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

Panagiotis Pantazopoulos **Ioannis Stavrakakis**

Department of Informatics and Telecommunications,
Advanced Networking Research Group
National & Kapodistrian University of Athens,
Email: {ppantaz, ioannis}@di.uoa.gr

HELLENIC REPUBLIC
National and Kapodistrian University of Athens

11th International Conference on
DRCN | 2015
Design of Reliable Communication Networks

Modeling Internet node attacks

- Internet malicious activities are mostly manifested through node attacks

- “Intelligent” attacks usually consider the highest-degree nodes (hubs) as favorable targets
  - Akamai : more than 60% of 497 attacks orchestrated in the State of California involved subsets of vulnerable (hub-node) servers

- Typical countermeasure against node attacks: enhance connectivity adding redundancy (in the form of extra links)

- Simple heuristics have been mostly introduced to drive link-additions
  - Connect minimum/maximum Degree Centrality (DC) nodes
  - Connect minimum/maximum Betweenness Centrality (BC) nodes

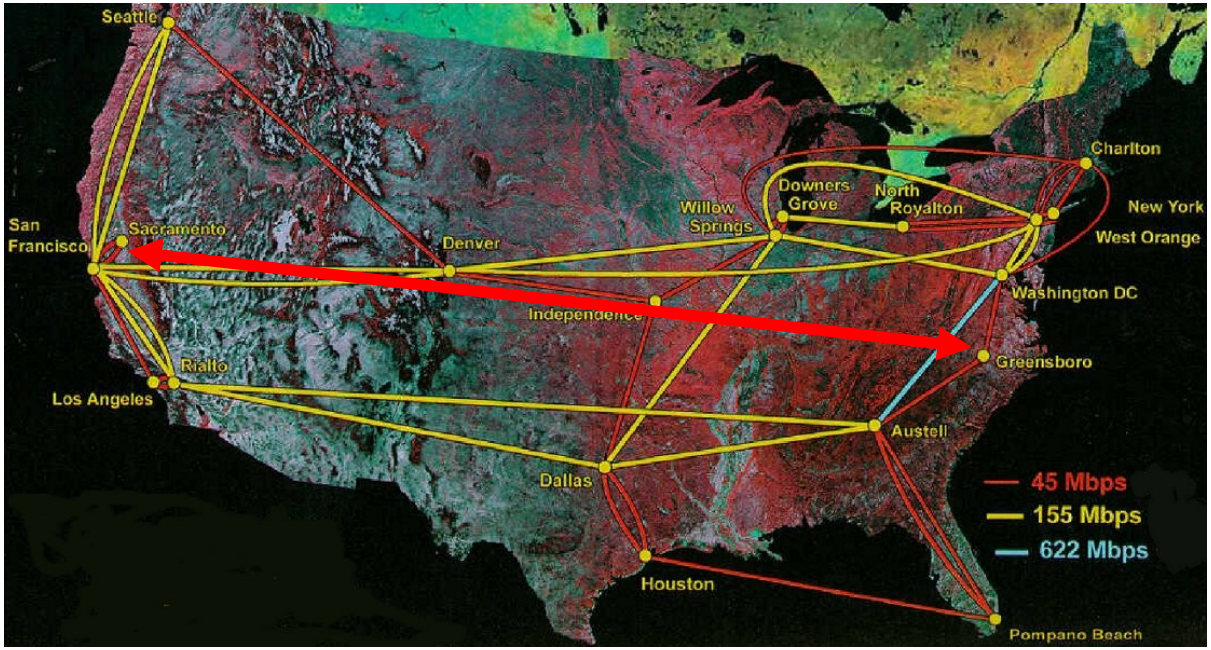
Improving the so-far best link-addition heuristic

- MinDC heuristic adds each of the extra links between the nodes exhibiting the lowest degree
  - The **most effective** approach in terms of connectivity of the enhanced network (for synthetic and real-world Internet maps)
  - However, no constraints posed on the distance of the linked node pairs!

- **MinDC link addition:**

Connect **Sacramento** with **Greensboro!**

- Can we lower that cost and at the same time preserve high connectivity?



