PHYSICAL REVIEW E **71**, 015103(R) (2005)

Vulnerability and protection of infrastructure networks

Vito Latora[1] and Massimo Marchiori[2]
[1]*Dipartimento di Fisica e Astronomia, Università di Catania and INFN, Via S. Sofia 64, 95123 Catania, Italy*
[2]*W3C and Laboratory for Computer Science, Massachusetts Institute of Technology, 545 Technology Square, Cambridge, Massachusetts 02139, USA*
(Received 19 July 2004; published 20 January 2005)

Infrastructure systems are a key ingredient of modern society. We discuss a general method to find the critical components of an infrastructure network, i.e., the nodes and the links fundamental to the perfect functioning of the network. Such nodes, and not the most connected ones, are the targets to protect from terrorist attacks. The method, used as an improvement analysis, can also help to better shape a planned expansion of the network.

DOI: 10.1103/PhysRevE.71.015103 PACS number(s): 89.75.Hc, 89.20.Hh, 89.40.−a, 89.75.Fb

The resilience of complex networks to the malfunctioning of its components and to external disturbances—simulated as the deletion of nodes or links—has been the subject of a great deal of attention in the recent literature [1]. The network structure and functions strongly rely on the existence of paths between pairs of nodes. When nodes or links are removed, the typical length of such paths will increase and eventually some couples of nodes will become disconnected. There are various ways in which nodes and links can be removed, and different networks exhibit a different level of resilience to such disturbances [2]. For instance, one can simulate errors as the deletion of nodes/links chosen at random, or intentional attacks as the targeted removal of a specific class of nodes/links. Attacks have been studied by sorting and removing progressively the nodes in descending order of degree [2–5] or betweenness [3,5,6], or the links in descending order of betweenness [3,7] or range [8]. The network robustness is usually measured by the size of the largest connected component and by the average node-node distance as a function of the percentage of nodes/links removed. In these works the main attention has been on the number of removals needed to observe the disappearance of a macroscopic connected component [1], while we are often interested in finding what are the *critical components* of the network, i.e., the nodes/edges really crucial for the functioning of the network. In this Communication we propose a method to evaluate the importance of an element of the network by considering the drop in the network's performance caused by its deactivation. In practice we check for the redundancy of an element by calculating the performance of the perturbed network and comparing it with the original one. Notice that the element can be either a single node or edge, or a group of nodes/edges in the case we want to simulate multiple attacks. We focus in particular on infrastructure networks, defining the vulnerability under various classes of attacks and producing a list of the points of the network that should be the first concern of any policy of protection from terrorist attacks. Analogously, we measure the importance of an improvement by the increase in the network's performance caused by such improvement.

The paper is organized as follows: We first present the general framework to define critical damages, critical improvements, structural vulnerability, and improvability of a network. We then show how the method works in practice on some examples of communication and transportation infrastructures.

We assume that a generic infrastructure S is characterized by a single variable $\Phi[S]>0$, the *performance* of S [9]. The definition and quantitative analysis of the *critical components* of S, we propose in this paper, uses, as reference observable, variations in the performance. We consider separately the study of damages and of improvements.

Attacks analysis. Let us indicate by D a set of possible damages on the infrastructure S, and with $\mathcal{D}(S,d)$ a map that gives the infrastructure resulting from S after the damage $d\in D$. We measure the importance of the damage d by the relative drop in performance $\Delta\Phi^{-}/\Phi$, with $\Delta\Phi^{-}=\Phi[S]-\Phi[\mathcal{D}(S,d)]\geqslant 0$, caused by d. In particular, the *critical damage* $d^{*}\in D$ is the damage of D that minimizes $\Phi[\mathcal{D}(S,d)]$. The *vulnerability* V of S under the class of damages D can be defined as

$$V[S,D]=\frac{\Phi[S]-W[S,D]}{\Phi[S]}, \tag{1}$$

where $W[S,D]=\Phi[\mathcal{D}(S,d^{*})]$ is the worst performance of S under the class of damages D. The vulnerability $V[S,D]$ is defined in the range $[0,1]$.

