R U T C O R
R ESEARCH
R E P O R T

NETWORK RESILIENCY

Michael Tortorella[a]

RRR 3-2010, FEBRUARY, 2010

RUTCOR
Rutgers Center for
Operations Research
Rutgers University
640 Bartholomew Road
Piscataway, New Jersey
08854-8003
Telephone: 732-445-3804
Telefax: 732-445-5472
Email: rrr@rutcor.rutgers.edu
http://rutcor.rutgers.edu/~rrr

[a] RUTCOR, Rutgers, the State University of New Jersey, 640 Bartholomew Road, Piscataway, NJ 08854-8003. mtortore@rci.rutgers.edu

RUTCOR RESEARCH REPORT
RRR 3-2010, FEBRUARY, 2010

NETWORK RESILIENCY

Michael Tortorella

Abstract. "Network resiliency" has been used in common language for many years in a variety of contexts. The intent of the usage seems to always have included the notions that a resilient network is one that continues to deliver services in a satisfactory fashion despite possible disruptions in the network infrastructure and/or one that returns to normal operation quickly after a disturbance. In this report, we consider the former interpretation in detail. The notion of "satisfactory" obviously involves a matter of degree, so a consistent quantitative framework for network resiliency is proposed for the first interpretation. This report organizes and subsumes current streams of network resiliency thought by providing a generalization of the various notions that have been used in the past to describe network resiliency and using this generalization to create a quantitative framework for resiliency characterization. The key concepts are the notions of a network with associated delivery function and of delivery importance. We discuss various design principles that promote greater network resiliency in the context of this framework. We introduce a network resiliency optimization scheme and briefly show how the network interdiction problem can be solved using the proposed network resiliency framework.

1 FOREWORD

1.1 Rationale

In the beginning there was the social network. From the earliest days of *Australopithecus* and *homo erectus*, individuals formed complex webs of social interactions that we today conceptualize as networks. It is no exaggeration to say that these networks enabled the species to survive and evolve into today's *homo sapiens* who are capable of such conceptualization. Until roughly the Renaissance, mostly that's all there was, when the advent of sailing ships and intrepid sailors enabled goods to be moved greater distances. This, and further advances during the nineteenth century, begat the supply chain network, although, again, no one used that name at the time. The invention of the telephone in 1876 begat the telecommunications network. The postal system, the electric power grid, and commodity networks such as oil and gas transport systems brought forth a new way of life. Indeed, civilization as we know it would be impossible without networks.

In the twentieth century, mathematicians began to study networks as abstract objects to do a better job of creating and improving them. Our position in this paper is that networks are not ends in themselves; they exist to perform certain functions. Society is interested in the continued performance of these functions to an adequate degree and at reasonable cost. The concept of *resiliency* is introduced to enable discussion of the degree to which the functions performed by a network continue to be performed when various disruptions occur. The purposes of this article are to describe a mathematical framework for discussing network resiliency in quantitative terms and to list principles and practices that can be employed during the design of a network to promote its greater resiliency.

1.2 Scope

The material presented in this article is generic in the sense that it applies to all types of networks and is not restricted to a particular application. We consider networks that transport continuum materials such as oil and gas, and networks that transport discrete physical objects such as letters and packages as well as notional objects such as data packets in telecommunications. Social networks, too, are included in that they exist to share something among the individuals comprising the nodes of the network: information, or status, or power relationships.

We reserve for future consideration further development of an alternate property characterizing resilient networks, namely that they return quickly to a normal functioning state after a disruption. Again, a matter of degree is involved, and it will be necessary to develop quantitative models to treat this phenomenon adequately.

1.3 Background

"Network resiliency" began to be used as a phrase because people wanted a way to express how well a network may continue to perform its intended functions when the inevitable degradation of operating conditions occurs, or how quickly it returns to normal operation after a disruption. Infrastructure degradation includes failures and/or loss of capacity of the network's constituent elements (switches, routers, valves, transport systems, etc.) that may be due to reliability

problems (hardware, software, operations) in the physical equipment (pipes, wires, towers, circuit cards, trucks, buildings, etc.) making up the network elements. These failures, in turn, may be due to "random" manifestation of broken or deteriorated components, or to natural disasters such as earthquakes and fires, or to deliberate attacks. The study of the effects of such failures on the network as a whole gives rise to the mathematical theory of network reliability [4], [6] which usually considers figures of merit like two-terminal connectivity and all-terminal connectivity as indicators of the purposes of the network. The literature on the mathematics of network reliability is extensive.

In prior studies, "network resiliency" was used to convey a sense of continued satisfactory fulfillment of some purpose of the network in the face of failures and disruptions and/or the rapid return of the network to normal operation after a disruption. Specific purposes studied have included communication among node-pairs [7], survivability [39], the ability to provide alternate communication paths [3], reachability [12], economic damage [35], and disconnection of the network [26], [27]. Occasionally the term is used without specific definition, assuming that resiliency would be equated with "good" properties implicitly known to the reader, as in [21]. In these cases, some specific measure of how well the network continues to operate, *i. e.*, provide its intended service(s), serves as the proximate interpretation of network resiliency. The generalization to delivery functions (Section 2.1) and delivery importance (Section 2.2) is natural.

1.3.1 Resiliency and Reliability Contrasted

Before proceeding further, it is worth clarifying the differences between resiliency and reliability. The primary difference is that resiliency applies to the functions or services performed by the network whereas reliability is usually used to describe the failure and repair behavior of network elements. Of course, a primary indicator of network resiliency is the reliability of the *services* that run on the network (see [37] and [38] for further clarification on service reliability), but service reliability is not the only reasonable delivery function for network resiliency characterization. It is but one of many useful delivery functions (Section 2.1) that describe a network's intended purpose. In this report, we will see that figures of merit and metrics for network resiliency depend on the particular delivery function associated with the network. The examples in Section 6 are specializations of the general theory presented in Sections 2 and 3.

1.3.2 Resiliency and Survivability Contrasted

The related notion of survivability has been studied primarily in the telecommunications context [25]. The earliest approaches to survivability were mainly concerned with the provision of geographically diverse alternate routes so that traffic that would normally be carried on facilities that could be lost due to natural disaster or deliberate attack would continue to reach its destination [15]. While the amount of traffic that is successfully rerouted and carried is a key indicator of network resiliency, survivability studies did not necessarily explicitly incorporate measures of how much traffic would be successfully rerouted. Later studies began to incorporate such considerations (for example, see [23]), and thereby resemble more comprehensive resiliency studies.

Some care is required in interpreting the literature regarding reliability, survivability, and resiliency because these terms are sometimes used interchangeably with no warning.

1.4 Synopsis

The approach taken in this article goes beyond that of the usual network reliability theory. To capture the notion that the network exists to perform a desired function, we define a *delivery function* associated with a network that tells what the network is being used for. The precise definition is given in Section 2.1; some examples of delivery functions here serve to provide a flavor of what this means. In the most general case, we may consider a delivery function to be the reliability of a service or application [37], [38] to a user in the sense that is used in telecommunications networks: for example, a user pays an Internet Service Provider (ISP) for access to the ISP's network so that the user may acquire page views from the World Wide Web, upload files, etc., and the reliability of those services as seen by the user/purchaser of the service is the delivery function. For a commodity network such as an oil or gas transport network, the delivery function may be the quantity of commodity delivered to specified destination nodes over a stated period of time. In other cases, the delivery function can be as simple as the two-terminal connectivity in an uncapacitated network. The key idea is that the delivery function captures the essence of what the network is being used for.

Delivery importance of a subset of the network is then defined in terms of the change in the value of the delivery function when conditions in that subset change (for example, increase or decrease in capacity or removal from service). A derivative comes into play here, and so delivery importance represents a generalization of the idea of reliability importance [1], [31] to this broader context. *Network resiliency* is defined as the degree to which the network is sensitive to changes in its infrastructure: a network is resilient if there are "few" subsets that have "high" delivery importance. We then discuss three examples particular delivery functions, in increasing order of complication: connectivity, reliability, and flow in a stochastic network. We give examples of each and discuss computational issues.

Some design principles promoting greater network resiliency are reviewed in Section 8. The two principles discussed are overprovisioning and network control. A straightforward network resiliency optimization problem is introduced in Section 7, and the network interdiction problem is discussed in Section 9 as an extension of the network resiliency problem.

