A geographic perspective on commercial Internet survivability

Tony H. Grubesic *, Morton E. O’Kelly, Alan T. Murray

Department of Geography, The Ohio State University, Columbus, OH 43206-1361, USA

Abstract

The earliest predecessor to the commercial Internet of today was ARPANET, a packet switched computer network developed by the US Defense Department’s Advanced Research Projects Agency. Designed to withstand a nuclear attack, ARPANET utilized a deurbanized, decentralized, and distributed network topology. As ARPANET gradually evolved into NSFNET and eventually the commercial Internet, increasing traffic, demands for interconnection, and growing private interests required the movement from a distributed network topology to a more economically viable network configuration, hub-and-spoke. Although transmission speeds and capacities of today’s commercial Internet clearly surpass those of its predecessors, the economics of network survivability and reliability have also become more relevant. With thousands of businesses, corporations, universities, and governments relying on the Internet for day-to-day functions, major disruptions in service have the potential to be economically catastrophic. This paper explores the network topology of the commercial Internet, with a focus on network survivability. GIS based approaches are used to simulate both nodal and link failure in the US commercial backbone system in order to assess potential impacts. Results suggest that many of the larger metropolitan benefit from robust network infrastructure, while smaller cities are more prone to service disruptions.

© 2002 Elsevier Science Ltd. All rights reserved.

1. Introduction

Recent explorations of the geographical characteristics of the US commercial Internet reveal a complex “network of networks” with scores of individual backbones linking hundreds of cities with fiber-optic cables (Wheeler and O’Kelly, 1999; Gorman and Malecki, 2000; Malecki and Gorman, 2001). This infrastructure not
only provides the means to exchange information via basic services such as email, but also facilitates the exchange of complex digital information such as audio and video. Although there is some debate as to the effects of information technologies on productivity, there is little doubt of the overall impact of information technology on the evolving digital economy (Leinbach, 2001). For example, a recent report by the US Census Bureau (2001) estimates that online purchases in the United States totaled $28 billion for 2000. This represents an increase of 62% over 1999. Moreover, evidence suggests that online retailing and e-commerce are not the only applications benefiting from advanced telecommunications infrastructure and technologies. Internet related applications for e-government are also growing rapidly. For example, a recent report from the Taubman Center for Public Policy (2001) at Brown University indicates that 93% of government publications are now available online, up from 74% in 2000. In addition, 54% of government databases are online, up from 42% from 2000. Results also reveal a modest increase in the number of applications that are fully executable on the web. Examples include filing taxes online, being able to order government publications online, filing complaints, registering vehicles and ordering hunting licenses.

Although the number of web-based applications enabled by advanced infrastructure continues to grow, the previous examples fail to convey the overall importance of telecommunication infrastructure in the day-to-day operations for both private and public sectors of the US economy. The recent tragedy in New York City has made clear the extent to which an increasing dependence on telecommunication networks permeates day-to-day functions. After the collapse of the World Trade Center towers, three New York counties lost their connection to the statewide computer system when a major telecommunication hub located at ‘ground zero’ failed. As a result, Nassau, Suffolk, and Westchester counties lost the ability to make social services entries and payments and were cut off from the statewide health network and most of the state police network (GovTech, 2001). With a combined population of 3.7 million, all three counties have significant interaction with the state, making the service failure a noteworthy interruption. Fortunately, Westchester County was able to redirect network applications through a newly constructed fiber-optic network that connects to the state’s new fiber backbone, NyeNet (GovTech, 2001). Nassau and Suffolk counties also utilized the Westchester network hub to bypass the crippled hub in Manhattan and access state computers located in Albany. Other telecommunication customers and facilities were not so lucky or flexible. Telecommunication ‘carrier hotels’ were forced to utilize backup generators to maintain service because the Con Edison power grid was completely knocked out in lower Manhattan. Due to problems such as fuel shortages and overheating, the generators were not much help. As such, several major Internet services and e-business providers including Electronic Data Systems, which uses the New York facility to clear transactions for the European financial system, were left without service for nearly two days (Miller, 2001).

