Proceedings of the 2017 Winter Simulation Conference
W. K. V. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page, eds.

Simulating DDoS Attacks on the US Fiber-Optics Internet Infrastructure

Sumeet Kumar

School of Computer Science
Carnegie Mellon University
5000 Forbes Ave
Pittsburgh, PA 15213, USA

Kathleen M. Carley

School of Computer Science
Carnegie Mellon University
5000 Forbes Ave
Pittsburgh, PA 15213, USA

ABSTRACT

Network-based attacks like the distributed denial-of-service (DDoS) attacks are not new, but we are beginning to see attacks of unprecedented scale. Recent examples of such attacks include the DDoS attack on DYN INC that crippled a part of the Internet for hours, and the attack on Liberia, which partially brought down the African nation. Limitations in identifying vulnerable Internet infrastructure and testing possible defense strategies are a part of the problem. We need a simulation testbed that can mirror the complexity of the Internet, yet allows to swiftly test attacks, providing insights that can apply to real-world attack scenarios. In this research, we have designed a test-bed that mirrors the Internet infrastructure of the US and can simulate the Internet traffic flow patterns for different attack targets. We also estimate the degradation in the QoS and the number of users impacted in two attack scenarios.

1 INTRODUCTION

Network based cyber-attacks like DDoS appear to be a growing phenomenon. However, there is no clear understanding of where the attacks are coming from, how the attacks are organized, and how the attack targets are identified. While it is known that a majority of these attacks are originating from bots (Kandula, Katabi, Jacob, and Berger 2005), there is lesser clarity on where the bots are located and how the bots are controlled (Stinson and Mitchell 2007). To add to the puzzle, the specific impact and maximum possible damage of such attacks are also not known, and we only estimate the impact after the attacks have happened. However, one thing is clear that the bandwidth used in these attacks are increasing with time (Inofsecurity 2016). In a recent example of the cyber-attack on the African nation of Liberia (Kumar 2016), thousands of bots targeted the fiber-optic cable exchange point, bringing down the Internet connectivity of the entire country for almost a day. Another example is the attack on Estonia (Ottis 2008), which crippled the Estonia's government web-services for a few weeks. These incidents highlight the serious nature of such attacks and call for strategies to counter them, and ways to assess and control them. Because of the complexity of the Internet, testing defense strategies require a realistic simulation environment. In this research, our focus is on designing a simulation test-bed to estimate the impact of such attacks. Our simulation test-bed allows to flexibly pick a target, select a distribution of botnet spread across the Internet, and determine the implications of attacking the target. In addition to simulating network flow attacks, the simulation model also allows finding more vulnerable and more resilient IXPs.

Because of the complexity of the Internet, simulating an Internet-scale network based attack is nearly infeasible. Most existing simulation environments are designed for small-scale tests and are usually conducted in a lab setting using a few systems. Results from a small-scale simulation may not be a good approximation to the Internet scale problem. The complexity is two-fold. First, it's hard to model agents (systems) that could resemble computer systems, but at the same time occupy reasonable computer memory for a large scale simulation. Second, openly available information on the Internet infrastructure

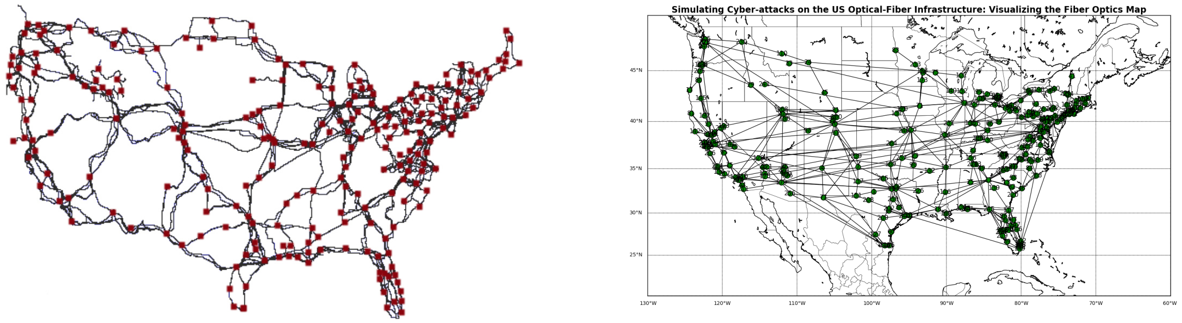
Simulating Cyber-attacks on the US Optical-Fiber Infrastructure: Visualizing the Fiber Optics Map
45°N
40°N
35°N
30°N
25°N
130°W
120°W
110°W
100°W
90°W
80°W
70°W
60°W

Figure 1: Left image shows the location of physical conduits obtained from study 'InterTubes: a study of the US long-haul fiber-optic infrastructure' . The right image visualizes the network built using the physical conduits data. In the right image, the green nodes are the IXPs located in different cities, and the conduit optical pipes connecting them are shown as black lines. The network shown in the right image is used to design a network simulation to test the resiliency of the US Internet infrastructure.