2 DELIVERY FUNCTIONS AND DELIVERY IMPORANCE

2.1 Delivery Functions

To enable discussion of network resiliency in quantitative terms, we associate with a network a function Ψ that we will call the *delivery function*. The intent is that this function will express quantitatively the purpose of the network or the service that the network is intended to provide. It may be a function, such as connectivity, that is already familiar in network reliability theory, or it may be a function, such as service reliability, that has not been previously been considered in this context. The delivery function associates a real number or a vector with the elements of the network.

A network $(\mathcal{H}, \Psi)$ with delivery function is an ordered pair consisting of a graph $\mathcal{H}$ and associated delivery function Ψ. $\mathcal{H} = (\mathcal{N}, \mathcal{L})$ where $\mathcal{N}$ is a set of elements, called nodes, and $\mathcal{L}$ is a subset of $\mathcal{N} \times \mathcal{N}$. We denote by $|\mathcal{N}| = N$ the number of nodes in the graph. If, for $i_1 \neq i_2$ belonging to $\mathcal{N}$, $(i_1, i_2) \in \mathcal{L}$, then (i_1, i_2) is a link; if $i_2 = i_1$, then (i_1, i_1) is another way of writing the node i_1. $\mathcal{N} \cup \mathcal{L}$ is the set of network elements and Ψ maps $\mathcal{N} \cup \mathcal{L}$ onto the real numbers or a real vector space, depending on the range appropriate to the particular case studied.

2.2 Delivery Importance

This Section discusses delivery importance, the influence of a network element, or a set of network elements, on the delivery function. The concept of delivery importance is introduced to enable us to see how changes in the network element(s) are reflected in the delivery function, that is, in how well the network continues to perform the functions expected of it when there are changes in the network infrastructure. We first consider delivery importance in deterministic capacitated networks, both continuum and discrete. Delivery importance is defined for uncapacitated networks by considering the presence or absence of nodes and/or links in the graph, and assigning probabilities in the case of a stochastic uncapacitated network; see Section 2.2.3. We then extend the definitions to stochastic capacitated networks, both continuum and discrete.

2.2.1 Delivery Importance in Deterministic Capacitated Networks

The *capacity* of a network element is a nonnegative real-valued function $c : \mathcal{N} \cup \mathcal{L} \rightarrow \mathbf{R}^+$. The capacity may represent the presence or absence of a network element (in which case the range of the capacity function is $\{0, 1\}$), the probability that a network element is in the graph, or the *reliability* of the network element (in which case the range is $[0, 1]$), or it may represent what is commonly understood as capacity in a flow network [14]. Capacity may be constant (static network capacity) or a function of time (dynamic network capacity). We also consider capacitated networks in which the capacities are random quantities (stochastic capacitated networks), as distinct from deterministic capacitated networks in which the capacities are not random.

We define the *capacity matrix* $C(\mathcal{H})$ to be the matrix whose (i, j) entry c_{ij} is the capacity of the link (i, j) if (i, j) is a link $(i \neq j)$, 0 if (i, j) is not a link $(i \neq j)$, and the capacity of the node i if $j = i$. In capacitated networks, we consider the domain of the network's delivery function is the space Γ of capacity matrices $\mathbf{R}^{N \times N}$. So, for example, a scalar-valued delivery function maps a capacity matrix $C(\mathcal{H})$ into $\mathbf{R}^+$. It is also possible to consider vector-valued delivery functions. These are important for applications, but the extension of the theory given here to vector-valued delivery functions is straightforward (see equation (1.8)); so a full discussion of vector-valued delivery functions is not undertaken here. A generic capacitated network with associated delivery function may be denoted by $(\mathcal{H}, C, \Psi)$.

2.2.1.1 Delivery Importance in Deterministic Capacitated Continuum Networks

In a continuum network, the range of the capacity function is the nonnegative real numbers, and a flow in the network [14] may take on nonnegative real values as well (subject, of course, to the

constraint that the value of the flow at any network element may not exceed the capacity of that network element). Continuum network models describe networks that deliver continuum materials such as oil, gas, water, electricity, etc., or social networks in which the weight on a link (its "capacity") is the strength of the relationship between the two nodes it connects. In a continuum network, we assume that the delivery function is a continuously differentiable function of the capacities except possibly on a set of Lebesgue measure zero. That is, for each network element $(i, j) \in \mathcal{N} \cup \mathcal{L}$ the derivative $\partial\Psi/\partial c_{ij}$ exists, except possibly on a set of Lebesgue measure zero in the space of capacity matrices, and is continuous as a function of the capacity wherever it exists.

The main idea of delivery importance is to see how the delivery function changes when the network capacity changes incrementally. That is, we would like to compare $\Psi(C_0 + hA)$ to $\Psi(C_0)$ where C_0 is a nominal capacity matrix, or study value of the capacity matrix, A is a matrix representing a "direction" in which capacity increases (or decreases) form the basis of the study, and h is a real number, typically close to zero. The basic delivery importance equation for the single network element (i, j) is, for fixed $a \neq 0$,

$$\Psi\left(c_{ij} + ha\right) - \Psi\left(c_{ij}\right) = \left.\frac{\partial\Psi}{\partial c_{ij}}\right|_{C_0} ha + o(h) \tag{1.1}$$

where the remainder term satisfies $o(h)/h \to 0$ as $h \to 0^+$, provided that Ψ is differentiable at $(C_0)_{ij}$. There is no loss of generality in taking a to be 1 or –1. $a = 1$ (resp., $a = -1$) indicates the delivery importance when the capacity of network element (i, j) increases (resp., decreases). The one-term Taylor series expansion (1.1) motivates the following definition.

Definition. For a differentiable scalar delivery function, the *delivery importance at C_0 in the direction a* ($a = \pm 1$) of a network element $(i, j) \in \mathcal{N} \cup \mathcal{L}$, denoted by $\Omega_a(i, j; C_0)$, is defined as the derivative of Ψ with respect to the capacity c_{ij} of the network element (i, j), evaluated at the capacity matrix C_0:

$$\Omega_a(i, j; C_0) = a \left.\frac{\partial\Psi}{\partial c_{ij}}\right|_{C_0} \tag{1.2}$$

when it exists.

Combining this with (1.1) yields $\Psi\left(c_{ij} + ha\right) - \Psi\left(c_{ij}\right) = \Omega_a\left(i, j; C_0\right)h + o(h)$ for $a = \pm 1$ and $h \to 0^+$.

Example 1. Consider the simple directed, capacitated network consisting of two links in series, shown in Figure 1.

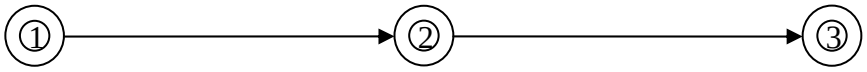

Figure 1.

The three nodes each will be considered to have infinite capacity and the capacities of the two links are x and y, respectively. The capacity matrix for this network is

$$C = \begin{pmatrix} \infty & x & 0 \\ 0 & \infty & y \\ 0 & 0 & \infty \end{pmatrix}. \tag{1.3}$$

The delivery function for this network will be defined as the maximum flow from node 1 to node 3, which in this case is $\Psi(C) = \min\{x, y\} = \frac{1}{2}(|x + y| - |x - y|)$, $x, y \geq 0$. Here Ψ is differentiable except on the line $x = y$. We have, for $x \neq y$,

$$D_1\Psi(x, y) = I\{x < y\} \text{ and } D_2\Psi(x, y) = I\{x > y\} \tag{1.4}$$

so the delivery importance in the positive direction of the link (1, 2) is $\Omega_1(1, 2; C) = 1$ if $x < y$ and 0 if $x > y$, and that of link (2, 3) is 0 if $x < y$ and 1 if $x > y$. Define the nominal capacity matrix

$$C_0 = \begin{pmatrix} \infty & c_{12} & 0 \\ 0 & \infty & c_{23} \\ 0 & 0 & \infty \end{pmatrix}. \tag{1.5}$$

For this nominal capacity matrix, $\Omega_1(1, 2; C_0) = 1$ if $c_{12} < c_{23}$ and 0 if $c_{12} > c_{23}$, and $\Omega_1(2, 3; C_0) = 0$ if $c_{12} < c_{23}$ and 1 if $c_{12} > c_{23}$.