This example of network telecommunications failure and its subsequent restoration (in stages) in the New York City area highlights the importance of network survivability for the Internet. Not only were city and county governments tempo-
rarily left without access to important information, global financial services that depend on the rapid transfer of information were also halted. Although most events stimulating a network failure are not as dramatic as those surrounding the World Trade Center collapse, the ability for networks to re-route, re-connect and to have redundancy is clearly an important asset for telecommunications services. To date, issues of telecommunications network survivability and reliability are largely ignored in the literature on city accessibility to the commercial Internet. In fact, the existing work on city accessibility assumes 100% reliability and survivability and focuses on relative network location, infrastructure equity and bandwidth availability for analysis (Wheeler and O’Kelly, 1999; Gorman and Malecki, 2000; Moss and Townsend, 2000). This is a potentially problematic considering the social and economic implications of network failure.

Albert et al. (2000) do provide a brief analysis and discussion on attack tolerance and failure of complex networks. Their results indicate that scale-free networks such as the Internet are very tolerant to random failures at nodes—providing a reliable system for information distribution. However, their results also suggest that targeted attacks at the most connected nodes have the potential to fragment scale-free networks, revealing a significant lack of survivability. Because of the mutual relationship between survivability and reliability, it is important to make an initial clarification of their differences (more detail will be provided in Section 3). Network reliability measures are concerned with the dependability of network equipment and the ability of a network to provide communication given component failure (Medhi, 1999). Network survivability measures are concerned with the smallest amount of damage that can disconnect a network or reduce its performance to unacceptable levels or total failure (Colbourn, 1999). Basic network failure can be categorized in the following ways: (1) arc failure (the severing of an optical cable); and (2) node failure (the failure of electronic equipment) (Lisser and Sarkissian, 1995). In this context, questions of network survivability are inherently geographic because component failures at an origin node (i.e. source node) have the potential to impact the entire network both downstream and upstream from the failure. This is particularly problematic for destination nodes (i.e. terminal node) on a network (Colbourn, 1999). Given the significance of reliable network performance and the implications of network failure, additional questions of city accessibility to the commercial Internet must be addressed. The purpose of this paper is to explore various dimensions of network reliability and survivability, stressing the latter, for the US commercial Internet from a geographic perspective. More specifically, questions of city accessibility as they relate to network topology are addressed using GIS based approaches that simulate node failure in the US commercial backbone system. Not only will these simulations provide a more complete picture of city accessibility to the commercial Internet, they will also highlight the advantages and disadvantages of fully meshed telecommunications networks versus the more cost effective hub-and-spoke topologies.

The remainder of this paper is organized as follows. Section 2 briefly explores the historical context of network survivability and reliability as it relates to the commercial Internet. Section 3 compares and contrasts the differing perspectives of telecommunication engineering and geography for network reliability and survivability.
Section 4 outlines a methodology for evaluating the geographic dimensions of network survivability as they relate to city accessibility. Section 5 presents the results of the application study. Finally, Section 6 provides a discussion and brief conclusion.

2. Evolution of a survivable telecommunications network

The development of the US commercial Internet is well documented (Abbate, 2000). Initially conceived as a communication system able to survive a nuclear attack, the Internet was first developed through a series of projects in the 1960s conducted by the US Department of Defense and the Advanced Research Project Agency. Numerous important technical innovations were realized from these early efforts, including packet switching and TCP/IP protocol. For a more thorough discussion on these topics see Abbate (2000).

Although packet switching and TCP/IP would serve as the technological foundations of the Internet, they did not necessarily guarantee survivable communications in the event of an attack at a node or a link. In conventional circuit switched communication systems, such as the public switched telephone network, switching is concentrated and hierarchical; providing a private, hardwired connection through a network between two subscribers (Sharma, 1990). For example, each call is routed through a local office, and then to a regional or national switching office if a long distance connection is required. Because each local office serves thousands of users, destroying a single office would leave many thousands without service. Therefore, the combination of packet switching, TCP/IP and a distributed network topology were recommended for constructing a survivable telecommunications system (Abbate, 2000).

Network topology defines the manner in which network nodes (cities) are interconnected. There are many different topological frameworks that are utilized in telecommunication systems. Fig. 1 illustrates several of the most basic system topologies, including centralized, decentralized, distributed, and partially connected. As Fig. 1a demonstrates, a partially connected mesh topology allows for a direct link between certain pairs of nodes in a network. From an economic perspective, the partially connected mesh is advisable when traffic flow between certain nodes is low. The interaction between low-flow nodes can be switched to longer paths resulting in a better economy of scale (Sharma, 1990). The centralized topology (Fig. 1b) utilizes a single switching node that directly connects subscribers. This type of topology is frequently used for university campus networks in order to connect all subscribers to a PABX voice or data switch. \(^1\) The hub-and-spoke (i.e. decentralized) network topology (Fig. 1c) is designed for serving information flows between multiple origins and destinations. As O’Kelly and Miller (1994) note, hubs allow for the construction of a network where direct connections between all origins and destinations (fully connected mesh) is replaced with fewer indirect connections. This reduces construction costs and allows for scale economies through the consolidation of flows.