is limited. Fortunately, we can approximate both the problems. For a network-based attack like DDoS, we can model the attackers, and the target as simple systems, with a limited capacity to send and receive network packets. Based on these simplifications, we modeled bots as sources of network packets and the target systems as sinks with limited network capacity. The other significant limitation of modeling a close to real test environment i.e. the Internet infrastructure could be approximated by using some recent work on mapping the optical fiber pipes of the US. The mapping of the Internet infrastructure project (Durairajan, Barford, Sommers, and Willinger 2015) (see Fig.1, left) has provided a practical knowledge of the US cyber infrastructure. By combining the cyber-infrastructure data with data on the Internet adoption by countries (Whitehouse.gov 2016), a simulation environment could be built to model a real attack environment. We expect a part of solving the DDoS attacks problem is to develop a test bed that allows simulating the Internet-scale attacks so that the impact of such attacks could be estimated, and resiliency of vulnerable infrastructure could be enhanced. Our simulation environment also allows to evaluate different attack scenarios and providing information on average degradation of the quality-of-service, and the approximate number of users impacted.

The important contributions of this research are: a) We design a simulation testbed mirroring the fiber-optics Internet architecture of the US. To the best of our knowledge, this is the first work that simulates DDoS attacks on realistic US cyber infrastructure. b) We present a model to estimate the degradation in the quality-of-service and the number of users impacted in different attack scenarios. To make our simulation more realistic, we use a dynamic flow model that changes the Internet traffic flow pattern with congestion. c) Our model allows finding cyber installations that are more vulnerable to attacks. We believe our approach could be very useful to cyber-infrastructure companies and homeland security.

The rest of this paper is organized as follows. Section 2 presents related work. We introduce the methodology of the simulation in section 3. In section 4, we model and discuss the experiments that we conducted. The results of the experiments are presented in section 5, along with a discussion on the outcomes of the simulations. Limitations of our approach are mentioned in section 6. We finally conclude and discuss future work in section 7.

2 RELATED WORK

Prior work on using simulation to model cyber-attacks can be grouped in two categories: a) Measuring resiliency of the Internet infrastructure b) Simulation of network based attacks. We discuss each of them in more details next.

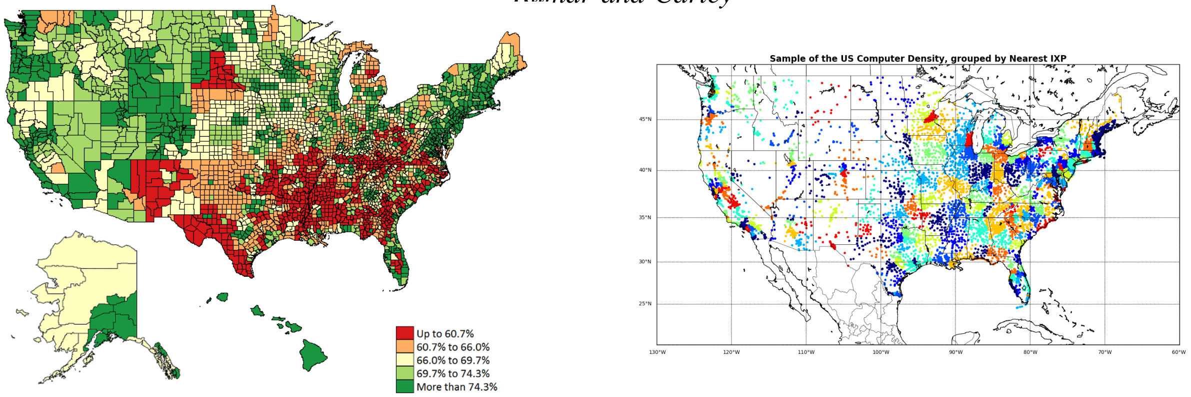

Figure 2: Left image shows the map of Internet adoption in United States by county from 2013 census. The right image shows the computer population randomly generated from population density distribution and Internet adoption map. This computer distribution is later used for modelling the Internet traffic flow. Computers in each cluster, shown in different color, are connected to one of the local (nearest) IXP.

2.1 Measuring Resiliency of the Internet Infrastructure

Mayada et. al. (Omer, Nilchiani, and Mostashari 2009, Omer, Nilchiani., and Mostashari 2009) used the submarine cable map, and countries data demand, capacity and flow information to model the resiliency of the global Internet infrastructure, and explore the node to node resiliency of the under-sea cables to a physical damage. The authors approached the problem as a network optimization problem with constraints, and solved it using linear and mixed integer programming. In this research, we are also interested in resiliency, but we don't model physical damage to the under-sea cables. Since most cyber-attacks don't entirely stop the flow of network packets but rather makes the Internet slower, this model is not best suitable for simulating DDoS attacks. Moreover this model did not account for the dynamic nature of the path used by Internet traffic, which is included in our model.