When $x = y$, the derivatives do not exist. Define

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{1.6}$$

Then we return to the basic idea of delivery importance to write, for $h > 0$,

$$\frac{1}{h}\left[\Psi\left(C_0 + hA\right) - \Psi\left(C_0\right)\right] = \frac{1}{h}\left[\min\left\{c_{12} + h, c_{12}\right\} - c_{12}\right] = 0\,. \tag{1.7}$$

For equal initial link capacities, the delivery importance of link (1, 2) is zero when its capacity increases and –1 when its capacity decreases (use $-A$ in place of A); a similar analysis shows the same is true for link (2, 3). Thus, when the two links have the same capacity, they are of equal delivery importance. If the initial link capacities are equal, making one of them larger has no effect on the delivery function. ■

For a vector delivery function $\Psi = (\Psi_1, \ldots, \Psi_v)$, the delivery importance of a network element $(i, j) \in \mathcal{N} \cup \mathcal{L}$ is the vector

$$\left(\frac{\partial \Psi_1}{\partial c_{ij}}, \ldots, \frac{\partial \Psi_v}{\partial c_{ij}}\right) \tag{1.8}$$

when it exists. While vector delivery functions are important for many applications, the fact that the delivery importance is computed element by element, as in equation (1.8), means that the delivery importance of a vector delivery function can be evaluated element by element, so we need not deal with vector delivery functions further in this article.

Because $\Psi : \Gamma \to \mathbf{R}$, $\Psi'(C)$ is a linear map from Γ into $\mathbf{R}$ for every C [11], and so is represented by a vector. At this point it is convenient to order the network's incidence matrix as a vector by writing out its rows one after another as a row vector of length N^2. Then the vector representing $\Psi'(C)$ is the vector of partial derivatives

$$\left(\frac{\partial \Psi}{\partial c_{11}}, \ldots, \frac{\partial \Psi}{\partial c_{1N}}, \frac{\partial \Psi}{\partial c_{21}}, \ldots, \frac{\partial \Psi}{\partial c_{2N}}, \frac{\partial \Psi}{\partial c_{31}}, \ldots, \frac{\partial \Psi}{\partial c_{NN}}\right)$$

where the (i, j) entry is zero if $(i, j) \notin \mathcal{L}$. Thus, we may write $\Omega_a(i, j; C_0)$ from equation (1.2) as $\Omega_a\left(i, j; C_0\right) = \left\langle \Psi'\left(C_0\right), \mathbf{1}_{(i,j)} \right\rangle$ where $\mathbf{1}_{(i,j)}$ is a vector containing 1 in the (i, j) position and zeros elsewhere, and $\langle , \rangle$ represents the ordinary inner product in $\mathbf{R}^{N^2}$.

We may also consider the delivery importance of larger subsets of $\mathcal{N} \cup \mathcal{L}$. Let $\mathcal{M} \subset \mathcal{N} \cup \mathcal{L}$. Define $\mathbf{1}_{\mathcal{M}}$ to be the matrix whose (i, j) entry is given by $I\{(i, j) \in \mathcal{M}\}$, written as a row vector (see preceding paragraph) by concatenating the rows as a row vector of length N^2. That is, $\mathbf{1}_{\mathcal{M}}$ contains a 1 in every position corresponding to a network element belonging to $\mathcal{M}$ and a zero everywhere else. The direction $\mathbf{1}_{\mathcal{M}}$ represents increasing capacity of each of the network elements in $\mathcal{M}$. Then the delivery importance of $\mathcal{M}$ in the positive direction at C is defined as the derivative of the delivery function at C in the direction $\mathbf{1}_{\mathcal{M}}$, *i. e.*,

$$\Omega_+\left(\mathcal{M};C\right)=\left\langle \Psi'(C),\mathbf{1}_{\mathcal{M}}\right\rangle \ . \tag{1.9}$$

$\mathcal{M}$ need not be a subgraph of $\mathcal{H}$ because we may wish to study delivery importance for sets of elements that may be unrelated to one another by the graph structure. The delivery importance of the empty set $\mathcal{M} = \varnothing$ is clearly zero.

Under suitable regularity conditions, the delivery importance may be computed by the Fréchet differential [36]

$$\left\langle \Psi'(C),\mathbf{1}_{\mathcal{M}}\right\rangle =\lim_{h\to 0^+}\frac{1}{h}\left[\Psi\left(C+h\mathbf{1}_{\mathcal{M}}\right)-\Psi\left(C\right)\right]. \tag{1.10}$$

We may consider delivery importance as in equation (1.10) with any arrangement of signs in the matrix $\mathbf{1}_{\mathcal{M}}$, allowing for any combination of increases and decreases in network element capacity to be studied. We have chosen the positive sign everywhere primarily for consistency with the classical definition of reliability importance [1]. For network resiliency (see Section 3), we will be primarily interested in what happens when capacities decrease, so in the definition of network resiliency we will use delivery importance in the negative direction $-\mathbf{1}_{\mathcal{M}}$. It is also reasonable for some applications to consider delivery importance in other directions which can be represented by matrices that contain $\pm I\{(i, j) \in \mathcal{M}\}$ where the 1 and –1 are chosen to represent growth or diminishment of the capacity of the network element (i, j). In this case, letting $A_{\mathcal{M}}$ stand for a matrix of this kind that contains $\pm I\{(i, j) \in \mathcal{M}\}$, we will represent the delivery importance in the direction $A_{\mathcal{M}}$ by $\Omega_A(\mathcal{M}; C) = \langle \Psi'(C), A_{\mathcal{M}}\rangle$; in case all – signs are chosen, we write $\Omega_-(\mathcal{M}; C) = \langle \Psi'(C), -\mathbf{1}_{\mathcal{M}}\rangle$.

Theorem 1. For an almost everywhere differentiable scalar delivery function, the delivery importance in the positive direction at C_0 of a set $\mathcal{M}$ of network elements is equal to the sum of the partial derivatives of Ψ with respect to the capacities of the network elements in $\mathcal{M}$, evaluated at the nominal capacity matrix C_0:

$$\Omega_+\left(\mathcal{M};C_0\right)=\sum_{(i,j)\in\mathcal{M}}\left.\frac{\partial\Psi}{\partial c_{ij}}\right|_{C_0} \tag{1.11}$$

when it exists.

Proof. This is a restatement of the inner product expression $\Omega_+\left(\mathcal{M},C_0\right)=\left\langle \Psi'\left(C_0\right),\mathbf{1}_{\mathcal{M}}\right\rangle$.

■

Clearly, when $\mathcal{M}$ consists of a single element (i, j), $\Omega_+(\mathcal{M}; C)$ reduces to $\Omega_1(i, j; C)$. Again, this definition is motivated by a one-term Taylor series expansion

$$\Psi\left(C_0 + hA\right) - \Psi\left(C_0\right) = \sum_{(i,j)\in\mathcal{M}} h_{(i,j)} \left.\frac{\partial \Psi}{\partial c_{ij}}\right|_{C_0} a_{ij} + o\left(|h|\right) \tag{1.12}$$

where h is now an $|\mathcal{M}|$-vector; (1.12) comes about from taking all elements of h to be equal (which amounts to modeling the postulate that the capacities of all network elements in $\mathcal{M}$ are changed at the same rate).

