Transport Network Reliability Measurement and Analysis

A. Nicholson
University of Canterbury, Christchurch, New Zealand

RESUMO: Este artigo revisa as pesquisas recentes sobre a confiabilidade de redes de transportes. São discutidas as fontes da falta de confiabilidade e a praticidade de melhorar a confiabilidade através do uso da capacidade ociosa. Discute-se as diversas definições e medidas de confiabilidade e também as técnicas para a modelagem da confiabilidade de redes de transporte. Descreve-se a aplicação de técnicas de avaliação e gerenciamento de risco a um estudo de caso dos efeitos de desastres naturais e suas remediações. Conclui-se que a importância da confiabilidade de redes provavelmente continuará a crescer e que existe a necessidade de conduzir pesquisas futuras para identificar as medidas e técnicas apropriadas de confiabilidade, levando em consideração as expectativas dos usuários do sistema de transportes.

ABSTRACT: This paper reviews recent research on the reliability of transport networks. The sources of unreliability and the practicability of improving reliability through providing spare capacity are discussed. Alternative definitions and measures of reliability are described, along with techniques for modelling transport network reliability. The application of risk evaluation and management techniques is described via a case study of the effects of natural hazards and their mitigation. It is concluded that network reliability is likely to continue growing in importance, and that there is a need for further research to identify appropriate measures of reliability and modelling techniques, taking account of the expectations of users of transport systems.

1 INTRODUCTION

Interest in transport network reliability is not new, with graph theory concepts being used over 40 years ago to assess the reliability of the USA Interstate Highway system (Garrison, 1960). During the last decade or so, however, there has been a growing interest in the topic. This has been prompted by studies (e.g. Parkhurst et al. (1992)) showing that transportation system users attach considerable importance to the quality of service, which embraces a wide range of service attributes, with one of the most important being reliability. Parkhurst et al. report that transportation system users commonly mentioned unreliability, and the consequent variability and unpredictability of travel times, as a negative service attribute. Transportation system users include the shippers and carriers of goods as well as travellers. To reach their destinations in time, travellers must commence their trips earlier than if the travel times were less variable and unpredictable. Given the trend towards just-in-time production methods, which involve reducing the space and investment associated with goods storage and relying upon fast and reliable delivery, the economies of developed countries are becoming increasingly dependent upon their transportation systems being reliable.

Some countries are particularly vulnerable to natural hazards (e.g. earthquakes, storms, volcanic eruptions), with the water supply, energy supply, sewage disposal, communication and transportation systems (commonly called lifeline systems) being prone to disruption due to infrastructure damage. This has led to the emergence of lifeline engineering, which involves assessing and reducing the vulnerability of lifeline systems to damage and disruption during major disasters. A study of the effect of earthquakes on lifelines in Wellington (the capital of New Zealand) considered the interdependence of lifeline systems, and found the transportation system to be the most important (Hopkins et al., 1991). This is because the restoration of virtually all other lifeline systems is very dependent upon people and equipment being able to move to the sites where damage has occurred, and damage to the transportation system thus inhibits repairs to the other lifeline systems.

The emphasis in lifeline engineering has been upon reducing the direct costs of repairs to lifeline systems...
by relocating services and strengthening components, with little attention to the increase in user costs during periods of disruption. However, a study of the costs of the failure of the road bridge over a river bisecting a town in New Zealand after a major storm found that the readily estimated indirect costs (e.g., construction of temporary bridges to carry services and pedestrians, operation of a ferry service, upgrading of an alternative route for diverted vehicle traffic) were 50% greater than the direct costs of replacing the bridge and its approaches (Works Consultancy Services, 1990). The study did not identify the increased user costs.

Yee et al. (1996) studied how the 1994 Los Angeles (Northridge) earthquake, which closed highways that carried some of the highest daily traffic volumes in the world, affected user costs. They found that the cost of motorist delay associated with the closure of just one facility (Interstate Highway 10) was almost US$1 million per day, even after detours, car-pool lanes, and rail and bus service enhancements had been established.

This estimate of the increased user cost excluded the socio economic costs due to the disruption of commercial traffic movements and business, and these costs are likely to be much greater than the direct cost of replacing damaged infrastructure. For instance, the business interruption losses resulting from the collapse of the World Trade Center buildings have been estimated (Münchener Rück, 2001) to “far exceed” the cost of the property losses (i.e., the cost of replacing the structures and the equipment and supplies in the structures).

There are various methods for improving network reliability (Nicholson and Du, 1994) including:

1. improving component reliability (e.g., replacing or strengthening bridges);
2. improving the network configuration (e.g., constructing new links);
3. having stand-by components, which are activated after degradation of the original component (e.g., Bailey bridges, emergency air-ferry services);
4. monitoring critical components, to detect degradation and advise users of alternatives;
5. undertaking regular preventive maintenance;
6. identifying priorities for repairing degraded components to minimise the socio-economic impacts, and optimally deploying resources for repair work.

Lifeline engineering has traditionally focused upon the first option, but Goodwin (1992) suggested pursuing the second option. He proposed the concept of a “quality margin” in transport, akin to a margin of safety, suggesting that transport planners “should deliberately allow for spare capacity in the system, some redundancy — some inefficiency, in a sense — in order to enjoy benefits that are not measured by maximum production”. The practicability of increasing the reliability of a transport system by providing spare capacity is discussed later.

Goodwin (1992) also noted that existing transportation planning models pay little or no attention to transportation system reliability, or many of the other attributes that determine the quality of service; they almost invariably focus on the travel time or distance, or a weighted sum of these. Such models generally assume that components in a transportation system never fail (i.e., they always operate at their initial capacities), and the network structure and the attributes of network components are fixed. Any change in network structure or component attributes requires the model to be run again, in order to assess the effect of the change. Goodwin suggested that to analyse transportation system reliability efficiently, models that allow explicitly for such changes are needed.

