Maginot Lines and Tourniquets: On the Defendability of National Cyberspace

Devashish Gosain[1], Madhur Rawat[1], Piyush Kumar Sharma[1], and H.B. Acharya[2]

[1]Indraprastha Institute of Information Technology Delhi (IIITD), New Delhi, India
Email: {devashishg,madhur15030,piyushs}@iiitd.ac.in
[2] Rochester Institute of Technology, NY, USA
Email: acharya@mail.rit.edu

***Abstract*—National governments know the Internet as both a blessing and a headache. On the one hand, it unlocks great economic and strategic opportunity. On the other hand, government, military, or emergency-services become vulnerable to scans (Shodan), attacks (DDoS from botnets like Mirai), etc., when made accessible on the Internet.**

How hard is it for a national government to effectively secure its entire cyberspace? We approach this problem from the view that a coordinated defense involves monitors and access control (firewalls etc.) to inspect traffic entering or leaving the country, as well as internal traffic. In several case studies, we consistently find a natural Line of Defense — a small number of Autonomous Systems (ASes) that intercept most ($> 95\%$) network paths in the country. We conclude that in many countries, the structure of the Internet actually makes it practical to build a nation-scale cordon, to detect and filter cyber attacks.

Keywords: **Cyber high ground, Internet Maps, Cyber defense**

I. Introduction

The Internet, computers and smartphones (which provide access to the Internet), are defining technologies of our generation [15]. As a consequence, practically all enterprises and even government agencies make extensive use of the Internet. As noted by Geer [26], Internet-dependence is transitive: if some entity A uses the Internet, other people have an incentive to use the Internet to interact with A.

Unfortunately, the ubiquity of the Internet also makes it a high-value target. Various kinds of network adversaries — script kiddies, hacktivists, corporate hackers, terrorists, and nation states — monitor, exploit, and take down online resources and services [17]. Some direct attacks take control of the resource itself. For instance, in the recent Russian attack on the Ukraine smart grid [53], the adversary targeted computers that control specific systems. Other attackers simply take over all available computers, to use their computation or communication power; a good example is the Mirai botnet [6].

The fact that governments and important service providers (banks, railways, airlines etc.) make extensive use of the Internet — for example, to schedule passport appointments, buy tickets, or even to vote in elections [51] — they become vulnerable to such attacks.

In response to the threat of online attacks, the security community has developed two lines of defense.

- The first direction, *endpoint* security, focuses on protecting end hosts on the Internet. For instance, the study and prevention of attacks on an online voting machine [21].
- The second direction, *network* security, includes early detection (and prevention) of attacks, such as denial-of-service, spoofing, man-in-the-middle etc. from network traffic.

Additionally, of course, there is a need to educate the user about security to guard against social engineering, phishing, and similar attacks.

Network security has historically been the first line of defense for many critical systems. This is because, Internet services make use of many kinds of systems (with various versions of software, patched and unpatched, various configurations, etc.); as a result, it is not feasible to expect *complete* endpoint security. Further, some attacks (such as DDoS) may not be defended by endpoint security; in these cases network security plays an important role. Thus, there has been considerable research in developing tools to monitor and filter traffic, ranging from flow tools [25] firewalls [13] and proxy servers [10], through signature-based [47] and event-based [12] Intrusion Detection Systems, up to complete Security Incident and Event Management tools [38].

However, the development of security solutions is not enough — in order to detect large scale infestations or attacks by botnets (e.g., Mirai [6]), it is necessary to *place* these solutions where they are effective. For instance, where would the defenders place the IDS devices? Where would firewalls be most effective? These questions are routinely asked by network administrators in the context of small networks, e.g., enterprise Local Area Networks.

The question of placement takes on a different scale when we consider the Internet as a whole, or in a country. Usual techniques to reduce the volume of data (looking at logs and flows rather than complete packet captures) are no longer sufficient; we need to make sense of "oceans of data". Thus, it becomes imperative to identify strategic locations in the network, where some sophisticated devices can be positioned (*e.g.,* NIDS) to analyze the country scale traffic. This brings us to our high-level question "how hard is it for a defender (nation-state) to secure its cyberspace against an adversary?"

We assume the defender has the capacity to install defences

(firewalls, NIDS, etc.) in the network; the problem is thus reduced to *intelligently* positioning these defences. To that end, we note that the Internet is known to have a scale-free network structure. A relatively small number of heavy-hitter Autonomous Systems (ASes) form an *Internet backbone*[1].

Moreover, it is already known that a few nation sates take *advantage* of these small number of heavy-hitter ASes, for instance, China implements stricter form of control through these ASes [59, 14, 45]. Whereas, others like Iran intentionally routes its Internet traffic through a centralized choke point [7, 45].

Thus, in this paper, we explore the question of whether the cyberspace of specific countries also has such heavy-hitters, where a well-placed defender can make a big difference. Our intuition is that such heavy hitters would form two groups:

- **Maginot Line:** the boundary ASes of national cyberspace, where Internet traffic passes to and from the country to external users or resources.
- **Tourniquet:** a small cutset of important ASes, which cover the paths internal to a country.

With several case studies, we study whether it is in fact possible to construct a small set of heavy hitters (Maginot Line $\cup$ Tourniquet) for National Cyberspace (refer figure 1). If such is the case, we suggest that the government make use of these ASes as a *Line of Defense* to deploy middleboxes, monitors, and other infrastructure.

Fig. 1: Sample Representation: Nodes in blue represent border ASes — a few of them would constitute Maginot Lines; and nodes in red represent internal ASes — a few of them would constitute Tourniquets.

We begin by *constructing* the nation-level Internet map, consisting of ASes (nodes) and their *business relationships* [23] (edges), using the BGP simulator C-BGP [44]. We also *annotate* the nodes with node weights, i.e., AS characteristics (or properties) [56] such as AS degree or cone size, which help define the strategic importance of the AS.

In our next step, we use the Ford-Fulkerson [57] method to determine min-cuts, with different source-sink pairs, and report a cut of the graph using the top-ranked nodes (i.e., we report a minimal set of top nodes, by AS degree, cone size, and betweenness, which cover all or almost all Internet paths in the country).

In practice, country-sized Internet graphs are very large (e.g., India's AS graph has $\approx$ 1200 nodes), so this direct approach does not scale to the data we require [57]. We therefore made our approach more tractable, using the technique of Vertex Splitting, as well as additional heuristics (details in Subsection III-C).

This approach allows us to find the "Line of Defense" (which, we expect, will consist of a "Maginot Line" and a "Tourniquet") for each country. Our objective is to check whether *the Line of Defense, for a country, is relatively small* (compared to the total size of the network, i.e., all ASes in the country). In cases where the Line of Defense is large, we also check if we can solve easier versions of the coverage problem: inspecting flows to the important sites (government, banks, emergency services etc.) in a country, allowing partial (say 95%) coverage instead of 100%, and so on.

We begin by discussing background and related work in the next section, then detail our approach and algorithm, followed by our experimental results. We then discuss our findings, concerns, limitations, and end with some concluding remarks.

II. Background and Related Work

In this section, we begin our exploration of lines of defense in cyberspace with a discussion of the adversary, literature review on DDoS attacks and of the network mapping techniques which we use to build Internet maps.

A. Adversary: Distributed Attacker

Our adversary is a distributed attacker, who aims to disrupt the availability of important resources, typically through a Distributed Denial-of-Service attack. In recent years, such an adversary typically makes use of a botnet [24, 36].