2.2 Simulation of Network Based Attacks

A number of researchers (Kotenko and Ulanov 2006, Li, Li, and Zhao 2008, Kotenko and Ulanov 2005, Kotenko 2005, Qwasmi, Ahmed, and Liscano 2011, Zhang, Xie, Zhang, and Zhang 2008, Grunewald, Lützenberger, Chinnow, Bye, Bsufka, and Albayrak 2011) have simulated DDoS attacks, albeit on a much smaller scale. The limitation in scale of simulation is because computer-systems are complex, and modeling their behavior makes agents complex and resource intensive. For example, Kotenko et al. (Kotenko and Ulanov 2005) investigated different attack scenarios and protection mechanisms for networks with different structures and security policies, and used OMNeT++ INET Framework for their development. If an agent is complex, it takes more computer memory to create and execute the simulation. The relative complexity of these models render them limited to small scale, with limited number of systems. We resolve this problem by proposing very simple agents, with only limited networking (send and receive data) capability.

The study by Kong et al. (Kong, Mirza, Shu, Yoedhana, Gerla, and Lu 2003) is the closest to our research. In this work, Kong et al. modeled general flow of computer network using NS2 based simulation engine. The simulator creates Internet traffic from regular systems (not infected) as Pareto distribution, and traffic from bot systems as uniform distribution. In evaluation, they present how changes in source-network and sink-network affect the quality of services on the network. However, their network is randomly generated and hence does not reflect the reality. Simulating network flow on the realistic Internet network topology of the US, is one of our important contributions.

3 METHODOLOGY

Distributed Denial of Service (DDoS) attacks are generally of two types a) Network-layer attacks b) Application layer attacks. In a network layer attack, the adversary tries to exhaust the available network bandwidth of the service provider or the server, where as, in an application layer attack the goal is to overwhelm the processing capability of the target-server or router. In this study, we focus on network layer attacks. A network layer attack is based on the assumption that a company's network resources like firewalls, routers, and servers have limited capacity, and the capacity is based on the network requirements of the company. These requirements are determined based on average usage and are not designed for attack scenarios. However, the attacker's capacity is not decided by the number of machines he owns, but rather decided by the strength of the botnet he controls. The strength of botnet could increase with time. Especially with an exponential growth in internet-of-things, and their weak security, often it is easy to compromise these IoT devices, and add them as a part of botnet. During an attack, these botnets send a large amount of network packets to saturate the limited bandwidth of the target company, so that genuine users are not able to access the services.

We use an agent-based network simulation approach to simulate cyber-attacks. Our simulation environment comprises of a network of connected Internet Exchange Points (IXPs) as nodes and Internet packet traffic as flow (see Fig. 1). The nodes and links used in this network mirrors the real fiber-optics infrastructure of the US, and are obtained from the data shared by Durairajan (Durairajan et al. 2015), on special request.

An attack scenario is modeled as flow of large traffic to a target node. In an experimental attack scenario, we first estimate the regular traffic in the Internet (without any attacks) flow using population density distribution and Internet adoption percentage (Fig.2). For estimating the attack traffic, we pick an attack target, and direct all bots to send data packets to the target and then superimpose the attack traffic on the regular traffic. We use the information on the bots locations shared by a website (Malwaretech 2016) to find the nearest IXP for all bots, and estimate the number for bots connected to each IXP (Fig. 3). The number of bots connected to each IXP leads to an estimate of the attack-traffic that an IXP generates and adds to the Internet. After estimating the attack-traffic, we remove the bots from the simulation. This way we avoid the complexity of using millions of systems in the simulation, but still use their attack impact. The superposition of attack traffic on regular traffic creates a congestion map of the entire network, and allows us to measure the drop in average quality-of-service. We also estimate the number of users impacted, which is indirectly approximated by the number of users connected to each IXP (see Fig. 3, right) and the estimate of congestion at each IXP. Note that this problem is not a simple network analysis problem, as the network-packets paths are dynamic, and change with congestion in the network (explained later in this section).

Before we go into more details, we first define a few important terms that are frequently used in this paper. Then we discuss the simulation environment model, the agent model, our simulation strategy and how we measure the output variables.

3.1 Definition of Terms

A *Botnet* is a computer program that allows an attacker to take control over a remote computer (victim machine). Botnets are a network of compromised machines that stretch across the Internet. Bots are either controlled directly by attackers or indirectly by handlers (Stinson and Mitchell 2007). Often botnets also communicate with each other to pass commands. An *Internet service provider* (ISP) uses the Internet backbone infrastructure to deliver traffic to consumers. These are the companies like Comcast Inc. that directly serve the end users. An *Internet exchange point* (IX or IXP) allows Internet service providers (ISPs) and *Content Delivery Networks* (CDNs) to exchange Internet traffic between their networks. They are physical infrastructure through which large part of the internet traffic flows. *Network Flood Attacks* are bandwidth saturation attacks, with a goal to saturate the bandwidth of a company. Network resources like firewall, routers and servers have limited capacity. If a network infrastructure is overwhelmed, the genuine users will not be able to reach the website/web-service. *Distributed Denial of Service* (DDoS) attack is a type of network flood attacks that are initiated by bots distributed over the Internet.