In recent years, several measures of reliability and several techniques for reliability modelling have been proposed. This paper describes and discusses these, after first discussing the sources of unreliability and the practicability of improving reliability via spare capacity.

2 SOURCES OF UNRELIABILITY

Nicholson and Du (1997) suggest unreliability can be considered to arise from two distinctly different sources: flow variations and capacity variations. Figure 1 shows that for an arc with capacity $x_{a0}$, the travel time varies as a result of arc flow variation. The travel time varies about $t_{a2}$ (corresponding to an arc flow $v_{a} = v_{a2}$), between a lower bound $t_{a3}$ (corresponding to the lower bound arc flow $v_{a3}$) and an upper bound $t_{a1}$ (corresponding to the upper bound arc flow $v_{a1}$). Figure 2 shows that for an arc with flow $v_{a}$, the travel time varies as a result of arc capacity variation. The travel time again varies about $t_{a2}$ (corresponding to an arc capacity $x_{a} = x_{a2}$), between a lower bound $t_{a3}$ (corresponding to the upper bound arc capacity $x_{a0}$) and an upper bound $t_{a1}$ (corresponding to the lower bound arc capacity $x_{a2}$).

![Figure 1. Arc Flow Variation](image-url)
In reality, travel time variation can arise from both sources, and it is not always an easy matter to identify the separate effects of flow and capacity variations.

The main focus of transportation network reliability research has been upon reducing the impact of arc capacity variations. This is probably because there are authorities that are responsible for managing transport networks and are expected to minimise the frequency and consequence of such events. Travel time variations associated with variations in travel demand are the result of decisions made by many individual travellers, and are thus less amenable to reduction via direct intervention.

3 PRACTICABILITY OF SPARE CAPACITY

Queueing theory has been applied to the analysis of traffic flow at intersections and along links for many years (e.g. Gerlough and Huber (1975)). However, the emphasis has been upon estimating the mean travel time or delay for users. One of the simplest queueing theory models is the M/M/1 model, where traffic arrives according to a Poisson process (i.e. the headways between vehicles are negative exponentially distributed), the service times (e.g. the times spent at the head of the queue before a suitable gap occurs in the priority flow) are also negative exponentially distributed, and there is one service channel (i.e. one approach lane for the non-priority flow). For such a system, the total travel time (i.e. the time from joining the queue until being able to join or cross the priority flow) will vary according to the negative exponential distribution, where the parameter is the mean service rate minus the mean arrival rate (Wolff, 1988). That is, the mean travel time equals

\[ \frac{1}{(1-\bar{e})} \]  

and the variance of the travel time equals

\[ \frac{1}{(1-\bar{e})^2} \]  

where \( \bar{e} \) and \( \bar{i} \) are the mean arrival and service rates, respectively, and \( \bar{e} < 1 \). As the mean arrival rate increases and approaches the mean service rate, both the mean travel time and the variance of the travel time increase. Figure 3 shows how the mean and variance of the travel time increase as the traffic intensity \( \bar{n} (= \bar{e} / \bar{i}) \) increases.

![Figure 3. Mean and Variance of Travel Time Versus Traffic Intensity](image)

It can be seen in Figure 3 that the mean travel time starts to increase much more rapidly once the traffic intensity \( \bar{n} \) exceeds about 80%, and it is thus common practice when designing transport facilities to define the practical capacity of such a system to be 80% of the theoretical capacity (e.g. Austroads (1988)). This implies a safety margin of 20%. It can also be seen in Figure 3 that for \( \bar{n} \) less than 60%, the variance of the travel time is very similar to the mean travel time, but the variance starts to increase much more rapidly once the traffic intensity \( \bar{n} \) exceeds about 60%. If one was to design to keep the ratio of the variance to the mean low (less than unity, say), to achieve a high level of reliability, the practical capacity would be only about 60% of the theoretical capacity. This implies that the safety margin should be about 40% (or about double that commonly used at present).

While it has been shown, for a simple M/M/1 queueing system, that the traffic intensity at which the travel time variance starts to increase rapidly is markedly lower than the traffic intensity at which the mean travel time starts to increase rapidly, it seems likely that this will also apply for the more complex queueing systems used for analysing some traffic flow situations (Gerlough and Huber, 1975).

The mean arrival and service rates reflect the demand and supply, respectively, while the travel time represents the average rather than the marginal travel cost. In the absence of a mechanism for ensuring the marginal travel cost equals the marginal utility, it is difficult to see how maintaining such a large safety margin as 40% could be achieved, as the provision of such a large margin will lead to lower mean travel times and thus greater usage. That is, designing for a large “quality margin”, as proposed by Goodwin (1992) is unlikely to lead to greater reliability, unless we can ensure equality of the marginal costs and utilities (via congestion pricing, say).
4 MEASURING RELIABILITY

4.1 Introduction
Wakabayashi and Iida (1992) defined reliability as “the probability of a device performing its purpose adequately for the period of time intended under the operating conditions encountered”. Such a definition does not highlight the importance of identifying and meeting the needs of users, who will consider a transport facility ‘reliable’ if their expectations are met, with the reliability increasing as the frequency and/or consequence of failing to meet user expectations decrease. Expectations can vary considerably between users, and can vary within users (i.e. can vary with time), with the spread of the ‘just-in-time’ philosophy indicating an increase in the expectations of users of the transport system.

A number of different quantitative measures of reliability have been proposed, but the relationship between them and user needs and perceptions is not clear. For instance, Chapman (1976) identified sixteen measures of bus reliability, some of which were very meaningful to bus operators (e.g. the ratio of actual to scheduled buses running) but are not well related to the needs and perceptions of bus users.