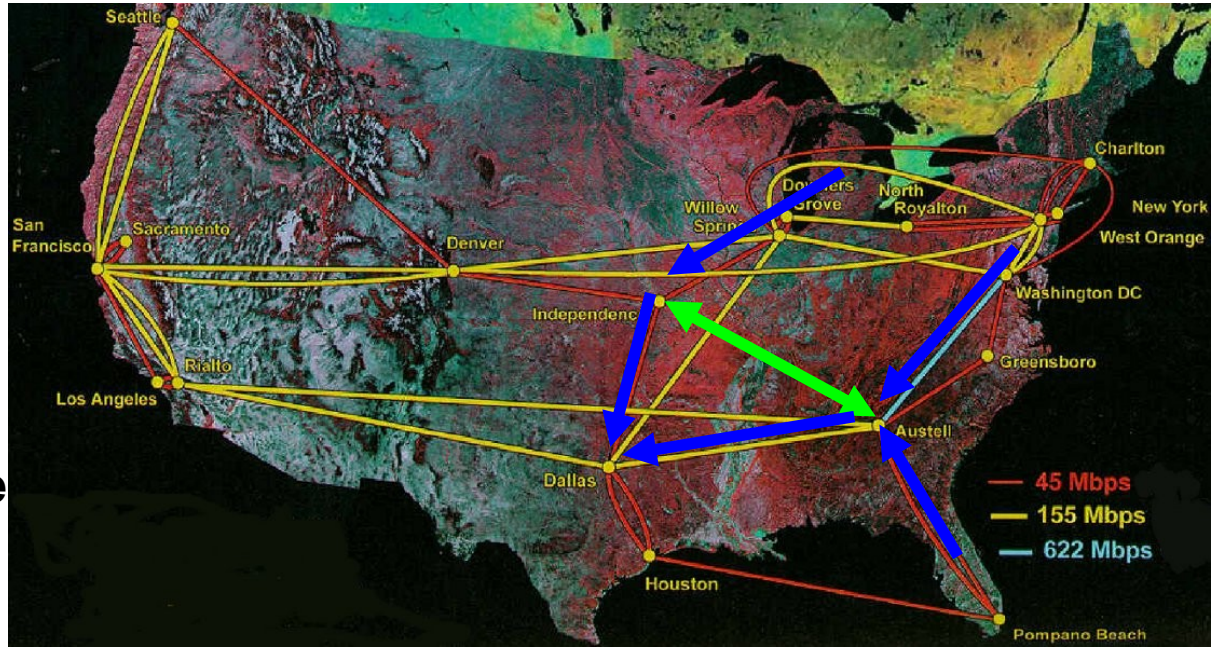
(Revisited) problem statement and relevant intuition

- Device a link-addition heuristic that adds *k* links in the network such as:
  - The connectivity of the enhanced network is improved in face of node attacks
  - The implementation cost associated with the link length be minimized

- **Idea:** Place each link only between the first neighbors of the network hubs (e.g., Dallas)

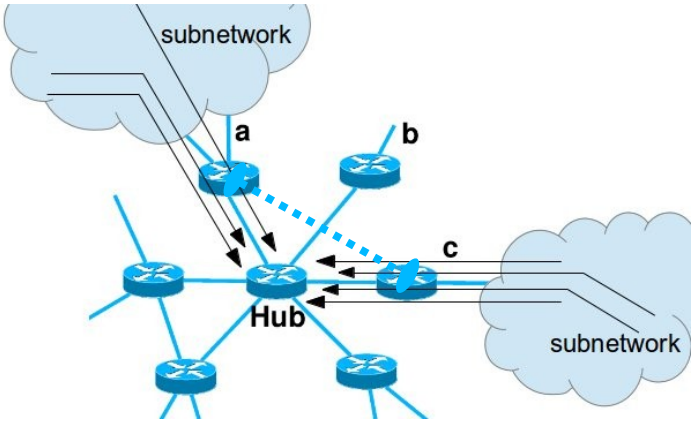
- Identify first neighbors that help establish many communication paths towards the hub

- Thus link **Independence** with **Austell!**



Link-utility metric to select which first neighbors to connect

- Which first neighbors of a hub to connect?
  - Those that aggregate the most shortest paths from the rest of the network



  - Establishing a link between nodes 'a' and 'c' a large number of nodes are expected to remain connected should the Hub be removed

