A transport network reliability model for the efficient assignment of resources

M. Sánchez-Silva a,*, M. Daniels b, G. Lleras a, D. Patiño b

a Department of Civil and Environmental Engineering, Universidad de los Andes, Bogotá, Colombia,
Cr. 1 No.18A-70, Ed W, P3, Bogotá, Colombia
b Department of Industrial Engineering, Universidad de los Andes, Bogotá, Colombia

Received 22 May 2002; received in revised form 5 August 2003; accepted 24 March 2004

Abstract

Recently, several papers have been published addressing transport network reliability acknowledging it as a priority for future research. A model for optimizing the allocation of resources based on the operational reliability of transport network systems is proposed. The optimum assignment of resources is carried out based on a set of possible actions described in terms of the failure and repair rates of every link. Thus, the model optimizes the assignment of resources so that the accessibility of a centroid or the total network is maximized. The methodology provides also an alternative to model the decisions of the user as he/she travels between two centroids. A case study in Colombia is used to illustrate the applicability and the benefits of the model. The results can be used for the optimum allocation of resources for road maintenance and rehabilitation.

© 2004 Elsevier Ltd. All rights reserved.

Keywords: Transport; Network; Reliability; Stochastic model; Optimization

1. Introduction

Safety, competitive travel times and cost are important conditions that a highway network must offer to its users. In countries where freight transportation is mostly carried out through its highway system, national economic competitiveness is highly correlated to the reliability of such
infrastructure. In many countries, resources to construct, maintain and rehabilitate roads are limited; furthermore, natural phenomena such as earthquakes and landslides frequently cause traffic interruptions. Therefore, optimizing resources becomes a priority to ensure the competitiveness of the transport network and whereby of the economy.

Most road network studies address preconditions for the failure of particular sectors of the network (e.g. landslides, bridge failure) but do not evaluate the effect of a specific failure scenario (i.e. failure of a set of links) over the whole network. Several papers have been published addressing transport network reliability through, for example, a game theoretic approach (Bell, 1999), using stochastic user equilibrium (Asakura, 1999), simulating traffic flow (Lam and Xu, 1999), based on measuring the capacity of the network (Chen et al., 1999) and in terms of the form of the network (Lleras and Sánchez-Silva, 2001). In addition, many other papers on network reliability, whose ideas can be applied directly to this problem, have been also published (see Malachy and Hendrickson, 1984; Sanso and Soumis, 1991).

Risk is the combined effect of the chances of occurrence of some failure or disaster and its consequences in a given context (Sánchez-Silva et al., 1996). Thus, the network’s reliability analysis requires considering three fundamental aspects: (1) criteria (e.g. strength, form); (2) expected losses (e.g. economic, social); and (3) definition of failure scenarios (e.g. most frequent, maximum losses) (Sánchez-Silva, 2001). The criteria considered in this model will be “form”, the consequences will be measured in terms of disutility (e.g. cost) and the failure scenarios as traffic interruption through a link. Appropriate reliability models are useful as long as they support the decision making process. Thus, the reliability model proposed will be used to define efficient actions for resource allocation.

2. Transport systems reliability

Reliability of transport systems is a complex issue that involves several factors that differ in nature. A systems analysis must combine physical and functional considerations, which are not necessarily independent. Physical aspects are related to the impossibility for the user to reach a destination due to damage of the infrastructure (e.g. collapse of a bridge). Functional aspects are concerned with level of service provided; for instance, excessive travel times. Although, in some cases damage to physical infrastructure may not cause a disconnection of the network, it might reduce substantially the level of service.

The measurement of network reliability involves both the physical infrastructure (e.g. bridges, tunnels) and the behavioral response of network users (Bell, 1999). The first issue is a resistance-based approach (Sánchez-Silva, 2001) and will be considered in this paper in terms of the mean value of failure occurrences. The impact of a failure of physical infrastructure on the function of the network depends not only on the direct impact in mobility, but also on how users respond to that failure. This in turn depends on factors such as the user’s knowledge of the network, the information provided to the user on the state of the network, the physical characteristics of the network to maneuver and the cultural context (e.g. organization and cooperation).

Bell and Iida (1997) argue that reliability is measured as the degree of stability of the quality of service, which a system normally offers. Traditionally, reliability of transport networks can be
studied from two different perspectives namely connectivity and travel time reliability. Connectivity is the probability that traffic can reach a destination at all. Travel time reliability is the probability that the destination can be reached in a time less than some threshold value. While connectivity reliability focuses on the analysis of the complete network, travel time reliability concentrates on a particular link or a set of links representing any path within the network.