Nicholson et al. (2003) suggest that transport network users and planners have quite different viewpoints. Network users might, before each journey, have questions such as:

(1) is it still possible to reach my destination by any route?
(2) is the route that I normally take likely to be closed?
(3) if it is open, am I likely to encounter an unusual event (e.g. a delay)?
(4) what is the likely delay on my usual route?
(5) should I choose a different route or mode?
(6) should I postpone the trip?
(7) should I choose a different destination?
(8) should I cancel the trip?

Transport network planners, however, will be more interested in questions such as:

(1) how many users will not travel to their destination?
(2) which links will be congested or closed (i.e. are weak links)?
(3) which are the important links in the network?
(4) which are the critical (important and weak) links?
(5) what are the expected economic costs of closures?

Given the range of questions focusing on different aspects of reliability, it is not surprising that a range of reliability measures have been proposed or used. The main attributes of those measures are now described.

4.2 Terminal reliability
Graph theory has been used by various researchers (e.g. Wakabayashi and Iida (1992), Bell and Iida (1997), Bell and Iida (2001)) to calculate the ‘terminal reliability’, which is generally defined as “the probability that nodes are connected”, and is the simplest, most fundamental reliability measure. It allows only for links being either open or closed (and thus locations being connected or disconnected), but in reality links may be only partly closed.

Focusing on connectivity may be appropriate when congestion is not an issue, and one is dealing with a relatively sparse network, so that the loss of a link is likely to result in long detours being necessary (e.g. the Garrison (1960) study of the USA Interstate Highway system). However, the concern over transport reliability is now largely focused on the performance of congested networks.

4.3 Flow decrement reliability
Du and Nicholson (1997) proposed measuring reliability using the probability that the reduction in flow does not exceed a specified threshold. They defined network reliability in terms of the reliabilities of the sub-networks connecting individual origin-destination (OD) pairs. When a link is degraded, the cost of travel between one or more OD pairs may be affected, and the flow between those OD pairs will be affected, as a result of supply-demand interaction. The greater the level of link degradation, the greater will be the reduction in flow between each of the affected OD pairs, in general. Du and Nicholson defined the ‘flow decrement’ for each sub-network to be the reduction in flow (between the OD pair connected by the sub-network) when a link is degraded, as a proportion of the flow when there is no degradation.

That is, the sub-network flow decrement can vary between zero (when there is no flow reduction) and unity (when the flow is reduced to zero). They suggested that a 'threshold flow decrement' could be set for each sub-network, with a sub-network being considered to have failed if the estimated flow decrement exceeds the threshold value for that sub-network. The sub-network reliability was defined to be the probability that the sub-network would not fail.

The network flow decrement was equated to the
flow-weighted average of the sub-network flow decrements, with the weight for a sub-network being the ratio of the flow in the sub-network to the flow in the whole network, when there is no degradation. The network reliability was defined to be the probability that the network flow decrement does not exceed a threshold value (i.e., no failure).

Du and Nicholson showed that using this approach, which is based on classical reliability analysis techniques, can lead to a network being considered to have not failed even though one or more sub-networks have failed. This is consistent with a well-known result in reliability engineering that a system may have high reliability even though some sub-systems have low reliabilities.

A variation on this measure is the “probability that the travel demand is not less than the demand satisfaction ratio” (Lam and Zhang, 2000). They considered the case of a network with varying demand, and noted that the demand is not only a function of the travel time but also of demand variations due to “latent demand events”. These include recurrent events like peak-period traffic, and “regular special events” (e.g., sporting events) that cause additional demand for a certain period of time. The demand satisfaction ratio is measured as the ratio between equilibrium travel demand and the “latent travel demand”. The equilibrium travel demand is the demand that is satisfied (i.e., the number of users that are travelling at the current network costs). The latent travel demand is the sum of the equilibrium travel demand and the non-satisfied travel demand (i.e., the suppressed demand). They define the reliability to be the probability that the latent travel demand is satisfied, and also note that ignoring variations in travel demand can lead to error in the estimation of network reliability.

### 4.4 Encountered reliability

Bell and Schmöcker (2002) defined encountered reliability to be “the probability of not encountering a link degradation on the path with least (expected) costs”. This measures the likelihood of users encountering a disruption on their preferred route (a matter not addressed by terminal reliability). Even though users might still be able to get to their destination by a circuitous route without experiencing a degraded link, it may be likely that they encounter degradation if all the cheaper routes are affected by the event. Encountered reliability involves identifying all users who have taken a link that is degraded (i.e., it does not include users who are delayed as a result of traffic diverting from degraded links).

A closely related measure is capacity reliability, which was defined by Chen et al. (1999) as “the probability that the network can accommodate a specific demand level”, and is another measure of the likelihood of users experiencing difficulties on their chosen path.

### 4.5 Travel time and cost reliability

Neither terminal nor encountered reliability identify the impact caused by network degradation. Travel time reliability has been defined as “the probability that a trip can be successfully finished within a specified time interval” and Cassir (2001) defines the acceptable level of travel time as the travel time in normal conditions, plus a safety margin. This measure can be used for link closure as well as for any capacity degradation, and focuses on whether the increased travel time exceeds the acceptable threshold. In some cases a slight capacity reduction might lead to large delays, while in other cases (e.g., an abundance of re-routing options) the impact on network users might be very small.

As noted in Nicholson et al. (2003), the level of information on network conditions that is available to the users will influence their travel behaviour, e.g.,

1. cancelling their trip (i.e., a change in trip generation);
2. postponing their trip (i.e., a change in the temporal distribution of trips);
3. choosing another destination (i.e., a change in the spatial distribution of trips);
4. travelling by another mode (i.e., a change in mode split);
5. choosing a different route (i.e., a change in the traffic assignment).

Such behaviour changes will affect the travel time reliability, which will reflect the impact of capacity reductions on network users.

The concept of travel time reliability can be extended to travel cost reliability (Schmöcker and Bell, 2002). In this case the acceptability threshold is defined in terms of generalised travel costs, including time, distance, vehicle operating costs and public transport fares. Whereas travel time and travel cost reliability are often very similar if only re-routing is considered, considerable differences might be observed if mode choice is included in the behavioural responses to link degradation.