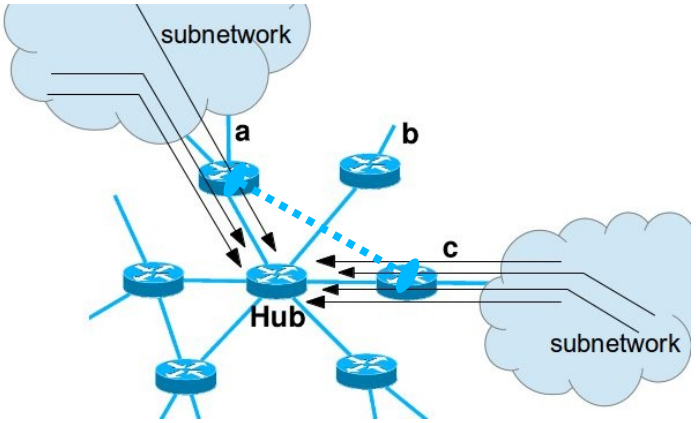
- Conditional betweenness assesses to what extend a node acts as a shortest path aggregator

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

ratio of all shortest paths σ_{st} towards target node t, over those that that pass through node n $\sigma_{st}(n)$

Link-utility metric to select which first neighbors to connect

- The sum of the CBC values of each pair of the first neighbors of a hub is a meaningful measure of the utility of the corresponding link



- Given the H top hubs, how to assign the available *k* links to each hub?

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

Proved to be constant for each network node

As attacks take place over nodes of decreasing degree, higher utility is assigned to relatively high-degree nodes

- Assignment: Use one link to connect each of the *k* top LU node pairs

Intra-domain network topologies

- `mrinfo` topologies (76-1240 nodes, 11 snapshots)
  - Snapshots correspond to Tier-1 and Transit ISPs
  - Collected during 2004-08 using a multi-cast discovering tool

- `Rocketfuel` topologies (41-2515 nodes, 6 snapshots)
  - Widely used in experimental studies
  - 800 vantage points as `traceroute` sources

- Topology-Zoo (27-92 nodes, 5 snapshots)
  - 2011 dataset reported directly by network operators of academic and research networks
  - Geographical coordinates for all their nodes

Experimentation methodology

- Compare the introduced LU-heuristic against the so-far most effective minDC (connects minimum degree nodes *regardless of* their location)

- Enhance each topology adding an extra 5% of its total links
- "Attack level": percentage of total node removed (set to 3%)
    - LU-heuristic: first we set "attack level" = *H* (as if the number of hubs can be estimated), then we relax this assumption

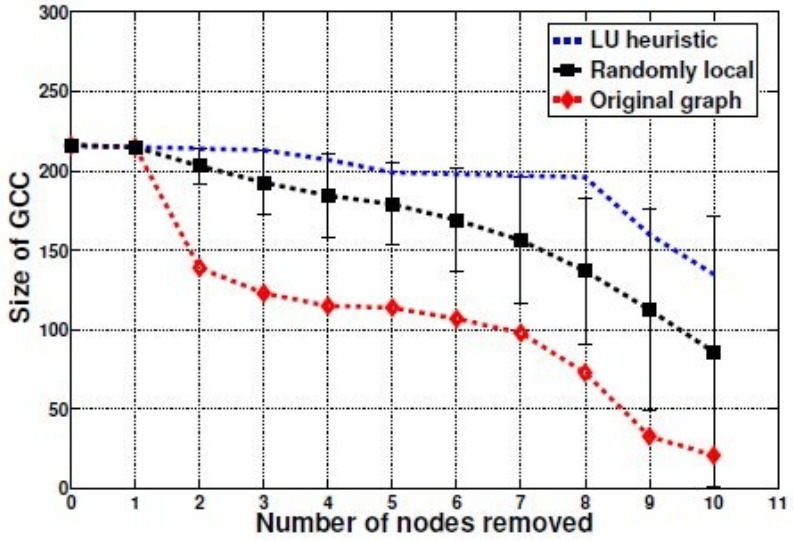
- Study ***connectivity*** of the enhanced topology as node attacks evolve in terms of:
    - Giant connected component (GCC)
    - Number of components
    - Average shortest-path length
- Study ***implementation cost*** in terms of the length of added links

Assumption: cost is proportional to length
Valid yet not always accurate

Linking high-utility first neighbors: proof-of-concept

- Does the LU-heuristic identify the appropriate first neighbors of a hub?
- Comparison with random selection of the first neighbors in terms of connectivity (GCC size)

- `mrinfo` snapshot (216 nodes)
- Attack level ~5%
- Add an extra 5% of the total links