Improvements analysis. We now turn our attention on how to improve an existing infrastructure [10]. Various improvements can be added to S, so given a set of improvements I we define, for any improvement $i\in I$, the map $IM(S,i)$ that gives the resulting infrastructure obtained after the improvement i. We measure the importance of i as the relative increase in the performance $\Delta\Phi^{+}/\Phi$, with $\Delta\Phi^{+}=\Phi[IM(S,i)]-\Phi[S]$, caused by i. In particular we define the *critical improvement* i^{*} as the best possible improvement in I, i.e., the improvement of I that maximizes $\Phi[IM(S,i)]$. Then, the *improvability* IM of S under the class of improvements I can be defined as

$$IM[S,I]=\frac{B[S,I]-\Phi[S]}{\Phi[S]}, \tag{2}$$

where $B[S,I]=\Phi[IM(S,i^{*})]$ is the best performance of S under the class of improvements I.

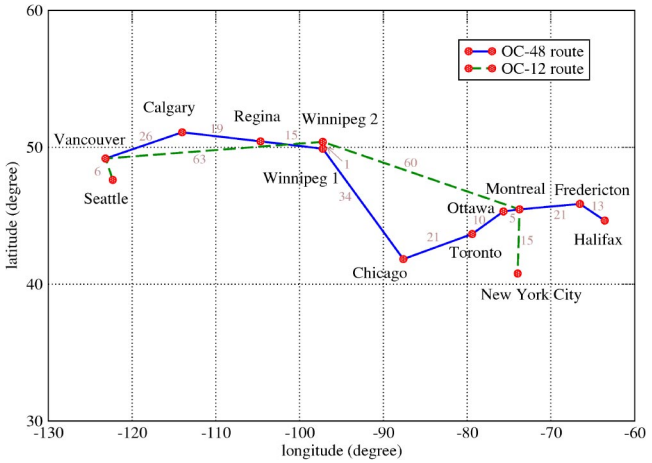

FIG. 1. Ca*net3 IS-IS routing network. The numbers reported are a measure of the latency associated to each link [16].

As a practical application of the method we consider communication-information (as the Internet [11]) and transportation infrastructure networks. We represent the infrastructure network S as a valued undirected [12] graph with N nodes (for instance the routers in the Internet, or the stations in a railway transportation system) and K links (the cables connecting two routers, or the lines connecting couples of stations). S is described by the $N \times N$ adjacency matrix $\{l_{ij}\}$. If there is a link between node i and node j, the entry l_{ij} is a positive number measuring the *link latency*, otherwise $lij = +\infty$. For instance, in the Internet (in the railway system) the larger l_{ij} is, the longer it takes for a unitary packet of information (a train) to go along the link from i to j. We have now different ways to measure the performance of S. In this paper we identify the performance of S with the *efficiency* of the network, i.e., we assume: $\Phi[S]=E[S]\equiv 1/N(N-1)\Sigma_{i\neq j\in S}(1/d_{ij})$, where d_{ij} is the smallest sum of the links latency throughout all the possible paths in the graph from a node i to a node j (in the particular case of unvalued graphs d_{ij} reduces to the minimum number of links traversed to get from i to j). The *efficiency* is a quantity recently introduced in Refs. [13] to measure how efficiently the nodes of the network communicate if they exchange information in parallel. A second possibility is to assume the performance $\Phi[S]$ to be equal to the inverse of the characteristic path length $L \equiv 1/N(N-1)\Sigma_{i\neq j\in S} d_{ij}$ [13,14]. An alternative possibility to avoid the shortest path assumption on which both E and L rely, is to identify $\Phi[S]$ with the mean flow rate of information over S [15].

*Ca*net3*. We show how the method works in practice by considering the Ca*net3 IS-IS routing network [16] represented in Fig. 1, a simple example of an Internet backbone, consisting of two main routes, OC-12 and OC-48, $N=13$ routers, and $K=14$ links. As the backbone has diverse routes of different bandwidths, the preferred path between any two routers is the path which presents the least amount of latency under normal router load conditions. We consider three different classes (sets) of damages D: the damage of a single cable connection, of a single Internet router, and of a couple of routers. $\mathcal{D}(S,d)$ is the network we obtain from S after the deactivation of the damaged component (respectively, the damaged link, node, or couple of nodes). The damage of single links allows to investigate the finer effects on the network, since the damage of a node implies the damage of a number of links equal to the node's degree. The entity of the damage d is given by the relative drop in the efficiency $\Delta\Phi^-/\Phi[S]$ caused by d.