Botnets, i.e., large remote-controlled collections of compromised devices, have been deeply studied; a wide range of host-based as well as network-based tools have been developed to detect and neutralize them. The first step for such tools is to gather data: this involves stateless and stateful header inspection, Deep Packet Inspection (DPI), and sometimes Deep Content Inspection (reassembly of files from packets). Single bots may be detected using standard signature-based tools such as Snort NIDS [48], if there is a specific pattern of command-and-control data. More specialized tools detect the group behavior of botnets, or sometimes, the behavior of a single bot. For instance,

- BotSniffer [30], looks for traffic patterns: commands from the botnet command server.
- BotMiner [29], detects clusters of communication that resemble malicious activity.
- BotTracer [39] uses virtualization to detect the three steps in the life of a bot: automatic startup (without user intervention), command channel establishment, and attack.

These lines of research focus on the question of how to identify a botnet from its traffic; they are, therefore, orthogonal to our

[1]There are more than $60,000$ ASes, but with the cooperation of only the top 30 ("heavy-hitter") ASes, an authority gets access to over 92.5% of traffic paths in the Internet [5].

research question, *i.e., where to place the sensors and firewalls to see this traffic*.

The closest related work to this paper is the study of "high ground in cyberspace" [46, 8, 54]. Sweeny et al. [54] demonstrate that an attacker who controls this cyber high ground (*e.g.,* nodes in a network that capture a large fraction of traffic), gains a superior capability to achieve his malign objectives. They assume that the defenders are randomly placed in the network topology, and an adversary could bypass them by tactically placing the bots by exploiting the topology knowledge. We approach this concept from the defenders' point of view: they occupy "high ground" so the attacker cannot sneak past them.

B. Adversary: DDoS Attacks

We attempt to identify strategic locations on the Internet (*e.g.,* a small set of ASes), where a national agency could deploy NIDS and other DDoS mitigation strategies, to safeguard the critical infrastructure of the country. While this general approach would help to defend against several attackers — detecting Internet wide scans (Shodan etc.), or botnet activity in general — we anticipate that our most common use case will be to detect and mitigate Denial-of-Service attacks, for example, to take down government websites.

A distributed denial-of-service (DDoS) attack occurs when the attacker exhausts the bandwidth or resources of a target system, by sending attack traffic from many different hosts on the Internet (sometimes as many as 500,000 hosts [37]).

Evidence of DDoS dates back to 1996, where the Panix ISP was subjected to a SYN flood attack that disrupted their operations for several days. More recent attacks like Crossfire [34] and Coremelt [52], instead of directly flooding victims, exhaust the backbone links of ISPs. (These attacks used a large number of low-rate and short-lived benign traffic flows to saturate the links that connect victims to the Internet.) Sadly, the impact of DDoS attacks have increased even faster than the ability of communication networks. For instance, botnets like Mirai [37], generated attack traffic peaking at the rate of 1.1 Tbps. As may be expected, given its long history and many forms over the years, DDoS attacks have been extensively studied, from several different perspectives.

- One standard approach is to focus on attacks through the lens of a specific *protocol or tool*. For instance, Czyz et al. [16] characterized DDoS attacks using Network Time Protocol, and concluded from the attacked port numbers that a large fraction of NTP DDoS attacks target gamers.
- Another way to approach such attacks is to start with a single target *vulnerability*: for example, Durumeric et al. [19] analyzed the scanning behavior triggered by vulnerabilities in OpenSSL, NTP and some routers. The two approaches may be combined, for example by Rossow [49], who examined UDP based network protocols, and identified those that are susceptible to amplification attacks.
- A third approach is to directly study the adversary, i.e. the agents seen in actual attacks. Chang et al. [11] present a botnet measurement study based on public data, and analyze the attacking capabilities of different types of botnets. (Among other findings, they note that some bots are employed by many different botnets.)

In response to the threat of such attacks, a range of prevention measures have also been proposed.

- Keromytis *et al.* [35] proposed a secure overlay service to proactively prevent DDoS attacks. Only authenticated users can use the overlay network to reach the protected target. Similarly, Francois et al. [22] proposed a system to detect flooding DDoS attacks within the ISP (closer to the attacker, but far away from the victim); in their distributed architecture, multiple IPSs form an overlay network to protect the subscribed customers.
- Giotsas et al. [27] developed a methodology to detect BGP *blackholing*, an approach to restrict the reachability of selected targets on the Internet; while this may be considered an abuse of the BGP protocol, blackholing is an effective technique to shield hosts when under a DDoS attack. They reported that, between 2014-2017, blackholed prefixes increased by a factor of six over 400 different ASes; if this is in fact a response to DDoS, it would indicate that attacks are on the rise.
- Passive DNS analysis techniques can detect specific domains that are involved in malicious activity (for instance, command servers for botnets, large websites known to be vulnerable to amplification attacks, etc.), as suggested by Bilge et al. [9].

The above is a brief overview; more detail can be found in recent surveys such as the work of Kalkan et al. and others [33, 60, 43]. We note that our work is, once again, complementary to such approaches; in identifying a small set of heavy-hitter ASes, we make it more likely that attacks will be detected and blocked, before reaching the target. We intend to explore whether alternate approaches (e.g. BGP blackholing) are also more effective when implemented by our "Line of Defense" ASes, in future work.

C. Techniques in Internet Cartography

We propose to use Internet mapping techniques, to identify the best places to put *defenders* against attackers such as botnets. Our work depends on finding the paths to a particular destination taken by Internet traffic; in this subsection, we explain our approach.

The Internet is a network of networks: it consists of independent entities called Autonomous Systems (ASes), which are themselves networks of devices called as routers and end hosts. ASes operate independently, but collaborate to route traffic among themselves. ASes can be customers, peers, or providers to other ASes; besides a physical connection, there must be an acceptable business relationship between two ASes, before they route traffic through each other[2].

[2]A customer AS routes traffic through its providers; but providers do not route transit traffic through their customers. The only traffic a provider sends a customer, is meant for that customer, or *its* customers, and so on [23].

Most existing projects, such as CAIDA Ark [1] and *iPlane* [41], map the Internet at router level, with the tool `traceroute`. Traceroute returns the IP addresses of each hop along the path from a source to a destination; a complete map can be built by running traceroute from distributed volunteer nodes to various targets. If we then map the IP addresses (returned by Traceroute) to AS numbers, the router level map can be "zoomed out" into an AS-level graph. However, such maps are incomplete owing to the limited network locations and availability of volunteer nodes. They may not provide the AS-level path between any two randomly chosen ASes, and even where they do, they may be inaccurate [42, 40].

In this paper, we used a different approach, generating country-specific BGP maps using the C-BGP simulator [44]. C-BGP takes as input AS relationships (provider-customer and peer-peer links) and IP prefix – to – AS mapping information, which we obtain from CAIDA [2] and the CIDR report [3], respectively. C-BGP runs actual BGP sessions (the various ASes in the model advertise the IP prefixes they host, and paths to their neighbors, just as real ASes do in the Internet). This allows us to build up a complete map of AS-level paths (as already demonstrated by Houmansadr et al. [31]).

III. Approach: The Line of Defense

A. Problem Definition

Our aim in this paper is to identify the cyber defense line of a country: a (hopefully small) collection of strategically important ASes that intercept all network paths leading to important destinations of the country.

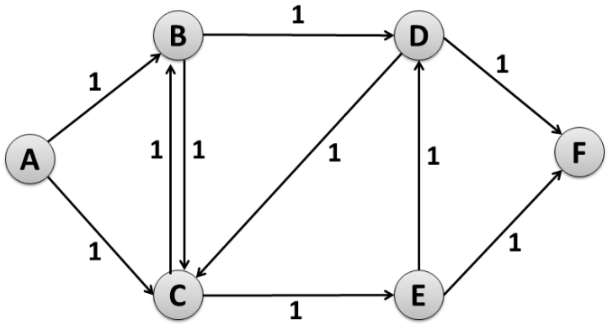


Fig. 2: Sample AS topology: Node A is source and Node F is target.

Problem: Our objective is to find a (graph) cut of minimum size, in the network of a given country, such that all communication paths (i.e., paths between all source (s) – target (t) pairs) transit this minimum cut.

