Low-cost Enhancement of the Intra-domain Internet Robustness Against Intelligent Node Attacks

Panagiotis Pantazopoulos and Ioannis Stavrakakis
Department of Informatics and Telecommunications,
National & Kapodistrian University of Athens,
Ilissia, 157 84 Athens, Greece
Email: {ppantaz, ioannis}@di.uoa.gr

***Abstract*—Internet vulnerability studies typically consider highly central nodes as favorable targets of intelligent (malicious) attacks. Heuristics that use redundancy adding k extra links in the topology are a common class of countermeasures seeking to enhance Internet robustness. To identify the nodes to be linked most previous works propose very simple centrality criteria that lack a clear rationale and only occasionally address Intra-domain topologies. More importantly, the implementation cost induced by adding lengthy links between nodes of remote network locations is rarely taken into account.**

In this paper, we explore cost-effective link additions in the locality of the targets having the k extra links added only between their first neighbors. We introduce an innovative link utility metric that identifies which pair of a target's neighbors aggregates the most shortest paths coming from the rest of the nodes and therefore could enhance the network connectivity, if linked. This metric drives the proposed heuristic that solves the problem of assigning the link budget k to the neighbors of the targets. By employing a rich Intra-domain networks dataset we first conduct a proof-of-concept study to validate the effectiveness of the metric. Then we compare our approach with the so-far most effective heuristic that does not bound the length of the added links. Our results suggest that the proposed enhancement can closely approximate the connectivity levels the so-far winner yields, yet with up to eight times lower implementation cost.

I. Introduction

Due to the Internet's high penetration in the society, the disruption of Internet services can have an enormous economical impact. This disruption may be the outcome of a malicious attack that is most of the times manifested through an Internet node attack. According to the Akamai global Internet monitor tool, the State of California, home to Silicon Valley technology giants, saw a massive 497 attacks in a single day of January 2014 [1]. More than sixty percent of the attacks were reported to involve single targets, most likely a subset of vulnerable servers.

Internet node attacks have been the focus of network research for some years. A well-known result suggests that scale-free topologies are highly robust to random failures but extremely vulnerable to targeted (*i.e.*, intelligent) node attacks [2]. The skewness of their degree distribution yields (a few) hub-nodes that can paralyze the network, if removed. Similar hub-nodes structure has been observed for topologies that represent snapshots of the intra-domain Internet [3]. The corresponding (router-level) topologies are the ones that a malicious user can practically access and orchestrate, there, a node attack [4].

To mitigate the risk of serious communication disruption when hub-nodes are taken down, simple modifications of the existing network structure have been sought, rather than redesigning the network topology. Along these lines, the idea of link redirection between two initially connected node pairs appears to be a promising countermeasure [5]. A number of studies propose adding redundancy in the form of k extra links and explore whether their appropriate addition in the network can effectively enhance connectivity. Simple topology-aware heuristics are used to determine where to add the available k links in the network. Adding links between nodes presenting the lowest Degree Centrality (DC) is shown to outperform all considered alternatives in keeping the topology of synthetic and real-world Internet maps connected [6]. In a more dynamic setting of repeated attack-and-defense phases, it has been shown that connecting low Betweenness Centrality (BC) nodes along the defense phase is the most efficient modification for a wide range of networks including synthetic Intra-domain graphs [7]. More sophisticated link addition methods, such as the one in [8] which seeks to create sequences of network links in the form of a cycle, suffer from significant scalability problems rendering their applicability over large real-world topologies problematic.

Motivation and objective: Common to almost all past studies is a lack of consideration of the cost associated with the implementation of the proposed redundancy. For instance, linking nodes of far-away network locations can be of prohibited cost. Moreover, topologies that represent real-world communication networks are only occasionally put under the microscope. Our paper seeks to *systematically* study link addition heuristics in Intra-domain Internet topologies. The primary objective is to devise a *cost-effective* solution that enhances the Internet robustness against intelligent attacks targeting the network hubs, and evaluate the approach using both connectivity and cost metrics.

Our contribution lies in addressing the link addition problem from a different angle, aiming at devising solutions of *low link cost*. To this end, we propose that the links are placed to connect directly two first neighbors of (expected to be) targeted nodes, expecting that the associated link length (cost) would be low. By doing so, the impact of the removal of the targeted

This work has been partially supported by EINS, the Network of Excellence in Internet Science through EC's Grant Agreement FP7-ICT-288021.

nodes (after being attacked) would also be mitigated, as the added link could provide for the recovery of (some of) the lost (local) connectivity. As intelligent node attacks target the top hubs of the topology, our approach amounts to placing the available k additional links to connect the "most appropriate" first neighbors of the top hubs of the topology.

Identifying the "most appropriate" first neighbors of a targeted node is not trivial. Another contribution of this work relates to the introduction of a centrality-based Link Utility (LU) metric based on which such neighbors would be identified: this LU metric expresses the extent to which a network node pair is part of shortest paths from the rest of the nodes towards a considered hub. Driven by this metric, the proposed LU heuristic assigns links among the (not directly connected) first neighbors of the hubs that induce the k top LU values. Extensive simulations with more than 20 intra-domain (router-level) graphs show that our heuristic achieves similar connectivity levels with previous approaches while inducing up to 8 times lower implementation cost.

The remainder of the paper is structured as follows: In Section II, we formally express the problem and introduce the link utility (LU) metric used subsequently to drive our link addition heuristic. The latter is assessed using three different datasets of intra-domain snapshots described in Section III. In Section IV we present extensive experimental results demonstrating the effectiveness of our heuristic in terms of both topological robustness and implementation cost. Related literature is summarized in Section VI and the paper concludes in Section VII with pointers to future directions.