TABLE I. Attacks and improvement analysis of Ca*net3. For each class of damage/improvement considered (see text) we report the cases having the highest effects on the performance of the network. The rank and name of the damaged link (node, or couple of nodes, respectively) and of the added link are listed in the first two columns. The relative drop or increase of the efficiency is in the third column.

	Damaged link	$\Delta\Phi^-/\Phi$
1	Winnipeg2 - Winnipeg1	0.358
2	Ottawa - Montreal	0.146
3	Montreal - Fredericton	0.123
4	Seattle - Vancouver	0.098
	Damaged node	$\Delta\Phi^-/\Phi$
1	Winnipeg1	0.466
2	Winnipeg2	0.408
3	Montreal	0.317
4	Ottawa	0.220
	Damaged couple of nodes	$\Delta\Phi^-/\Phi$
1	Winnipeg1 + Montreal	0.792
2	Winnipeg1 + Ottawa	0.723
3	Winnipeg2 + Montreal	0.702
4	Winnipeg2 + Ottawa	0.700
5	Winnipeg2 + Toronto	0.633
	Added link	$\Delta\Phi^+/\Phi$
1	Toronto - NYC	0.01237
2	Ottawa - NYC	0.00770
3	Winnipeg1 - Toronto	0.00587
4	Fredericton - NYC	0.00546
5	Winnipeg2 - Toronto	0.00514
6	Seattle - Calgary	0.00508

As a class of improvements I we consider the effect of adding a new link (the addition of groups of links will be considered in [17]). $IM(S,i)$ is the network we obtain from S after the addition of the new link. The results shown in Table I indicate that the connection Winnipeg2-Winnipeg1 is by far the most important one since it is crucial for the correct interplay of the OC-12 and OC-48 routes. The routers Winnipeg1 and Winnipeg2 are, respectively, the first and the second in the list of the most important nodes. Conversely, when two nodes are removed at once, the couple Winninipeg1 + Montreal produces a larger effect than the couple Winnipeg1 + Winnipeg2, which is only the tenth in the list (not in Table I) with $\Delta\Phi^-/\Phi=0.570$. Concerning the improvement analysis, the best links to add are long cables bridging two different parts of the network, as for instance, Toronto-NYC or Winnipeg1-Toronto.

TABLE II. Attacks and improvement analysis of Infonet 2001 [18] as of September 2001. Same as in Table I. In the last column we report the betweenness b of the removed edge, the degree k of the removed node, and the sums of the degrees of the two removed nodes.

	Damaged link	$\Delta\Phi^-/\Phi$	b
1	NYC-New Jersey	0.379	2205
2	New Jersey-Chicago	0.229	1185
3	NYC-Washington	0.197	1185
4	Washington-Atlanta	0.183	1120
5	New Jersey-San Jose	0.179	984
6	New Jersey-Dallas	0.122	609
	Damaged node	$\Delta\Phi^-/\Phi$	k
1	New Jersey	0.573	9
2	NYC	0.530	9
3	Chicago	0.280	15
4	Amsterdam	0.241	9
5	Atlanta	0.227	14
6	Washington	0.203	2
	Damaged couple of nodes	$\Delta\Phi^-/\Phi$	k_1+k_2
1	NYC + New Jersey	0.723	17
2	New Jersey + Amsterdam	0.710	18
3	New Jersey + Atlanta	0.707	23
4	New Jersey + Frankfurt	0.689	20
5	NYC + Chicago	0.685	24
6	New Jersey + Washington	0.673	11
	Added link	$\Delta\Phi^+/\Phi$	
1	New Jersey-Atlanta	0.0522	
2	Chicago-Atlanta	0.0481	
3	NYC-Atlanta	0.0437	
4	San Jose-Atlanta	0.0395	
5	Dallas-Atlanta	0.0341	
6	Chicago-Amsterdam	0.0339	
7	NJersey-Amsterdam	0.0329	
8	NYC-Chicago	0.0326	
9	Atlanta-Amsterdam	0.0318	