A high-level overview of our approach is as follows.

1) We collect information about the Internet structure in the nation under study (consisting of AS relationships, and what ASes host what IP Prefixes).
2) Using C-BGP, we reconstruct the network (i.e., build a graph of the AS-level routes).
3) The Line of Defense is identified as a minimal cutset of the AS graph, which captures 100% of the traces. (In this Line of Defense, we call the border ASes the Maginot Line, while the inner ASes form the Tourniquet.)

B. Algorithm: Finding a Cut-Set

We construct an AS-level map for a country using the popular mapping tool, C-BGP. Its AS level traceroute[3] reports all-to-all paths between ASes is the country; the union of these paths gives the complete AS-level topology of the country.

Our next step is to find the *minimum cut* of the graph. The Ford-Fulkerson method gives the cut for one source-sink pair. The union of all such min-cuts is the cut we want. Our algorithm is as follows.

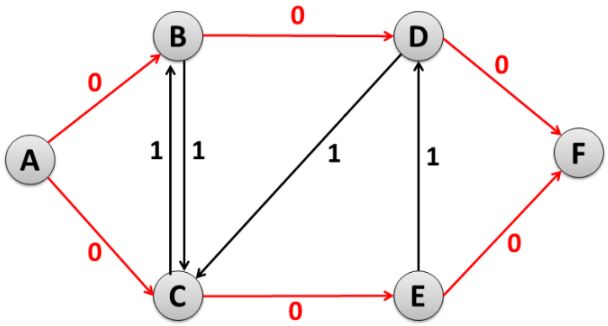


Fig. 3: Residual AS topology: Paths A-B-D-F and A-C-E-F are residual paths.

1) Construct the AS-level map of National Cyberspace, using data from CAIDA and CIDR report.
2) Annotate the map: set the weights of all edges (in the AS-level graph) to 1. (Figure 2.)
3) For every $s-t$ *pair*:
   a) Find the *residual graph* of the AS topology, using Ford-Fulkerson. Figure 3 represents the *residual graph* of A-F as $s-t$ pair.
   b) Find the *zero-weight paths* from source s to sink t in the residual graph.
      - Starting with source s, we use BFS/DFS along edges of zero weight, to find zero-weight paths to the sink node t.
      - The number of such paths in the residual graph is the *cardinality* of the cut set (for residual graph in figure 3, A-B-D-F and A-C-E-F are such two paths; thus cut size is two).
   c) Find all possible cuts in the graph. (In figure 3, for the given $s-t$ pair the possible cuts were {B,C}, {B,E}, {D,C}, {D,E}).
   d) Out of these cuts, select the best cut, i.e., the one that maximizes *value metric*.

[3]AS level traceroute is an inbuilt utility provided by C-BGP simulator. It finds the AS level path between a source and destination AS, based on the BGP sessions generated in the simulator itself.

4) The union of cuts for all $s-t$ pairs in the graph, is the complete cut of the graph.

The question immediately follows as to what we mean by the *best* cut, i.e., what the value metric is that we aim to maximize. We choose five characteristics [56] of an AS which intuitively define its strategic importance:

1) Provider Degree : number of providers an AS has.
2) Customer Degree: number of customers an AS has.
3) Peer Degree: number of peers an AS has.
4) Cone Size: size of the set consisting of the AS, its customers, customers of customers etc.
5) Betweenness: a graph-theoretic measure that quantifies the centrality of a vertex in terms of its involvement in connecting pairs of vertices in a graph.

The value metric of a cut is the sum of the value metrics of the nodes in the cut.

C. Vertex Splitting: a Practical Approach

The approach above, requires that we inspect *all* cuts to find the best cut. Unfortunately, the number of cuts grows exponentially with the number of paths. The reason is that a cut should include one node from each zero-weight path of the residual graph; the number of choices for a node is the length of the path (not counting the source and target), and this choice has to be made once for each path. In our example, the paths are ABDF and ACEF; the choices are $(B,D) \times (C,E) = \{B,C\},\{B,E\},\{D,C\},\{D,E\}$.

In practice, even a small country has a much larger graph than our example. For instance, with Israel (214 ASes), one $s-t$ pair yields 30 residual paths (of varying path length ref. Table I), but $7^2 \times 6^5 \times 5^8 \times 4^{10} \times 3^5 = 3.79 \times 10^{19}$ cuts.

Path Length	No. of Paths
7	2
6	5
5	8
4	10
3	5

TABLE I: Paths: Residual graph from one $s-t$ pair

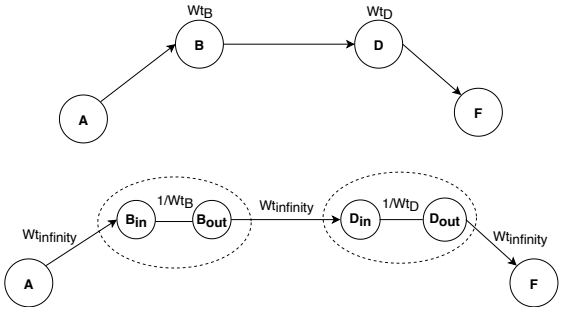


Fig. 4: Vertex Splitting

In order to make the problem (i.e., identifying the min-cut) computationally tractable, we reduced the number of zero edges. Rather than setting all edges to 1, we transformed node weights to edge weights, using the *vertex splitting* transformation, as follows.

1) Construct the AS-level map of National Cyberspace, using data from CAIDA and CIDR report.
2) Annotate the map: set the weights of all edges (in the AS-level graph) to ∞, and set the node weights to the desired metric (cone size, etc.)
3) Split each node into two – one with all the incoming edges, and one with all the outgoing edges. Connect these two nodes with an edge annotated with the reciprocal of the weight of the original node (as shown in Figure 4).
4) For every $s-t$ pair:
   a) Find the *residual graph* of the AS topology, using Ford-Fulkerson.
   b) Find the *zero-weight paths* from source s to sink t in the residual graph.
   c) Find the cut for the $s-t$ pair.
5) The union of cuts for all $s-t$ pairs in the graph, is the complete cut of the graph.
6) To minimize the cut we adopt the following heuristic
   a) Sort the cut nodes by descending frequency of occurrence in cuts of different $s-t$ pairs.
   b) Greedily select the top $1, 2...$ nodes until the selected nodes touch all the paths.

The above algorithm succeeds in reducing the number of zero edges in the residual graph, and therefore step 4(c) usually finds a single path, not a combinatorial explosion of paths. But it raises two questions. Why the *reciprocal* of the node weight? And why do we not directly choose the nodes with largest weight, as we did in the previous algorithm?

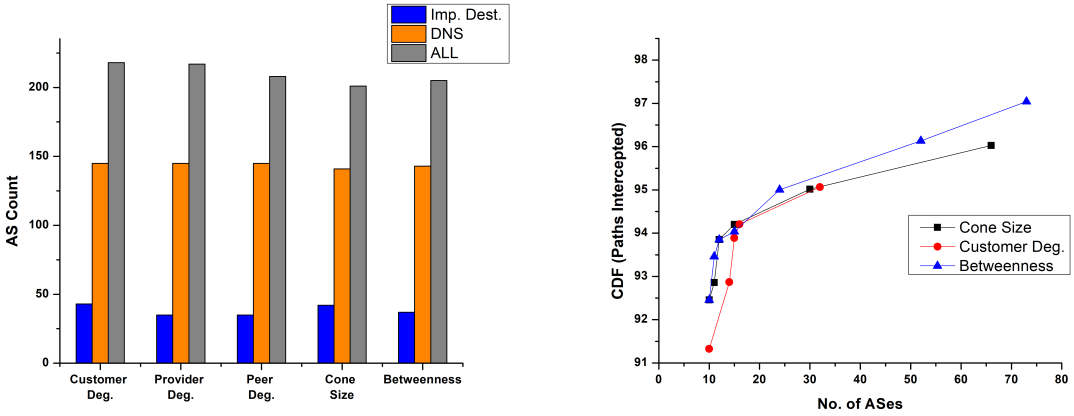
The answer is that, as Ford-Fulkerson is a *min-cut* algorithm, it assigns the smallest-weight edges of the original graph to zero weight edges in the residual graph i.e., the nodes with the smallest values get chosen in the cut.