The setting of an acceptability threshold for travel time or cost is rather difficult, as expectations depend on the trip purpose and traveller characteristics.
4.6 Variance

Classical statistical measures of the variability of network performance indices (e.g., the variance of trip times) can be used to measure reliability, which is maximised when the variance is zero and decreases as the variance increases. If a risk-neutral traveller wants to get somewhere by the time $x_2$ and has a choice between two routes (A and B) with equal expected travel times, but the travel time variance for route A is less than for route B (see Figure 4), then the route with the smaller travel time variation (i.e., route A) should be chosen, because the probability of the travel time exceeding $x_2$ is less than for route B. However, if the traveller has already been delayed and the time threshold is $x_1$ (i.e., it is unlikely that the destination will be reached in time), the traveller should choose route B with the larger travel time variation, to improve their (low) chances of arriving on time. That is, the lower variance option is not always the better option.

An increase in the variance relative to the mean generally means a reduction in reliability, but a doubling of the variance does not necessarily mean a halving of the reliability (i.e., the relationship between the reliability and variance is not necessarily linear). Although perceived reliability is high when the variance is low, and is low when the variance is high, the relationship between the perceived reliability and variance is not well-defined for intermediate levels of variance, because the perceived level of reliability of an option will depend upon whether the user has an attractive alternative (see Figure 5).

Where users have no attractive alternative, they may well be quite insensitive to low levels of variation in the performance measure and may perceive the reliability as being very high, with the perceived reliability falling slowly as the variance increases. If, however, the user has attractive alternatives, they may be very sensitive to even low levels of variation in the performance measure, with the perceived reliability falling quickly as the variance increases.

Bonsall (2000) noted that when users of transport systems are confronted with an unpredictable situation, they appear to adopt strategies that cannot be explained in terms of the variance of the performance measure. In such circumstances, it is necessary to consider the full probability distribution. It seems that while the variance of travel time, say, is likely to influence users' perceptions of reliability, the variance alone is not always a good reliability measure.

![Figure 4: Two probability density functions of travel time](image)

Figure 4: Two probability density functions of travel time

![Figure 5: Perceived reliability depending on available route alternatives](image)

Figure 5: Perceived reliability depending on available route alternatives

It is interesting that the procedure for evaluating transport projects in New Zealand (Transfund NZ, 1997) has recently been amended to include reductions in the standard deviation of journey times. The value of a one minute reduction in the standard deviation has been valued at 80% and 120% of the value of a one minute reduction in the mean journey time, for cars and commercial vehicles respectively.

4.7 Vulnerability

Two quite different definitions of ‘vulnerability’ have been proposed. D’Este and Taylor (2003) define vulnerability to be the likelihood of severe adverse consequences if a small number of links or a single link is degraded. They distinguish between ‘connective vulnerability’ and ‘access vulnerability’. The former considers a pair of nodes and the generalised cost of travel between them; if the loss or substantial degradation of one or more network links leads to a substantial increase in that cost, then the connection between those nodes is deemed vulnerable. Access vulnerability considers a single node and the overall quality of access from that node to all other parts of the network; a node is vulnerable if the substantial degradation of a small number of links results in a significant reduction in the accessibility of that node, as measured by a standard index of accessibility.
D’Este and Taylor note that a network is not necessarily most vulnerable where the link flows are greatest, because alternative routes may be available nearby, allowing a new equilibrium flow pattern with little reduction in overall network performance. They suggest that a link in a path is vulnerable if the probability of that path being used for travel between OD pairs is much higher than the probabilities of alternative paths being used. For instance, if there are three paths (A, B and C, say), carrying 50%, 40% and 10% of the traffic respectively, then the closure of path A is not as disruptive as it would be if the proportions were 90%, 5% and 5% respectively. In the former situation, there is a similarly attractive alternative to path A, but this is not the case in the latter situation. That is, the greater the impact of a link being degraded, the greater is the link vulnerability.

The above definition of vulnerability is a measure of just the consequence of degradation, and does not take account of the probability of degradation. This is in contrast to the concept of vulnerability used by Nicholson and Du (1994), for whom vulnerability includes both the probability and consequence of degradation. They argue that links are ‘weak’ if the probability of degradation is high and ‘important’ if the consequences of degradation are high, with the ‘critical’ links (i.e. the links which should be treated) being those that are both important and weak. The danger with focusing on links that are only important or only weak is that resources will be wasted on non-critical links, while critical links are deprived of the resources needed to improve them and the reliability of the network.

4.8 Risk

The ‘risk’ of a hazard is the product of the probability of the hazard occurring and the consequence of its occurring (i.e. the expectation of the hazard or threat). Risk evaluation and management is a well-established process for identifying hazards, identifying their probabilities and consequences, assessing the acceptability of the risks, and taking action to address unacceptable risks. It includes low and high probability hazards and minor and major consequence hazards, and seeks to identify the critical links. Risk is a most comprehensive, complex and demanding reliability indicator, as it requires information about the probabilities of hazards occurring and links being degraded, and such information can be difficult to obtain.

A risk evaluation and management study of a road network is described later, after a description and discussion of the main reliability models proposed or used to date.

5 RELIABILITY MODELLING

5.1 Graph theory methods

Wakabayashi and Iida (1992) developed a graph theory approach for assessing the reliability of travel between a pair of nodes in a transportation network. This approach traditionally requires identification of all minimal paths and cut sets between the node pair, and the use of Boolean absorption for computing “terminal reliability” (i.e. the probability that the nodes remain connected). This is very time-consuming (especially for large networks), as the number of minimal path and cut sets (and thus the computation effort) increases rapidly as the number of nodes and links increases.