Infonet. As a second example we study the Internet backbone of Infonet [18] as of September 2001. The network of Infonet has $N=94$ nodes and $K=96$ cable connections and carries about 10% of the traffic over U.S. and Europe. It consists of two main parts, the U.S. and the European backbone, respectively, with $N_1=66$ and $N_2=28$ nodes, connected by three overseas cables. In Table II we consider the same classes of damages and improvements as in the previous example. The vulnerability of Infonet under single link damages is $V=0.379$, with NYC-New Jersey being the critical link damage. Such a link plays in the network a role similar to red bonds in percolation [19]. In fact, the removal of such a link will result in a breakup of the network into two disconnected parts of about the same size, with a decrease of the 38% in the performance of the network. Notice that the second highest link damage produces only a drop of 23% in the performance. Other important links are those connecting New Jersey with Chicago, with San Jose, and with Dallas, and some links in the east cost as NYC-Washington and Washington-Atlanta. The links in Table II, ordered according to $\Delta\Phi^-/\Phi$, have also a decreasing betweenness b, another measure of link centrality defined as the number of times the link is in the shortest paths connecting couples of nodes [7]. Nevertheless, the correlation between $\Delta\Phi/^-\Phi$ and b is not perfect: for instance the link NYC-Amsterdam, with the second highest betweenness, ranks only 14th according to $\Delta\Phi^-/\Phi$. The vulnerability under damages of single nodes (couples of nodes) is $V=0.573$ ($V=0.723$). New Jersey and NYC are by far the two most important nodes: the damage of either one would disconnect the U.S. from the European backbone, reducing by more than 50% the performance of the network. The damage of both nodes at once reduces by more than 70% the network performance. The damage analysis of other networks [20] shows that the link NYC-New Jersey and the nodes NYC and New Jersey play an important role also in other Internet backbone maps. Such a result might explain the significant drop in performance, marked by increased packet loss and difficulty in reaching some websites (in particular in the connection from U.S. to Europe), experienced by the Internet in the aftermath of the September 11 terrorist attacks. In fact, the stress the U.S. Internet infrastructure was subjected to was the greatest encountered over its 32-year history and was probably related to the damages

TABLE III. Attacks and improvement analysis of the MBTA. The same as in Table I. The letters in parentheses indicate the line/lines the stations belong to: R=red, G=green, G_B=green B, G_C=green C, O=orange, B=blue.

	Damaged link	$\Delta\Phi^-/\Phi$
1	Park Street(RG)- Boylstone(G)	0.275
2	Boylstone(G) - Arlington(G)	0.270
3	Arlington(G) - Copley(G)	0.270
4	Copley(G) - Hynes(G)	0.256
	Damaged node	$\Delta\Phi^-/\Phi$
1	Kenmore(G)	0.343
2	Copley(G)	0.333
3	Park Street(RG)	0.331
4	Boylstone(G)	0.285
	Damaged couple of nodes	$\Delta\Phi^-/\Phi$
1	Down. Cross.(RO) + Kenmore(G)	0.508
2	Park Street(RG) + Kenmore(G)	0.495
3	Down. Cross.(RO) + Copley(G)	0.465
4	Boylstone(G) + Kenmore(G)	0.444
	Added link	$\Delta\Phi^+/\Phi$
1	Mount Hood(G_B)- Dean (G_C)	0.0390
2	Mount Hood(G_B)- Tappan(G_C)	0.0370
3	Washington(G_B)- Tappan(G_C)	0.0369
4	Washington(G_B)- Dean (G_C)	0.0368

of Internet routers and cables in the south of NYC [21].