But, we would like to choose a small number of nodes, with *high values* for our heuristics (cone size, betweenness etc.). Thus, we compensate by setting the reciprocal of these values as the edge weight: now the first edges to go to zero, correspond to high values of the heuristics. As we wish to choose the nodes with high values for the metric itself (not the ones with high value for the reciprocal!), we now have a separate step where we greedily choose a small set of nodes that cover all paths.

IV. Results

This section presents some case studies of nations, and the *"Line of Defense"* we find for each: a small set of strategic ASes, located in those nations[4], that can monitor, shape, and filter all traffic.

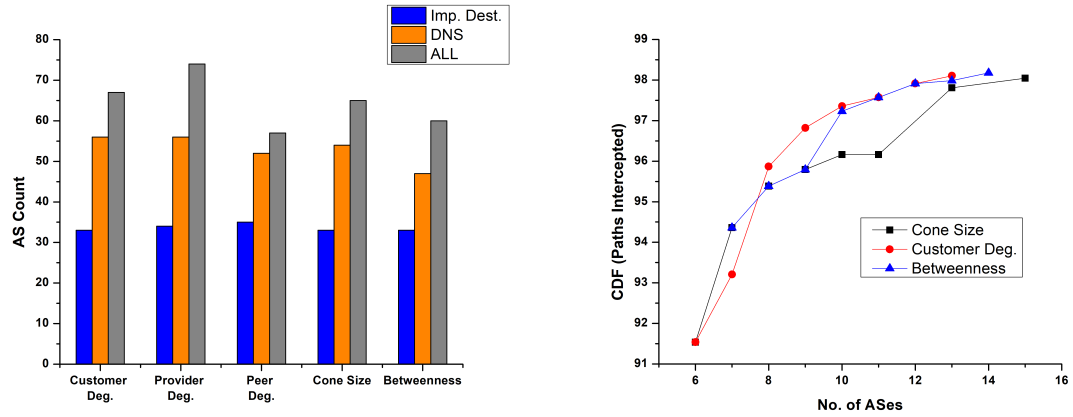
[4]We assume that the country in which an AS is registered, is the country it is actually located in.



(a) Line of Defense

(b) Impact of Heavy Hitters

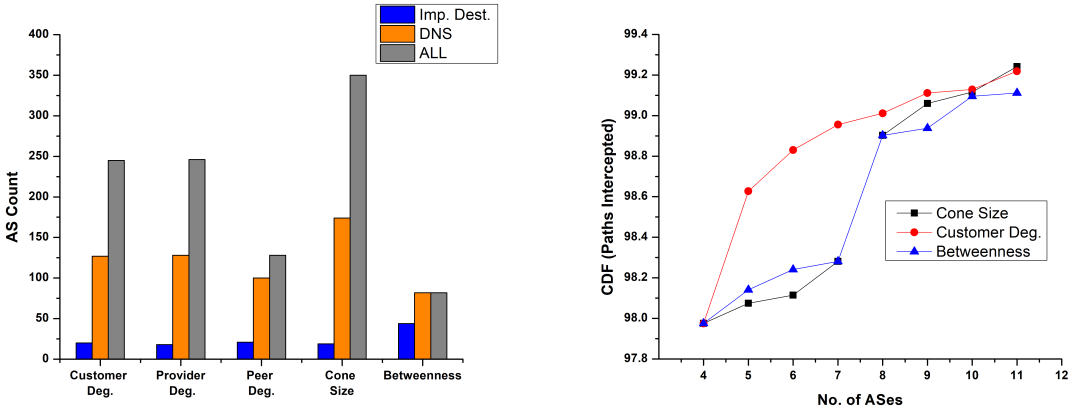
Fig. 5: China: Network Size 345 ASes; Total Border ASes 97.
All heavy hitters in all countries are border ASes.



(a) Line of Defense

(b) Impact of Heavy Hitters

Fig. 6: Vietnam: Network Size 205 ASes; Total Border ASes 18.



(a) Line of Defense

(b) Impact of Heavy Hitters

Fig. 7: South Korea: Network Size 711 ASes; Total Border ASes 37.

We consider eight countries. We mainly selected countries in small, antagonistic groups, i.e., neighbors with tensions, as the relative difficulty of securing their cyberspace would be of interest in case of cyber-warfare.

- China, Vietnam, and South Korea.
- India, China, and Pakistan.
- Israel, Iran, and Egypt.

For each country, we find the cut-set of heavy-hitting ASes (as mentioned in Section III-A), and note the border ASes of the country (Section V). These ASes constitute the defense line when the adversary attacks from internal and external ASes, respectively. We consider three cases, with respect to the location of the target.

- Any ASes in the country can be the target.
- The adversary targets specific high-value ASes, i.e., ASes hosting important websites. (We treat the important Government, Banking, and Transport websites in each country as high-value targets.)
- The adversary targets the Domain Name Service, i.e., attacks ASes that host DNS resolvers. (We obtained all open DNS servers in the country using the Censys project [18, 20].)

Area I (South China Sea Zone):
Figures 5(a), 6(a) and 7(a) represent the Line of Defense (i.e., min-cuts) for China, Vietnam and Korea respectively. Each figure represents five Lines of Defense corresponding to the five AS characteristics (ref. Section III-B) viz., customer degree, provider degree, peer degree, cone size and betweenness of an AS. Further, corresponding to each AS characteristic, the three bars represent min-cut size for three different sets of targets — DNS resolvers, ASes hosting important destinations and *all* ASes of the country.

It is evident from the figures that, South Korea and Vietnam require roughly one-third (35% and 33.8%) of the total ASes as (min-cuts) when target was all ASes of the country (e.g., for Vietnam, Line of Defense (based on cone size) consists of about 65 ASes, whereas total ASes in Vietnam are 206). However, China requires roughly two-thirds (63.7%) of the total Chinese ASes as Line of Defense.

Further, if we relax the requirement of 100% *coverage, we find that a small fraction (*$< 10\%$*) of ASes intercept a large fraction (*$> 90\%$*) of network paths.*

Our tests resulted in two interesting observations:

1) All "heavy-hitter" ASes in the cut-set are also border ASes (ref. Figures 5(b), 6(b) and 7(b)). It is evident from the figures that, a few border ASes, cumulatively capture $> 90\%$ of the total AS level paths, within a country. For instance, in South Korea, four ASes capture $> 95\%$ of the AS level paths (ref. Figure 7(b)).
2) There is no meaningful difference in the trends, as we vary the AS characteristic used to choose strategic ASes. (The only outlier is South Korea, where cone size performs poorly and results in a large number of ASes.)

Area II (Kashmir Zone):
Among India, Pakistan and China (ref. Figures 8(a), 9(a) and 10(a)), India requires the smallest fraction ($< 35\%$ of ASes) as a Line of Defense; it is a comparatively high 51.2% for Pakistan and 63.7% for China. However, the actual number of ASes needed for, India to build a Line of Defense with 100% coverage is more than for China and Pakistan; India requires > 300 ASes, whereas Pakistan requires about 40 and China requires around 200. This is likely an artifact of the very large number of ASes in India. In comparison, China has much better developed Internet within the country [55] but fewer ASes (this may indicate the existence of an official policy to control the growth of networks in the country).