3.2 Agent Modeling

We use an agent-based model for this simulation. We model computer-systems as active agents, and the fiber-optics network as a passive environment. For simplicity, an agent is any system (including laptops, mobile devices, Internet of Things, and Servers) that is connected to the Internet. We model the location and density of agents based on population and computer ownership data (Fig.2, left). For simulation, agents send packets that travel from agent (source) to target (sink). Agents can act both as source or sink of packets, and to simulate a more realistic environment both the source and the sink have limited bandwidth. Flow through links also have a bandwidth limitation.

We have two types of agents. First agent type is 'Un-compromised Agents' i.e. the systems that are not compromised. The second agent type is bots, i.e. the systems that are compromised and are used in attacks. For location of 'un-compromosed agents' we use the data from US census on population and percentage population with active Internet connection (Fig.2, left). For location of botnets (infected computers), we us botnet tracker data from Malwaretech website (Malwaretech 2016). Since we know the approximate location of each of the agents (shared by Malwaretech), we could approximate the route the packets from an agent could take to reach a target (using a shortest path algorithm as explained later). We can do the same for all the bots based in the US, to understand the network congestion these bots can create.

3.3 Environment Modeling

As explained earlier, we used systems as agents to simulate the DDoS cyber attacks on the US Internet infrastructure. The environment in our case is the network generated by the optical fiber cables and the Internet exchange points. We model it as a network with nodes and links. Nodes represent the Internet exchange points (IXP) and links are optical cables connecting different IXPs. We use optical fibers (Durairajan, Barford, Sommers, and Willinger 2015) data for modeling the links. The authors (Durairajan, Barford, Sommers, and Willinger 2015) shared their data on nodes, and their connections between cities (Fig:1, left), but they do not have data on bandwidth capacity of these optical fibers. We approximate the bandwidth limitation of the long-haul pipes to 8 Tbps (Betker et al. 2014). Fig:1 (right) shows the network built using nodes and edges data from (Durairajan, Barford, Sommers, and Willinger 2015).

3.4 Simplified Model of Packet Flow on the Internet

Because the Internet is very complex with billions on systems and millions of routers, simulating an attack with so many components is unrealistic. We simplify the packet flow architecture such the simplified model contains the necessary characteristics needed for a realistic simulation. Fig.3 shows a simplified version of packet flow through the Internet that has been used in this work. In the figure on the left, compromised systems (shown in light green) send a stream of packets to a target system (shown in red). As a number of compromised systems (bots) send packets at the same time, the packet flow (shown as purple arrows) adds up as the flow moves towards the target. To keep the simulation simple, we have entirely removed the systems by their contribution made to an IXP (shown in the right image). This simplification is possible because the number of systems connected to an IXP does not change dynamically (on an average) and hence the network topology is mostly static (flow is not). Thus for this simulation, we replace all systems with their weight on the nearest IXP, similarly we replace all bots with their equivalent weight on the nearest IXP. The count of bots and systems at an IXP enables us to estimate the attack traffic the IXP can generate, and also enables the count of users that get impacted when an IXP is congested.

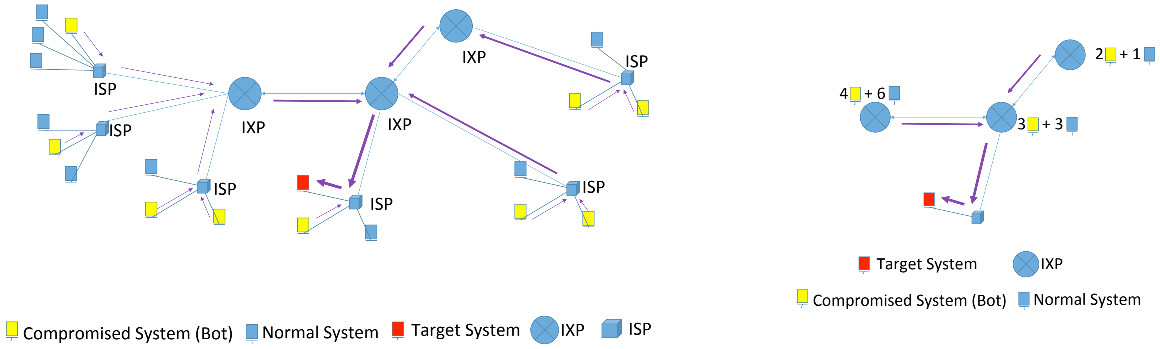

Figure 3: A Network flow representing an attack scenario. Left figure shows a scenario in which the red node is the target of attack. Yellow nodes are bots or attacking systems that generates network flow (traffic). Purple arrows show the direction of traffic flow, and the width of arrows shows the volume of traffic. Figure on the right is a simplified version on the left network, where we have replaced the individual systems by their counts on IXPs as attributes. The simplification speeds up the network simulation as the simplified network only includes IXPs (with equivalent system weights) and a target.