Classical systems reliability theory supposes that whether the system functions or not is determined solely by the knowledge of which components are functioning. The description of a transport system is made in terms of a combination of series and parallel arrangement of its components. For classic system reliability analysis, if $x_k$ is the state variable of link $k$, the state vector of the system $x = \{x_1, x_2, \ldots, x_n\}$ indicates which components are functioning and which have failed. The structure of the system $\Phi(x)$ is a function that shows if the system is functioning or not when the state vector is $x$. On the other hand, travel time reliability is computed in terms of the probability that traffic can reach a destination within a given period of time.

A more elaborated analysis for modeling transport network reliability is the vulnerability theory of systems based on the structural vulnerability analysis. It has been proposed as a way in which a structure is connected together. The theory enables the identification of particular ways in which it might fail. The network is subdivided in rings which are defined as a balance loop of driver and change for the general case. In systems dynamics they are called “across variables” and “through variables” (Shearer et al., 1967). In the case of a transport system, the across variables (drivers) are the needs (movement of passengers or freight) and the through variables are the flow. Failure is associated to separation which is the disassociation of a cluster from the rest of the system (Wu et al., 1993; Lu et al., 1999).

3. Network reliability model

3.1. General description

A transport network system can be thought of as a stochastic dynamic system, where the state of links (i.e. failed or not failed) and the users’ decisions change permanently. Thus, in order to assess the reliability of a network system, it is required to consider the network’s variation with time. This can be looked at from two perspectives: (1) the decisions that the user has to make as he/she goes along through a route from one node to another; and (2) the average failure and repair rates of a link within a route between two nodes. These two aspects are considered in this paper within a single model that will be explained in the following sections.

The main steps of the proposed methodology are as follows:

- Basic assumptions of the model and notation.
- Definition of the network.
- Continuous time Markov chain modeling of failure states.
- Calculation of expected operation conditions.
- Optimization and resource allocation.
3.2. Basic assumptions of the model and notation

3.2.1. Basic assumptions

The model proposed in this paper has been built on the following assumptions:

- At the start of the trip, the traveler chooses the route based on full knowledge of the state of the network.
- The user re-evaluates the route choice at every node based on the state of the network.
- The user selects the route with the minimum disutility among all available routes. Perfect information is available.
- Total interruption of the traffic over the link is always considered.
- The user has a waiting policy within which, once the failure occurs (total interruption), he/she waits until the interruption in the link is repaired.
- At most, there is only one interruption in every link at a time.

3.2.2. Notation

In order to make the following sections clearer, the notation that will be used is:

- \( \lambda_i \) failure rate of link \( i \)
- \( \mu_i \) repair rate of link \( i \)
- \( N_{i \rightarrow j} \) number of vehicles traveling from node or centroid \( i \) to \( j \)
- \( C_{i \rightarrow j} \) cost of traveling from node or centroid \( i \) to \( j \)
- \( r_n \) route \( n \) to travel between any two centroids
- \( r_i^c \) route \( r_i \) is not available
- \( K \) expected cost of not taken any route
- \( k_w \) value of waiting time per time unit
- \( k_t \) value of traveling per time unit
- \( K_0 \) cost of opportunity associated with the impossibility of traveling

3.3. Definition of the network

A road network is defined as a system, which can be represented mathematically as a graph \( G(N,A) \) made up of a finite set of \( N \) nodes and \( A \) links. The network includes a set of roads selected on the basis of any technical, functional or administrative criteria. Centers of special interest such as cities are designated as centroids and should be clearly defined within the network. All centroids define a subset of the \( N \) nodes (Larson and Odoni, 1981). Links are represented by a set of attributes (i.e. location, distances, travel time, average operational speed, volume delay functions, capacity).

The function of a network is to manage the demand for transportation from one centroid, or node, to another. Reliability analysis requires a detailed definition of centroids because it is intended to measure how their accessibility decreases under a particular failure scenario.
3.4. Continuous time Markov chain model of failure states

Markov chains model the transition of a system between all possible states it can take, in order to calculate the probability that the system is in state \( i \) at time \( T \) (i.e. \( P_i(T) \)). The probability that the system moves to a given state depends only on the current state. A continuous-time Markov chain is a stochastic process with the following properties: (1) the amount of time the system spends in a given state before making a transition into a different state is exponentially distributed with rate, \( \nu_i \); and (2) when the process leaves a given state \( i \) it will next enter a state \( j \) with some probability \( P_{ij} \) where the sum over all \( i, j \) must be equal to 1 (Ross, 2000).