Example 1 (cont'd). We compute the delivery importance of the two links (1, 2) and (2, 3) together. In this example, this is the delivery importance of the entire network with respect to itself. We have

$$\Omega_+\left(\mathcal{H}; C_0\right) = I\left\{x < y\right\} + I\left\{y < x\right\} = I\left\{c_{12} \neq c_{23}\right\} \tag{1.13}$$

which is 1 as long as $c_{12} \neq c_{23}$. Again when $c_{12} = c_{23}$, (1.11) is not defined and we use the $|\mathcal{M}|$-dimensional analog of (1.1) instead, with $\mathbf{1}_{\mathcal{H}}$ as the 3×3 matrix containing 1 in the (1, 2) and (2, 3) positions and zeros elsewhere:

$$\Omega_+\left(\mathcal{H}; C_0\right) = \lim_{h\to 0^+} \frac{1}{h}\left[\Psi\left(C_0 + h\mathbf{1}_{\mathcal{H}}\right) - \Psi\left(C_0\right)\right] = \lim_{h\to 0^+} \frac{1}{h}\left[\min\left\{c_{12} + h, c_{12} + h\right\} - c_{12}\right] = 1. \tag{1.14}$$

In this example, the delivery importance of the entire network with respect to itself is 1 regardless what C_0 might be. ■

2.2.1.2 *Delivery Importance in Deterministic Capacitated Discrete Networks*

In a capacitated discrete network, the capacity matrix has integer entries and the flows also take on only integer values. Derivatives of the delivery function are not defined in the classical sense. There are (at least) two alternatives. First, one may approximate the discrete network by a sequence of continuum networks for which the delivery importance is defined as in Section 2.2.1.1 and then define the delivery importance for the discrete network as the limit of the delivery importance values in the approximating sequence of continuum networks, if the limit exists. Some conditions for the execution of this scheme are discussed in [13]. The second alternative is to define delivery importance in a purely discrete fashion, as follows. Let $(\mathcal{H}, C, \Psi)$ be a deterministic capacitated discrete network with associated delivery function Ψ. The delivery importance at C of a network element (i, j) in the positive direction is given by

$$\Omega_+\left(i, j; C\right) = \Psi(C + \mathbf{1}_{ij}) - \Psi(C) \tag{1.15}$$

where $\mathbf{1}_{ij}$ is a matrix with a 1 in the i, j position and zeroes everywhere else. Similarly, for the delivery importance in the positive direction of a subset $\mathcal{M}$ of $\mathcal{N}\cup\mathcal{L}$, replace in (1.15) the $\mathbf{1}_{ij}$ matrix by $\mathbf{1}_{\mathcal{M}}$.

2.2.2 Delivery Importance in Stochastic Capacitated Networks

In a stochastic capacitated network, the network element capacities are random variables. Consequently, the delivery function (provided it is measurable), delivery importance, and network resiliency in such a network are also random variables. As such, figures of merit for delivery importance and network resiliency are constructed from moments and percentiles of these random variables. Thus, we speak of, for example, expected delivery importance as

$$E\Omega_{+}\left(\mathcal{M};C\right) = E\left\langle \Psi'(C), \mathbf{1}_{\mathcal{M}} \right\rangle = \int \left\langle \Psi'(C(\omega)), \mathbf{1}_{\mathcal{M}} \right\rangle P(d\omega) \tag{1.16}$$

where the integration is over the sample space of the capacity random variables. In a similar fashion, we may employ the variance of delivery importance, the median delivery importance, the 90th percentile delivery importance, and so on

2.2.2.1 Static Networks

In a static stochastic capacitated network, the capacities are fixed random quantities. These may represent the reliability of the network elements (in which case the range of the capacity function is [0, 1] and the capacity represents the probability that the element is present in $\mathcal{H}$; see Example 2 for an illustration of this case), or the load-carrying capacity of the network element in a model in which the capacity is indeterminate for any reason (for example, lack of precise information on the part of the network manager about the status of recently added elements to the network). It is required that the joint distribution of the network element capacities be known.

2.2.2.2 Dynamic Networks

In a dynamic stochastic capacitated network, the capacities are represented by stochastic processes with time as their parameter space. These networks model realistic networks having network element capacities that change at random because of, say, failures and repairs to parts of the network element. For example, the capacity of a network element in a deterministic capacitated network could be modeled with a continuous-time, discrete-state semi-Markov process. For dynamic stochastic capacitated networks, the figures of merit for delivery importance, like those described in Section 2.2.2, become functions of time as well.

2.2.3 Delivery Importance in Uncapacitated Networks

2.2.3.1 Delivery Importance in Deterministic Uncapacitated Networks

In an uncapacitated network, proper functioning of the network is affected by only the presence or absence of a node or link, so "capacities" 0 or 1 are assigned to each element of the network. The delivery function in such a network is restricted to be one of the connectivity functions: 2-terminal, *k*-terminal, or all-terminal. Delivery importance in the positive direction, as defined in Section 2.2.1.2, makes sense only for entries in the nominal capacity matrix where there are zeros. Similarly, delivery importance in the negative direction makes sense only for entries in the nominal capacity matrix where there are ones. For applications, probably the most useful delivery importance measure is delivery importance in the negative direction starting from a fully operational network, as in

$$\Omega_{-}\left(\mathcal{M};H\right)=\Psi\left(H-\mathbf{1}_{\mathcal{M}}\right)-\Psi\left(H\right) \tag{1.17}$$

where H is the full incidence matrix (*i. e.*, including the nodes) of $\mathcal{H}$. As usual, a vector delivery function is handled term-by-term.

2.2.3.2 Delivery Importance in Stochastic Uncapacitated Networks

In a stochastic uncapacitated network, the basic setup is as in Section 2.2.3.1 with the additional feature that network elements are each assigned a number representing the probability that the element is in the graph or not. In case these probabilities are fixed, the network is static in the sense of Section 2.2.2.1; in case they vary over time in some known fashion, the network is dynamic in the sense of Section 2.2.2.2. Moments, percentiles, etc., of the delivery function and various delivery importance values can then be computed from these probabilities.

Example 1 (cont'd): Consider again the network shown in Figure 1, now with unreliable links. We set $p_{ij} = \mathrm{E}I\{(i, j) \in \mathcal{H}\} = P\{(i, j) \in \mathcal{H}\}$, $i, j = 1, 2, 3$. For economy's sake, we will assume $p_{11} = p_{22} = p_{33} = 1$, and $p_{13} = p_{31} = 0$ because the network does not contain a link from node 1 to node 3. $p_{21} = p_{32} = 0$, because the network is directed, and we will assume that the remaining probabilities are neither zero nor one. Then the expected value of the delivery function is $\mathrm{E}\Psi = p_{12} + p_{23} - p_{12}p_{23}$. The expected delivery importance in the negative direction of the links are $\mathrm{E}\Omega_{-}(1, 2; H) = \mathrm{E}\Omega_{-}(2, 3; H) = p_{12}p_{23} - p_{12} - p_{23}$. ■

3 NETWORK RESILIENCY

3.1 Deterministic Networks

The first essential idea of network resiliency is that a resilient network is one that has few subsets that have large delivery importance, or conversely most subsets have small delivery importance. This expresses the idea that removal, or decrease in capacity of, most network elements has little effect on the delivery function of the network if the network is resilient. The remainder of this Section is devoted to formalizing this idea in the context of the framework we have created for delivery functions and delivery importance.

In the most general formulation, we define network resiliency to be the proportion of subsets of the network whose absolute delivery importance values in the negative direction do not exceed a given value. Formally, this is

$$\rho\left(\mathcal{H},C_0;x\right)=\frac{1}{2^{|\mathcal{N}\cup\mathcal{L}|}}\sum_{\mathcal{M}\subset\mathcal{N}\cup\mathcal{L}}I\left\{\left|\Omega_{-}\left(\mathcal{M};C_0\right)\right|\le x\right\},\quad x\ge 0\,, \tag{3.1}$$

a nondecreasing function. A resilient network is one for which this function remains close to zero, only increasing sharply to 1 for large values of x.

The computation of $\rho(\mathcal{H}, C_0; x)$ by equation (3.1) can be daunting for even modest-sized networks, and in practical cases there is usually little reason to consider very large subsets of $\mathcal{N}\cup\mathcal{L}$ unless events that disturb all but a small number of network elements are common.[1] Therefore, there is interest in restricted network resiliency figures of merit based on only certain subsets of $\mathcal{N}\cup\mathcal{L}$, especially if study of only a particular subset is of interest or if it is possible to identify in advance subsets that have less influence on the delivery function. It is useful to denote restricted network resiliency, say with respect to a subset Z of $\mathcal{N}\cup\mathcal{L}$, by $\rho(\mathcal{H}, Z, C_0; x)$. Then we would have

$$\rho\left(\mathcal{H}, \mathcal{Z}, C_0; x\right) = \frac{1}{\left|\mathcal{Z}\right|} \sum_{\mathcal{M}\subset\mathcal{Z}} I\left\{\left|\Omega_{-}\left(\mathcal{M}; C_0\right)\right| \le x\right\}, \quad x \ge 0\,. \tag{3.2}$$

For instance, we may be interested in the resiliency of the network when disruptions to only single network elements are considered; we call this the 1-resiliency of the network and compute it with $\mathcal{Z}$ equal to the set of all single-element subsets of $\mathcal{N}\cup\mathcal{L}$. More generally, we may consider the k-resiliency of the network,

$$\rho_k\left(\mathcal{H}, C_0; x\right) = \binom{\left|\mathcal{N}\cup\mathcal{L}\right|}{k}^{-1} \sum_{\mathcal{M}\subset\mathcal{Z}_k} I\left\{\left|\Omega_{-}\left(\mathcal{M}; C_0\right)\right| \le x\right\}, \quad x \ge 0\,, \tag{3.3}$$

where $\mathcal{Z}_k$ represents the set of all subsets of $\mathcal{N}\cup\mathcal{L}$ that contain k elements.