II. Problem statement and proposed heuristic

The network robustness to node attacks may be of interest to various parties; a potential adversary would like to know which node attacks cause the most significant impact on the network performance. From the network operator's side which we herein adopt, one seeks appropriate countermeasures to maintain high levels for the underlying network connectivity and accordingly the provided QoS.

A. Enhancing the topological robustness with link additions

To enhance the topological robustness a commonly used method is to purchase and install a small number of k extra links. These links directly connect pairs of nodes that are selected in such a way as to ensure the highest possible network connectivity after certain nodes are attacked (and removed). Besides connectivity though, a frequently neglected issue is that of the implementation complexity and ultimately the *cost* of installing those links (*e.g.*, [6], [7]). We therefore revisit the enhancement of the Internet robustness against intelligent node attacks by posing the following problem: "*Given a set of links L with $|L|$=k and a network represented by the connected graph G=(V, E) of $|V|$ nodes and $|E|$ links, find the appropriate k node pairs to be linked in a way that a) enhances network robustness against attacks over its hubs and b) incurs the minimum possible implementation cost*".

In this paper we try to address this problem by making certain assumptions which are either standard, intuitively meaningful, or backed by experimental data, and their validity is assessed in this work based on the induced experimentation results. Experimental data suggest that intelligent attackers select the hubs (nodes of highest DC) as the targeted nodes. A standard and intuitively meaningful (though not always accurate) assumption is that the implementation cost of a link is proportional to its length (*i.e.*, physical distance between the connected nodes); if this assumption is not strictly true then still the problem posed above is important, as it seeks to minimize the physical length of the added links. The other assumption made (whose overall validity is assessed through the results derived in this paper) is that low hop count corresponds (on the average) to low physical distance and, thus, the length of a link is minimized if it connects unconnected nodes of hop count two (*i.e.*, the minimum hop count of any unconnected nodes).

Based on the aforementioned assumptions and in order to minimize the implementation cost, we propose adding links only between nodes of hop count two. Among all possible links of hop count two, we propose that those that are first neighbors of the top hubs be considered only. The rationale behind this is the assumption (to be tested through the experimentation results) that such links would help preserve some of the connectivity lost after the hub node is removed (attacked). Finally, among all first neighbors of the hubs, links will be placed among the k not randomly selected node pairs but among pairs presenting the highest values of an appropriate link utility metric introduced next.

B. The Link-Utility metric and the associated LU-heuristic

With the link utility metric we seek to identify the pair of nodes whose connection with a direct link would have the greatest impact in terms of connectivity. Intuitively, high utility should be assigned to those first neighbors of a given hub that aggregate the most communication paths from the rest of the nodes towards that hub. By establishing a direct link between them we expect that a large number of network nodes would remain connected should the corresponding hub be removed. We have implemented this idea for the case of shortest path communications which is the typical Internet practice by employing appropriate centrality metrics.

Betweenness Centrality (BC) is a common metric of the Complex Network Analysis toolbox that reflects to what extent a node lies on the shortest paths linking other nodes. In [9] we have proposed the Conditional BC (CBC), as a way of capturing the centrality of a network node with respect to a specific node t. If σ_{st} denotes the number of shortest paths between any two nodes s and t in a connected graph $G = (V, E)$ and $\sigma_{st}(n)$ is the number of shortest paths passing through node $n \in V$, then CBC is defined as

$$CBC(n;t) = \sum_{s \in V \setminus t, n \neq t} \frac{\sigma_{st}(n)}{\sigma_{st}} \quad (1)$$

with $\sigma_{st}(s) = 0$. The summation is over all node pairs (x, t) $\forall x \in V$ destined at node t rather than all possible pairs, as in the BC definition. Effectively, CBC assesses to what extent a node n acts as a *shortest path aggregator* towards the target t. Consequently, a meaningful measure of the utility of a link to be placed between two neighboring nodes of a hub can be obtained through the sum of the CBC values of these nodes, computed with respect to the hub. Before formally introducing

the Link Utility metric, we prove the following proposition regarding the sum of the CBC values that the first neighbors of a hub attain.

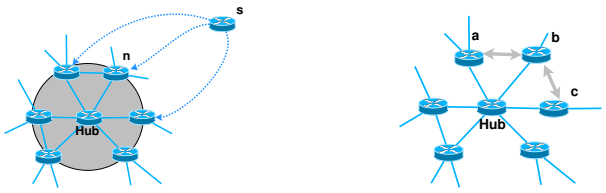


Fig. 1. **Left**: The different shortest paths from node s towards the Hub shape the CBC value of each of the Hub's first neighbor. The sum of these values over every first neighbor n is constant for all hubs and equals $|V|$-1. **Right**: If added links are placed between nodes a-b and b-c (due to their high utility), then it is redundant to connect a with c.

Proposition 2.1: The sum of $CBC(n; Hub)$ values over every node n that belongs in the set of the first neighbors Ng of a hub-node Hub, is equal to $|V|$-1.