3.5 Approximating the Congestion Control Algorithms Used on the Internet

We approximate the Border Gateway Protocol (BGP) algorithms commonly used by gateway routers by Dijkstra's shortest path algorithm. BGP is commonly used as the protocol between Internet service providers and is designed for packets to take the most efficient route. In simple terms, BGO uses cost metrics for each path to find the most optimum path, which also makes it complex. To keep the simulation simple, but still include the impact of cost of traversal through congested paths, we used Dijkstra's shortest path algorithm, where cost in a path is determined by the congestion in the path. Including Dijkstra's algorithm makes the simulation dynamic and more realistic, e.g. as we increase the attack bandwidth sent by a bot, the packets are likely to take different paths, and hence creates congestion in other paths influencing the quality of service in areas which may not be close to the target.

3.6 Problem Modeling

Given a flow network $G = (V,E)$, many non-negative sources $f : (V_i) \rightarrow Z+$, a negative sink $h : (V_j) \rightarrow z-$, links with limited capacity $E_i(V_i,V_j) < E_{(}max)$, there exists a equilibrium state of flow using any weighted shortest path algorithm. The equilibrium state allows estimation of congestion at each node C_i, which enables to estimate the number of users impacted $\sum_i U_i$.

$$B \sim \text{Bot Density} \tag{1}$$

$$U \sim \text{User Density} \tag{2}$$

$$f : (V_i) \propto B_i \propto \text{count of bots connected to node } V_i \tag{3}$$

$$U_{impacted} = \sum_{i \in c} U_i, \text{where c is set of congested nodes} \tag{4}$$

To estimate $f : (V_i)$, i.e. the flow generated by node V_I (an IXP), we use bot density function (Eqn. 1), and group the bots to their nearest node which implies B_i is the total number of bots connected to i^{th} node. To estimate the number of users impacted ($U_{impacted}$), we sum the number of users $\sum_i U_i$ connected to i^{th} congested node. U_i is the number of systems (or users) connected to V_i (IXP) which is determined by the systems density map (Fig. 2, right), where users grouped to nearest node (explained in Fig. 3). Similarly B_i is the number of bots connected to V_i (IXP). Here B_i is determined by bots density map from (Malwaretech 2016) grouped by nearest node (IXP). Given an attack target, the average reduction in quality of service could be measured by estimating the congestion in the network, which could be approximated by $\sum_{i,j} E(V_i,V_j)_{avg} / \sum_{i,j} E(V_i,V_j)_{congestion}$. In an attack situation, the users trying to access the targeted server (or IXP) get impacted the most. Also the users connected to other nodes experiencing congestion get impacted. As explained in Eqn. 4, we estimate the number of users impacted by counting all the users connected to the congested IXPs. Given an attack scenario with an attack target, we measure the trend of 'number of users' impacted, with increase in attack bandwidth (e.g. each infected system sends attack bandwidth from 1 Mbps to 5 Mbps).

Table: 1 summarizes the variables used in the experiment. We have two parameters in the model, a target and the bandwidth of attack initiated by each bot. The constants are maps that we obtained from different data sources including 'optical-cable' map from (Durairajan, Barford, Sommers, and Willinger 2015), population and computer ownership from census, and map of Mirai botnet (Malwaretech 2016). The output variables are the number of users that get impacted in an attack scenario, and the degradation on QoS (sometimes referred as congestion). To validate, we used the data from downdetector website.

Table 1: Experiment Table

Factors	**Values**	**# of Values**
Inputs		
Attack Target (A web-server or an IXP)	One of the nodes in the Network	1
Bandwidth of Attack	1 to 5 Mbps per bot	10
Constants		
Long-haul optical-fiber Map	Map from (Durairajan et al. 2015)	
Population density	Number from census	
Computer Ownership	Percentage from census	
Map of Botnets	Mirai Bot family from (Malwaretech 2016)	
Outcomes		
Number of Users Impacted	In thousands	
Degradation in Quality of Service	Percentage	1-100
Validation		
AT&T Attack on 28th of Oct, 2016	Map of Impacted Users	
DYN Attack on 21st of Oct, 2016	Map of Impacted Users	

3.7 Why is Network Simulation Needed?

We would like to clarify that the problem that we are solving is different from the network resiliency problem as solved in (Omer, Nilchiani, and Mostashari 2009, Omer, Nilchiani., and Mostashari 2009), and the network robustness problem as done in (Durairajan, Barford, Sommers, and Willinger 2015). The important difference is the dynamic nature of traffic in the Internet. In the Internet, the flow of packets change path as a portion on the Internet gets congested, and hence assuming the Internet network to be a static network is not right. In fact, two packets meant for same target may take different routes. The dynamic nature of the Internet results in impacting people that are far away from the actual target, as further explained in the results and conclusion section.