In applications, these expressions may be difficult to work with because they are functions, rather than scalars, so comparisons between different networks are not always possible.[2] Also, significant computation might be required to evaluate them. Thus, there is interest in simpler network resiliency characterizations; Section 4 is devoted to some explorations of these ideas.

3.2 Stochastic Networks

In a stochastic network, $\Omega_{-}(\mathcal{M}, C_0)$ is a random variable, and the expected value of (3.1) contains its distribution:

$$\mathrm{E}\rho\left(\mathcal{H}, C_0; x\right) = \frac{1}{2^{\left|\mathcal{N}\cup\mathcal{L}\right|}} \sum_{\mathcal{M}\subset\mathcal{N}\cup\mathcal{L}} P\left\{\left|\Omega_{-}\left(\mathcal{M}; C_0\right)\right| \le x\right\}, \quad x \ge 0\,. \tag{3.4}$$

Similar expressions for the expected values of (3.2) and (3.3) may be developed.

[1] Certain kinds of cyberattacks can disturb large subsets of a network at once.
[2] A partial order, similar to stochastic order, can be developed based on these expressions.

4 FIGURES OF MERIT FOR NETWORK RESILIENCY

4.1 Deterministic Networks

We use the concepts of delivery function and delivery importance of (sets of) network elements to define network resiliency in quantitative terms. Note that because delivery importance is keyed to a specific delivery function and nominal capacity matrix, network resiliency is specifically defined only with respect to the chosen delivery function and nominal capacity matrix.

Definition. For a deterministic capacitated network ($\mathcal{H}$, C, Ψ) with associated delivery function, *scalar network resiliency* $\rho^*(\mathcal{H}; C_0)$ at a nominal capacity matrix C_0 is defined as the largest of the absolute delivery importance values in the negative direction (capacities decrease) across all subsets of the elements (nodes and links) in the network:

$$\rho^*\left(\mathcal{H};C_0\right) = \max_{\mathcal{M}\subset\mathcal{N}\cup\mathcal{L}} \left|\Omega_-\left(\mathcal{M};C_0\right)\right| \tag{4.1}$$

.

The largest delivery importance value indicates the subset of $\mathcal{N}\cup\mathcal{L}$ to which the delivery function is most sensitive or that has the largest "lever effect" on the delivery function. Note that smaller is better in (4.1) because we want a resilient network to be one that is insensitive to changes in its network elements, *i. e.*, one that makes the right-hand side of (4.1) small. That is, we say a network is resilient if $\rho^*(\mathcal{H}, C_0)$ is small. Computation of network resiliency for a scalar delivery function by this definition requires only $|\mathcal{N}\cup\mathcal{L}|$ delivery importance computations and at most $|\mathcal{N}\cup\mathcal{L}|\cdot 2^{|\mathcal{N}\cup\mathcal{L}|}$ additions because the delivery importance for a subset $\mathcal{M}$ of $\mathcal{N}\cup\mathcal{L}$ is obtained by summing the delivery importance values of each of the individual network elements contained in $\mathcal{M}$; see Theorem 1.

Clearly, the same kinds of restricted scalar network resiliency values as, for example, in (3.2) and (3.3) are possible here as well; unless otherwise specified, further discussion of scalar network resiliency in this article will refer to the full definition (4.1).

Scalar network resiliency is not an absolute scale quantity but it does induce a partial order on the set of networks with associated delivery functions. Therefore, it is mainly useful for making comparisons in a set of networks rather than for associating a meaningful number with a fixed network.

4.2 Stochastic Networks

In a stochastic network, scalar network resiliency (4.1) is a random variable, and may be a function of time also if the network is dynamic. As before, figures of merit for network resiliency are constructed based on moments and percentiles of the network resiliency random variable. If the network is dynamic, these figures of merit will be functions of time as well.

5 METRICS FOR NETWORK RESILIECY

A figure of merit, in the sense used in Section 4, represents an "ideal" description of some phenomenon in a model. A figure of merit is computed from assumptions and a model structure and represents a quantitative description of some inherent quality of the phenomenon under study. For example, "maintainability" is the inherent ability of a system to be repaired and restored to service when maintenance is conducted by personnel using specified skill levels and prescribed procedures and resources [2]. Many figures of merit are associated with maintainability, such as maintenance downtime, logistic delay, etc. These may be developed from a model of the system under study (for example, an alternating renewal process) together with assumptions about location of spares, skill of repair personnel, etc. When the system is in operation, data may be collected relating to these figures of merit, and estimators of the population value of the figure of merit may be constructed from these data. These estimators are called metrics. For example, during system operation we may record the amount of time the system is out of service for maintenance reasons (preventive, corrective, opportunistic, etc.) and these data may be used to estimate the mean and variance of the maintenance downtime across a population of similar systems. For further discussion of figures of merit and metric, see [37].

In the case of network resiliency, we have introduced several figures of merit in Section 4. Metrics may be constructed for these from data relating to delivery importance. For instance, we may record the change in the delivery function between two specified network nodes when there is a failure of one element somewhere in the network. Aggregation of these data would permit estimation of the 1-resiliency of the network. Clearly, a lot of detail has been omitted from this description, and several questions would need to be settled before satisfactory metrics for network resiliency could be developed. We reserve detailed discussion of metrics for network resiliency for further treatment elsewhere.

6 THREE EXAMPLES OF DELIVERY FUNCTIONS

6.1 Connectivity

Consider a network $(\mathcal{H}; C)$ in which all capacities are either zero or one (that is, $c_{ij} = I\{(i, j) \in \mathcal{L}\}$). Associated delivery functions consistent with this example are 2-terminal, k-terminal ($2 < k < N$), or all-terminal connectivity of $(\mathcal{H}; C)$. We define the connectivity of some subset of nodes in the network to be 1 if there are paths connecting all the nodes in the subset to one another (that is, every node in the subset can be reached from other node in the subset), and 0 otherwise. Note that some authors define connectivity to be the probability that paths connecting the nodes exist when the probability that a network element is in the graph is given; this is the expected connectivity in our formulation and is studied in the next Section under Reliability. Computation of all-terminal connectivity in the latter sense (*i. e.*, all-terminal reliability) using a reduction algorithm is reviewed in [40].

The following simple example conveys the essential points of network resiliency involving connectivity.

Example 2. Consider the simple three-terminal directed network $\mathcal{H}$ shown in Figure 2.

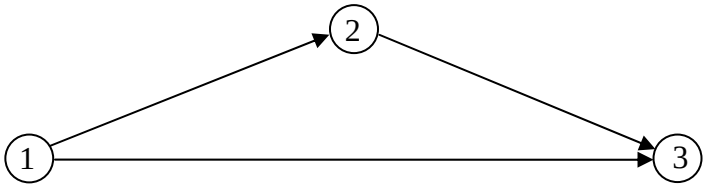

Figure 2.