\(^1\) PABX stands for “Private Automatic Branch Exchange”. Today, they are usually referred to as PBXs and are automatic, meaning that an operator is no longer needed to place a call (Newton, 2000).
(O’Kelly and Miller, 1994). Considering that the cost of transmission in most telecommunication networks is predominately influenced by link costs versus switching, the hub-and-spoke system for telecommunication can make good sense if capital is restricted. Finally, the distributed network (Fig. 1d) is the configuration that most closely resembles the design recommended by Baran (1964b) for the Internet.

2.1. The US commercial Internet

Recent studies exploring the spatial characteristics of the US commercial Internet reveal complex interplays between market demand, network economics, and network topology. For example, Gorman and Malecki (2000) and Moss and Townsend (2000) suggest that patterns of Internet backbone development have created an accessibility hierarchy, with the most accessible cities located at major network access points (NAPs). O’Kelly and Grubesic (in press) note that while network complexities are increasing, individual network connectivity and redundancy appear to be decreasing. This is significant for a number of reasons. First, and most obvious, a trend toward sparser networks suggests that providers are building network configurations that are not fully meshed or redundant. Although this has the potential to be problematic in a telecommunication system served by a single provider, peering agreements between commercial Internet service providers (ISPs) partially ensure that some form of redundancy between major nodes exists. However, locations...
where redundancy is minimal and peering with other network providers is of negligible value, can pose a serious threat to robust network service. Second, given the competitive nature of the telecommunication industry, O’Kelly and Grubesic (in press) suggest that providers have moved to hub-and-spoke topologies. Consolidating flows using a hub-and-spoke system allows ISP networks to increase the geographic footprint of their service areas through the provision of lower capacity spokes in smaller markets. The information flows from these smaller markets are then consolidated at a larger hub and routed along a major trunkline to other major hubs for distribution. Clearly, hub-and-spoke topologies combined with peering agreements allows ISPs to operate at substantially lower costs than scenarios where direct, high-capacity connections are required between all nodes. This is directly analogous to the current air-passenger system (O’Kelly and Wheeler, 2002).

The movement to hub-and-spoke topologies has clear and serious implications on network survivability, however, as illustrated in the introduction, the failure of a major node such as that in lower Manhattan has the potential to seriously disrupt Internet traffic flows. Consider, for example, the simultaneous failure of NAPs in Chicago, New York, Washington, DC, and San Francisco. Although the example is somewhat extreme, there is no doubt that Internet traffic and performance would be impacted. Recently, network reliability and survivability were tested when a train carrying hazardous materials derailed in a tunnel outside of Baltimore, Maryland (Associated Press, 2001). Because fiber-optic backbones from major providers such as WorldCom, PSINet, and AboveNet were also located in the tunnel, significant Internet slowdowns, beginning in the Middle Atlantic States, rippled across the country as companies diverted web traffic to other paths (Associated Press, 2001). Although larger cities have displayed an ability to recover from the catastrophic failure of telecommunication components because of infrastructure redundancy, questions of survivability remain for smaller second and third tier nodes (cities) where network redundancy and provider alternatives are limited. This is especially important in smaller spoke cities, where the only connection to the Internet might be from a single network provider.

The following section compares and contrasts the concepts of survivability and reliability in telecommunications network engineering and geography. This discussion will serve as a foundation for the examination of network survivability from a geographic perspective in Section 4, where simulations emulating node failures in a spatial context are explored.

3. Perspectives on network survivability and reliability

3.1. Telecommunications engineering

As previously outlined, communication networks are comprised of nodes and links. A wide variety of software and hardware components are required to enable communication through these networks. Network reliability refers to the ability of the overall network to provide communication in the event of component failure.
(Medhi, 1999). Fault-tolerant refers to how reliable a particular component (element) of the network is. For example, switches and routers are elements of a telecommunication network. Alternatively, fault-tolerant network refers to how resilient the entire network is against the failure of a component. Overall network reliability is also contingent on a variety of external factors, which may or may not cause a portion of the network to fail. For example, environmental factors such as heat stress, construction, or attack may create significant disruptions in operation.