4 EXPERIMENTS

Our experiment has different scenarios, and each scenario has a target. Though it is possible to use multiple targets, to keep the results easy to understand, we have used a single target in each scenario. For example, we can pick a server hosted in New York city as an attack target, and simulate the experiment. In any experiment, we generate network traffic from all nodes (except the target node) directed towards the chosen target. The simulation will result in a map of users that get impacted because of attack on the target. We

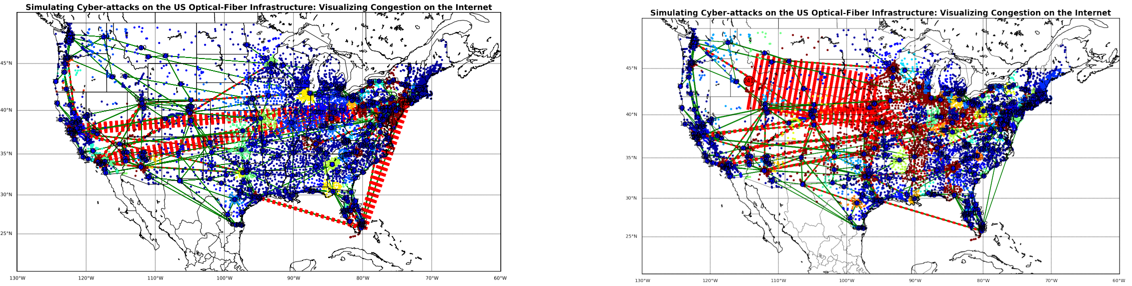
Simulating Cyber-attacks on the US Optical-Fiber Infrastructure: Visualizing Congestion on the Internet
45°N
40°N
35°N
30°N
25°N
130°W
120°W
110°W
100°W
90°W
80°W
70°W
60°W
Simulating Cyber-attacks on the US Optical-Fiber Infrastructure: Visualizing Congestion on the Internet
45°N
40°N
35°N
30°N
25°N
130°W
120°W
110°W
100°W
90°W
80°W
70°W
60°W

Figure 4: (Left) Simulation Scenario 1: Visualizing the congestion in the network while an attack on the DNC INC New York city service. (Right) Simulation Scenario 2: Visualizing the congestion in the network while Attack on an AT&T server in Chicago. The width of edges indicate the network flow through an optical pipe. Color of an edge indicates congestion. More red an edge is, the more it is congested.

can also determine the trend the number of users that get impacted as the attack bandwidth (by each bot) is varied. Besides, we can also estimate the average degradation (using congestion in the network) in the quality of service with the increase in attack bandwidth. We simulated two different attack scenarios and evaluated the relative impact of attacks.

4.1 Scenario 1 - Attack on Dyn Inc on 21st of Oct, 2016

Targeting a server hosted in New York City, we try to mimic the attack on DYN Inc (DYN INC 2016). In this attack simulation, all bots target the New York city server. Since there is some known data on how people in different regions got impacted because of the DYN attack, this attack could be used to approximately validate the simulation. Note that the attack on Dyn Inc. is more complex than just a simple ddos attack on a target, as the attack was done in three waves and the target of attack varied throughout the day. However, because of absence of data on the location of the exact server that was attacked in each wave, we assume the New York server to be the only target. Even we this simplification, we get a close estimate of users impacted by this attack.

4.2 Scenario 2 - Attack on AT&T on 28th of Oct, 2016

In this scenario, we try to model an attack on the AT&T Internet infrastructure in the Chicago area. This scenario attempts to model the network outage reported by Downdetector website (http://downdetector.com/) on 28th of October 2016. The exact cause of the problem with AT&T servers is unknown, but a visualization map of the users impacted was obtained from the website.

In the next section, we describe the results obtained each of the experiment.

5 RESULTS AND DISCUSSIONS

In this section, we present the simulation results of the two attack scenarios, and compare them with their known impact map for validation. We also describe the impact of attacks on the QoS for each target.

5.1 Scenario 1 - Model Attack on Dyn Inc on 21st of Oct

For scenario 1, we use New York city IXP as the target of attack. The attack tries to mirror the DYN server attack that happened on Oct 21, 2016. We first discuss the congestion in the fiber-optics cables as observed in the simulation (Fig:4, left). In Fig.4, the width of edges indicate the network flow through an optical pipe, and the color indicates the congestion level. The figure is the result of the final stage of simulation (max bandwidth of attack), and the mid-stage results are not shown. As the bandwidth of attack is increased in each iteration, more and more edges (optical-fibers) showed congestion. This is as expected in an attack. However, the edges that got more congested were not always close to the target.