To overcome this difficulty, Wakabayashi and Iida proposed a method using partial minimum path and cut sets. This involves excluding all paths that are circuitous and are not perceived as feasible alternatives by users; such paths contribute very little to the reliability. Using an algorithm for calculating the upper and lower bounds for terminal reliability (based on partial minimal cut and path sets, respectively), they show that the bounds converge rapidly towards the exact reliability (based on the complete minimal cut and path sets) as the number of partial minimal path and cut sets increases. Their results suggest that the bounds are close (and close to the exact value) for eight or more minimal path and cut sets.

The graph theory approach has been applied to the Kyoto City network, comprising 95 nodes and 290 links, to estimate the terminal reliability with and without a traffic management scheme (involving converting some links to one-way operation); the results indicate a substantial improvement in terminal reliability for some node pairs and a small improvement for others (Wakabayashi and Iida, 1993). The graph theory approach has also been used for a study of the Shikoku region highway network (Asakura and Kashiwadani, 1992), with the reliability measure being the coefficient of variation of travel time.

More recently, Bell and Iida (2001) used a shortest path algorithm (a variant of Floyd’s algorithm), not to find the best path between origins and destinations, but to determine whether there is at least one functional path.

5.2 Game theory approach

Bell (1999) considered the situation where a network ‘spoiler’ seeks to disrupt the network to maximise user costs (by choosing the link that causes the maximum impact), while users try to minimise their costs (by adjusting their routes according to the expected link costs). The expected link cost for a link \( i \) is

\[
r_i C_{i,c} + (1-r_i) C_{i,o} \quad (3)
\]
where \( r_i \) is the link failure probability, and \( C_{lc} \) and \( C_{lo} \) are the costs when the link is closed (or failed) and open (not failed), respectively. The probability of the ‘spoiler’ choosing a link reflects the consequence of failure of that link (i.e. the sum of the increased costs to all users if that link fails).

In such a situation (a ‘minimax’ game), the user behaviour depends upon the estimated link failure probabilities, while the ‘spoiler’ behaviour (and hence the link failure probabilities) depends upon the user behaviour. The link failure probabilities must be determined iteratively, using the Method of Successive Averages, say, until the ‘mixed strategy Nash equilibrium’ is reached, where neither the network spoiler nor the network users can further improve their route choice probabilities or link failure probabilities, respectively.

The results of the game are therefore worst case link failure probabilities, which can be used to calculate the above mentioned reliability measures, to find the upper-bound impact of link degradation. Applications of this approach are described by Cassir (2001) and Cassir et al. (2003).

The method can be modified as shown by Bell (1999), who assumed the spoiler is somewhat random in the selection of which link to fail and used the logit model to find the link failure probabilities. In this case, the approach does not necessarily produce an upper-bound estimate of the impact of link degradation.

5.3 Absorbing Markov chains

The absorbing Markov chain approach (Bell and Schmöcker, 2002) focuses on the encountered reliability. This approach involves defining transition probabilities for each possible movement within the network. If no link exists between two nodes the transition probability will be zero, and if the transition is part of the best path to the destination the probability will be unity (if doing an all-or-nothing assignment) or a value between zero and unity (if using stochastic user equilibrium). Bell and Schmöcker also assigned a transition probability to a ‘bin’ for each node, representing the probability the user encounters a network problem (e.g. degraded link), and so were able to estimate the proportion of users who arrive at their destination without encountering any such problem. They did not allow for different levels of degradation (i.e. whether completely or partly closed) and the effects of degradation on travel time or cost were not calculated, so they could not estimate travel time or cost reliability.

Estimates of the actual failure probabilities, as well as worst-case failure probabilities, can both be input to the model. The transition probabilities are calculated according to the least cost paths, with the link cost being

\[
c_{ij} = d_{ij} - \beta \ln(r_i)
\]

(4)

where \( d_{ij} \) is the cost ignoring unreliability, \( \beta \) is a risk averseness factor, and \( r_i \) is the reliability of node \( i \). If \( \beta \) is zero the link reliability is not considered, but if \( \beta \) is large and the user has information on the reliability of the link, then the reliability is an important factor in the user’s cost estimation. They showed that the encountered reliability increases with increasing \( \beta \).

5.4 Monte Carlo Simulation

This technique is often used to identify the effects on a dependent variable, as a result of variations in those variables upon which the first variable depends, especially when the relationship between the dependent variable and the other variables is complex, and/or the variations in the other variables is not described by a mathematically-convenient distribution (e.g. are not Normally distributed). It is also used when the dependent variable depends upon many other variables, even though the relationship is fairly simple and the latter variables are Normally distributed (say). Monte Carlo simulation involves simulating the random behaviour of the other variables, to identify the random behaviour of the dependent variable.

Monte Carlo simulation was used extensively by Dalziell and Nicholson (2001) to develop closure duration probability distributions for hazards that may cause link closure. Chen et al. (1999) also used the technique to identify the distribution of the largest OD demand multiplier that does not cause link capacity violations, given a set of random link capacities. Given the complexity of the interactions between variables describing or affecting transport system performance, and the complexity of the hazard mechanisms affecting link capacity or availability, it is likely that Monte Carlo simulation will receive increasing use in future network reliability studies.

5.5 Microsimulation

While the use of equilibrium-based models may be appropriate when assessing the effect of long-term degradations, as is generally the case with natural disasters, degradations will frequently be for short durations (e.g. lanes being blocked by a truck during loading and unloading). In such circumstances, it is very unlikely that an equilibrium situation will be achieved. If one is concerned about increases in the demand for travel (e.g. before and after major sporting events), the increases are unlikely to be sustained long enough for equilibrium to be achieved. For both short-term degradations and demand increases, the use of an equilibrium-based method is likely to lead to error in assessing the effect. Even if equilibrium is achieved, the effect during the period when the traffic situation is adjusting from an equilibrium situation to another cannot be accurately assessed (Berdica, 2000).
A recent study into the effects of short-term blockages of a link in an urban network indicates that microsimulation models are more sensitive to disturbances (including disturbance duration) than equilibrium models, which underestimate the effects of the short-term disturbances (Berdica et al., 2003).

