*
27	0.923	0.901	*2.4*	0.94	0.918	*2.3*	0.943	0.921	*2.3*
28	**0.922**	**0.901**	***2.3***	**0.939**	**0.918**	***2.2***	**0.942**	**0.921**	***2.2***

6. CONCLUDING REMARKS

It has been demonstrated that it is possible to formulate a cause-related version of the travel time reliability problem, relaxing the usual assumptions of independent link failures and fully adaptive driver response. A potentially efficient algorithm has been identified for estimating bounds on a reliability measure, avoiding the combinatorial overhead of enumerating all possible states, and avoiding the estimation uncertainty of Monte Carlo methods. At present, only a small network test has been presented, clearly the real computational test of the method will be in larger scale problems. Additionally, efficiency comparisons should be made with alternative solution methods, ranging from alternative 'most probable states' algorithms to a fully Monte Carlo based method. From a modelling perspective, a useful comparison could be made of the impact on reliability of alternative response assumptions, eg partial UE, standard UE, and risk-related UE. Finally, the method could be extended in a number of ways, for example to accommodate O-D demand variation as one of the causal scenario elements.

ACKNOWLEDGEMENTS

This research is supported by the UK Department of Transport under the New Horizons research programme.

REFERENCES

Ang, A.H.S and Tang, W.H. (1990) **Probability Concepts in Engineering Planning and Design, vol.1 Basic Principle; vol. II Decision, Risk and Reliability**, Wiley, New York.

Asakura Y, Hato E & Kashiwadani M (2001) Stochastic Network Design Problem: An Optimal Link Improvement Model for Reliable Network, **1st INSTR**, Kyoto, Japan.

Asakura Y (1996). Reliability measures of an origin and destination pair in a deteriorated road network with variable flows, in Bell M.G.H. (Eds.), **Transportation Networks: Recent Methodological Advances**, Pergamon Press, Oxford, 273-288.

Bazaraa, M.S., Sherali, H.D. and Shetty, C.M. (1993) **Nonlinear Programming: Theory and Algorithms**, John Wiley&Sons, Inc, New York

Bell, M.G.H. and Schmöcker, J.-D. (2002) Public transport network reliability: Topological effects. **3rd ICTT**, Guilin, China.

Bell MGH, Cassir C, Iida Y & Lam WHK (1999) A sensitivity-based approach to network reliability assessment, **Proceedings 14th ISTTT**, Jerusalem, 283-300.

Bell MGH & Iida Y (1997) **Transportation Network Analysis**, John Wiley & Sons, Chichester, UK.

Berdica K (2002) An introduction to road vulnerability: what has been done, is done and should be done, **Transport Policy 9(2)**, 117-127.

Cassir, C. (2001) A Flow Model for the Analysis of Transport Network Reliability, PhD Thesis, University of Newcastle, United Kingdom.

Chen A, Yang H, Lo HK & Tang WH (2002) Capacity reliability of a road network: an assessment methodology and numerical results, **Transportation Research 36B(3)**, 225-252.

Chiou, S.N. and Li, O.K. (1986) Reliability Analysis of a Communication Network with Multimode Components, **IEEE Journal on Selected Areas in Communications, 4(7)**, 1156-1161.

D'Este GM & Taylor MAP (2001) Network vulnerability: an approach to reliability analysis at the level of national strategic transport networks, **1st INSTR**, Kyoto, Japan.

Du ZP & Nicholson AJ (1997), Degradable transportation systems: sensitivity and reliability analysis, **Transportation Research 31B(3)**, 225-237.

Iida, Y., Kurauchi, F., and Shimada, H. (2000) Traffic Management System Against Major Earthquakes, **IATSS Research, 24(2)**.

Iida, Y. and Wakabayashi, H. (1989) An approximation method of terminal reliability of a road network using partial minimal path and cut set, **Proceeding of the Fifth WCTR**, Yokohama, Japan.

Kurauchi F, Iida Y & Shimada H (2001) Evaluation of road network reliability considering traffic regulation after a disaster, **1st INSTR**, Kyoto, Japan.

Larsson, T., Lundgren, D., Patriksson, M., and Rydergren, C. (1993) An algorithm for finding the most likely equilibrium route flows, presented at **TRISTAN III**, San Juan, Puerto Rico.

Le, K.V. and Li, O.K. (1989) Modeling and Analysis of Systems with Multimode Components and Dependent Failures, **IEEE Transactions on Reliability, 38(1)**.

Liu HX, Ban X, Ran B & Mirchandani P (2002) An Analytical Dynamic Traffic Assignment Model with Probabilistic Travel Times and Perceptions, **81st annual TRB meeting**, Washington, USA.

Lo H & Tung YK (2000) A Chance Constrained Network Capacity Model. In **Reliability of Transport Networks** (eds. M. Bell & C. Cassir), Research Studies Press Ltd., 159-172.

Nicholson A & Du ZP (1997) Degradable transportation systems: an integrated equilibrium model, **Transportation Research 31B(3)**, 209-224.

SACTRA (1999) **Transport and the Economy,** The Stationary Office, UK.

Sheffi Y, Mahmassani HS & Powell W (1982) A transportation network evacuation model, **Transportation Research 16A(3)**.

Smith, M.J. (1979) The existence, uniqueness and stability of traffic equilibrium, **Transportation Research 13B**, 295-304

Wakabayashi, H. and Iida, Y. (1992). Upper and lower bounds of terminal reliability of road networks: an efficient method with Boolean algebra. **Journal of Natural Disaster Science**, **14**, 29-44.

Yang H, Lo HK & Tang WH (2000). Travel time versus capacity reliability of a road network. In *Reliability of Transport Networks* (eds. M. Bell & C. Cassir), Research Studies Press Ltd.

Yin Y & Ieda H (2001). Assessing peformance reliability of road networks under nonrecurrent congestion, **80th annual TRB meeting**, Washington, USA.