Much of telecommunications engineering is concerned with the probability of component failure and its ramifications on network performance. Consider the following example forwarded by Medhi (1999) for a basic telephone network. Fig. 2 illustrates two telephones connected by distribution segments (“A”) to local switches (“S”), with switches connected by the facility (“B”). Each element of this simple network is assigned an outage/downtime percentage: S, 0.01%, A, 0.01%, B, 0.03%. The availability of this connection is \( \left( \frac{1}{0.001} \right)^4 \left( 1 - 0.0003 \right) = 99.93\% \). This translates to a maximum of 368 min of downtime per year for the system.

Contrary to the circuit switched telephone network illustrated in Fig. 2, which provides a private, hardwired connection through a network between two subscribers, the Internet uses packet switching technology, permitting the transfer of information over multiple pathways simultaneously. In practical terms, therefore, a distributed network topology (Fig. 1d) with many redundant pathways offers significant advantages over decentralized and centralized networks having aggregating nodes that are vulnerable to attack (Baran, 1964b).

With respect to the previous example, it is clear that network engineers must be concerned with both the causes and consequences of component failure. Clearly, a catastrophic failure causing network disconnection is the greatest concern, but the more frequent consequence of component failure is degraded performance (Colborn, 1999). Therefore, from the perspective of network engineers, survivability issues can be categorized in the following manner (Medhi, 1999). First, what is the smallest amount of damage that can disconnect the network or reduce its performance to an unacceptable level? Second, environmental failures caused by extrinsic factors, and failures associated with network overload, lead to probabilistic measures involving statistical dependence of failure. Finally, random failures such as “wearout” lead to probabilistic measures with statistical independence.

3.2. Geography

Geographic perspectives on network survivability and performance are quite different than those outlined above. Although component failure must be considered indirectly, the spatial ramifications of element failure and its impact on accessibility or service are of greater concern to geographers. For example, what are the impacts

\[ \text{It is important to note that recent technological advancements in synchronous transmission products such as SONET/SDH, include the development of self-healing rings. Self-healing rings allow for service to be restored very rapidly in the event of a single link or node failure (Armony et al., 2000). For a more detailed discussion see Wasem et al. (1994) or Felekis and Milis (1996).} \]
on city accessibility if the MAE-East node completely fails? Does the entire East Coast of the United States lose Internet service? Philadelphia and Baltimore only? It is probable that Internet service to the major cities will suffer in terms of performance (latency and network delays), but service will not completely shut down. Because the Internet is a packet switched network, traffic can be rerouted through the NAPs in New York or Washington, DC, or private peering points located on the East Coast. So, the larger networks such as Qwest, Sprint, and WorldCom have the ability to dynamically reroute packets within minutes of component failure. However, smaller networks with limited transit or peering partners may not be able to restore service as quickly. Cities serviced by smaller network providers or a limited number of networks may be left without service until MAE-East is repaired. From a geographic perspective, it is clear that these scenarios are closely linked to the network topology of the commercial Internet.

Consider the Multacom backbone illustrated in Fig. 3a and the IDT backbone shown in Fig. 3b. Both networks provide service to a number of major cities in the

Fig. 3. Internet backbone networks: (a) Multacom network; (b) IDT network; (c) combined network.
United States. First, let us consider the failure of a single node in the network, with a major Multacom component failing in Chicago, effectively disrupting service (Fig. 3a). Because Multacom fails to provide redundant connections between cities, all information originating in Los Angeles, Dallas, Miami, Tampa, Atlanta, Washington, and New York destined to Seattle, Portland, San Jose, and Denver cannot be delivered (and vice versa). Now, consider a single link failure on the IDT network, with the fiber-optic cable between Phoenix and Denver suffering a cut (Fig. 3b). Because IDT maintains multiple connections between Denver and other major cities, traffic destined for Phoenix can be rerouted from Denver to Los Angeles or Houston for delivery to its final destination.

Given the size of the commercial Internet in the United States, it is clear that no network operates in complete isolation. As illustrated in Fig. 3c, simply combining the IDT and Multacom network substantially increases service coverage and redundancy—especially if the two networks have a peering agreement. However, several nodes, including Portland and Seattle, remain vulnerable to node or link failure in this combined network. In reality, there are scores of networks that provide service to larger cities such as Portland and Seattle in the US, however, cities further down the hierarchy of Internet accessibility (Townsend, 2001; O’Kelly and Grubesic, in press; Grubesic and O’Kelly, 2002) do not benefit from high connectivity and thus remain vulnerable to service disruptions. This is particularly problematic in smaller “spoke” cities.