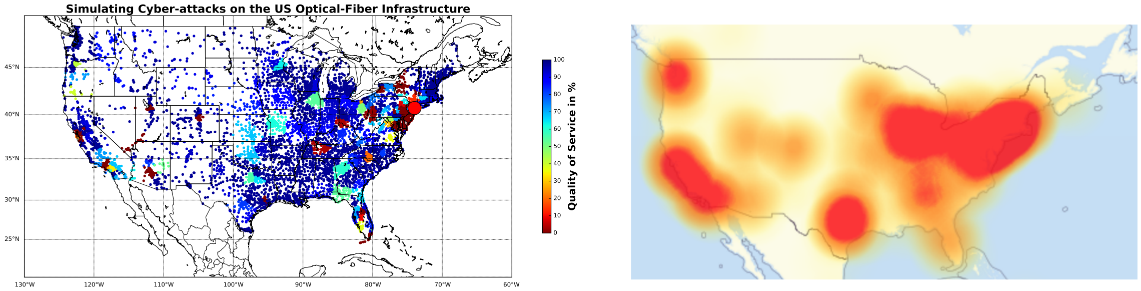

Figure 5: DYN attack Simulation: For DYN outage on 21th Oct, we compare the result of simulation (left) to actual impact as reported by DownDetector website (right). The red areas are most adversely impacted in the right image. The color bar in the left image shows indicates the relative impact in different areas. As we can observe, an attack on the New York IXP not only creates problems in the the east coast regions, but also affects areas in the west coast. Also, it can be observed many centers of the impacted areas have been correctly predicted in the simulation.

In fact, two of the most congested links are actually connecting the west coast areas, and one of the links is connecting the southern part of the US. This is only possible because of dynamic nature of simulation. Because Dijkstra's shortest path algorithm chooses path with the least cost from a source to a target, this changes the packet flow routes as current routes get more congested. Moreover, we can see that the most congested links are mostly linking high density areas and are long-haul links. This can be explained by the fact that these long haul links are likely to be the best route from far-off areas, as the length of optic-fibers do not affect the edge weight.

In Fig.5 , we compare the result of simulation (left) to actual impact as reported by downdetector website (right). The left image shows the simulation result of regions impacted by the attack. The QoS not only degraded near the New York city (as observed by red color dots) but also in some areas far away. The image on right was downloaded from wikipedia, and shows the impact of the DYN attack as measured by 'downdetector.com'. The red areas are most adversely impacted regions. The image shows that the east coast of the US was primarily impacted because of the attack, but outages were also recorded in some western and southern parts of the US. If we compare the simulation result (Fig:5, left) and the actual impact (Fig:5, right), we can observe that both of them highlight the eastern areas as mostly impacted. This is expected as the target of the attack was based in the New York city. What is interesting to see that some areas in the central, and western parts of the US were impacted, and the simulation also predicted similar areas.

Lastly, we use the trend plot (Fig:7, left) to discuss the number of users impacted, and how the number changes with increase in the bandwidth of attacks. As we can observe, the increase is steeper in the beginning but the rate of increase slows down with increase in attack. The estimate of the number of users impacted is based on the assumption that 10000 users are actively connecting to their server at any point of time.

5.2 Scenario 2 - Attack on AT&T on 28th of Oct

For scenario 2, we use Chicago city AT&T server as the target of attack. The attack tries to mirror the AT&T server problem that happened on Oct 28, 2016. We first discuss the congestion in the fiber-optics cables as observed in the simulation (Fig:4, right). As described earlier, the width of edges indicate the network flow through an optical pipe, and the color indicates congestion level. The figure is the result of the final stage of simulation, and the mid-stage results are not shown. As the bandwidth of attack is increased in each iteration, more and more optical-fibers (edges) registered congestion.

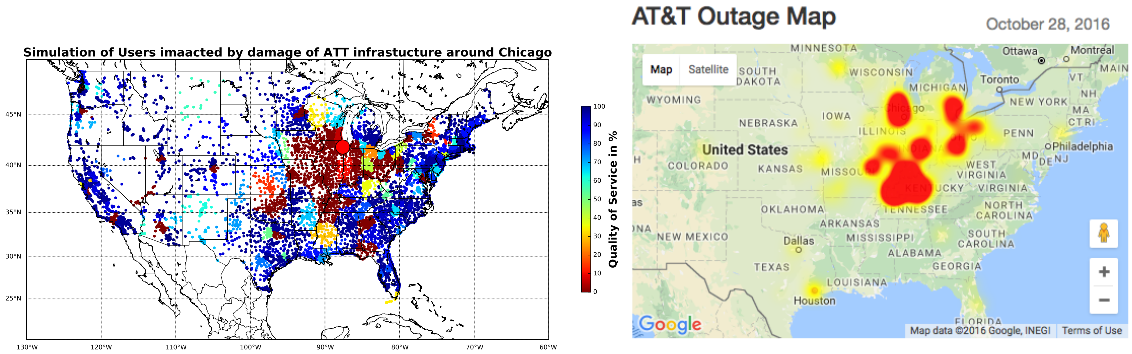

Figure 6: AT&T attack Simulation: For AT&T outage on 28th Oct, we compare the result of simulation (left) to actual impact as reported by DownDetector website (right). The red areas are most adversely impacted in the right image. The color bar in the left image indicates the relative impact in different areas. As we can observe, an attack on the Chicago IXP not only creates problems in north-central regions, but also affects areas in south and west coast. Also, it can be observed many centers of the impacted areas have been correctly predicted in the simulation.