Proof: The sum of the CBC values is

$$\sum_{n \in Ng} CBC(n, Hub) = \sum_{n \in Ng} \sum_{s \in V \setminus Hub} \frac{\sigma_{sHub}(n)}{\sigma_{sHub}}$$

$$= \sum_{s \in V \setminus Hub} \sum_{n \in Ng} \frac{\sigma_{sHub}(n)}{\sigma_{sHub}}$$

$$= \sum_{s \in V \setminus Hub} \frac{1}{\sigma_{sHub}} \sum_{n \in Ng} \sigma_{sHub}(n) \quad (2)$$

$$= \sum_{s \in V \setminus Hub} 1 = |V| - 1 \quad (3)$$

Note that the second term of Eq. 2 is the sum over all nodes $n \in Ng$ of all shortest paths that stem from s, pass through n and reach the Hub (Fig. 1 left); this quantity equals the number of shortest paths linking s with the Hub *i.e.*, σ_{sHub}. ■

Proposition 2.1 allows for the direct comparison of link utilities between node pairs that are first neighbors to different hub nodes. The ratio between the sum of $CBC(i; Hub)$ and $CBC(j; Hub)$ values of nodes i, j over the constant sum of proposition 2.1 reflects the portion of the overall paths towards the hub that pass through i and j. Another remark is that since intelligent attacks are carried out over a sequence of hub-nodes with decreasing degree, the utility of a link potentially placed between neighbors of a high-degree hub should be higher than the corresponding one of a lower-degree hub. To factor this in, we employ the ratio of the degree of the considered hub $DC(Hub)$ over the maximum degree DC_{max} across the network. Finally, the utility of a link added between any two nodes i and j of the hub-node Hub is given by:

$$LU^{ij}_{Hub} = \frac{CBC(i; Hub) + CBC(j; Hub)}{\sum_{n \in Ng} CBC(n; Hub)} \cdot \frac{DC(Hub)}{DC_{max}} \quad (4)$$

The pseudocode for the corresponding LU-heuristic appears in Algorithm 1. The parameter of the number of top hubs H that appears in the pseudocode can in principle be set equal to $|V|$, implying that all nodes are considered as targets and potentially could be assigned links to connect their neighbors. In practice, the value of H can be set much lower than $|V|$ (to reduce the computational complexity of the algorithm), as dictated by the expected number of attacks that the network operator estimates based on possessing logs and global network statistics. The LU-based approach that distributes the available k links to the k top-utility pairs of non-connected neighbors of the H hubs, is intuitively more appealing compared to one that would naively assign a fix number of links to each hub.

Algorithm 1 Link-Utility heuristic over *G=(V,E)*

```
1.  compute utilities :
2.  for all u ∈ H do
3.    identify the set Ng of first neighbors
4.      for all v ∈ Ng do
5.        R_k ← compute the Rank k of v w.r.t. CBC(v; u)
6.      end
7.      for n = R_1 down to R_Ng do     #from top to bottom of the ranking
8.        if node n not connected to n + 1 then compute LU_u^{nn+1}
9.      end
10. end
11. add links :
12. for all links of the budget k do
13.   identify the top LU_u^{ij} value across all computed LU_u^{nn+1}
14.   add one link between nodes i and j
15. end
```

To avoid adding links (around a hub) that will not bring about further robustness improvements (see Fig. 1 right), we do not actually consider all the non-connected pairs of the hub's first neighbors. We rather identify top utility pairs by carrying out a simple one-pass (pseudocode line 7) over the ranking of all neighbors in decreasing order of their CBC values. Computing the LU metric of two consecutive in the ranking (*i.e.*, n - n+1), non-connected neighbors of a hub (line 8) prevents the addition of such almost "useless" links. Then, we use the link budget k to connect the identified pairs with the k top utilities (lines 13-14).

III. Intra-domain network topologies

All our experiments are carried out over datasets collected in the context of three projects. The first two (Table I) relate to measurement projects and are referred to as *mrinfo* [10] and Rocketfuel [11]. They report binary router-level graphs[1] for different Internet ASes. The third dataset (Table II) called Topology Zoo, contains also topologies at the router- and PoP-level [13]. It is collected directly by the operators of primarily academic/research networks and therefore contains some extra information such as the nodes coordinates. Next, we briefly describe each dataset:

Mrinfo datasets: The dataset was collected during 2005-2008 and contains numerous Tier-1, Transit and Stub ISP network topology files [10]. To cope with `traceroute` inaccuracies, snapshots were extracted by the new `mrinfo` tool which silently crawls IPv4 addresses. The tool efficiently discriminate interconnections between ASes without suffering from IP alias resolution problems. In our study we considered only the largest available snapshots that correspond to Tier-1 and Transit topologies leaving aside the small-sized Stub topologies.

Rocketfuel dataset: The Rocketfuel dataset [11] is the chronologically oldest dataset, drawn with the help of the

[1]Many of the original network topology files, as released in a raw trace-based format, miss some edges. We have therefore used a well-known linear-time algorithm [12] to retrieve the giant connected component (GCC).

TABLE I. PROPERTIES OF THE MRINFO AND ROCKETFUEL SNAPSHOTS

DataSet	ID	ISP(AS number)	Links	Diameter	Nodes	<degree>
	36	Global Crossing(3549)	141	10	76	3.71
	35	-//-	189	9	100	3.78
	33	NTTC-Gin(2914)	318	11	180	3.53
m	21	Sprint(1239)	332	12	216	3.07
r	13	Level-3(3356)	849	25	378	4.49
i	12	-//-	1087	28	436	4.98
n	20	Sprint(1239)	827	16	528	3.13
f	9	-//-	1220	13	741	3.29
o	39	Iunet(1267)	1228	13	711	3.45
	44	Telecom Italia(3269)	1816	13	995	3.65
	50	TeleDanmark(3292)	1896	15	1240	3.06
R	60	VSNL(4755)	68	6	41	3.32
O	61	Ebone(1755)	544	13	295	3.68
C F	62	Tiscali(3257)	653	14	411	3.18
K U	63	Exodus(3967)	820	14	353	4.65
E E	64	Telstra (1221)	3045	15	2515	2.42
T L	66	Level-3(3356)	6743	10	1620	8.32

`traceroute` active measurement tool. The Rocketfuel engine collected raw `traceroute` data from public BGP tables, processed them and extracted router-level networks by mapping diverse ISP routers to ASes. ISPs across the world were mapped utilizing approximately 800 `traceroute` sources hosted by nearly 300 servers.