The next section outlines a methodological approach for simulating node failure for the US commercial Internet. The spatial implications of such failures are explored, with attention paid to the topological characteristics of individual backbone providers and the “downstream” impacts of node and link failure on city accessibility.

4. Methodology for estimating network survivability

A viable option for evaluating the survivability of an Internet backbone network is to run a series of scenarios examining failure of nodes in a telecommunication system. Table 1 displays a sample connection matrix for six network backbones and ten cities utilized in O’Kelly and Grubesic (in press). Consider a set of nodes (cities), each with a unique identification number from \( i = 1, \ldots, N \), and a group of backbones with unique identification numbers, \( k = 1, \ldots, B \) which serves a subset of \( N \) nodes. Let \( A_{ik} = 1 \) when node \( i \) is in network \( k \). Following the notation outlined in O’Kelly and Grubesic (in press):

- \( N_k \) indicates the set of nodes in the \( k \)th network.
- \( N_k \) is the number of nodes in the \( k \)th network.
- \( B_i \) indicates the set of networks in which node \( i \) appears.
- \( B_i \) is the number of networks in which node \( i \) appears.
- \( N_k = \{ i | A_{ik} = 1 \} \) is the index set of nodes in network \( k \).
- \( B_i = \{ k | A_{ik} = 1 \} \) is the index set of the backbone networks in which node \( i \) appears.
Relating this to Table 1, backbone 5 has $N_5 = \{1, 3, 6, 9, 10\}$ and thus $N_5 = 5$. Each node may appear in one or many networks, and each network may have between two and ten nodes. For example, node 4 appears in the smallest number of networks, $B_4 = 1$, and is listed in $B_4 = \{1\}$. The number of nodes in network $k$ is $\sum_i A_{ik} = N_k$. Similarly, the number of networks that node $i$ occurs in is $\sum_k A_{ik} = B_i$.

Now that the various sets of nodes and backbones are defined, the connection matrix for the $k$th network backbone is defined as $X_{ij}^k$, indicating the number of direct links between $i$ and $j$. Similar to Table 1, the connection matrix for each backbone consists of zeroes and ones. A zero in the cell means that there is no linkage between that particular origin node, $i$, and destination node, $j$, for network backbone $k$. The row sums represent the total number of other nodes in the network that are connected to $i$ by direct (one-step) connections. Therefore, $\sum_i X_{ij}^k$ represents the degree of a node. This measure will be utilized for simulating one aspect of network survivability. In this context, the degree of node describes two conditions. A degree of 1 indicates that a city is a spoke city, with only one direct connection to a different node. As the degree increases from 1, 2, ..., $n$, the more closely a city resembles a hub. For example, preliminary calculations on a backbone network used for analysis in this paper indicates the degree of node measure for the city of Chicago on the AT&T network in 2000 is 23. Meaning that a total of 23 direct connections between Chicago and other cities exists. In contrast, the degree of node measure on the AT&T backbone for Salt Lake City was 3. In this example, Chicago is clearly more qualified as a hub than Salt Lake City. As such, loss of functionality in Chicago, for whatever reason, is likely to have more significant ramifications than the loss of Salt Lake City because 23 cities are directly impacted. This is not to suggest that service becomes unavailable in these 23 cities, but there is a high probability that performance would be degraded.

<table>
<thead>
<tr>
<th>Node</th>
<th>Network</th>
<th>Row sum</th>
<th>#</th>
<th>Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td></td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Column sum</th>
<th>#</th>
<th>5</th>
<th>6</th>
<th>5</th>
<th>5</th>
<th>5</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set</td>
<td>N1</td>
<td>N2</td>
<td>N3</td>
<td>N4</td>
<td>N5</td>
<td>N6</td>
<td></td>
</tr>
<tr>
<td>1, 4, 7,</td>
<td>2, 3, 5,</td>
<td>1, 2, 7,</td>
<td>1, 2, 5,</td>
<td>1, 3, 6,</td>
<td>3, 5, 7, 9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9, 10</td>
<td>7, 8, 9</td>
<td>9, 10</td>
<td>6, 8</td>
<td>9, 10</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Given the accounting system outlined above, several scenarios can be explored through simulation. First, if an entire node is knocked out (e.g. complete system failure), an entire row of Table 1 disappears. This means that service to the city in question is impossible for any backbone. Clearly, the probability of a complete loss is low, particularly in large cities where providers maintain multiple points of presence (Grubesic and O’Kelly, 2002). However, the recent events in New York City showed the impact of an important node sustaining damage. A second scenario might include the loss of a backbone provider. In this case, a column would disappear. This means that peering between the disabled backbone and any other provider is not possible, but more importantly, it means that cities serviced by a single backbone (such as node 4) are completely without Internet service. Again, this is particularly problematic for spoke cities. Service is possible, however, if a second or third backbone operates connections to a node (such as node 10). Thus, backbone variety and the presence of multiple providers is a key element for network survivability. A third possibility is the failure of a single network node. In this case a cell disappears. For example, the router at node 1 for backbone 1 fails. Although this eliminates service to node 1 from a single provider, other backbones remain operational and traffic continues to flow. Similar to the loss of a backbone provider, the failure of a single network node is most problematic for spoke cities because it effectively eliminates the delivery of all traffic destined for the node in question. In contrast, cities served by multiple providers or larger hub cities can reroute and take advantage of peering agreements in the delivery of data traffic, if there is some type of localized component failure. Finally, if select links in a network are severed, isolated nodes suffer loss of service, but nodes with multiple links remain functional. This is another example where the importance of the degree of node measure is evident. The availability of multiple connections, whether redundant (from a single ISP) or unique, the more survivable a node is.