The full incidence matrix of $\mathcal{H}$ is

$$H = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \tag{6.1}$$

and the delivery function Ψ for this example will be the two-terminal connectivity for terminals 1 and 3. That is, $\Psi = 1$ if there is a path connecting node 1 to node 3 and is zero otherwise. The capacity of element $(i, j) \in \mathcal{H}$ is $c_{ij} = I\{(i, j) \in \mathcal{H}\}$, $i, j = 1, 2, 3$. For economy's sake, let us ignore the contribution of the nodes; it is clear enough how to incorporate their contribution should it be desirable. Then

$$\Psi = I\left\{\left\{(1,3) \in \mathcal{H}\right\} \cup \left[\left\{(1,2) \in \mathcal{H}\right\} \cap \left\{(2,3) \in \mathcal{H}\right\}\right]\right\} = c_{13} + c_{12}c_{23} - c_{12}c_{23}c_{13}\,. \tag{6.2}$$

The delivery importance of link (1,3) in the negative direction is

$$\begin{aligned} \Omega_{-}\left((1,3); C\right) &= \Psi\left(C - \mathbf{1}_{(1,3)}\right) - \Psi\left(C\right) = c_{12}c_{23} - \left[c_{13} + c_{12}c_{23} - c_{12}c_{23}c_{13}\right] \\ &= c_{12}c_{23}c_{13} - c_{13}\,. \end{aligned} \tag{6.3}$$

The following table gives the delivery importance values for each nonempty subset of $\mathcal{H}$.

SUBSET OF $\mathcal{H}$	**DELIVERY IMPORTANCE**
$\{(1, 3)\}$	$c_{12}c_{23}c_{13} - c_{13}$
$\{(2, 3)\}$	$c_{12}c_{23}c_{13} - c_{12}c_{23}$
$\{(1, 2)\}$	$c_{12}c_{23}c_{13} - c_{12}c_{23}$
$\{(1, 2) \cup (2, 3)\}$	$c_{12}c_{23}c_{13} - c_{12}c_{23}$
$\{(1, 3) \cup (2, 3)\}$	$c_{12}c_{23}c_{13} - c_{12}c_{23} - c_{13} = -\Psi(\mathcal{H})$
$\{(1, 3) \cup (1, 2)\}$	$c_{12}c_{23}c_{13} - c_{12}c_{23} - c_{13}$
$\mathcal{H}$	$c_{12}c_{23}c_{13} - c_{12}c_{23} - c_{13}$

Table 1. Delivery Importance Values for Example 2.

Choose as a nominal capacity matrix the network's incidence matrix H (equation (6.1)). This represents a situation in which all the network's elements are initially in an operating condition. Then, with respect to this nominal capacity matrix, the delivery importance values from equation (1.17) are $\Omega_-(\mathcal{M}, H) = 0$ for $\mathcal{M} = \{(1, 3)\}$, $\{(2, 3)\}$, $\{(1, 2)\}$, and $\{(1, 2) \cup (2, 3)\}$, and $\Omega_-(\mathcal{M}, H) = -1$ for $\mathcal{M} = \{(1, 3) \cup (2, 3)\}$, $\{(1, 3) \cup (1, 2)\}$, and $\mathcal{H}$. The network resiliency function (3.1) at the nominal capacity matrix H is a right-continuous step function with jumps at 0 and 1; the height of the jump at zero is 5/8 and the height of the jump at 1 is 3/8. ■

6.2 Reliability

Consider a network $(\mathcal{H}; C)$ in which all capacities are either zero or one (that is, $c_{ij} = I\{(i, j) \in \mathcal{L}\}$ as before) and, in addition, $P\{c_{ij} = 1\} = p_{ij}$ for all $(i, j) \in \mathcal{L}$. Such a model represents a network in which the network elements are unreliable, *i. e.*, may fail to function in the network with some given probability. Suitable delivery functions in this case include 2-terminal, k-terminal, and all-terminal reliability, defined as the probability that there are working paths connecting a given subset of the nodes of the network (two nodes, k nodes, or all nodes, as appropriate). We continue with Example 2 to illustrate network resiliency in this case.

Example 2 (cont'd): The capacity of element $(i, j) \in \mathcal{H}$ is $c_{ij} = I\{(i, j) \in \mathcal{H}\}$, $i, j = 1, 2, 3$, with $P\{c_{ij} = 1\} = p_{ij}$ for all $(i, j) \in \mathcal{L}$. We assume the network elements are mutually stochastically independent. For economy's sake, we again ignore the contribution of the nodes. The following table gives the delivery importance and expected delivery importance values for each nonempty subset of $\mathcal{H}$.

SUBSET OF $\mathcal{H}$	DELIVERY IMPORTANCE	EXPECTED DELIVERY IMPORTANCE
$\{(1, 3)\}$	$c_{12}c_{23}c_{13} - c_{13}$	$p_{12}p_{23}p_{13} - p_{13}$
$\{(2, 3)\}$	$c_{12}c_{23}c_{13} - c_{12}c_{23}$	$p_{12}p_{23}p_{13} - p_{12}p_{23}$
$\{(1, 2)\}$	$c_{12}c_{23}c_{13} - c_{12}c_{23}$	$p_{12}p_{23}p_{13} - p_{12}p_{23}$
$\{(1, 2) \cup (2, 3)\}$	$c_{12}c_{23}c_{13} - c_{12}c_{23}$	$p_{12}p_{23}p_{13} - p_{12}p_{23}$
$\{(1, 3) \cup (2, 3)\}$	$c_{12}c_{23}c_{13} - c_{12}c_{23} - c_{13} = -\Psi(\mathcal{H})$	$p_{12}p_{23}p_{13} - p_{12}p_{23} - p_{13}$
$\{(1, 3) \cup (1, 2)\}$	$c_{12}c_{23}c_{13} - c_{12}c_{23} - c_{13}$	$p_{12}p_{23}p_{13} - p_{12}p_{23} - p_{13}$
$\mathcal{H}$	$c_{12}c_{23}c_{13} - c_{12}c_{23} - c_{13}$	$p_{12}p_{23}p_{13} - p_{12}p_{23} - p_{13}$

Table 2. Expected Delivery Importance Values for Example 2.

As before, when the nominal capacity matrix is H, the delivery importance values from equation (1.17) are $\Omega_-(\mathcal{M}, H) = 0$ for $\mathcal{M} = \{(1, 3)\}$, $\{(2, 3)\}$, $\{(1, 2)\}$, and $\{(1, 2) \cup (2, 3)\}$, and $\Omega_-(\mathcal{M}, H) = -1$ for $\mathcal{M} = \{(1, 3) \cup (2, 3)\}$, $\{(1, 3) \cup (1, 2)\}$, and $\mathcal{H}$. The network resiliency function is zero

with probability $p_\infty = P[\{(1, 3) \notin \mathcal{H}\} \cup \{(2,3) \notin \mathcal{H}\} \cup \{(1, 2) \notin \mathcal{H}\} \cup [\{(1, 2) \notin \mathcal{H}\} \cap \{(2, 3) \notin \mathcal{H}\}]]$ and is -1 with probability $1 - p_\infty$. The expected network resiliency is $p_\infty - 1$. ■

6.3 Stochastic Flow

Consider a continuum network that delivers *e. g.* oil from node A to node B. The node and link capacities in this network are modeled as continuous time Gaussian processes $\{X_{ij}(t) : t \geq 0\}$ having almost surely continuously differentiable sample paths ([9], section 4.3), indicating the change in capacity of the network elements from time to time due to failures and repairs, degradation, corrosion, or like phenomena. The delivery function will be the volume of oil delivered from node A to node B over a stated interval of time $[0, T]$, $T < \infty$. The capacity of a network element is the maximum flow (volume per unit time) permitted in that network element. Set $X(t) = (X_{ij}(t): i, j = 1, \ldots, N)$ and $C_0 = X(0)$. Let $\varphi(t)$ denote the flow in the network at time t assuming a maximum-flow scenario [14], and $\varphi_{AB}(t)$ denote the flow from node A to node B under this scenario. Then the delivery function consistent with this model is

$$\Psi(X;T) = \int_0^T \varphi_{AB}(t)\, dt\,. \tag{6.4}$$

Then the delivery importance of a subset $\mathcal{M}$ of network elements is given by the random variable

$$\Omega_{-}(\mathcal{M}, X) = -\sum_{(i,j)\in\mathcal{M}} \int_0^T \frac{\partial \varphi_{AB}}{\partial x_{ij}}(t)\, dt\,. \tag{6.5}$$

If, as is usually reasonable to assume, the flow φ_{AB} is a nondecreasing function of the network element capacities, then $\Omega_{-}(\mathcal{M}, X) \leq 0$, and the scalar network resiliency with respect to the given delivery function is

$$\rho^*(\mathcal{H}; X) = \max_{\mathcal{M}\subset\mathcal{N}\cup\mathcal{L}} \sum_{(i,j)\in\mathcal{M}} \int_0^T \frac{\partial \varphi_{AB}}{\partial x_{ij}}(t)\, dt\,. \tag{6.6}$$

Assuming the processes $\{X_{ij}(t)\}$ are mutually stochastically independent, we obtain

$$P\{\Psi(X;T) > V\} = \int_0^\infty \cdots \int_0^\infty P\left\{\int_0^T \varphi_{AB}(t)\, dt > V \,\middle|\, X_{ij}(t) = x_{ij}\right\} dx_{ij}\,, \tag{6.7}$$

and then the methods of [29] or [30] may be used to calculate $P\{\Psi(X; T) > V\}$, where V is a desired volume of oil.