6 A RISK EVALUATION AND MANAGEMENT
CASE STUDY

6.1 Introduction

This study involved using risk evaluation and management methods to identify, for the New Zealand ‘Central North Island’ network:

(1) the expected cost (i.e. the risk) of various road closure scenarios;

(2) appropriate actions for reducing the risk.

This network includes the Desert Road section of State Highway 1, which is the main route between Wellington (the capital) and Auckland (the largest urban area), and carries the bulk of the traffic between those two cities (and between other urban areas). It is a two-lane two-way highway, carrying about 4000 vehicles/day, with about 15% being heavy commercial vehicles. The Desert Road is New Zealand’s highest section of State Highway, reaching an elevation of 1130 metres above sea level as it crosses the Central Plateau region of the North Island (see Figure 5). This region lies largely within an ecologically-sensitive area (a National Park) and contains three active volcanoes, and is also subject to earthquakes.

6.2 Risk evaluation method

The first step was to identify the hazards, which were found to be:

(1) snow and ice formation;

(2) ash fall during volcanic eruptions;

(3) lahar damage to roads and bridges;

(4) earthquake damage to roads and bridges;

(5) motor vehicle accidents.

The next step was to estimate, for each hazard, both the probability and the consequence of its occurrence. The consequence includes the cost of any remedial work (e.g. bridge repairs or replacement) and the economic cost borne by users whose travel is affected.

The user costs will depend on the closure duration, as well as the availability of alternative routes and cost of diverting (or loss of utility if users cancel or postpone their travel).

It was necessary to derive a frequency distribution of closure duration for each hazard. This was done using a mixture of historical information about the probability of events and an understanding of the generating mechanisms of each of the hazards, plus Monte Carlo simulation. Details of the techniques are given in Dalziell (1998).

The closure costs depend on the reason for closure, because of correlation between route conditions (i.e. whether open or closed) in the network, particularly for events (e.g. major earthquakes or volcanic eruptions) where the effects are likely to be geographically wide-spread. It was therefore necessary to allow for the simultaneous closing of the Desert Road and the nearby alternative routes (Figure 6). There is only one alternative north-south route nearby (State Highway 4), but there are three more distant alternative north-south routes, which were included in the network used for estimating the cost of link closures (Figure 6), but the probability that they will also be closed was deemed negligible. A total of 22 closure scenarios (i.e. combinations of the Desert Road with zero or more of the nearby alternative roads being closed) were considered.

In addition to considering the dependence of route conditions, it was necessary to consider the interactions between the different hazards (i.e. whether the to
occurrence of one hazard affects the occurrence of another hazard). Interactions can be ‘two-way’ or ‘one-way’ (e.g. snow and ice may increase the probability of a traffic accident occurring, but a traffic accident will have no effect on the probability of snow and ice forming). They can affect probabilities and/or consequences (e.g. snow or ice may have no effect on the probability of a volcanic eruption, but may make a lahar larger and more destructive).

The SATURN computer model (Van Vliet, 1995) was used to estimate traffic flows within the network (Figure 7), which included all towns in the region and roads used for travel through the region and/or between the towns in the region. All the roads are two-lane two-way highways, carrying between 500 and 5000 vehicles/day, so the network is not congested. A user-optimal equilibrium traffic assignment was used.

\[ T_{Cij} = T_{0ij} (C_{0ij} / C_{Cij})^p \]  

where \( T_{Cij} \) and \( C_{Cij} \) are the number of trips and the travel cost (respectively) between an origin zone \( i \) and a destination zone \( j \) when one or more road links are closed, \( T_{0ij} \) and \( C_{0ij} \) are the corresponding number of trips and travel cost when all links are open, and \( P \) is the elasticity parameter. Now \( C_{Cij} \) will be greater than \( C_{0ij} \) for an un congested network, and hence \( T_{Cij} \) will be less than \( T_{0ij} \) provided \( P \) is greater than zero.

The route choice was assumed to depend on the time and distance costs, but accident costs (based on accidents being proportional to the vehicle-km of travel on each link) were included in the total user cost, along with the travel time and vehicle operating costs. Standard values for economic appraisal in New Zealand (Transfund, 1997) were used.

The total cost of road closures was assumed to be the sum of:

1. the change in the vehicle operating and occupant time cost;

\[ \sum_i \sum_j C_{ij} T_{Cij} - \sum_i \sum_j C_{0ij} T_{0ij} \]  

2. the lost user benefit from those trips that are cancelled or suppressed;

\[ \sum_i \sum_j 0.5 \left( C_{0ij} + C_{Cij} \right) \left( T_{0ij} - T_{Cij} \right) \]  

3. the change in the accident cost.

The traffic model assumes that travellers have perfect knowledge of the road network and its characteristics (including which roads are closed) when commencing a trip, and will always choose the least-cost path. If a link is closed after a journey has commenced, however, a vehicle may have already passed the junction to the best alternative route, and will therefore either have to go back to that route, or select the next best alternative. This would mean some under-estimation of the cost of road closure, especially for short closures. Countering this is the tendency of drivers to wait for the preferred route to open, if this is likely to occur soon.

6.3 Risk evaluation results

Setting the elasticity parameter \( P \) to 3.0 gave a good agreement between the predicted flows (with the Desert Road closed) and the observed flows during a nine-day closure. Table 1 shows the total cost and the cost components, for four of the 22 closure scenarios. All costs are expressed in July 1997 NZ$ (equivalent
US$0.66 in July 1997). It can be seen that closure of the Desert Road alone costs the New Zealand economy nearly $8,000 per hour. When the nearby major north-south route (State Highway 4) is also closed, detour lengths are greater and the loss of user benefits (due to the suppression of travel) increase substantially, with closure costs reaching nearly $23,000 per hour. Simultaneous closures of other nearby roads have less impact, as they do not create a major barrier to the north-south traffic flow.