TABLE II. PROPERTIES OF THE TOPOLOGY ZOO SNAPSHOTS

Network	Geo Location	Date of snapshot	Links	Diameter	Nodes
Kentman	Kent, UK	8/2005	29	8	27
Geant	cross-Europe	2009	52	7	34
Telcove	USA	2010	70	7	71
Uninett II	Norway	2010	101	9	74
VTLWavenet	cross-Europe	2011	96	31	92

Topology Zoo dataset: Whereas previous studies employ a number of route discovery tools to reveal the Internet connectivity, the Topology Zoo gathers the maps of more than 140 real-world topologies directly from the network operators. As the resulting maps (topologies and associated attributes) come from the owner and/or manager of the network, they are claimed to reflect an accurate network view circumventing any errors due to biases of measurement techniques. We have selected a subset of the largest router-level snapshots (Table II) retrieved during 2008-11. We will use them in Section IV-C to evaluate the cost of link additions in terms of their length.

IV. EXPERIMENTATION RESULTS

We now proceed with a systematic assessment of our heuristic. First, we demonstrate the effectiveness of the link utility metric in identifying the most appropriate first neighbors (of a hub) to be connected, according to the LU heuristic. Then, we compare the introduced LU-heuristic with the so far most efficient link-addition method in connectivity terms [6] that we refer to as minDC heuristic; the latter connects nodes of minimum degree without posing constraints on their distance. The connectivity of the enhanced (through the k links addition) topologies is assessed in terms of the giant connected component (GCC) size which is extensively used in literature [6]–[8], as well as the total number of connected components. We present the values that these performance metrics attain as the hub-nodes (together with their attached links) are removed from the topology. The third metric considered is the average shortest path, which seeks to capture the network's communication efficiency as path lengths may serve as delay indicators [8]. Finally, the term "attack level" is used to denote the percentage of the total nodes that are removed. Unless otherwise stated, the value of nodes H in the pseudocode for which the LU metric is calculated is also set equal to the "attack level".

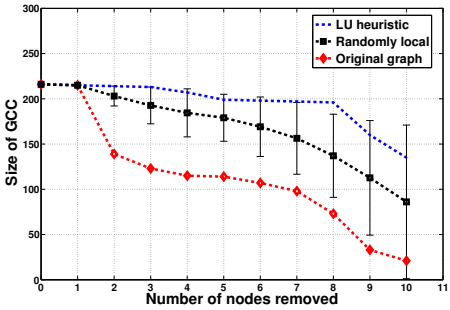


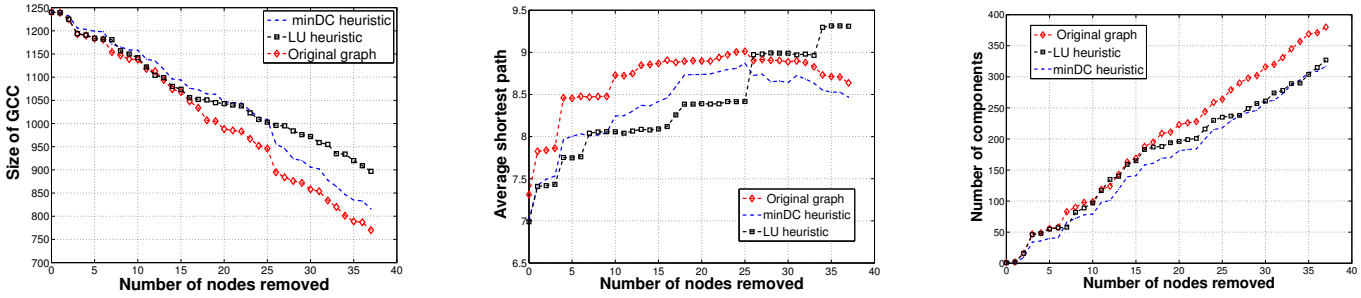
Fig. 2. Comparing methods of link-addition between the first neighbors of a hub: the connectivity levels in terms of GCC that the LU and a random link-addition heuristic achieve over a Sprint snapshot (dataset 21).

A. Connecting first neighbors based on the link utility metric: proof-of-concept experiments

To show that the link utility metric and, accordingly, the LU heuristic effectively identifies the most appropriate first neighbors around each hub, we compare it against a simple method called "Randomly local", that randomly selects which neighbors to connect. To this end we employ a representative `mrinfo` snapshot *i.e.*, dataset 21, and enhance it by adding an extra 5% of its total links considering the 10 top hub-nodes as targets (*i.e.*, attack level approx. 5%). As our focus is on the node pair that each method selects to connect, we consider a common additional link assignment vector to the hub-nodes; inline with the concept of intelligent attacks, this is done proportionally to the degree of each hub. For the "Randomly local" approach we generate a sufficiently large set of 10000 different enhanced topologies and measure the average GCC size values along with their 95% confidence intervals.