When the user has to make the decision on the route to travel between any two centroids, he/she considers the availability of such a route (i.e. set of consecutive links or the so called a simple path). This is performed by analyzing the state vector of the system, in which each state is described by the state vector of the system \( x = \{x_1, x_2, \ldots, x_n\} \) that indicates which links are functioning and which have failed. The model considers only two possible states for every link: operation \( (x_i = 1) \) and no operation \( (x_i = 0) \). In the proposed model, Markov chains are used to calculate the probability that the links composing a given route are available.

It is assumed that the probability that any two events (failures or repairs) occur at the same time is null. Therefore, the probability that the system moves from one state to another depends upon a series of events that may develop a new scenario. These events can be represented, for instance, in terms of the failure and repair rates \( \lambda \) and \( \mu \). Therefore, once the system has moved to a given state, it can move further to other state if an additional link fails to operate as designed, or if the operation of a failed link is restored. The failure rate (i.e. \( \lambda \)) can be computed as the number of hours that a link is interrupted per time unit (i.e. year), and the repair rate (i.e. \( \mu \)), as the time required to repair the link, also per time unit. This approach can also be used for modeling the effect of natural phenomena such as earthquakes or floods over the network. Finally, in some cases it may be possible to model events (e.g. earthquake, flood) that might affect several links at the same time.

3.5. Calculation of expected operating conditions

3.5.1. Expected cost of traveling through a particular link

A route is defined as a set of consecutive links connecting two nodes/centroids. In order to compute the expected cost of a route, it is necessary to calculate the expected cost of traveling through a link. The concepts discussed in this section will be used later for computing the expected cost of a route.

The expected cost of traveling through a particular link between two consecutive centroids \( i \) and \( k \), \( E[C_{i,k}] \), is given by

\[
E[C_{i,k}] = E[C_{i,k}|E]P(E) + E[C_{i,k}|E]P(E)
\]  

where \( E \) is the event “there is an interruption” and \( C_{i,k} \) is the estimated travel cost without congestion. If the user has to wait until the link is repaired, the expected cost of traveling is given by

\[
E[C_{i,k}] = C_{i,k}(1 - P(E)) + (C_{i,k} + (1/\mu_{i,k})k_w)P(E)
\]  

where \( k_w \) represents the cost of waiting per time unit while the link is repaired and \( 1/\mu_{i,k} \) is the expected value of the waiting time. This equation leads to
The model assumes that there is at most one failure during the trip from \( i \) to \( k \). Also, the occurrence of the failure is a random variable \( T \) that follows an exponential distribution \( f_T(t) \) with an average time between occurrences \( \lambda_{i,k} \). If \( t \) is the time at which the interruption occurs; and \( T_{i,k} \) the total time required to travel from \( i \) to \( k \) (Fig. 1), the failure does not affect the user if it occurs after he/she has arrived to centroid \( k \)—that is, \( t > T_{i,k} \). Therefore,

\[
P(E) = \int_0^{T_{i,k}} P\{\text{Wait} \mid t\} f_T(t) \, dt = \int_0^{T_{i,k}} P(\text{Wait} \mid t) \lambda_{i,k} e^{-\lambda_{i,k} t} \, dt
\]

(4)

If the failure may occur anywhere with the same probability, the probability that the user has to wait, given that a failure has occur at time \( t \) \((t < T_{i,k})\), is equal to the proportion of time left before he/she reaches the centroid \( k \):

\[
P(\text{Wait} \mid t) = \frac{T_{i,k} - t}{T_{i,k}}
\]

(5)

Therefore, the expected cost of traveling between two consecutive centroids can be expressed as follows:

\[
E[C_{i,k}] = C_{i,k} + \frac{k_w}{\mu_{i,k}} \int_0^{T_{i,k}} \frac{T_{i,k} - t}{T_{i,k}} 2^{\lambda_{i,k} e^{-\lambda_{i,k} t}} \, dt
\]

(6)

where \( k_w \) is the value of the waiting time. If \( C_{i,k} = T_{i,k} k_t \), where \( k_t \) is the cost of traveling per time unit, the expected cost of traveling through a link can be expressed as

\[
E[C_{i,k}] = C_{i,k} + \frac{k_w k_t}{\mu_{i,k} \lambda_{i,k} C_{i,k}} \left[ C_{i,k} \lambda_{i,k} k_t + \exp \left( \frac{-C_{i,k} \lambda_{i,k}}{k_t} \right) - 1 \right]
\]