Table 1: The Costs ($/hr) of Various Road Closure Scenarios

<table>
<thead>
<tr>
<th>Closure Scenario</th>
<th>Operating plus Time</th>
<th>Accidents</th>
<th>Lost User Benefit</th>
<th>Total Cost</th>
<th>Total Cost of Closure</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Roads Open</td>
<td>180,590</td>
<td>43,017</td>
<td>N/A</td>
<td>223,607</td>
<td>N/A</td>
</tr>
<tr>
<td>Desert Rd Closed</td>
<td>172,610</td>
<td>40,858</td>
<td>18,129</td>
<td>231,597</td>
<td>7,990</td>
</tr>
<tr>
<td>Desert Rd and SH 4 Closed</td>
<td>164,412</td>
<td>38,570</td>
<td>43,498</td>
<td>246,480</td>
<td>22,870</td>
</tr>
<tr>
<td>Desert Rd and SH 47 Closed</td>
<td>165,220</td>
<td>38,846</td>
<td>33,887</td>
<td>237,953</td>
<td>14,350</td>
</tr>
<tr>
<td>Desert Rd and SH 49 Closed</td>
<td>169,944</td>
<td>39,783</td>
<td>26,058</td>
<td>235,785</td>
<td>12,180</td>
</tr>
</tbody>
</table>

Table 2 shows the effect of closure of the Desert Road on travel within the whole network. The total travel and total cost are predicted to increase by about 2.2% and 2.6% respectively, with no change in the number of trips, if the elasticity of demand is ignored (i.e. $P=0$). If the elasticity of demand is taken into account (i.e. $P=3$), the number of trips, the total travel and the total cost are predicted to decrease by 3.3%, 4.5% and 4.4% respectively.

Table 2: The Effect of Closure on Travel Within the Study Network

<table>
<thead>
<tr>
<th>Measure of Effect</th>
<th>Desert Rd Open</th>
<th>Desert Rd Closed ($P=0$)</th>
<th>Desert Rd Closed ($P=3$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Trips (veh-trips/h)</td>
<td>1956</td>
<td>1956</td>
<td>1891</td>
</tr>
<tr>
<td>Total Travel (veh-km/h)</td>
<td>299890</td>
<td>306610</td>
<td>286250</td>
</tr>
<tr>
<td>Total Cost (veh-$/h)</td>
<td>180580</td>
<td>185210</td>
<td>172610</td>
</tr>
</tbody>
</table>

Table 3 shows that closure of the Desert Road causes a 7.4% decrease in travel across the screenline (Figure 6), when allowing for the elasticity of demand. This is a larger reduction than for trips in the whole network, reflecting the fact that many trips do not involve using the Desert Road. It can also be seen that 80-90% of the traffic diverting from the Desert Road diverts to State Highway 4 if it is open, rather than the more distant alternative routes.

The value of $P$ indicates high demand elasticity. This probably reflects the perception that if the Desert Road is closed due to snow and ice, then the driving conditions on the nearby alternative routes will be too dangerous to travel. The Central Plateau is a popular recreational area, and it is likely that the weather will be unsuitable for many recreational pursuits when the Desert Road is closed due to snow and ice, so fewer trips will be generated. Table 1 shows that the operating-plus-time costs and the accident costs reduce for the four closure scenarios; this is also true for the other 18 closure scenarios (Dalziell, 1998). It is only after the lost travel benefits due to trips being cancelled or postponed have been taken into account that one gets an economic loss due to road closure.

Table 3: The Effect of Closure on Flows (veh/h) Across the Screenline

<table>
<thead>
<tr>
<th>Crossing Point</th>
<th>Desert Rd Open</th>
<th>Desert Rd Closed ($P=0$)</th>
<th>Desert Rd Closed ($P=3$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SH1 (Desert Rd)</td>
<td>155</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SH4</td>
<td>100</td>
<td>225</td>
<td>185</td>
</tr>
<tr>
<td>Other Routes</td>
<td>555</td>
<td>585</td>
<td>565</td>
</tr>
<tr>
<td>Total</td>
<td>810</td>
<td>810</td>
<td>750</td>
</tr>
</tbody>
</table>

A point-estimate of the closure cost does not reflect the uncertainty regarding the probabilities and consequences of the hazards, and Monte Carlo simulation was used to obtain the probability distribution of the annual cost of closures for each hazard (Figure 8).

Figure 8. Closure Cost Exceedance Probabilities for Each Hazard
Snow and ice is the most important hazard, the expected annual cost being $1.9 million, compared to $1.5, $0.3 and $0.2 million for earthquakes, traffic accidents and volcanic events, respectively.

6.4 Risk management

After risk evaluation was completed, the acceptability of the risks was assessed. The criterion for assessing acceptability was that a risk is not acceptable if the benefit of mitigating that risk (i.e. the reduction in the expected cost of road closure) is sufficiently greater than the cost of mitigating that risk (in New Zealand, the benefit/cost ratio must exceed about four).

Mitigation options may affect the probability and/or cost of road closure. For instance, the threshold event size required for road closure may increase, and/or the time to repair and re-open the road may reduce. It was necessary to re-calculate the risk of road closure (by re-calculating the probability and cost of closure) with each mitigation option in place.

The benefit of a mitigation option is the risk (or expected cost) of road closure without the mitigation in place, minus the risk (or expected cost) of road closure with the mitigation in place. This was compared to the sum of the capital and maintenance costs of the mitigation, to assess the worth of the mitigation.