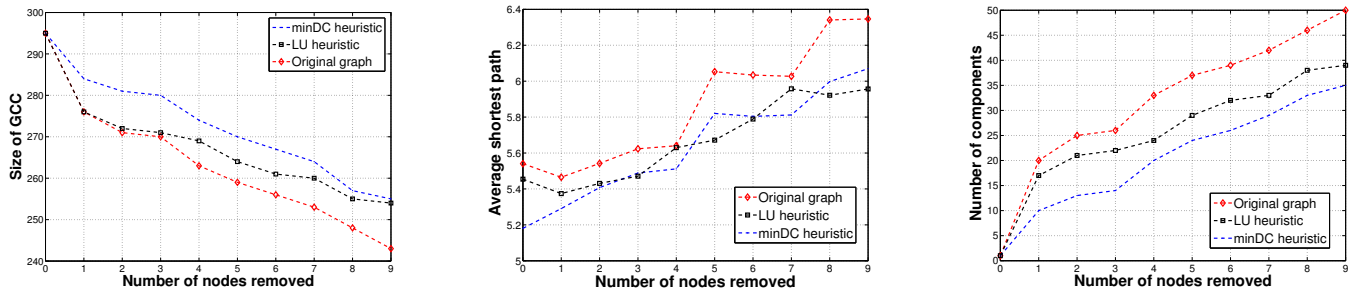
Fig. 2 shows the size of the GCC of the original topology, the one enhanced according to the LU heuristic and the one enhanced under the "Randomly local" approach. The results show that the LU heuristic can preserve high connectivity levels for a considerable number of removals (up to 4% of the original network size) and clearly outperforms the "Randomly local" approach. This suggests that the LU heuristic is effective in identifying the most appropriate node pairs out of all the first neighbors of a hub validating both our intuition about the criticality of the neighbors that establish many paths towards the hub and the effectiveness of the introduced link utility metric in capturing this notion. The random link addition cannot guarantee that large network areas will remain connected after the attack. Even worse, the variance of the GCC size increases with the number of removals *i.e.*, a large number of enhanced network instances that had many of their hubs removed faced severe fragmentation. Finally, the original network appears highly vulnerable as more than half of its nodes become disconnected when the seventh hub is removed.

Our last comment relates to the potential use of other centrality-driven heuristics, to drive the link addition process between the first neighbors of a hub. A number of well-known centrality metrics introduced in the literature [14] could be



a. b. c.

Fig. 3. Comparison of link-addition heuristics with respect to (a) the GCC size (b) the average path length and (c) number of components for dataset 50.



a. b. c.

Fig. 4. Comparison of link-addition heuristics with respect to (a) the GCC size (b) the average path length and (c) number of components for dataset 61.

employed to this end. It is important to identify the most appropriate to capture the notion of significance in a given context. That is what we have sought to do with the introduction of the link-utility metric based on the (CBC) centrality. Finally it should be mentioned that while ignoring any centrality consideration leads to a poor performance ("Random local" approach), it is the simplest solution to implement.

B. Comparing the robustness of the enhanced networks

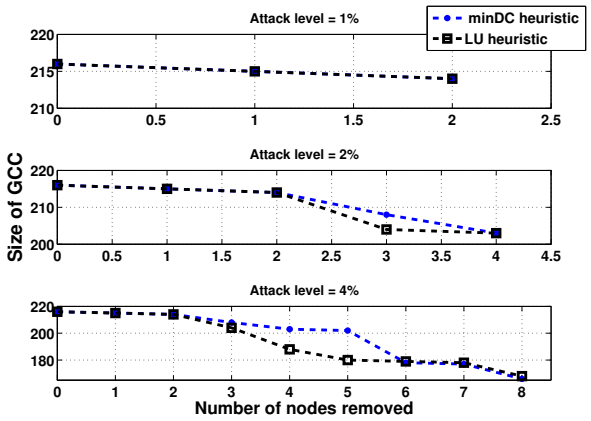


Fig. 5. The GCC size that the heuristics achieve over dataset 21 when a few hubs are removed. The topology is enhanced with 5% of its original links.

In this section we compare the performance of the introduced LU heuristic with that of the most effective link addition heuristic, with respect to the GCC metric. The latter heuristic simply adds the k additional links between the network nodes with the lowest degree (to be referred to as the minDC heuristic), without attempting implicitly or explicitly to keep the link length low [6]. Indicative results are presented in Figs. 3 and 4 for an `mrinfo` and a Rocketfuel snapshot, respectively. Similar results have been observed across all considered datasets. In these experiments the attack level is set to 3% and each topology is enhanced by adding 5% of the total number of its links.

The connectivity of the network under attack is of our primary interest. The GCC size and the number of connected components measure the impact of the hubs' removal. Regarding the first metric, Fig. 3.a shows that the two heuristics perform similarly with respect to the maximum GCC size, while the original network suffers from rapid fragmentation as the number of attacks increases. Our results over the snapshots of the two datasets suggest that no heuristic is consistently more effective than the other. Fig. 4.a depicts a less favorable result for our approach over the Rocketfuel dataset. However, the difference in the GCC size the two approaches achieve is no larger than 3.5%. We have identified only one network topology where the GCC size difference between the two approaches reaches up to 9% for certain node removals. This result implies a welcome property for our heuristic; it remains effective regardless of the size of the considered network.

On a more practical note, realistic malicious attacks especially over large networks would target no more than two or three strategic hubs rather than a set sizing up to tens of hubs. In this case, the two competing heuristics yield almost

identical impact in terms of the GCC size for the vast majority of the studied networks. As Fig. 5 suggests, it takes a higher number of node removals for the GCC size differences to become significant. Furthermore, Fig. 5 illustrates another positive result regarding the LU-heuristic performance. As the number of removed nodes increases (*i.e.*, attack level), the number of hubs H considered in the pseudocode calculating the corresponding LU metrics increases similarly. This leads to the consideration of more LU values and a change in the distribution of the k links to the H hubs. The proposed heuristic manages to effectively distribute the available link budget across the H hubs and perform very close to the minDC heuristic, that links the same node pairs regardless of the number of removed hubs.