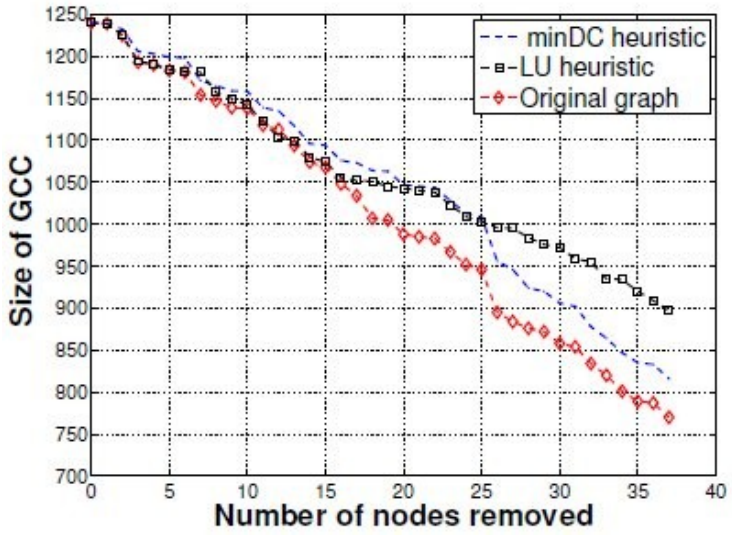


- Results validate:
  - the intuition about the criticality of neighbors establishing many paths towards a hub
  - the effectiveness of the link-utility to capture this notion

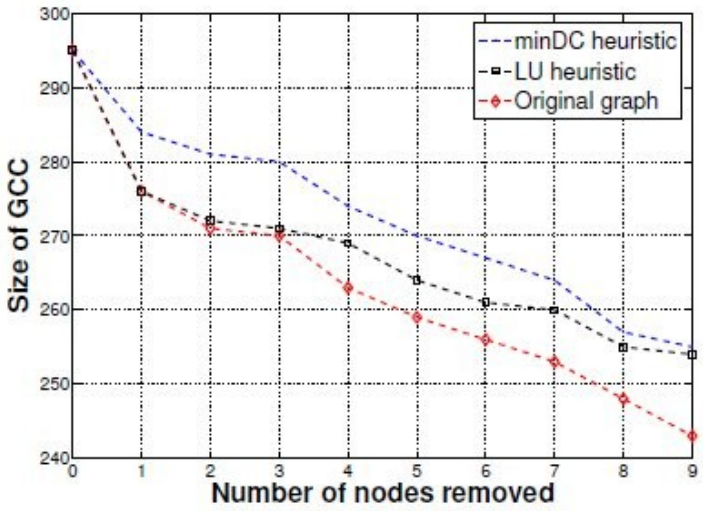
Robustness comparison of enhanced networks (1/3)

- GCC size of the enhanced networks by minDC and LU-heuristic
- The two heuristics perform almost similarly (difference is no larger than 3.5%)
- The original network suffers from rapid fragmentation

`mrinfo` snapshot (ID50, 1240 nodes)



`Rocketfuel` snapshot (ID61, 295 nodes)



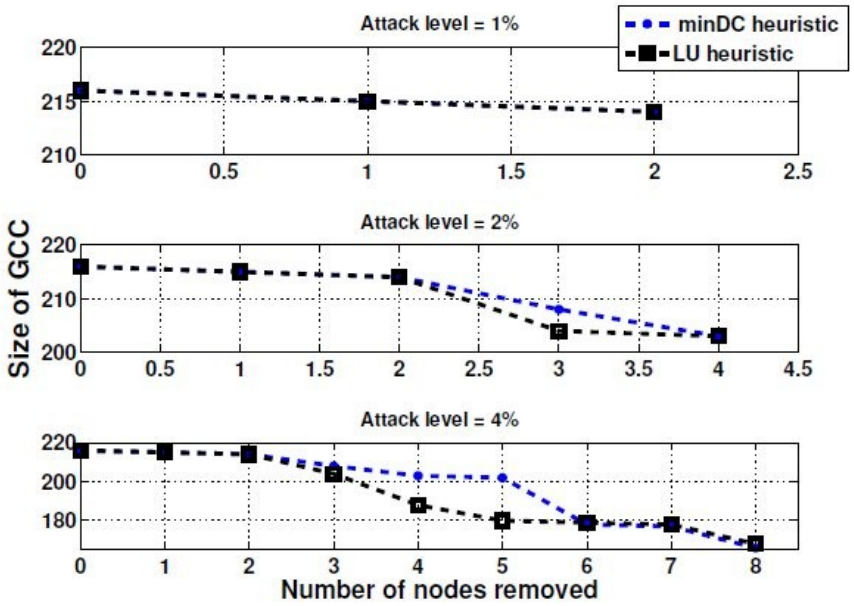
- The LU-heuristic remains effective regardless the network size!