4.1. Simulating network survivability

This section, and the remainder of this paper, will explore the spatial ramifications of node failure for several Internet backbone networks. We will also attempt to highlight pertinent trends in our analysis by concentrating on the impacts of complete node failure on city accessibility. Network backbone information from the 12th Edition of the Boardwatch Directory of Internet Services Providers will be utilized in this analysis (Boardwatch, 2000). These data are compiled annually by Boardwatch Magazine, one of the leading sources of information on Internet backbone providers in the United States. The 12th Edition contains information on 41 commercial backbone providers. Additional confirmatory data were acquired from the Cooperative Association for Internet Data Analysis (CAIDA) (based on the Boardwatch data) for comparative purposes.

In order to simulate node failure, a select group of cities, common to a majority of the 41 networks, were selected for analysis (Table 2). The subset of cities selected represents a mix of coastal and interior locations. Malecki and Gorman (2001) note that this is an important distinction because of the general trend for coastal cities to
be relatively closer together in terms of network links than interior cities. In other words, this pattern represents a ‘coastal wrapping’ where cities such as Los Angeles, New York, San Francisco, and Washington, DC, are topologically nearer to each other than to most interior cities (Malecki and Gorman, 2001). The major exceptions to this general trend are the interior cities of Chicago, Atlanta, and Dallas, which all contain major public or private NAPs.

A scheme for simulating network survivability with a node failure is detailed as follows:

(a) Select a network for analysis.
(b) Select a node for failure (this node may or may not appear in all networks).
(c) Close all links into or out of that node.
(d) Calculate the new degree of node values for each node in the network.
(e) Calculate the availability of network paths between nodes (for all nodes) in the network.

This scheme provides several interesting measures pertaining to the impacts of node closure for individual networks as well as spatial impacts of node closure on city accessibility. First, it allows one to evaluate the centrality of a city (node) to an individual network. In essence, the degree to which a node is central to the network is a measure of its importance. Hub cities provide the majority of switching functions for networks on the Internet (Grubesic and O'Kelly, 2002). Therefore, in terms of functional hierarchies, hub cities are more important and central to a network than spoke cities. As a result, the loss of a hub city has significant impacts on city accessibility. Second, this simulation scheme will also provide basic measures on the overall vulnerability of a network. If networks utilize hub-and-spoke topologies, the loss of certain nodes is more dangerous than the loss of others to the overall functionality and performability of network. Thus, networks with several hub cities are

<table>
<thead>
<tr>
<th>Cities (nodes) selected for stimulated failure</th>
</tr>
</thead>
<tbody>
<tr>
<td>1    Atlanta</td>
</tr>
<tr>
<td>2    Boston</td>
</tr>
<tr>
<td>3    Chicago</td>
</tr>
<tr>
<td>4    Dallas</td>
</tr>
<tr>
<td>5    Kansas City</td>
</tr>
<tr>
<td>6    Los Angeles</td>
</tr>
<tr>
<td>7    Minneapolis</td>
</tr>
<tr>
<td>8    New York</td>
</tr>
<tr>
<td>9    Palo Alto</td>
</tr>
<tr>
<td>10   Phoenix</td>
</tr>
<tr>
<td>11   Salt Lake City</td>
</tr>
<tr>
<td>12   St. Louis</td>
</tr>
<tr>
<td>13   Tampa</td>
</tr>
<tr>
<td>14   Washington</td>
</tr>
</tbody>
</table>
more likely to become disconnected than networks that utilize meshed or redundant topologies.