(7)

3.5.2. Measure of accessibility

Accessibility is the ability to command the transportation facilities that are necessary to reach desired destinations at suitable times. It is the most important relationship emerging from the interaction between the elements of the network. This concept is widely used in transport studies under different contexts. Different authors such as Moseley (1979), Halden (1996), Geertman and
Van Eck (1995) agree that accessibility depends on two factors: (1) an activity or motivation based on the opportunities available in a location; and (2) a resistance factor based on generalized cost of traveling (e.g. efficiency, low cost). Others (Shen, 1998) have also included concepts such as the demand for the foregoing mentioned opportunities. Accessibility is strongly related to the location and relevance of centroids, the willingness to move and the opportunities and benefits of moving in accordance with the attributes of the network.

In the proposed model, the operation condition of the network is evaluated through the Accessibility Index \( A \), which is used for describing the efficiency of the network to communicate centroids. Accessibility can be defined in terms of any variable of the network system, however, disutility, which is the cost of the trip as network users perceive it, is a usual parameter. Disutility encompasses all factors that affect the cost for the user and the way he/she integrate them. That includes aspects such as travel time, speed limit, quality of the road, safety, landscape, congestion, and so forth. Although the disutility considers every important factor in the decision making process, travel time and direct cost are usually the most relevant components (Bell and Iida, 1997).

The accessibility to centroid \( i \) is defined here by

\[
A_i = \sum_{j=1, j\neq i}^{n} N_{j\rightarrow i} f(E[C_j\rightarrow i])
\]

where \( f(E[C_j\rightarrow i]) \) is a monotonic decreasing function, i.e., \( f(E[C_j\rightarrow i]) = 1/E[C_j\rightarrow i] \), \( n \) is the number centroids in the network, \( N_{j\rightarrow i} \) is the number of vehicles (traffic) and \( E[C_j\rightarrow i] \) is the expected cost of traveling from every centroid \( j \) to centroid \( i \). This definition implies that as the cost of traveling increases, accessibility decreases.

### 3.5.3. Methodology for computing accessibility

The main input for computing accessibility is the expected cost of traveling, which depends primarily on the actual cost of traveling between nodes, i.e., \( C_{i,k} \). The traffic assignment problem follows an all-or-nothing process. Hence, neither capacity nor the stochastic nature of travel time under normal operation conditions is considered. The proposed methodology is reasonable in sparse and uncongested networks (Ortuzar and Willimsen, 1994). Under these assumptions, determination of minimum paths is required.

Deterministically, minimum paths from each centroid to every other define the network’s optimum operating conditions. There are several well-known algorithms to calculate minimum paths, for instance, Dijkstra’s or Floyd’s are well-known procedures (Wright and Ashford, 1989; Larson and Odoni, 1981). In the proposed model, the procedure to computing the expected cost of traveling between any two centroids can be summarizes as follows:

1. Identify all “reasonable” routes to travel between any two centroids.
2. Classify all “reasonable” routes, in order of preference (i.e. less disutility).
3. Determine the probability that the route is operating.
4. Compute the expected value of the cost (disutility) of traveling between any two centroids.

A route is defined as a set of consecutive links connecting any two centroids. In a road network system there may be many possible routes between any two centroids, some of which differ only in a few links. A “reasonable” route between \( i \) and \( j \) is defined as a route for which the travel cost is
smaller than the sum of the minimum cost of traveling between \( i \) and \( j \) and a proportion of the cost of waiting; this is:

\[
(C_{i-j})_{ra} \leq \min\{(C_{i-j})_{\text{all routes}}\} + \xi w
\]

where \((C_{i-j})_{ra}\) is the cost of a “reasonable” route for traveling from \( i \) to \( j \); \( k_w \) is the cost of waiting and \( \xi \) is an arbitrary constant. This means that a user would not take a route which is more expensive than waiting until the route with the minimum cost is available.

Once a set of “reasonable” routes has been selected for the analysis, such routes have to be classified in order of preference by the user. The criteria used for ranking the routes could be the generalized travel cost, or more generally, disutility, which is a true measure of value of the decision maker. The order of preference of the \( n \) “reasonable” routes to travel from centroid \( i \) to \( j \) can be expressed as \( r_1 > r_2 > \ldots > r_n \), which means that the route \( r_1 \) is first in the order of preference, \( r_2 \) second, and so forth. A detailed discussion on the aspects involved in the travelers’ route choice decision process can be found in Bovi and Stern (1990).