In terms of the number of components, Figs. 3.c and 4.c show that the minDC heuristic generally enhances the network in a way that node attacks cause the lowest fragmentation. On the other hand, the original graph easily breaks into multiple components, as expected. Our heuristic offers a somewhat intermediate solution that occasionally coincides with the minDC. Regarding the average shortest path, the trend is to increase with the number of node attacks. As nodes and associated links are removed, node pairs communicate over paths of more hops. A differentiation from this general trend is the twofold behavior observed for the original graph and the minDC heuristic in Fig. 3.b. First, the average shortest path increases and then, suddenly follows a fast decay. This fluctuation seems to mainly depend on how fast the node removals lead to the total network fragmentation. Potentially, there exists an upper bound of removals that permits GCC to maintain a relatively large size before its connectivity has been significantly diminished. Consequently, as long as GCC maintains a significant size, node removals result in increasingly longer paths between its node pairs. When the network has been broken down to several small clusters, further removals tend to create single isolated nodes and therefore, decrease the average shortest path.

TABLE III. MEAN AND MAX OF THE (%) GCC SIZE DIFFERENCES BETWEEN MINDC- AND LU-HEURISTIC UNDER $H = |V|$

	Dataset ID					
	33	20	9	61	62	63
$mean\{S_{DIF}\}$ (%)	0.68	3.92	4.1	5.8	7.9	0.66
$max\{S_{DIF}\}$ (%)	1.14	19.2	7.2	13.9	14.5	1.98

In this paragraph we do not limit the parameter H, in the pseudocode computing the LU metrics, to be equal to the attack level, but allow it to be equal to the total number of network nodes (*i.e.*, it holds that H=$|V|$); in this case the LU metric computations involve the neighbors of all the network nodes. Equivalently, it is like assuming that the number of hubs to be attacked is not known to or estimated by the network operator[2]. To evaluate how the LU heuristic compares with the minDC under this scenario we have fixed the attack level to 3% and the added links to 5% of the total network links. For each topology we compute the relative (%) difference Δ^l=$|GCC^l_{minDC}$-$GCC^l_{LU}| \cdot 100/|V|$ between the GCC size that the minDC- and LU-heuristic achieve, as node l is removed. Thus, we have a set S_{DIF}=$\{\Delta^1, ..\Delta^m\}$ (with m=3%$|V|$) for each considered network topology. To obtain a summarizing view, we compute the mean and maximum value over each S_{DIF} and present indicative results for both `mrinfo` and Rocketfuel topologies in Table III. Our heuristic appears on average less effective than the minDC one, yet with relatively small differences. In all cases the mean GCC size difference does not exceed 7.9%. As now every network node l is considered a candidate target, it is the ratio $DC(l)/DC_{max}$ (see Eq. 4) that helps the LU-heuristic assign the k links appropriately *i.e.*, dedicate few links to nodes of small degree. Studying the full set of results, we again find that in several topologies the heuristics achieve (almost) the same connectivity levels for the first few node removals where the practical interest is concentrated, and exhibit larger differences with subsequent removals. An extreme yet welcome example of this behavior appears in dataset 20. The maximum value of the GCC size difference is kept to 8.1% across fifteen node removals and reaches 19.2% on the (final) sixteenth one. The achievable performance of the minDC heuristic, however, comes at a large implementation cost as we show next.

[2]The relevant problem that relates to the extent to which this estimation is accurate as well as how any introduced error affects the performance of the LU-heuristic, is clearly worth a separate study.

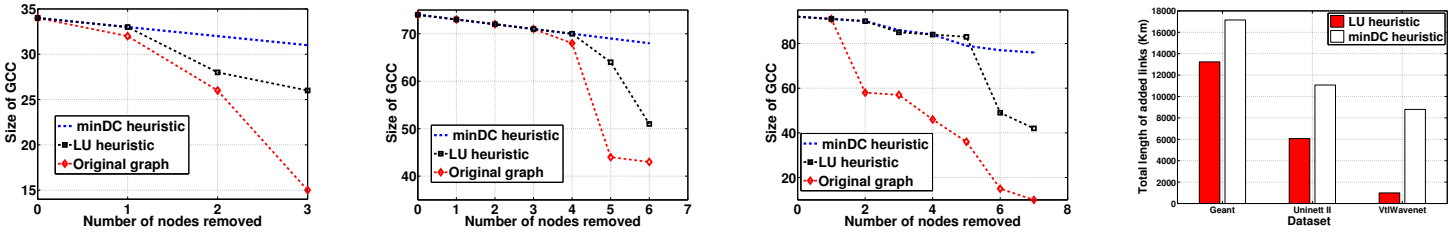
C. Measuring the implementation cost of the heuristics

We have seen how the LU heuristic compares with the minDC heuristic in terms of how robust the topologies become after the k link additions. We now turn our attention to the implementation cost of adding network links; as the two heuristics select different node pairs to connect, we seek to compare how different is the total length (and, thus, implementation cost) of the added links. Since the LU heuristic always connects node pairs with the minimum possible distance (for not directly connected nodes) of two hops, its cost is expected to be lower than that under the minDC heuristic. As the connected nodes under the minimum DC may lie in any network location, the total length of the added links can be very high, depending on the network topology. We measure implementation costs using a carefully selected subset of the largest Topology Zoo snapshots that contain geographical coordinates for all of their nodes. As the Zoo topologies are relatively small, we set the attack level to 8% of the total number of nodes and enhance each topology by adding 8% of its original number of links.