7 NETWORK RESILIENCY OPTIMIZATION

In creation of real networks, it is clearly of interest to be able to optimize network resiliency. In this Section, we introduce an optimization problem for network resiliency in a capacitated network when a cost function for provisioning specified node and link capacity is known. $\gamma(C)$ denotes the cost of deploying a network whose capacity matrix is C.

Solving the following optimization problem leads to a network deigned for maximum scalar resiliency:

$$\textbf{Maximize } \rho^*(\mathcal{H}; C) \textbf{ subject to } \gamma(C) \leq \gamma_0$$

where γ_0 is the allowable maximum cost. The design variables that may be manipulated to complete the optimization include the network topology and the capacity matrix.

8 DESIGN FOR NETWORK RESILIENCY PRINCIPLES AND PRACTICES

This Section discusses specific actions that can be taken by network designers to increase network resiliency. We evaluate the efficacy of design actions by determining whether the scalar network resiliency (Section 4) decreases (smaller is better) as a result of the action. In a capacitated network, it is reasonable to assume that the delivery function is a nondecreasing function of the capacity matrix (that is, an increase in one or more entries in the capacity matrix causes the delivery function to increase or remain the same).

8.1 Systematic Overprovisioning

In a capacitated network, one of the most common ways of promoting greater network resiliency is by overprovisioning, that is, making the capacities of the network elements be larger than would be required to just meet stated service reliability objectives. For example, in early telecommunications networks, certain trunk groups (collections of circuits connecting two switching offices) were designed to block no more than 1% of the offered traffic during the busy hour [33]. For an offered load of 100 erlangs, the number of trunks required to meet this goal is 117. To overprovision, a larger number of trunks, say 144, would be installed, and then, with the same 100 erlang load, the blocking probability would be reduced to 6.7×10^{-6}. This same idea can be replicated throughout a capacitated network; while the computation of delivery importance and network resiliency might be more complicated, the basic idea is the same as the simple circuit-switched telecommunications trunk group example.

If the delivery function is a nondecreasing function of the capacity matrix, then overprovisioning (that is, providing more capacity on a route than is required to just meet a stated service objective for that route) leads to increased scalar network resiliency with respect to a nominal capacity matrix representing exact provisioning because all delivery importance values are nonnegative. Whether overprovisioning is an efficient means of promoting network resiliency is another matter that is beyond the scope of this article.

8.2 Failure Decoupling

When a network's associated delivery function is monotone, the larger the number of network elements that experience malfunction (or decrease in capacity), the larger the decrease in the value of the delivery function. It follows that network resiliency will be improved if there can be employed schemes to prevent failures in network elements from cascading or propagating to other network elements. We refer to such schemes broadly as *failure decoupling*. Failure to properly decouple network elements may lead to widespread outages and greatly impaired delivery functions. For example, in January 1990, the AT&T Signaling System 7 (SS7) transaction management data network suffered an extensive outage because a software failure in one of the SS7 switching nodes led to a situation in which nodes connected to the failed node went out of service because they received improper status information from the failed node. The failures cascaded in less than 10 minutes to over 100 SS7 nodes, resulting in a nine-hour outage that affected 60,000 customers and cost AT&T approximately $60 million in lost revenue [18]. Had flawed status information not been passed on to neighboring nodes, the scope of the outage would have been greatly reduced.

Measures that can be taken as part of network design to minimize the possibility of propagating a network element failure to other network elements include implementation of a supervisory network monitoring system such as that employed in undersea cable telecommunications systems [34], implementation of a "heartbeat" status monitoring and communication system [28], and implementation of a capability to notify network operations managers of network element failure in real time.

The latter scheme points toward the role of network management and control in containing the propagation of network element failures. As many large network operations implement a centralized network monitoring and intervention capability (for example, the AT&T Global Network Operations Center [17]), we discuss this idea in more detail below.

8.2.1 Reactive Control

Reactive control refers to the actions that are taken by network managers (human or automatic) in response to a disruption in the network. A disruption is by definition something that interferes with the satisfactory functioning of one or more network elements, leading to a degradation in the performance of the network's functions, and (with the assumption about monotonicity of the delivery function given in Section 8) is reflected in a decrease in the value of the network's delivery function. Disruptions may be routine and manageable, or they may be sudden and unexpected.

Routine disruptions are those that can be anticipated and planned for in advance of their occurrence. For example, the chronic process of failures and repairs that takes place in all network elements can be thought of as a routine disruption because network managers know that these will occur even if it is not possible to predict exactly when and where they will occur. In a sense, they can be predicted statistically; indeed, about 3% of the routers in the Internet are out of service at any time because of routine failures. Consequently, managers can anticipate these failures and put in place operational tactics that mitigate their effects. Indeed, the TCP/IP

network protocol is specifically designed to automatically route around failed routers and/or links. This, together with the hub-and-spoke topology of the Internet, allows us to conjecture that the resiliency (4.1) of the Internet has a logistic shape: it is small for increasing numbers of failed routers and/or links until a threshold is reached, and then it increases abruptly as the number of failed routers and/or links continues to increase. An example of an automatic reactive control is given by the AT&T FASTAR [5] system created to enable automatic rerouting around failed DS3 links in the TDM public switched telephone network.

Sudden and unpredictable disruptions are due to events like natural disasters and deliberate attacks. These are usually large in scope and extended in time. In circuit-switched telecommunications networks, long experience helped codify a set of network management principles that were employed to mitigate the effects of losing large portions of a network due to, *e. g.*, earthquakes, floods, and other natural disasters. That technology had advanced to a point that it was encapsulated in an expert system [10].

8.2.2 Smart Networks

There has been a recent trend toward exploiting information technology for improved management of network operations. The ability to keep track of and archive large quantities of data makes it possible to obtain a richer picture of network operations and to run analyses in real time that allow more intelligent decisions to be made regarding management of a network for continued satisfactory operation of a delivery function when changes to the network occur. Two approaches have been common in recent literature: a top-down approach and a bottom-up approach. We briefly discuss each of these in turn.

8.2.2.1 Macro-Smarts

A macro-smart network is one that uses global, or network-wide, data to inform network managers about current status and potential evolution of network operation. Internet Protocol (IP) telecommunications networks encapsulate this sort of network management by the way routing tables change in response to failed links and routers. Updating routing tables in real time allows an IP network's open-shortest-path-first (OSPF) routing scheme to automatically bypass failed routers and links. That is, a router in an IP network will never even try to send a packet over a failed link or to a failed router when its routing tables are up to date with current information on the location of failures in the network.

8.2.2.2 Micro-Smarts

A micro-smart network is one that uses local data to inform network managers about current status and potential evolution of network operation. For example, electric power networks are beginning to use so-called "smart metering" to enable the network operator to control the flow of power to specific end users at specific times so that focused overloads, such as might occur because of heavy air conditioning demands on hot summer days, can be avoided for the benefit of a larger number of users who will not experience a blackout that might otherwise occur due to the overload. This is an example of an active intervention that is possible because information that was not previously available at the end-user level of detail is now made visible by smart metering. It is anticipated that the smart grid market will reach $200 billion by 2015 [19].