A given route \( r_i \) is available when all its component links are operating. Thus, the probability that the route \( r_i \) is available is the sum of the stationary probabilities of all links, \( q_i \), within that route. This is computing by obtaining the steady state vector from a classical Markov analysis.

\[
p(q_i) = \sum_{x_1=0}^{1} \sum_{x_2=0}^{1} \cdots \sum_{x_n=0}^{1} p(x_1, x_2, \ldots, x_{i-1}, 1, x_{i+1}, \ldots, x_n)
\]

and therefore, the probability that a route is available is

\[
p(r_n) = p(q_1)p(q_2)\cdots p(q_m)
\]

where \( m \) is the total number of links in the route. Any route will be taken only if it is available and all routes that precede it in the order of preference are not. Thus, the probability of taken any route can be expressed as

\[
p(r_k) = p(r_1^c, r_2^c, \ldots, r_{k-1}^c, r_k)
\]

where \( r_i^c \) denotes the complement (i.e. route \( r_i \) is not available). The cost of not taking any route, \( K \), can be computed as

\[
K = K_0 p(r_1^c, r_2^c, \ldots, r_{n-1}^c, r_n^c)
\]

Once the routes are ordered, the expected cost of traveling between \( i \) and \( j \) can be computed as

\[
E[C_{i-j}] = \left[ \sum_{k=0}^{n} C(r_k)p(r_k) \right] - K
\]

where \( C(r_k) \) is the cost of traveling through route \( r_k \); \( n \) is the number of routes available; \( p(r_k) \) is the probability that route \( r_k \) is taken; and \( K \) is the expected cost of not taken any route.

In some cases and depending upon the network characteristics, it may be required for the user to change his/her original route as he/she goes along. This only occurs if changing the route does not imply a significant additional cost. If this is the case, the expected cost of traveling between any two centroids \( i \) and \( j \), \( E[C_{i-j}] \), depends on a succession of decisions that take into account both the cost of taking a complete route and the availability of the next link in this route. The
expected value of the cost of traveling from centroid \( i \) to \( j \), given that the user has chosen the route \( r_m \) (includes nodes \( i \), \( k \) and \( j \)) is

\[
E[C_{i\rightarrow j}|r_m] = E[C_{i,k}] + E[C_{k\rightarrow j}]
\]

where \( C_{i,k} \) is the cost of traveling from centroid \( i \) (starting centroid of the route) to the immediate next centroid \( k \) in route \( m \); and \( C_{k\rightarrow j} \) is the cost of traveling from centroid \( k \) to centroid \( j \) (end of the route \( m \)). This implies that the selection of the first link is governed by the decision on the route but it does not mean that the user has to stick to that route all the way from \( i \) to \( j \).

The expected cost of traveling from centroid \( i \) to \( j \) is then computed by conditioning on the route chosen. This is,

\[
E[C_{i\rightarrow j}] = \sum_m E[C_{i\rightarrow j}|r_m]P(r_m) + P(r_1^i \cap r_2^i \cap \cdots \cap r_{m-1}^i)K_0
\]

For instance, in the road network presented in Fig. 2, two reasonable routes to travel between centroids 1 and 7 are: \( r_1 = 1\to 2\to 4\to 7 \), or \( r_2 = 1\to 3\to 6\to 7 \); thus, the expected cost of traveling from 1 to 7 is given by

\[
E[C_{1\rightarrow 7}] = (E[C_{1,2}] + E[C_{2,7}])P(r_1) + (E[C_{1,3}] + E[C_{3,7}])P(r_2) + P(r_1^i \cap r_2^i)k_w
\]

Since the expected cost is expressed recursively, a system of equations has to be solved in order to obtain the expected cost of traveling between any two centroids.

3.6. Optimization of resource allocation

The appropriate assignment of resources depends on an effective use of the resources available to induce a change of the accessibility as function of the changes in the failure and repair rates \((\lambda_k \text{ and } \mu_k)\). If the repair rate \((\mu_k)\) is increased, the response capacity of any interruption of link \( k \) improves; similarly, a decrease of the failure rate \((\lambda_k)\) means that prevention measures have been
successful. The approach using \( k \) and \( l \) as the main parameters of the model, facilitates the
definition of performance indexes and the decision making process.