In Table IV we show in detail the computation of the length for each added link over the GEANT network. The dataset is parsed to retrieve the coordinates of each node and then an online tool [15] provided by the US weather service is queried to determine the distance of the nodes over the globe (measured in kilometers). Inline with intuition the minDC heuristic connects nodes of longer distance; assuming a constant implementation cost per unit of link length, it yields approximately 1.3 times larger implementation cost than the LU heuristic. Similar trend is observed in Fig. 6.d for two other Topology Zoo networks *i.e.*, the Uninett and VtlWavenet. Especially for the latter topology which represents a cross-European network of almost 100 nodes, it is worth mentioning that the proposed heuristic yields a welcome cost reduction of 8.8 times compared to minDC. Finally, for the sake of completeness, we present in Figs. 6.a-c a comparison of the LU-heuristic over the three Zoo snapshots with the minDC in terms of the GCC

TABLE IV. COMPUTATION OF THE TOTAL LENGTH OF THE LINKS THE MINDC- AND LU-HEURISTIC ADD OVER THE GEANT NETWORK

minDC heuristic				**LU-heuristic**			
Node pair connected	Longitude	Latitude	Link length (km)	Node pair connected	Longitude	Latitude	Link lenght (km)
27	-21.89541	64.13548	2270	4	16.96667	1.0308	2733
32	22.26869	60.45148		13	34.75	31.5	
13	34.75	31.5	4063	13	34.75	31.5	3195
28	-6.26719	53.34399		3	12.56553	55.67594	
19	-9.13333	38.71667	2094	14	14.42556	35.90917	1712
23	14.50513	46.05108		12	33.36667	35.16667	
27	-21.89541	64.13548	1492	1	4.88969	52.37403	2147
28	-6.26719	53.34399		26	37.61556	55.75222	
13	34.75	31.5	3778	12	33.36667	35.16667	2142
30	25.46816	65.01236		24	14.28611	48.30639	
25	26.8	53.76667	727	1	4.88969	52.37403	710
26	37.61556	55.75222		6	14.42076	50.08804	
7	6.13	49.61167	2727	6	14.42076	50.08804	596
12	33.36667	35.16667		7	6.13	49.61167	
			Total length: 17151 km				Total length: 13235 km



a. Geant b. Uninett II c. VtlWavenet d. Implementation cost

Fig. 6. a-c) The connectivity of the original and two enhanced Topology Zoo snapshots as reflected by the size of the giant connected component (GCC) for each node removal. d) Corresponding total cost (in terms of link length) for the link addition heuristics over the three Topology Zoo snapshots.

size. As the hub-nodes are attacked (*i.e.*, removed), the LU-heuristic maintains the network connectivity at almost the same GCC level as the minDC, at least under the removal of the few top-hubs where the practical interest is.

V. DISCUSSION

This paper has studied the topological robustness of ISP router-level networks and introduced a heuristic that cost-effectively employs link additions to enhance it. In what follows we briefly comment on a number of assumptions made and justify why we believe that our insights are still valuable for the real-world Internet.

The proposed heuristic adds links between the first neighbors of the hubs. To identify the key neighbors of each hub, where the additional links should be placed, we have employed global topological information that may seem cumbersome to obtain; this information is required to compute the link utility and in particular, the involved CBC values. However, as with previous works we have adopted the network operator's point of view. The latter (*i.e.*, an ISP) typically possesses global information about its network topology and thus is able to compute the required link utility values; the CBC metric computations estimated with respect to the hubs H ($H \leq |V|$) can actually be carried out off-line without placing extra burden to the link addition process. The corresponding complexity[3] is $H \cdot O(|E|)$. On the other hand, acquiring the number of connections of each node (*i.e.*, DC value) is of no concern; it then takes $O(|V|log|V|)$ time to sort the nodes in decreasing DC order and implement the intelligent attack.

[3]The work in [16] shows that both the length and number of all shortest paths from a source node to all others can be determined in $O(|E|)$ for unweighted graphs. Further comments on the CBC complexity appear in [17].

A final note relates to the extend to which our results realistically reflect the Internet robustness. This is mainly a question of accuracy for the network discovery tool used each time to extract the underlying topology. We have employed a broad set of Intra-domain topologies extracted by different tools presenting our study with the highest possible credibility. However, there is always some hidden redundancy in the network topology that remains unnoticed. It then follows that our results of Section IV should be viewed as worst-case results for the Internet robustness.

VI. RELATED WORK

The work that concerns attacks over network nodes can be grouped along two threads. The first includes papers that study attacks directed towards the most central nodes and try to assess their impact. Most works concern synthetic graphs while the attack impact is measured through topological measures. A well-known result here suggests that scale free topologies appear highly vulnerable to high-degree nodes [2]. Over real-world AS-level topologies, attacks that target high DC and BC nodes have been found equally harmful in terms of the inverse geodesic length and the number of connected components in the residual network [18]. Regarding the intra-domain router-level topologies, evidence about the catastrophic impact of intelligent attacks has been provided in [4]. Recently, we experimentally assessed how vulnerable these topologies are to various centrality-driven node attacks considering both their topological properties and their traffic-carrying capacity [14].

This paper however falls under the second thread of research that seeks to draw on the conclusions of the above studies and propose effective countermeasures aiming to preserve high connectivity levels in face of the attacks. Adding

extra links is one of the most common ways. The work in [6] experimentally compares the effectiveness of placing extra links in the network as well as rewiring present ones over synthetic and real-world AS level topologies under intelligent attacks. In all cases the authors investigate how these modifications increase network robustness with respect to typical connectivity metrics, same as those employed here. They have found that linking the lowest degree (*i.e.*, minDC) nodes achieves the best connectivity levels. The efficient link addition problem is also considered in [8]. A new robustness metric that essentially relies on harmonic centrality captures both connectivity and network efficiency (in terms of path length). The question then is to identify the network locations where the addition of new links will maximize the robustness metric. Link additions seek to give rise to redundant paths forming cycles around candidate target nodes. Finding such protecting cycles however, requires heavy computations and thus the method does not scale with the network size. More recently, the cost of the added links in terms of physical length has been taken into account; the objective of the relevant heuristics is to add links in a way that maximizes the algebraic connectivity [19] or the path diversity [20]. To assess the resilience of the enhanced topologies against targeted attacks, the authors have used a single connectivity metric. Compared to all above works, we have devised a simple/scalable centrality-based link addition heuristic that combines the efficient enhancement of the Internet robustness (captured by various connectivity metrics) with notably lower implementation cost.