5. Application results

Node failure simulations were computed for the 14 nodes listed in Table 2 on each of the 41 networks. To more clearly illustrate this process, the Multacom Internet backbone illustrated in Fig. 4a will be used for discussion. Fig. 4b displays the derived network survivability matrix for the Multacom backbone after node failure is simulated for all 14 nodes selected for analysis. Each row represents an actual node on the Multacom backbone \( (n = 12) \). Row totals indicate the relative accessibility of a city in the Multacom backbone after failure.\(^3\) In this case, lower totals indicate a more accessible network location because they maintain network access to more cities after a node failure. For example, the city of Washington is the most accessible node on the Multacom network, with a row total of 28. If Tampa fails, Washington only loses access to Tampa and Miami. However, if New York fails, Washington loses access to six cities (New York, Chicago, Denver, San Jose, Portland, and Seattle). Clearly, the New York failure is a worst-case scenario for Washington. However, the loss of peripheral nodes such as Miami, Seattle, or Los Angeles will have very little effect because of their location on the network, relative to Washington. Column totals indicate the relative centrality or importance of a city on the Multacom network. In this case, higher totals suggest a higher centrality. For example, Atlanta is the most important node on the Multacom backbone, with a centrality score of 75. This suggests that the loss of Atlanta and its hubbing functions limits the overall nodal accessibility for other cities on the network at a greater magnitude than the loss of any other node. For example, if Atlanta fails, the cities of Dallas, Los Angeles, Miami and Tampa cannot reach ten of the Multacom network nodes (in different combinations). Similarly, the remaining Multacom network cities (Chicago, Denver, New York, Portland, San Jose, Seattle, and Washington) cannot reach five other cities. Again, this illustrates the importance of hub cities on many networks and the major impact of hub failure on network survivability. Individual city pairs and the impact of node failure can also be evaluated. For example, node failure in Dallas is the worst-case scenario for Los Angeles, eliminating access to 91\% (11 of 12) of the cities. Similarly, the loss of Tampa eliminates access to 11 of 12 cities on the network from Miami.

In addition to individual network analysis, results illustrating the relative importance of each city across all 41 networks are informative. Fig. 5 illustrates the aggregate impact of node failure on network accessibility for each of the 14 nodes in the analysis. In other words, Fig. 5 highlights node failures that disrupt the largest

---

\(^3\) Nodes not appearing on the Multacom backbone are assigned a value of 1. Column totals for nodes not appearing on the network will equal the number of nodes on the network being analyzed, which in the case of Multacom gives \( n = 12 \).
number of networks. As the Y-axis indicates, for the 14 cities analyzed, the most lethal node failure is Los Angeles. This relative measure of lethality indicates that more networks become disconnected with the failure of Los Angeles \((n = 34)\) than any other city in the analysis. The role of Los Angeles as an important interconnection point for the US commercial Internet (Grubesic and O’Kelly, 2002; O’Kelly and Grubesic, in press) is a primary reason such disconnects might occur. For many networks, Los Angeles represents an important hub location for connecting other major cities in the southern tier of the US. Therefore, the loss of this node not only impacts other larger cities in California, but also Las Vegas, Phoenix, Tucson, Denver, Dallas, and Houston. In fact, each of the top 6 cities illustrated in Fig. 5 are important hub cities for the commercial Internet—each containing a major NAP or metropolitan area exchange (Grubesic and O’Kelly, 2002). Again, this suggests the relative importance of these hub cities for many of the 41 networks analyzed. The loss of these nodes negatively impacts both network accessibility and survivability.
A final measure of network survivability is illustrated in Fig. 6. Given the 14 nodes utilized in the analysis of 41 networks, Fig. 6 displays the overall vulnerability of each network. The most susceptible network to disconnection is AT&T. Eight of the 14 nodes (when failed) effectively disrupt and disconnect the AT&T network. Not surprisingly, many of the cities utilized for analysis are hub cities for AT&T (e.g. Chicago, Los Angeles, Dallas). Not only does the loss of these nodes have the potential to impact the performance of the network, it also has the potential to leave many smaller spoke cities without service. Therefore, the use of a hub-and-spoke system does have real implications for network survivability in the case of AT&T. Similarly, the GTE network also makes extensive use of hub-and-spoke topological configurations. In this case, failure in 7 of the 14 cities utilized for analysis result in the disconnection of the GTE network.