A simple point-estimate of the benefit-cost ratio (i.e. the best estimate) does not reflect the uncertainty that exists regarding the probabilities and consequences of the hazards, and the effects of the mitigation options. Therefore the probability distribution of the benefit-cost ratio for each mitigation option was estimated (using Monte Carlo simulation), and Figure 9 shows the distribution obtained for the application of chemicals, salt or calcium magnesium acetate (CMA), with and without a road weather information system (RWIS), to mitigate ice formation on the Desert Road. From this one can establish, for each mitigation option, the likelihood that the benefit-cost ratio will exceed some threshold value for implementation.

It should be noted that the benefit-cost ratio for applying salt does not take account of the adverse ecological effects. Despite the ratio for applying salt being much greater than for applying CMA, the use of salt within a National Park was deemed unacceptable. The use of a RWIS in conjunction with applying salt or CMA enhances the benefit-cost ratio substantially, because it enables the prevention of road closure with a lower rate of application of the chemical. Indeed, the cost of CMA is so high that the cost of installing and running a RWIS is expected to be less than the savings in purchasing CMA.

6.5 Discussion

The use of Monte Carlo simulation enables the identification of probability distributions for the cost of closure for each type of hazard, and probability distributions for the benefit-cost ratio for each mitigation option. This information assists assessment of the relative importance of each type of hazard and the relative merits of mitigation options. The results reveal considerable uncertainty regarding the close costs for the hazards, but the probability distribution for snow and ice dominates that for the other hazards, showing that snow and ice is the most important hazard. This result shows that frequent, small-consequence events can cost more than rare, large-consequence events, and demonstrates the need to not focus on the latter.

The study has shown the importance of allowing for both the elasticity of demand for travel and re-routing of traffic within the network. If this is not done, one can get a quite erroneous estimate of the socio-economic impact of road closures. This is consistent with the results of a study of the effect of the 1989 Loma Prieta earthquake on the freeway network in the San Francisco Bay Area (Wakabayashi and Kameda, 1992). This study found that allowing for changes in the freeway traffic trip matrix gave a 28% larger post-earthquake network capacity (defined to be “the maximum number of car trips that the network can deal with without over-saturated traffic”) than if there was no change in the matrix. It was also found that ignoring such changes in travel behaviour resulted in travel times during the post-earthquake period being over-estimated by about 30%.

7 CONCLUSION

Interest in transport network reliability is likely to
continue increasing in the future, due to the increased expectations of the users of transport systems (e.g. the increased prevalence of the ‘just-in-time’ philosophy with its dependence upon transport reliability). Transport network congestion is increasing as investment in increasing network capacity fails to keep up with the increasing use of the network. Secondly, there is evidence that some natural disasters (e.g. flooding and hurricanes) are becoming more frequent. Finally, transport systems are increasingly the target of malevolent acts (or attacks) by those seeking to cause hardship to society by disrupting transport systems.

Nicholson et al. (2003) have outlined how the impact of a disruption depends upon the level of malevolence and the level of knowledge of the transport network possessed by users. They conclude that the malevolence of an intervention (or event) affects the location, level of degradation and duration of the event, and the more malevolent the intervention and the lower the level of user knowledge, the higher will be the impact. Network planners have no control over the level of malevolence, but may be able to reduce the impact of disruptions (i.e. improve network reliability) by making more information available to users.

The various reliability measures and analysis methods highlight different aspects of reliability (e.g. either the user or planner viewpoint) and thus have their advantages and disadvantages. A major difference between the measures is whether they require the probabilities of the adverse event to be estimated. Deterministic approaches do not, making their use substantially easier, but they estimate upper or lower bounds effects rather than actual effects. Risk is the most comprehensive measure available, and provides a good guide as to which links are critical (i.e. weak and important) and should be made more robust. However, risk assessment involves estimating the probabilities of links being degraded and entails considerable effort.

There is a lack of quantitative information regarding how users judge reliability and how travel behaviour is affected by unreliability, and further research on the human factors aspects is required. Some behavioural responses (especially postponing a trip, choosing a different destination, cancelling a trip, route choice changes after travel has commenced) are difficult to allow for using most reliability measures and analysis techniques.

The primary focus of traffic models has been the analysis of normal, steady-state conditions, using equilibrium-based methods. These are not well-suited to modelling dynamic conditions, such as occur at the start and finish of a disruption, and modelling short-duration disruptions, where traffic conditions may not be even approximately steady-state. Microsimulation seems to be the best modelling approach for assessing the impact of short-term disruptions. Whatever model is used needs to be calibrated, to ensure it accurately predicts what actually happens when closures occur. There is little information on actual traffic behaviour during disruptions, to enable calibration of a model, and this deficiency needs to be remedied.

The use of a risk evaluation and management approach has been demonstrated via a case study. The effort involved in accurately estimating event probabilities was considerable, but it was practicable (including allowing for inter-dependence between the states of nearby links). It is usually assumed that component states are statistically independent, but for a transport network, this assumption of independence does not always hold. For example, if there are two bridges over a river and the upstream bridge is washed away or damaged, the debris will move downstream, increasing the probability of the downstream bridge being damaged. If the two bridges are over different rivers with catchments with highly correlated rainfall patterns, then the bridge degradation probabilities may be highly correlated. If two links span the same seismic fault, then their degradation probabilities may also be strongly correlated. Du and Nicholson (1997) showed that network reliability can be sensitive to such correlations.

The risk evaluation and management study results highlights the importance of allowing for both traffic re-routing within the network and the elasticity of travel demand, if one is to get accurate estimates of the economic impact of link closures. The study has also highlighted the importance of not focussing on rare, major-consequence events (e.g. earthquakes) at the expense of frequent, minor-consequence events, as the risk (or expected cost) associated with the latter can exceed the risk for the former.

REFERENCES


Berdica, K. (2000). Analysing Vulnerability in the Road...
Transportation System: Putting Theory into Practice Using Sweden as an Example. Working Paper TRITA-IP FR 00-76, Department of Infrastructure and Planning, Royal Institute of Technology, Sweden.