Any change in these rates has a cost associated. Therefore, every action has to be looked at in
terms of the relationship between the cost and the benefit obtained in the operation of the
network if such action is taken. The objective of optimization is to maximize the change in the
accessibility, which may be looked at from two perspectives: (1) improving accessibility to a
particular centroid; or (2) enhancing the accessibility of the whole network. The first analysis
focuses on assessing the change in the accessibility of a given centroid as function of modifi-
cations of the parameter \( \lambda \) and \( \mu \). The second alternative considers an increase in the accessi-

For the centroid analysis, the optimization can be expressed as

Maximize:

\[
A_i(\lambda_1, \lambda_2, \ldots, \lambda_n, \mu_1, \mu_2, \ldots, \mu_n) = \sum_{j=1, j \neq i}^{n} N_j \cdot f(E[C_{j-i}])
\]  

Subject to:

\[
\sum_{i=1}^{n} C_{\lambda_i}(\lambda_i) + \sum_{i=1}^{n} C_{\mu_i}(\mu_i) = C_L, \quad \lambda_i \geq 0, \quad \mu_i \geq 0
\]  

where \( A_i(\lambda_1, \lambda_2, \ldots, \lambda_n, \mu_1, \mu_2, \ldots, \mu_n) \) is the accessibility as defined in Eq. (8); \( C_L \) is the maximum
amount of resources (e.g. US$) available for investing in the road network; \( C_{\lambda_i}(\lambda_i) \) and \( C_{\mu_i}(\mu_i) \) are
the cost of modifying the failure, \( \lambda \), or repair, \( \mu \), rates of link \( i \). The results of the optimization are
the changes on the failure and repair rates of every link such that the accessibility to centroid \( i \) is
maximized. Note that \( \partial A_k / \partial \lambda_i > 0 \) and \( \partial A_k / \partial \mu_i < 0 \); also \( \partial C_{\lambda} / \partial \lambda < 0 \) and \( \partial^2 C_{\lambda} / \partial \lambda^2 > 0 \). Similarly,
the cost associated to an increase of the repair rate \( \mu \), \( C_{\mu_i}(\mu_i) \), is a function such that \( \partial C_{\mu} / \partial \mu > 0 \)
and \( \partial^2 C_{\mu} / \partial \mu^2 > 0 \).

Eq. (8) does not provide enough information about the behavior of the whole network.
Therefore, the Network Reliability Index (\( F_N \)) is proposed to measure the change in the accessibility
of the network and it is defined as

\[
F_N = \sum_{j=1}^{m} w_j A_j
\]  

where \( m \) is the number of centroids in the network and \( w_j \) the weight of centroid \( j \) in accordance to its importance for the network (e.g. economic, social). Every centroid has a weight \( w_i \) calculated
according to the evaluation objectives, for instance, amount of freight or passengers generated
or attracted. In order for the values of \( w_j \) to guarantee completeness, it is necessary that the sum of all \( w_j \) be equal to 1 (Lleras and Sánchez-Silva, 2001). Thus, the optimization can be expressed as

Maximize:

\[
F_N(\lambda_1, \lambda_2, \ldots, \lambda_n, \mu_1, \mu_2, \ldots, \mu_n) = \sum_{j=1}^{m} w_j A_j(\lambda_1, \lambda_2, \ldots, \lambda_n, \mu_1, \mu_2, \ldots, \mu_n)
\]
Subject to:

\[
\sum_{i=1}^{n} C_{\lambda_i}(\lambda_i) + \sum_{i=1}^{n} C_{\mu_i}(\mu_i) = C_L, \quad \lambda_i \geq 0, \quad \mu_i \geq 0
\]  

(22)

The proposed model requires a nonlinear optimization of \(2n\) variables, note that the constrains \(\lambda_i \geq 0\) and \(\mu_i \geq 0\) will always be unbinding. It is stressed that the cost restriction in the optimization suggests that the costs are separable, however, this maybe modified if necessary.