Robustness comparison of enhanced networks (2/3)

- Realistic malicious attacks would target 2 or 3 network hubs

- It takes a higher number of node removals for the GCC differences to become significant

- As the number of removed nodes (*e.g.,* attack level) increases, the number of the considered LU values increases as well



`mrinfo` snapshot (ID21, 216 nodes)

- However the distribution of the *k* links to the H hubs performed by the LU-heuristic, remains effective

Robustness comparison of enhanced networks (3/3)

- General case: parameter *H* is not equal to the attack level, but to the total nodes (*i.e.,* the number of attacked hubs is not known/estimated)

- Compute the relative GCC difference as node *l* is removed

$$\Delta^l = |GCC^l_{minDC} - GCC^l_{LU}| \cdot 100/|V|$$

- For each topology we have $S_{DIF} = \{\Delta^1, .. \Delta^m\}$ with $m = 3\%|V|$

	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

- Link utility uses the ratio DC(*l*)/DCmax to appropriately assign the *k* links

(*i.e.,* few links to small degree nodes)

- Looking closer: heuristics achieve similar connectivity levels for the first few removals

Below 7% across the 10 first removals

How much do the link-addition heuristics cost?

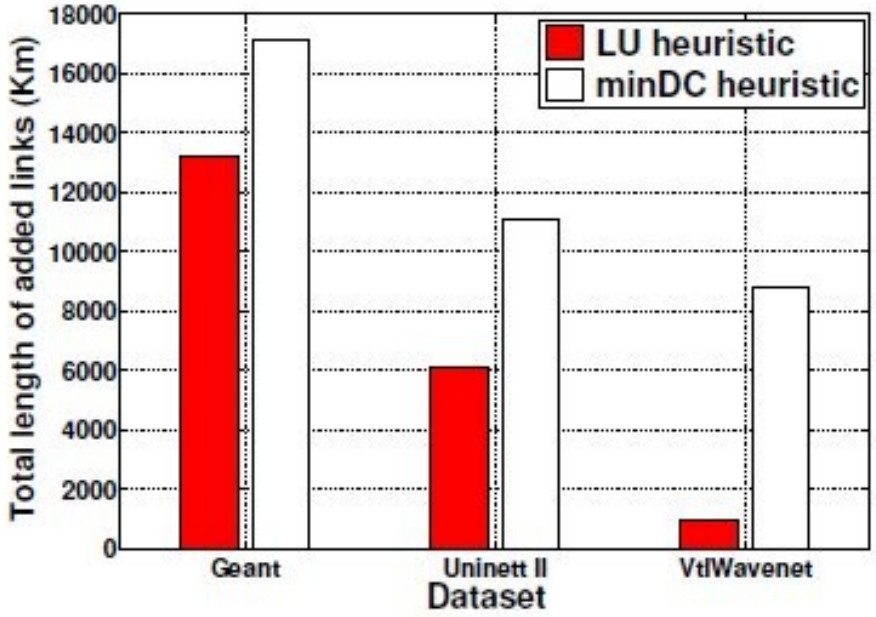
- Topology-Zoo dataset: parse it to retrieve each nodes coordinates
- Distance of node pairs over the globe determined by an online tool*
- Detailed results over the GEANT research network:
  - The minDC heuristic connects nodes of longer distance
  - LU-heuristic yields 1.3 times lower cost than minDC

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

*U.S. National Weather Service (NWS) http://www.nhc.noaa.gov/gccalc.shtml

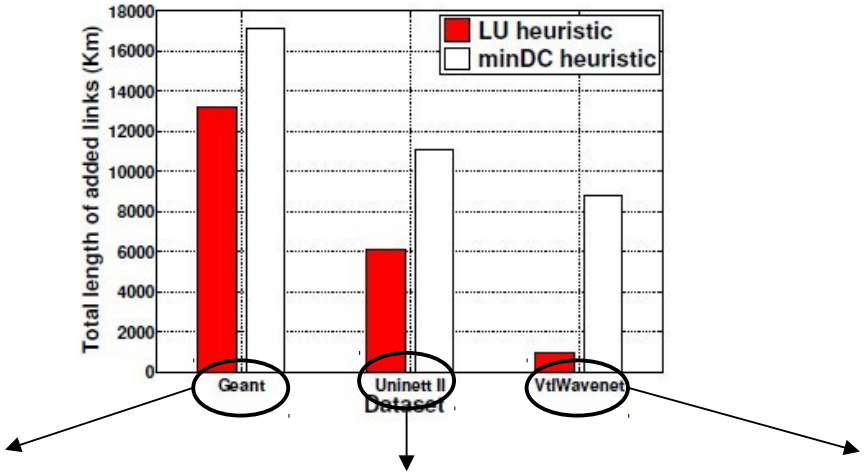
How much do the link-addition heuristics cost?