All three countries require < 10 nodes to intercept more than 90% of the paths, and as Figures 8(b), 9(b), and 10(b) show, our "interesting observations" from Area I hold for Area II as well.

Area III (Palestine Zone):
In the final set consisting of Iran, Egypt and Israel (ref. Figures 11(a), 12(a) and 13(a)), Egypt and Israel, which have small networks, require — 16.36% and 18.6% of the total ASes respectively for a Line of Defense. These may be the only countries where we would say that 100% path coverage is realistically possible (the total number of ASes to cover 100% paths, are relatively lower than for other countries: 40 for Israel, and only 9 for Egypt). The figure is comparatively higher for Iran, at 39.2%, i.e., 160 ASes.

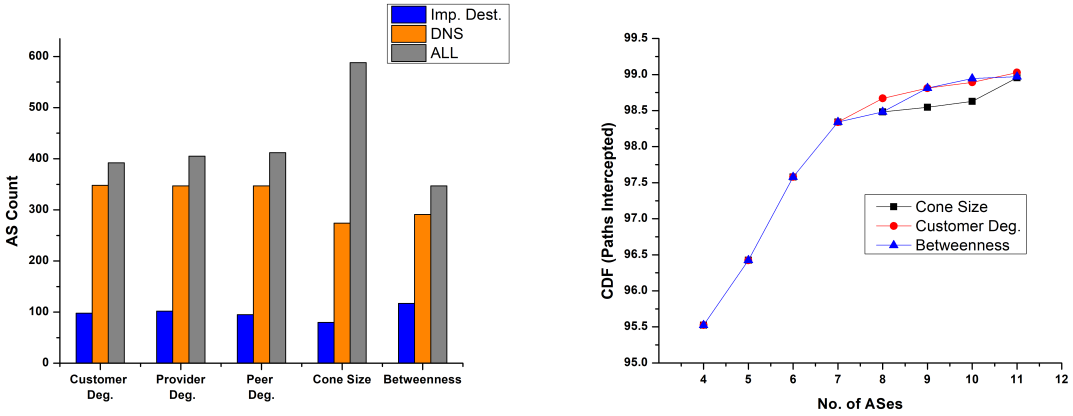
As Figures 11(b), 12(b) and 13(b) show, the phenomenon of heavy-hitter ASes at the border, is quite general. Among our heuristics, Israel yields another data point that cone size is a poor metric, but we cannot see a trend in general.

General Trends (across our case studies):
Our experiments show that countries differ significantly in their network size as well as topologies; both these factors affect the Line of Defense. For instance, for complete coverage Pakistan's Line of Defense requires 45 ASes (out of a total of 87). Egypt requires only 9 ASes (out of 55). This would suggest that cyberspace in Egypt has a more hierarchical structure, while in Pakistan it is relatively flat. However, two main features remain common across countries:

- The Line of Defense required to cover paths to *all* ASes, is quite large (a considerable fraction of the complete network). The set required to cover paths to DNS resolvers is smaller, and to cover paths to important websites, smaller still; but *in all cases,* 100% *coverage is costly*.
- Border ASes are very powerful: they see not only external paths, but almost all internal paths. $> 95\%$ *coverage of paths can be obtained using border ASes alone*.

V. Insights and Observations

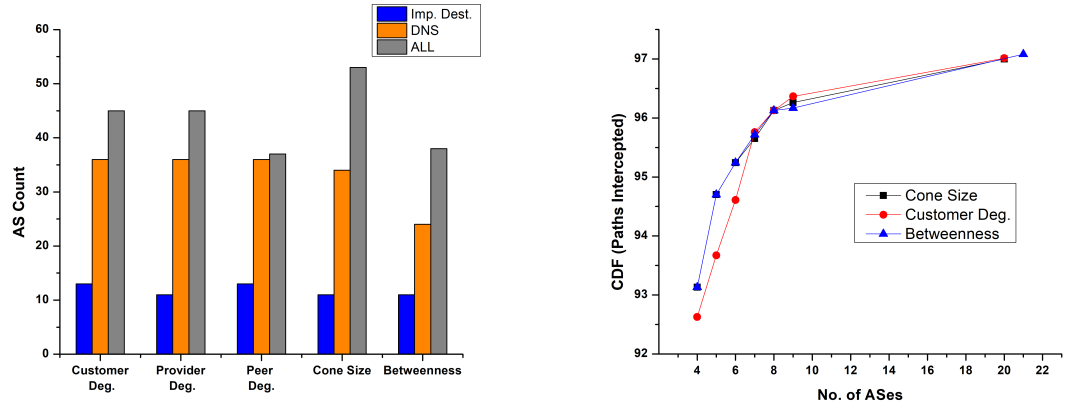
Our original hypothesis in this project was that a small set of autonomous systems, in a country, would form a Line of Defense. Such a set would consist of some border ASes (to protect against attacks from end points outside the country) and some additional cut-set ASes (for adversary end points



(a) Line of Defense

(b) Impact of Heavy Hitters

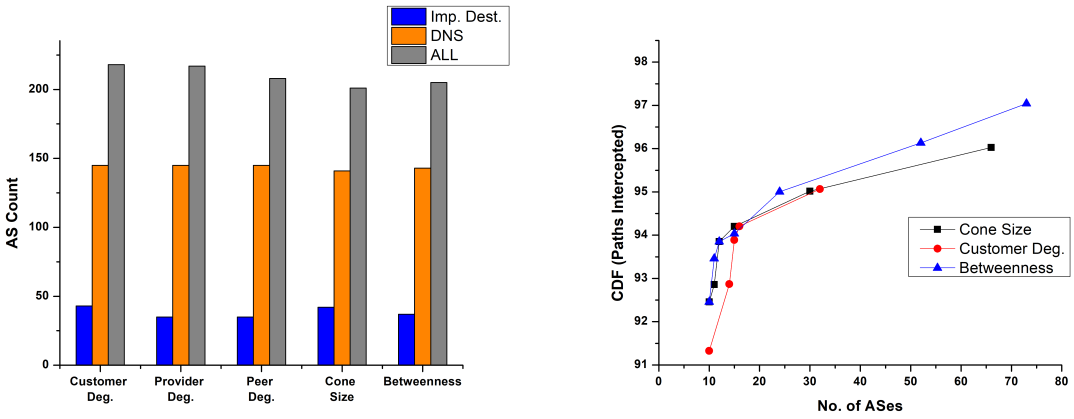
Fig. 8: India: Network Size 1176 ASes; Total Border ASes 112.



(a) Line of Defense

(b) Impact of Heavy Hitters

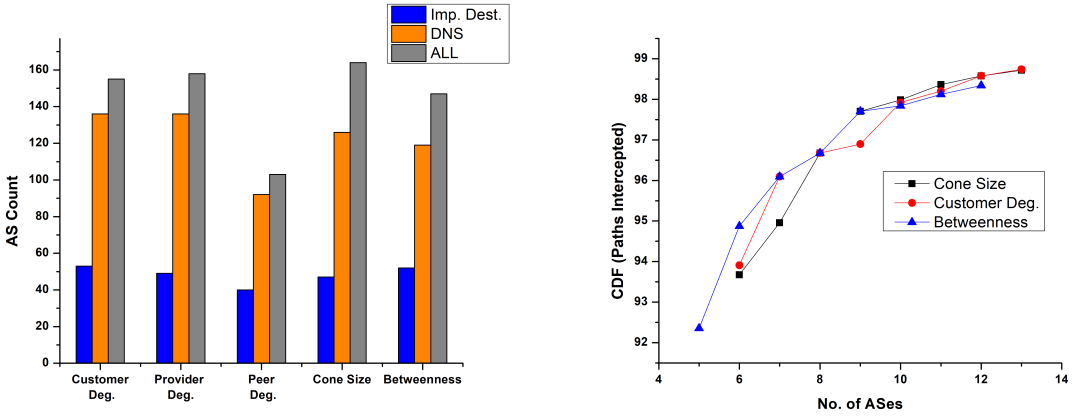
Fig. 9: Pakistan: Network Size 87 ASes; Total Border ASes 17.