4. Accuracy and computational considerations

Since transport networks may have a significant number of links, the computational complexity becomes an important aspect of the solution. The proposed algorithm is based on a Markov chain model which requires identifying all possible states of the system, for example, an average network with 50 links leads to a \(2^{50} = 1.12 \times 10^{20}\) possible states. However, the probability of having states of the system with a significant number of failed links is very unlikely. If the failure probability of every link is small, say 0.01, and it is assumed independence between the failure of links, a state with six links failed will have a probability of occurrence of 0.01^6 = 1 \times 10^{-12} which is extremely low and does not influence much the results. Therefore, the error is related to the failure rate and the maximum number of links that might fail at the same time. Thus,

\[
\text{err} = 1 - \sum_{i=0}^{l} \binom{n}{i} p^i (1-p)^{n-i}
\]

(23)

where \(n\) is the total number of links within the network, \(l\) is the maximum number of links that can fail at the same time and \(p\) is the average failure probability of all links. The error is plotted in Fig. 3 for different values of \(l, n\) and \(p\). If, on average, all links within a network of 100 elements are interrupted 24 h a year, the failure rate will be 2.74 \times 10^{-3} and an error of almost 0 is obtained if only three links are considered to fail simultaneously. In some very specific roads of Colombia, where interruptions are frequent due to landslides, the police reports interruptions of 350 h/year which correspond to an annual average failure probability of 0.04. A network of 50 links with

![Fig. 3. Error of the model as function of the number of failed links simultaneously.](image-url)
such extremely high failure rate will lead to an error of 12.5% for three and 1.5% for six simultaneous failed links.

As an alternative, Lleras and Sánchez-Silva (2001) suggested a weighting function for the probability of occurrence of a given state, in terms of the number of failed links $n$ per scenario, $f(n) = 1/n$ or $f(n) = e^{-n}$. The validity of this approach can be easily shown following the methodology proposed in the aforementioned paper.

5. Illustrative example

As an example of the proposed model, a part of the main transport network in central Colombia was considered (Fig. 4). The relevant information of the parameters, in suitable units for the study, is also presented in Fig. 4.

The Markov model is defined by identifying all possible states of the system. A vector of 10 components within which each position represents the possible state of a link is used to define the state of the system. For instance, a state defined by $(0, 1, 1, 1, 1, 1, 1, 1, 1, 1)$ means that the only link that is not operating is the first link. Since the system consists of 10 elements, $2^{10}$ possible states can be identified. The probability that the system will be in a given state is computed as the fraction of time the system remains in that state. For instance, the probability that the system will be in a state where all links are operating (i.e. $(1, 1, 1, 1, 1, 1, 1, 1, 1, 1)$) is 0.9396. All other states have very low probabilities, for instance $p(1, 1, 1, 1, 1, 1, 1, 1, 0) = 2.192 \times 10^{-4}$.

In order to compute accessibility, the analysis requires the identification of all “reasonable” routes to travel between any two centroids and the corresponding volume of traffic. Then, by using Eq. (7) it is possible to compute the cost of traveling through every link (Table 1). Table 1 is used to compute the total cost of traveling by every reasonable route between any two centroids.

Fig. 4. Transport network considered for the illustrative example.
For this example the three routes with the lower cost between any two centroids were considered. The decision criterion for allocating resources is the change in the accessibility; in particular, investment should lead to a reduction in the total accessibility of the network (Eq. (20)). Therefore, in first place it is necessary to compute the accessibility to every centroid independently. This is performed by using Eq. (8) with the results shown in Table 2.

By considering the weights shown also in Table 2, which where obtained based on traffic demand and especial socioeconomic characteristics of each centroid, it is possible to compute the Network Reliability Index, $F_N$, as

$$F_N = \sum_{j=1}^{m} w_j A_j = 0.224$$

This value corresponds to the current accessibility of the network. Thus, the optimization consists on defining the assignment of resources (fix budget) to modify these parameters such that the Network Reliability Index is maximized. This requires defining a cost function for every possible action, which in this case are the changes of the failure and repair rate of every link. Defining this cost functions depends on the context of the problem and have to be carefully structured. For this example, the cost function for every link has the form $C(x) = k(x - x_0)^4$ where $x$ represents $\lambda$ or $\mu$ and $x_0$ is the corresponding original value. The parameter $k$ for every link and for every possible

<table>
<thead>
<tr>
<th>Link</th>
<th>Cost US$ (x1000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.418</td>
</tr>
<tr>
<td>2</td>
<td>3.869</td>
</tr>
<tr>
<td>3</td>
<td>2.737</td>
</tr>
<tr>
<td>4</td>
<td>5.062</td>
</tr>
<tr>
<td>5</td>
<td>3.711</td>
</tr>
<tr>
<td>6</td>
<td>1.894</td>
</tr>
<tr>
<td>7</td>
<td>0.467</td>
</tr>
<tr>
<td>8</td>
<td>1.281</td>
</tr>
<tr>
<td>9</td>
<td>1.485</td>
</tr>
<tr>
<td>10</td>
<td>3.189</td>
</tr>
</tbody>
</table>