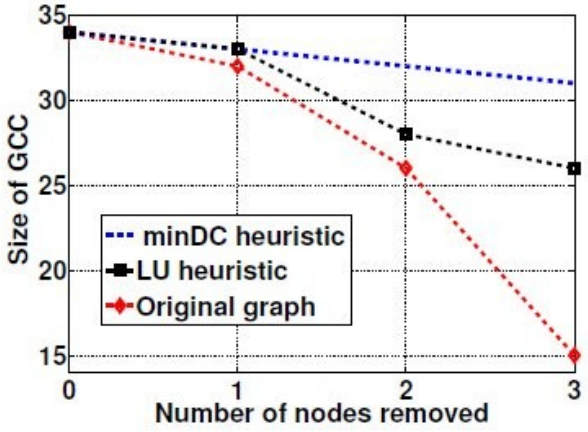
- Similar trend over the rest snapshots: LU-heuristic yields considerably lower cost
- Noteworthy result: 8.8 times lower cost than minDC over the cross-European VtlWavenet network (100 nodes)

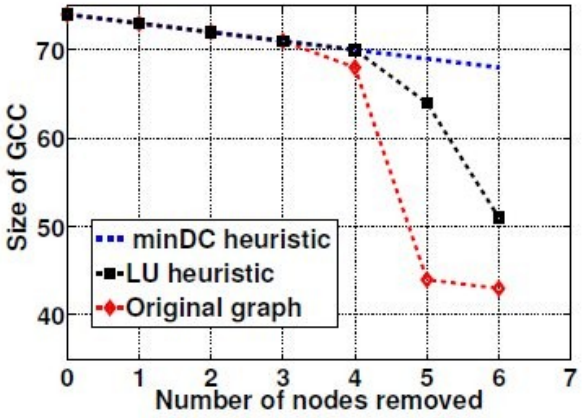


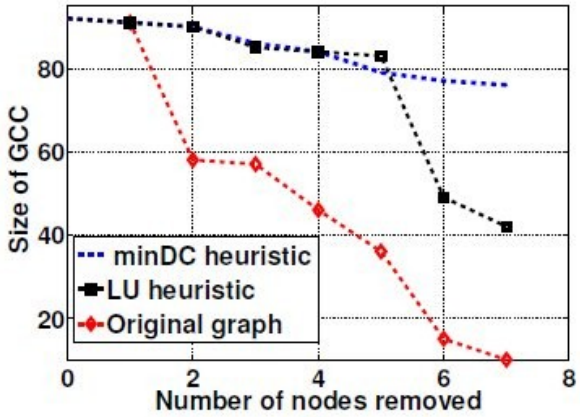
Comparison in terms of cost and connectivity

- The LU-heuristic keeps the network connectivity at (almost) the same GCC level as the minDC









Noteworthy remarks

- Link-utility requires global network information: Is it feasible?
  - We have adopted a network operator's view (*e.g.,* an ISP)
  - Typically an ISP possess global topological knowledge

- What is the CBC computational cost?
  - Offline CBC computations with respect to the H hubs in order not to place extra burden to the link addition
  - $H \cdot O(|E|)$ time complexity, $H \leq |V|$ (length and # of shortest paths from a source to all nodes takes O(|E|) for unweighted graphs)

- Do the presented results realistically reflect Internet robustness?
  - A question of how accurate are the network discovery tools
  - Highest credibility achieved by using three different datasets
  - Our results are “worst-case” due to the underlying hidden redundancy

Take-home results

- We *revisit* the mitigation of Internet hub-node attacks to account for costs : with a budget of *k* links, identify the network node-pairs that their connection
  - Preserves high connectivity levels
  - Yields low cost (in terms of link length)
- Contrary to previous approaches we *by-design* restrict the nodes-to-be-linked to the first neighbors of each hub
- We introduce a *novel* link-utility metric that uses centrality insights to quantifies whether a node pair aggregates many paths towards a hub
- Employing more than 20 real-world Internet topologies we show that the proposed link-addition heuristic:
  - achieves similar connectivity levels to the so-far winner method
  - induces up to 8 times lower cost

Thank you!

Questions ?

Back up slides