(a) Line of Defense

(b) Impact of Heavy Hitters

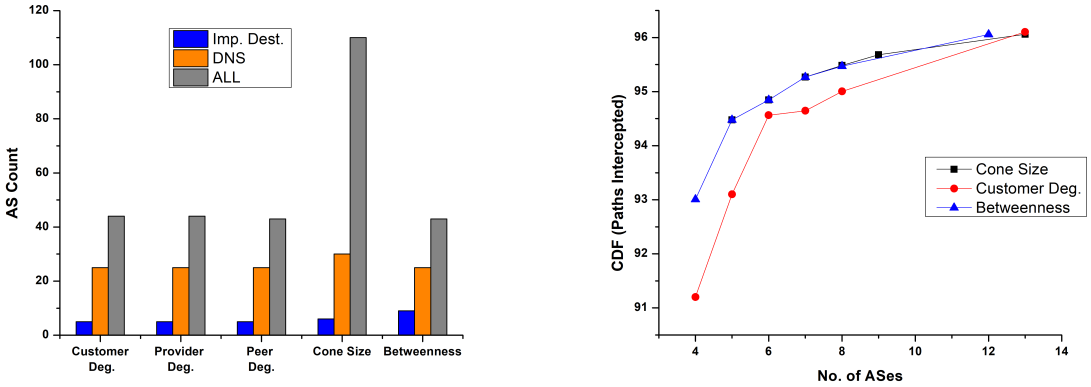
Fig. 10: China: Network Size 345 ASes; Total Border ASes 97.



(a) Line of Defense

(b) Impact of Heavy Hitters

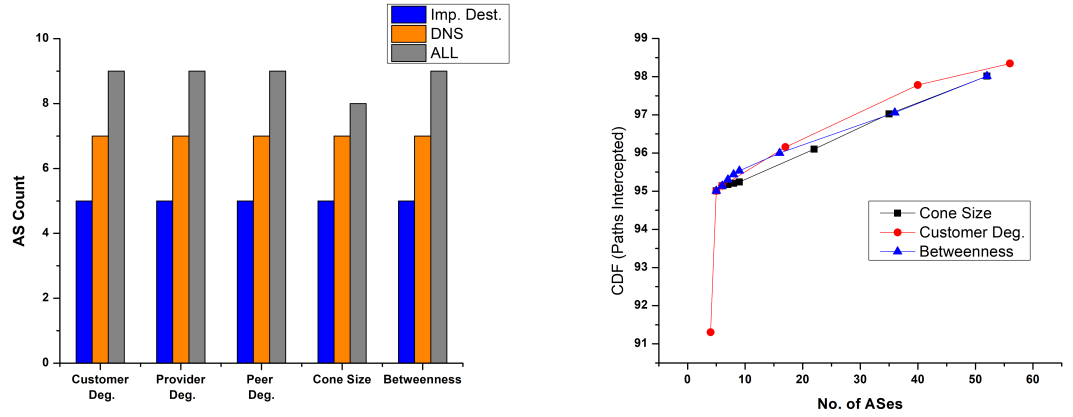
Fig. 11: Iran: Network Size 407 ASes; Total Border ASes 20.



(a) Line of Defense

(b) Impact of Heavy Hitters

Fig. 12: Israel: Network Size 214 ASes; Total Border ASes 54.



(a) Line of Defense

(b) Impact of Heavy Hitters

Fig. 13: Egypt: Network Size 55 ASes; Total Border ASes 7.

inside the country). Our results present a somewhat different picture.

- For each of the eight countries studied, *complete* coverage requires substantial effort. In China, for instance, the government would need the co-operation of over 200 ASes (out of a total 345 ASes) to form a cyber-defense line.
- At the same time, the distribution of AS-level paths is heavily biased towards a few ASes. In general, less than 10 ASes capture more than 90% of the country's AS level paths. In China, 8 (out of 200) cut ASes capture 91.2% of total AS paths.
- The cut-set ASes, that capture $< 99\%$ of all AS level paths, are boundary ASes of the country (i.e., they have peering relationship with foreign ASes).

Note: Our results suggest that a Line of Defense *does not need both a Maginot Line (border ASes) and a Tourniquet (cut-set ASes)*. Border ASes are not only important for intercepting foreign traffic, but are also the "heavy-hitter" ASes for paths within the country, so it is possible to achieve a good 99% path coverage using the right border ASes. We only need to also use internal ASes if the government is determined to achieve 100% path coverage (and in this case, a *large number* of such ASes will be required).

VI. Design Decisions, Limitations, and Future Work.

Mapping, Structure, and Border ASes.

One major design decision, in our study, is how we locate an AS in a country. An AS can have a presence in multiple countries [28], but we count an AS as belonging to the country where it is headquartered.

We model the world AS map on the C-BGP emulator. For a given country, we find the prefixes hosted in its ASes (as per CIDR [3]); C-BGP then runs a model BGP session among agents (which emulate the given ASes), where they share route advertisements and slowly populate their routing tables. For our purposes, we accept the map of relationships constructed by the C-BGP session. C-BGP also allows us to run AS-level `traceroute`, from all ASes worldwide, to the ASes of a target country. If A is the first AS on a path trace to be headquartered in a particular country, A is a border AS.

It may be noted that C-BGP is not perfect: it takes AS-relationships and IP prefix – to – AS mapping as input, and generates an AS-level map. While we used well-respected datasets (the AS relationship dataset from CAIDA [2] and IP-prefix-AS mapping from CIDR report [3]), this does not capture fine detail, such as BGP community values and local policies used by the actual ASes in the Internet. Finding exact AS relationships and IP prefix – AS mapping is an open problem; we use a standard approach, and accept its limitations.

Location of the Adversary.

In our model, we aim to achieve good coverage of paths so as to detect traffic traces from attack activities (scans, botnet activity, DDoS). We note we do not make any assumptions about the adversary; it can be located in *any* AS of the Internet – inside or outside the country.

We originally considered a more restricted model, where the adversary was restricted to the $11,100$ known malicious ASes from BGPRank [4]. However, we noted that many attacks involve reflection: non-malicious AS users can still be used by attackers to launch attacks [58]. We therefore removed any assumptions regarding where the attack comes from.

Choice of important AS.

In our study, we used the metrics of provider/customer/peer degree, cone size, and betweenness, to select important ASes. Each metric has its own intuitive importance, and in fact both attackers and defenders would like to choose such ASes as "high ground".

- *Attacker's perspective.* A rational attacker would most likely choose to attack through an AS with high provider/peer degree, i.e., an AS with a large number of provider/peer ASes. This would be because, if one provider fails (or detects and blocks the attack), the attacker could continue its attack through other providers or peers.
- *Defender's perspective.* A transit AS with high customer degree/cone size is a good choice for defender placement: malicious ASes may hide behind the upstream transit AS, to be hard to observe [32].
  ASes with high betweenness value could also be used for defender (e.g., NIDS) placement. (The *betweenness* of an AS, the number of AS pairs it connects, is a good heuristic for traffic flows through the AS.) Such defenders are in a position to analyze large volume of traffic for potential threat analysis and mitigation.
  In practice, for most of our case studies, choosing a Line of Defense based on betweenness as the AS characteristic, resulted in the smallest set of ASes.

How general are our results?

Our case study is of limited size: we analyzed a few countries in some regions of particular interest (viz., the South China sea, Kashmir, and Palestine zones). However, our *approach* for finding Line of Defense does not use any features specific to the aforementioned countries. In future, we plan to extend this study to other regions of the globe as well e.g., America, Australia and Europe etc. At present, our claim is simply that we found consistent patterns in the line-of-defense ASes for the countries we studied, despite variations in the complexity and structure of their AS-level topologies. We also report the general trends and the variations we found.