Table 2

<table>
<thead>
<tr>
<th>Centroid</th>
<th>Weight</th>
<th>Accessibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.2775</td>
<td>0.1213</td>
</tr>
<tr>
<td>2</td>
<td>0.0648</td>
<td>0.0853</td>
</tr>
<tr>
<td>3</td>
<td>0.0056</td>
<td>0.0383</td>
</tr>
<tr>
<td>4</td>
<td>0.1388</td>
<td>0.1559</td>
</tr>
<tr>
<td>5</td>
<td>0.1850</td>
<td>0.4614</td>
</tr>
<tr>
<td>6</td>
<td>0.0925</td>
<td>0.4566</td>
</tr>
<tr>
<td>7</td>
<td>0.0046</td>
<td>0.1579</td>
</tr>
<tr>
<td>8</td>
<td>0.2313</td>
<td>0.1497</td>
</tr>
</tbody>
</table>
action is shown in Table 3. These values were obtained from a regression analysis and reflects the actual socioeconomic conditions of the region.

The optimization is clearly a nonlinear problem which can be solved using standard methods such as the projected gradient. The restrictions on the optimization are the amount of resources available and the final value of the parameters $\lambda$ and $\mu$, which have to be positive. For a budget of US$1867 million, the results of the optimization are shown in Table 4. Columns 2 and 3 correspond to the final values of $\lambda$ and $\mu$; columns 4 and 5 are the cost associated to the change in these parameters and the last column is the total cost (sum of columns 4 and 5) to be invested in every link.

Fig. 5 shows the relative investment in every link for the failure and repair rates. It can be observed that investments on improving the repair rate are higher and more even throughout the links than those required for enhancing the failure rates. The investment in the failure rate is highly concentrated in the link 9 (Fig. 4) and is followed by the investment in links 5 and 6. Actions directed to reduce $\lambda$ are related to physical interventions such as construction of retaining wall structures, retrofitting bridges and so forth. In terms of the repair rates, links 3 and 4 are significant for the network since they have the smallest cost of repair and provide the redundancy.
for the route from centroids 1 to 8. A failure of link 3 makes links 2 and 5 to become extremely critical and any alternate route very expensive. In general, it was observed that the investment directed to improve repair rates has a lesser impact on the network since the time a link is out of service is very low.

Sensibility analysis showed that the trends presented in Fig. 5 are kept for similar cost functions, which makes the model reliable as tool for decision making. Considering an average failure probability for all routes of, \( p = 0.0216 \), the expected error in the calculations, taken only three failed links is \( \text{err} = 4.118 \times 10^{-5} \), which is acceptable.

If the interest of the Colombian government is to enhance the accessibility to the country’s capital (i.e. Bogotá) only, with the same budget, the optimization will concentrate in the centroid analysis approach. In this case, \( F_N \) (Eq. (20)) is not computed and only the accessibility of to centroid 1 is considered. The results for Bogotá are presented in Fig. 6 in terms of the relative change of the parameters \( \lambda \) and \( \mu \). It is interesting to observe that since the main traffic attracted comes from cities such as Manizales, Armenia and Pereira links 4, 5 and 6 become more important. It is also interesting to observe that the access through link 1 needs

![Fig. 5. (a) Relative investment in the failure rate with respect to the maximum; and (b) relative investment in the repair rate with respect to the maximum.](image1)

![Fig. 6. Distribution of investment to improve the accessibility to Bogotá.](image2)
to be improved. This is very consistent with the real traffic conditions in this part of the network.

6. Conclusions

Allocation of resources for enhancing the reliability of a transport system is a priority and a very controversial issue due to the differences in criteria used for that purpose. In this paper, an approach to computing the transport systems reliability of a network based on an entirely probabilistic view is proposed. As suggested by many authors this approach comprises two key elements of transport systems reliability: the state of the infrastructure and the behavior of the network users. It considers the state of the network through the relationship between the failure and repair rates of every link comprising the network. These rates are directly related to physical characteristics of the road, such as condition of the road, or frequency and size of landslides. The behavior of the user may be taken into account by modeling the decision making process as he/she goes along through a route between any two centroids. The proposed model provides a very solid framework for optimizing the assignment of resources to enhance the reliability of any transport network system.

References