Contrasting the extreme examples of AT&T and GTE are 11 other networks where the loss of a node does not result in a disconnection. This suggests a more robust, meshed network topology, with many alternative pathways for data transmission. Although meshed topologies are more expensive to construct, they clearly have advantages over hub-and-spoke configurations where survivability is concerned. In addition, meshed topologies are clearly more similar to the original topological structure recommended by Baran (1964a,b) than hub-and-spoke topologies.
6. Discussion and conclusions

A valuable function of the literature on city accessibility to the commercial Internet is its documentation of infrastructure diffusion availability throughout the United States (Wheeler and O’Kelly, 1999; Gorman and Malecki, 2000; Malecki and Gorman, 2001; Grubesic and O’Kelly, 2002). Moss and Townsend (2000) note that the information infrastructure of today plays a very similar role to the transportation networks of the past two centuries (rail, road, air, water); where the Internet transports the “valuable goods” of the digital economy—information, knowledge, and communications. Although the availability of such infrastructure is clearly an important consideration, the ability for such infrastructure to perform without interruption is also a significant concern. As discussed previously, infrastructure failure has the potential to be economically catastrophic for businesses, cities, and regions.

The analysis of network survivability from a geographic perspective illustrates a number of important points. The competitive nature of the Internet backbone provider industry has created a situation where many backbones are prone to disconnection if there is a major failure in a hub city for a given network. This is partially attributable to the economic pressures of network construction. In extreme cases, the cost of laying fiber-optic cable can reach nearly $1 million per mile—including the costs associated with digging trenches and obtaining rights-of-way.

Fig. 6. Network vulnerability.
(Kharif, 2001). Therefore, providers are in search of the most economically efficient network topology possible, one that does not sacrifice performance. This has motivated many providers to move to a hub-and-spoke topology (O’Kelly and Grubesic, in press). Hub-and-spoke topologies allow for economies of scale by consolidating flows from a set of geographically dispersed service points onto a fiber-optic trunk extending to a hub city where switching functions are performed. O’Kelly and Wheeler (2002) note that this is very similar to the existing commercial air-passenger system in the United States. Unfortunately, the movement to hub-and-spoke topologies for Internet backbones suggests a significant departure from the original topological structure outlined for the Internet. The consolidation of flows to hub cities creates a topological hierarchy where hub cities maintain a central role in all Internet traffic. Therefore, the loss of a hub city through a massive node failure has the potential to create significant disruptions in service quality and city accessibility. Further, as illustrated in Section 4, node failure at a hub has the potential to completely disconnect a network. This can leave many cities completely without Internet access (e.g. smaller spoke cities). At the very least, the loss of a major hub limits data traffic between origins and destinations on the network. 4 It was also shown that certain networks are more prone to disconnection than others, with GTE and AT&T being the two extreme cases. Both utilize the hub-and-spoke topology widely. However, these results are tempered by the fact that other networks may utilize a different subset of hub cities. Future work will incorporate a more extensive analysis and a wider selection of closures in order to remedy this limitation.

There are other directions in which future work should also will be directed. Although this paper employed a deterministic framework for modeling a series of node failures, there is the potential for application extensions that include probabilistic approaches. For example, Daskin (1983) and ReVelle and Hogan (1989) utilize probabilistic versions of the set covering model and the maximum covering location model to ensure adequate coverage within a given time standard. This approach helps determine the likelihood that the nodes in a covering problem would be covered with a given confidence level (e.g. 95%). Our analysis can be adapted to this framework and we are currently devising such a model. The key difference is that instead of looking deterministically at the effect (one-by-one) of a series of proposed node failures, we tabulate “what-if” parameters under a wide variety of failures, each of which occurs with probability. In this regard, it is very similar to the approach employed by telecommunication engineers (Colbourn, 1999). Thus, instead of all 14 nodes failing on an all or nothing basis, we might expect that the mesh of cities have a certain probability of failure. More importantly, future work will also account for link failure in the network. Finally, it is important to note that our data are a snapshot of Internet infrastructure. We have evidence that networks such as AT&T are constantly being re-engineered to configure the system for better survivability. Thus, a probabilistic analysis could very well prove even more useful than our

---

4 Recall that peering and transit agreements also play an important role in network survivability.
deterministic cases in spotting vulnerability, and consequently building sufficient redundancy to ensure survival.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Award no. BCS-9876455.

References