In order to check *stability* of our results, we collected both AS relationship and Prefix-to-AS information several times, at intervals of three months, and ran our entire algorithm. The chosen cut-set of ASes (i.e., the Line of Defense for various countries) remained roughly stable, for all countries in our study.

How can we build a watchdog AS in practice?

Our study simply asks which ASes to use when building a line of defense for a country. An AS is not a monolithic entity – it consists of hundreds or thousands of routers, middleboxes, and so on. So the question remains, exactly how we can build a Line of Defense using the ASes we identify. What firewalls, flow meters, Network Intrusion Detection Systems (NIDS) etc. can we use? And where in the AS can we deploy them?

This is a major question, though out of scope for the current paper; we intend to make a start, by identifying the important routers and middleboxes in Line of Defense ASes, in future work. For instance, we may make use of tools such as Rocketfuel [50] to identify such routers.

VII. Concluding Remarks

How might a nation effectively secure its cyberspace against internal and external attackers? This paper studies one approach: using network cartography to identify key locations, and installing *defenders* at these points to detect and intercept attack traffic. To be practical, such a system would require that a relatively small number of Autonomous Systems (ASes) intercept all network paths (and thus traffic). We call these key ASes in a country, its "Line of Defense".

Our case studies of eight different countries (including China and India) show that *national cyberspace consistently shows a hierarchical structure*. We require very few ASes to intercept over 90% of all intra-country AS paths. For example, in India only 4 ASes capture more than 95% of the network paths. Moreover, in the countries we studied, the boundary ASes (that have peering relationship with foreign ASes of the country) capture over 99% of the paths! – This would be a good explanation of, for instance, the Great Firewall of China. *A country with a strong "Maginot Line" i.e., security at the boundary ASes, has little need for a "Tourniquet" i.e., a cut-set of internal ASes with security measures.*

It is clearly feasible to get very good coverage of traffic paths in a nation; however, as an issue of *security*, it would be best to reach 100% coverage. Unfortunately, complete coverage requires a very large number of ASes – for example, in China 9 ASes intercept over 90% of the paths, and 90 ASes over 99% of paths, but 213 ASes are needed to cover 100% of paths. Indeed, complete coverage is hard to achieve even when we reduce the scope of protection (paths to important web resources in the country). Further, these results were consistent across the different heuristics we used in choosing important ASes (AS degree, cone size, etc.)

In our future work, we intend to build on these results, develop metrics to measure a nation's vulnerability to cyber attack, and formalize the notion of "Maginot Lines". We are particularly interested in the idea that Maginot Lines in cyberspace may not be restricted to *national* boundaries: borders between different Internet regimes (US, Europe, China) may well be more significant than national borders between similar neighbors (inside Europe, US-Canada, and so on).

References

[1] "Archipelago (ark) measurement infrastructure," http://www.caida.org/projects/ark/.

[2] "Caida as relationship dataset," https://www.caida.org/data/as-relationships/.

[3] "Cidr report," http://www.cidr-report.org/as2.0/.

[4] "Malicious ases list." [Online]. Available: http://bgpranking.circl.lu/

[5] H. Acharya, S. Chakravarty, and D. Gosain, "Few throats to choke: On the current structure of the internet," in *2017 IEEE 42nd Conference on Local Computer Networks (LCN)*. IEEE, 2017, pp. 339–346.

[6] M. Antonakakis, T. April, M. Bailey, M. Bernhard, E. Bursztein, J. Cochran, Z. Durumeric, J. A. Halderman, L. Invernizzi, M. Kallitsis *et al.*, "Understanding the mirai botnet," in *26th {USENIX} Security Symposium ({USENIX} Security 17)*, 2017, pp. 1093–1110.

[7] S. Aryan, H. Aryan, and J. A. Halderman, "Internet censorship in iran: A first look," in *3rd {USENIX} Workshop on Free and Open Communications on the Internet ({FOCI} 13)*, 2013.

[8] D. J. Betz, *Cyberspace and the State: Towards a Strategy for Cyber-power*. Routledge, 2017.

[9] L. Bilge, S. Sen, D. Balzarotti, E. Kirda, and C. Kruegel, "Exposure: A passive dns analysis service to detect and report malicious domains," *ACM Transactions on Information and System Security (TISSEC)*, vol. 16, no. 4, pp. 1–28, 2014.

[10] R. K. Chang and K. P. Fung, "Transport layer proxy for stateful udp packet filtering," in *Proceedings ISCC 2002 Seventh International Symposium on Computers and Communications*. IEEE, 2002, pp. 595–600.

[11] W. Chang, A. Mohaisen, A. Wang, and S. Chen, "Measuring botnets in the wild: Some new trends," in *Proceedings of the 10th ACM Symposium on Information, Computer and Communications Security*, 2015, pp. 645–650.

[12] B. Chen, J. Lee, and A. S. Wu, "Active event correlation in bro ids to detect multi-stage attacks," in *Fourth IEEE International Workshop on Information Assurance (IWIA'06)*. IEEE, 2006, pp. 16–pp.

[13] W. R. Cheswick, S. M. Bellovin, and A. D. Rubin, *Firewalls and Internet security: repelling the wily hacker*. Addison-Wesley Longman Publishing Co., Inc., 2003.

[14] R. Clayton, S. J. Murdoch, and R. N. Watson, "Ignoring the great firewall of china," in *International Workshop on Privacy Enhancing Technologies*. Springer, 2006, pp. 20–35.

[15] J. Curran, N. Fenton, and D. Freedman, *Misunderstanding the internet*. Routledge, 2016.

[16] J. Czyz, M. Kallitsis, M. Gharaibeh, C. Papadopoulos, M. Bailey, and M. Karir, "Taming the 800 pound gorilla: The rise and decline of ntp ddos attacks," in *Proceedings of the 2014 Conference on Internet Measurement Conference*, 2014, pp. 435–448.

[17] D. E. Denning, "Activism, hacktivism, and cyberterrorism: The internet as a tool for influencing foreign policy," *Networks and netwars: The future of terror, crime, and militancy*, vol. 239, p. 288, 2001.

[18] Z. Durumeric, D. Adrian, A. Mirian, M. Bailey, and J. A. Halderman, "A search engine backed by internet-wide scanning," in *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, 2015, pp. 542–553.

[19] Z. Durumeric, M. Bailey, and J. A. Halderman, "An internet-wide view of internet-wide scanning," in *23rd {USENIX} Security Symposium ({USENIX} Security 14)*, 2014, pp. 65–78.

[20] Z. Durumeric, E. Wustrow, and J. A. Halderman, "Zmap: Fast internet-wide scanning and its security applications," in *Presented as part of the 22nd {USENIX} Security Symposium ({USENIX} Security 13)*, 2013, pp. 605–620.

[21] A. J. Feldman, J. A. Halderman, and E. W. Felten, "Security analysis of the diebold accuvote-ts voting machine," 2006.

[22] J. François, I. Aib, and R. Boutaba, "Firecol: a collaborative protection network for the detection of flooding ddos attacks," *IEEE/ACM Transactions on networking*, vol. 20, no. 6, pp. 1828–1841, 2012.

[23] L. Gao, "On inferring autonomous system relationships in the internet," *IEEE/ACM Transactions on Networking (ToN)*, vol. 9, no. 6, pp. 733–745, 2001.

[24] S. García, A. Zunino, and M. Campo, "Survey on network-based botnet detection methods," *Security and Communication Networks*, vol. 7, no. 5, pp. 878–903, 2014.

[25] C. Gates, M. P. Collins, M. Duggan, A. Kompanek, and M. Thomas, "More netflow tools for performance and security." in *LISA*, vol. 4, 2004, pp. 121–132.