VII. Conclusions

We have studied topology-aware link addition heuristics to enhance the Internet robustness against intelligent attacks that target strategic hub-nodes. With a given budget of k links, the problem is to identify appropriate network node pairs that their connection would preserve high connectivity levels despite the node attacks. Contrary to the so-far approaches that neglect the location of the linked nodes and thus the cost of the link length, we have *by-design* restricted the candidate nodes-to-be-linked to the first neighbors of each hub, *i.e.*, to nodes of distance of two hops. To implement this approach we rely on a new link utility metric that quantifies the extent to which the corresponding node pair can aggregate shortest-paths towards a given hub. By employing this metric we devise a heuristic based on which all first neighbor pairs of the considered top hubs (or even all network nodes) are ranked, and the top k ones (according to the available link budget) are assigned a link. A proof-of-concept study over 20 different Intra-domain topologies has been carried out demonstrating the effectiveness of connecting high utility pairs of the hubs' first neighbors. Then we have compared the so far winner link-addition approach with our heuristic showing that the two can preserve similar connectivity levels, with the latter inducing up to eight times lower implementation cost (measured in terms of the length of the added links).

As link additions in dense network topologies are expected to yield negligible robustness improvements against node attacks, an interesting future direction is to relate the effectiveness of the link-addition heuristics class with the density of the Intra-domain Internet topologies.

References

[1] [Online]. Available: http://www.akamai.com/html/technology/dataviz1.html

[2] R. Albert, H. Jeong, and A.-L. Barabasi, "Error and attack tolerance of complex networks," *Nature*, vol. 406, no. 6794, pp. 378–382, Jul. 2000.

[3] A.-L. Barabâsi and E. Bonabeau, "Scale-free networks," *Scientific American*, vol. 288, no. 60-69, 2003.

[4] D. Magoni, "Tearing down the internet." *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 6, pp. 949–960, 2003.

[5] C. M. Schneider, A. A. Moreira, J. S. Andrade, S. Havlin, and H. J. Herrmann, "Mitigation of malicious attacks on networks," *Proceedings of the National Academy of Sciences*, vol. 108, no. 10, Mar. 2011.

[6] A. Beygelzimer, G. Grinstein, R. Linsker, and I. Rish, "Improving network robustness by edge modification," *Physica A*, vol. 357, pp. 593–612, 2005.

[7] H. Kim and R. Anderson, "An experimental evaluation of robustness of networks," *Systems Journal, IEEE*, vol. 7, no. 2, pp. 179–188, 2013.

[8] L. Li, Q.-S. Jia, H. Wang, R. Yuan, and X. Guan, "A systematic method for network topology reconfiguration with limited link additions," *Elsevier Journal of Network and Computer Applications*, vol. 35, no. 6, pp. 1979 – 1989, 2012.

[9] P. Pantazopoulos, I. Stavrakakis, A. Passarella, and M. Conti, "Efficient social-aware content placement for opportunistic networks," in *IFIP/IEEE WONS*, Slovenia, Feb 2010, pp. 17–24.

[10] J.-J. Pansiot *et al.*, "Extracting intra-domain topology from mrinfo probing," in *Proc. PAM Conference*, Zurich, Switzerland, April 2010.

[11] N. T. Spring *et al.*, "Measuring ISP topologies with rocketfuel." *IEEE/ACM Trans. Netw.*, vol. 12, no. 1, pp. 2–16, 2004.

[12] R. M. Karp and R. E. Tarjan, "Linear expected-time algorithms for connectivity problems (extended abstract)," in *ACM STOC '80*, Los Angeles, California, 1980, pp. 368–377.

[13] S. Knight, H. X. Nguyen, N. Falkner, R. A. Bowden, and M. Roughan, "The internet topology zoo." *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 9, pp. 1765–1775, 2011.

[14] G. Nomikos, P. Pantazopoulos, M. Karaliopoulos, and I. Stavrakakis, "Comparative assessment of centrality indices and implications on the vulnerability of ISP networks," in *26th International Teletraffic Congress (ITC 2014)*, Karlskrona, Sweden, Sep. 2014.

[15] [Online]. Available: http://www.nhc.noaa.gov/gccalc.shtml

[16] U. Brandes, "A faster algorithm for betweenness centrality," *Journal of Mathematical Sociology*, vol. 25, pp. 163–177, 2001.

[17] P. Pantazopoulos, M. Karaliopoulos, and I. Stavrakakis, "On the local approximations of node centrality in internet router-level topologies," in *the 7th IFIP IWSOS, Palma de Mallorca, Spain.*, May 2013.

[18] P. Holme, B. J. Kim, C. N. Yoon, and S. K. Han, "Attack vulnerability of complex networks," *Phys. Rev. E*, vol. 65, no. 5, May 2002.

[19] M. J. Alenazi, E. K. Çetinkaya, and J. P. Sterbenz, "Cost-efficient algebraic connectivity optimisation of backbone networks," *Optical Switching and Networking*, vol. 14, Part 2, no. 0, pp. 107 – 116, 2014.

[20] ——, "Cost-efficient network improvement to achieve maximum path diversity," in *6th International Workshop on Reliable Networks Design and Modeling (IEEE RNDM), 2014*, Nov 2014, pp. 202–208.