9 NETWORK INTERDICTION

Network interdiction can be considered as the study of methods for making a network *less* resilient. Network interdiction studies aim to find the network element, or set of network elements, that maximally disrupts the operation of the network. In the language introduced here, maximal disruption to the operation of the network means the largest decrease in the network's delivery function. The set $\mathcal{M}_0$ of $\mathcal{N} \cup \mathcal{L}$ that realizes the scalar network resiliency value, $\rho^*(\mathcal{H};C) = \left|\Omega_-(\mathcal{M}_0;C)\right|$, is the subset of network elements whose removal from the network decreases the delivery function (when the capacity matrix is *C*) more than the removal of any other subset and therefore is the set sought by network interdiction.

Network interdiction studies began in a military context; for some examples see [24], [16]. Other network interdiction studies include [32], [20], [22], and [8].

10 CONCLUSION

This article presents a quantitative framework for studying network resiliency. The concept of network resiliency has gained increased importance recently as executives and managers realize the possibilities for disruption of critical network infrastructures because of not only natural disasters but also malicious actors and deliberate attacks. We introduce the concepts of a network with associated delivery function and of delivery importance to enable quantitative characterization of network resiliency. We discuss the application to discrete and continuum capacitated and uncapacitated networks, both deterministic and stochastic. Three examples of delivery functions are provided, and we briefly summarize how to solve the network interdiction using the proposed network resiliency framework.

REFERENCES

1. Birnbaum, Z. W., Esary J. D., and Saunders, S. C. (1961), Multi-Component Systems and Their Reliability. *Technometrics* **3**, 55-77.
2. Blanchard, B. S., D. Verma, and E. L. Peterson (1995), *Maintainability: A Key to Effective Serviceability and Maintenance Management*. New York: John Wiley and Sons.
3. Bu, T., Norden, S., and Woo, T. (2004), Trading Resiliency for Security: Model and Algorithms. *Proceedings of the 12th IEEE Conference on Network Protocols*, 218-227.
4. Burr, S. A. (ed.) (1982), *The Mathematics of Networks*. Volume 26 in the Proceedings of Symposia in Applied Mathematics. Providence, RI: American Mathematical Society.
5. Chao, C.-W., Fuoco, G., and Kropfl, D. (1994), FASTAR™ Platform Gives the Network A Competitive Edge. *AT&T Tech. J.* **73** no. 4, 69-81.
6. Colbourn, C. (1987), *The Combinatorics of Network Reliability*. New York: Oxford University Press.
7. Colbourn, C. (1987), Network Resilience. *SIAM Journal on Algebraic and Discrete Methods* **8** no. 3, 404-409.
8. Cormican, K. J., Morton, D. P., and Wood, R. K. (1998), Stochastic Network Interdiction. *Operations Research* **46** no. 2, 184-197.
9. Cramer, H. and Leadbetter, M. R. (1967), *Stationary and Related Stochastic Processes: Sample Function Properties and Their Applications*. New York: John Wiley and Sons.
10. Cronk, R. N., Callahan, P. H., and Bernstein, L. (1988), Rule-based Expert Systems for Network Management and Operations: An Introduction. *IEEE Network* **2** no. 5, 7-21.
11. Dieudonne, J. (1968), *Foundations of Modern Analysis*. New York: Academic Press.
12. Dolev, D., Jamin, S., Mokryn, O., and Shavitt, Y. (2006), Internet Resiliency to Attacks and Failures under BGP Policy Routing. *Computer Networks* **50** no. 16, 3183-3196.
13. Driscoll, P. J. and Tortorella, M. (2009), Another Type of Continuum Approximation for Discrete Networks. In preparation.
14. Ford, L.R. and Fulkerson, E. (1962), *Flows in Networks*. Princeton, NJ: Princeton University Press.
15. Frank, H. (1974), Survivability Analysis of Command and Control Communication Networks. *IEEE Transactions on Communications* **22** no. 5, 589-605.
16. Golden, B. (1978), A Problem in Network Interdiction. *Naval Research Logistics Quarterly* **25** no. 4, 711-713.
17. http://www.corp.att.com/history/nethistory/management.html
18. http://www.informit.com/library/content.aspx?b=Signaling_System_No_7&seqNum=19
19. http://telephonyonline.com/residential_services/news/smart-grid-market-1229/
20. Israeli, E. and Wood, R. K. (2002), Shortest-Path Network Interdiction. *Networks* **40** no. 2, 97-111.
21. Lang, J. P. and Drake, J. (2002), Mesh Network Resiliency Using GMPLS. *Proceedings of the IEEE* **90** no. 9, 1559-1568.
22. Lim, C. and Smith, J. C. (2007), Algorithms for Discrete and Continuous Multicommodity Flow Network Interdiction Problems. *IIE Transactions* **39** no. 1, 15-26.

23. Liu, Y. and Trivedi, K. S. (2004), A General Framework for Network Survivability Quantification. In: *Proceedings of the 12th GI/ITG Conference on Measuring, Modelling and Evaluation Of Computer and Communication Systems (MMB) together with 3rd Polish-German Teletraffic Symposium*.
24. McMasters, A. W. and Mustin, T. M. (1970), Optimal Interdiction of a Supply Network. *Naval Research Logistics Quarterly* **17** no. 3, 261-268.
25. Medhi, D. (1994), A Unified Approach to Network Survivability for Teletraffic Networks: Models, Algorithms, and Analysis. *IEEE Transactions on Communications* **42** no. 2/3/4, 534-548.
26. Najjar, W. and Gaudiot, J.-L. (1990), Network Resilience: A Measure of Network Fault Tolerance. *IEEE Transactions on Computers* **39** no. 2, 174-181.
27. Najjar, W. and Srimani, P. K. (1991), Network Resilience of Star Graphs: A Comparative Analysis. *Proceedings of the 19th Annual Conference on Computer Science*, 349-357.
28. Nguyen, H. T., Griffiths, M., and LaPierre, S. (2006), Systems and Methods for an Infrastructure Centralized Heartbeat. U. S. Patent no. 7,120,688.
29. Ramirez-Marquez, J. E., Coit, D., and Tortorella, M. (2005), Multi-State Two-Terminal Reliability: A Generalized Cut-Set Approach. *IEEE Transactions on Reliability*
30. Ramirez-Marquez, J., Coit, D., and Tortorella, M. (2006), A Generalized Multi-State Based Path Vector Approach for Multi-State Two-Terminal Reliability. *IIE Transactions on Quality and Reliability Engineering* **38**, no. 6, 477-488.
31. Ramirez-Marquez, J. E., Rocco, C. M., Gebre, D. A., Coit, D. W., and Tortorella, M. (2006), New Insights on Multi-State Component Criticality and Importance. *Reliability Engineering and System Safety* **91** no. 8, 894-904.
32. Ramirez-Marquez, J. E. and Rocco, C. M. (2009), Stochastic Network Interdiction Optimization via Capacitated Network Reliability Modeling and Probabilistic Solution Discovery. *Reliability Engineering and System Safety* **94** no. 5, 913-921.
33. R. F. Rey (ed.) (1983), *Engineering and Operations in the Bell System*, 2nd edition. Murray Hill, NJ: AT&T Bell Laboratories.
34. Runge, P. K. (1992), Undersea Lightwave Systems. *AT&T Technical Journal* **71** no. 1, 5-13.
35. Sohn, J., Tschangho, J. K., Hewings, G. J. D., Lee, J. S., and Jang, S.-G. (2003), Retrofit Priority of Transport Network Links under an Earthquake. *J. Urban Planning and Development* **129** no. 4, 195-210.
36. Tapia, R. A. (1971), The Differentiation and Integration of Nonlinear Operators. In *Nonlinear Functional Analysis and Applications*, ed. L. B. Rall. University of Wisconsin Mathematics Research Center Pub. no. 26. New York: Academic Press.
37. M. Tortorella (2005), Service Reliability Theory and Engineering, I: Foundations. *Quality Technology and Quantitative Management* **2** no. 1, 1-16.
38. M. Tortorella (2005), Service Reliability Theory and Engineering, II: Models and Examples. *Quality Technology and Quantitative Management* **2** no. 1, 17-37.
39. Touvet, F. and Harle, D. (2001), Network Resilience in Multilayer Networks: A Critical Review and Open Issues. *Lecture Notes in Computer Science* no. 2093, 829-837. New York: Springer-Verlag.
40. Wood, R. K. (1986), Factoring Algorithms for Computing *K*-Terminal Network Reliability. *IEEE Transactions on Reliability* **R-35** no. 3, 269-278.