[26] D. Geer, "Resolved: the internet is no place for critical infrastructure." *Commun. ACM*, vol. 56, no. 6, pp. 48–53, 2013.

[27] V. Giotsas, G. Smaragdakis, C. Dietzel, P. Richter, A. Feldmann, and A. Berger, "Inferring bgp blackholing activity in the internet," in *Proceedings of the 2017 Internet Measurement Conference*, 2017, pp. 1–14.

[28] E. Gregori, A. Improta, L. Lenzini, L. Rossi, and L. Sani, "Discovering the geographic properties of the internet as-level topology," *Networking Science*, vol. 3, no. 1-4, pp. 34–42, 2013.

[29] G. Gu, R. Perdisci, J. Zhang, W. Lee *et al.*, "Botminer: Clustering analysis of network traffic for protocol-and structure-independent botnet detection." in *USENIX security symposium*, vol. 5, no. 2, 2008, pp. 139–154.

[30] G. Gu, J. Zhang, and W. Lee, "Botsniffer: Detecting botnet command and control channels in network traffic." in *NDSS*, vol. 8, 2008, pp. 1–18.

[31] A. Houmansadr, E. L. Wong, and V. Shmatikov, "No direction home: The true cost of routing around decoys." in *NDSS*, 2014.

[32] R. Howard, *Cyber fraud: tactics, techniques and procedures*. CRC press, 2009.

[33] K. Kalkan, G. Gür, and F. Alagöz, "Filtering-based defense mechanisms against ddos attacks: A survey," *IEEE Systems Journal*, vol. 11, no. 4, pp. 2761–2773, 2016.

[34] M. S. Kang, S. B. Lee, and V. D. Gligor, "The crossfire attack," in *2013 IEEE symposium on security and privacy*. IEEE, 2013, pp. 127–141.

[35] A. D. Keromytis, V. Misra, and D. Rubenstein, "Sos: An architecture for mitigating ddos attacks," *IEEE Journal on selected areas in communications*, vol. 22, no. 1, pp. 176–188, 2004.

[36] S. Khattak, N. R. Ramay, K. R. Khan, A. A. Syed, and S. A. Khayam, "A taxonomy of botnet behavior, detection, and defense," *IEEE communications surveys & tutorials*, vol. 16, no. 2, pp. 898–924, 2014.

[37] C. Kolias, G. Kambourakis, A. Stavrou, and J. Voas, "Ddos in the iot: Mirai and other botnets," *Computer*, vol. 50, no. 7, pp. 80–84, 2017.

[38] I. Kotenko and A. Chechulin, "Attack modeling and security evaluation in siem systems," *International Transactions on Systems Science and Applications*, vol. 8, pp. 129–147, 2012.

[39] L. Liu, S. Chen, G. Yan, and Z. Zhang, "Bottracer: Execution-based bot-like malware detection," *Information Security*, pp. 97–113, 2008.

[40] M. Luckie, A. Dhamdhere, B. Huffaker, D. Clark, and K. Claffy, "Bdrmap: Inference of borders between ip networks," in *Proceedings of the 2016 Internet Measurement Conference*, 2016, pp. 381–396.

[41] H. V. Madhyastha, T. Isdal, M. Piatek, C. Dixon, T. Anderson, A. Krishnamurthy, and A. Venkataramani, "iplane: An information plane for distributed services," in *Proceedings of the 7th symposium on Operating systems design and implementation*. USENIX Association, 2006, pp. 367–380.

[42] A. Marder, M. Luckie, A. Dhamdhere, B. Huffaker, k. claffy, and J. M. Smith, "Pushing the boundaries with bdrmapit: Mapping router ownership at internet scale," in *Proceedings of the Internet Measurement Conference 2018*, 2018, pp. 56–69.

[43] O. Osanaiye, K.-K. R. Choo, and M. Dlodlo, "Distributed denial of service (ddos) resilience in cloud: Review and conceptual cloud ddos mitigation framework," *Journal of Network and Computer Applications*, vol. 67, pp. 147–165, 2016.

[44] B. Quoitin and S. Uhlig, "Modeling the routing of an autonomous system with c-bgp," *Netwrk. Mag. of Global Internetwkg.*, vol. 19, no. 6, pp. 12–19, Nov. 2005.

[45] R. Ramesh, R. S. Raman, M. Bernhard, V. Ongkowijaya, L. Evdokimov, A. Edmundson, S. Sprecher, M. Ikram, and R. Ensafi, "Decentralized control: A case study of russia," in *Network and Distributed Systems Security (NDSS) Symposium 2020,*, 2019.

[46] D. Raymond, T. Cross, G. Conti, and M. Nowatkowski, "Key terrain in cyberspace: Seeking the high ground," in *2014 6th International Conference On Cyber Conflict (CyCon 2014)*. IEEE, 2014, pp. 287–300.

[47] M. Roesch *et al.*, "Snort: Lightweight intrusion detection for networks." in *Lisa*, vol. 99, no. 1, 1999, pp. 229–238.

[48] ——, "Snort: Lightweight intrusion detection for networks." in *Lisa*, vol. 99, no. 1, 1999, pp. 229–238.

[49] C. Rossow, "Amplification hell: Revisiting network protocols for ddos abuse." in *NDSS*, 2014.

[50] N. Spring, R. Mahajan, and D. Wetherall, "Measuring isp topologies with rocketfuel," *ACM SIGCOMM Computer Communication Review*, vol. 32, no. 4, pp. 133–145, 2002.

[51] D. Springall, T. Finkenauer, Z. Durumeric, J. Kitcat, H. Hursti, M. MacAlpine, and J. A. Halderman, "Security analysis of the estonian internet voting system," in *Proceedings of the 2014 ACM Conference on Computer and Communications Security*, ser. CCS '14. New York, NY, USA: ACM, 2014, pp. 703–715.

[52] A. Studer and A. Perrig, "The coremelt attack," in *European Symposium on Research in Computer Security*. Springer, 2009, pp. 37–52.

[53] J. E. Sullivan and D. Kamensky, "How cyber-attacks in ukraine show the vulnerability of the us power grid," *The Electricity Journal*, vol. 30, no. 3, pp. 30–35, 2017.

[54] P. Sweeney and G. Cybenko, "Identifying and exploiting the cyber high ground for botnets," in *Cyber Warfare*. Springer, 2015, pp. 37–56.

[55] Y. Tian, R. Dey, Y. Liu, and K. W. Ross, "China's internet: Topology mapping and geolocating," in *2012 Proceedings IEEE INFOCOM*. IEEE, 2012, pp. 2531–2535.

[56] M. E. Tozal, "Autonomous system ranking by topological characteristics: A comparative study," in *Systems Conference (SysCon), 2017 Annual IEEE International*. IEEE, 2017, pp. 1–8.

[57] V. Vatter, "Graphs, flows and the ford-fulkerson algorithm," *Tersedia di: http://www. math. ufl. edu/˜ vatter/teaching/flow. pdf.[diakses 10 maret 2013]*, 2004.

[58] C. Wagner, J. François, R. State, A. Dulaunoy, T. Engel, and G. Massen, "Asmatra: Ranking ass providing transit service to malware hosters," in *Integrated Network Management (IM 2013), 2013 IFIP/IEEE International Symposium on*. IEEE, 2013, pp. 260–268.

[59] X. Xu, Z. M. Mao, and J. A. Halderman, "Internet censorship in china: Where does the filtering occur?" in *International Conference on Passive and Active Network Measurement*. Springer, 2011, pp. 133–142.

[60] S. T. Zargar, J. Joshi, and D. Tipper, "A survey of defense mechanisms against distributed denial of service (ddos) flooding attacks," *IEEE communications surveys & tutorials*, vol. 15, no. 4, pp. 2046–2069, 2013.