The comparison of our measure with the node degree k, i.e., with the number of links incident with the node (see Table II) shows that the damage of the most connected nodes, the hubs [2], is not always the worst damage. In fact, the damage of Chicago, the node with the highest k, produces only a drop of 28% in the performance of the network, while the damage of Chicago and Atlanta, the couple with the highest number of links (29) gives $\Delta\Phi^{-}/\Phi = 0.476$ (the 187th damage in the list). This has deep consequences on the best strategy to adopt in a protection policy. In fact, a node with a large degree is immediately recognized as a major channel of communication, being very visible since it is in direct contact with many other nodes. On the other hand, Infonet is a typical example in which the crucial components, i.e., the nodes to protect from the attacks, are not the hubs, but less visible and apparently minor nodes.

Our results imply either an intense policy of protection of the critical links/nodes from attacks, or a strategic expansion of the network with the addition of new links. We now investigate the best strategies to increase the performance of the network by the addition of a new link. The improvability of S under such a class of improvements is $IM = 0.052$. In the highest positions we find two different classes of links: links connecting two IP presences in the U.S., and links connecting the U.S. and Europe. A new link between the U.S. and Europe, namely the link Washington-Geneva, was in fact planned in the expansion of Infonet 2001. Our method predicts that the inclusion of such a link increases by 2.5% the network performance.

MBTA. As a final example we consider a transportation system, the Boston subway (MBTA), consisting of four lines, $N = 124$ stations, and $K = 125$ tunnels [22]. Here the links latency has been taken to be proportional to the time it takes to go from a station to the next one. The results of the analysis are in Table III. The vulnerability V is equal to 0.275, 0.343, 0.508, respectively, for damages of single links, single nodes, or couples of nodes. The critical link is Park Street - Boylstone. IM is equal to 0.039 with best links to be added to those connecting stations on the green line B with stations on the green line C.

Summing up, in this paper we have proposed a general method to spot the critical components of a network. With this method we are able to identify the points of an infrastructure network that are crucial to the functioning of the system, i.e., those nodes and connections whose protection from terrorist attacks must be assumed as the first concern of any national policy. The method, used as an improvement analysis, can also help to better shape an expansion of the network. Other real and artificially generated networks are currently under study [17].

[1] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[2] R. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **406**, 378 (2000).
[3] P. Holme, B. J. Kim, C. N. Yoon, and S. K. Han, Phys. Rev. E **65**, 056109 (2002).
[4] P. Crucitti, V. Latora, M. Marchiori, and A. Rapisarda, Physica A **320**, 622 (2003).
[5] R. Albert, I. Albert, and G. L. Nakarado, Phys. Rev. E **69**, 025103(R) (2004).
[6] P. Crucitti, V. Latora, and M. Marchiori, Phys. Rev. E **69**, 045104(R) (2004).
[7] M. Girvan and M. E. J. Newman, Proc. Natl. Acad. Sci. U.S.A. **99**, 8271 (2002).
[8] A. E. Motter, T. Nishikawa, and Y. Lai, Phys. Rev. E **66**, 065103 (2002).
[9] The method can be extended to the case in which the performance is a combination of two or more variables.
[10] There are two main improvement strategies: We can better shape the expansion of a given infrastructure in order to increase its performance, or in order to decrease its vulnerability. The most general strategy is an appropriate combination of the above two strategies, to get a good mixture of performance and low vulnerability. In this paper we adopt the first of the two strategies.
[11] V. Rosato and F. Tiriticco, Europhys. Lett. **66**, 471 (2004).
[12] The formalism presented in this paper can be easily extended also to *directed* graphs.
[13] V. Latora and M. Marchiori, Phys. Rev. Lett. **87**, 198701 (2001); Eur. Phys. J. B **32**, 249 (2003).
[14] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[15] L. R. Ford and D. R. Fulkerson, *Flows in Networks* (Princeton University Press, Princeton, NJ, 1962).
[16] http://205.189.33.72/optical/pdf/canet3routing.pdf
[17] P. Crucitti, V. Latora, and M. Marchiori (unpublished).
[18] http://www.infonet.com
[19] H. E. Stanley, J. Phys. A **10**, 1211 (1977).
[20] http://navigators.com/isp.html
[21] http://www.cnn.com/2001/TECH/industry/09/12/telecom.operational.idg/
[22] V. Latora and M. Marchiori, Physica A **314**, 109 (2002).