In Figure:6, we compare the result of simulation (left) to actual impact as reported by downdetector website (right). The left image shows the simulation result of regions impacted by the attack. The QoS not only degraded near the Chicago city but also in some areas far away. The image on right was obtained from 'downdetector.com' website. The red areas are most adversely impacted regions. The image shows that areas near Chicago and south of Chicago were mostly impacted because of the attack, but minor outages were also recorded in some far areas. If we compare the simulation result (Fig:6, left) and the actual impact (Fig:6, right), we can observe that both of them highlight same areas as primarily impacted. This is expected as the target of the attack was based in the Chicago city.

Figure. 7 (right) shows the trend of 'number of users' impacted as the bandwidth of attack was increased assuming 10000 users were trying to actively connect to the server. As we can observe, the increase is steeper in the beginning but the rate of increase slows down and converges.

If we compare the result of attack on 'New York IXP' to the result of attack on 'Chicago IXP', we find that the plot of 'number of users' impacted is more steep in case of the attack on the 'New York IXP'. There could be many reason for his, including the the density of population near New York is higher compared to density of population near Chicago, especially on the western sides of Chicago.

6 Limitations

We designed a simple network simulation model for DDoS, yet sophisticated enough, to mimic complex nature of cyber-attacks. The benefit of simplicity is that we can efficiently simulate attacks that are generated by millions of systems. To keep the model simple yet realistic, we have made a few assumptions. First, The Quality of Service (QoS) as observed by a user is determined by the congestion of traffic at the nearest IXP. This may not always be true, especially when the traffic is local to an ISP, e.g. a client is watching video streaming data from a server located in proximity. In some other cases, rather than passing through nearest IXP for all their traffic flow, ISPs engage in peering. Second, we approximate the botnet data from the Mirai botnet population obtained from a website (Malwaretech 2016), which may or many not be an accurate representation for many attacks. Mirai has recently initiated only some of the known attacks. Also, these botnets have a dynamic nature, so they may grow or reduce in size with time. All these factors limit our estimations. Third, we assumed that the bot locations are known and do not vary. In reality, bots become alive when an infected system connects to the Internet (i.e. switched on) and disappear when the system is disconnected (i.e. switched off). Finally, we used Dijkstra's shortest path algorithm for path estimation, which is again a crude approximation of the BGP routing algorithm used on the Internet.

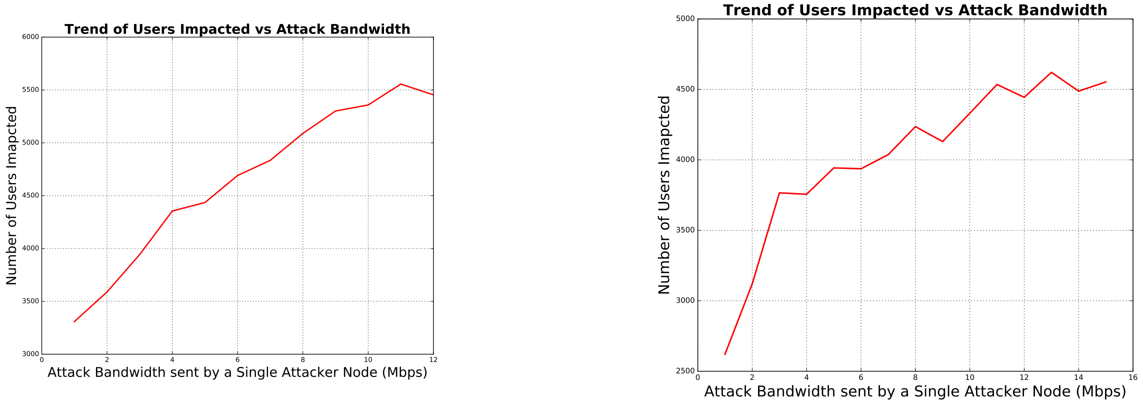

Figure 7: (Left) Simulation Result for attack on DNC server in the city of New York. (Right) Simulation Result for attack on AT&T server in Chicago. The plots how the trend of users impacted with increase in attack bandwidth assuming 10000 users are actively trying to connect.

7 CONCLUSIONS AND FUTURE WORK

In this research, we designed and implemented a network simulation model to understand the Internet traffic flow pattern in a DDoS attack situation. To keep the simulation simple, yet mirror the complexity of the Internet, we made certain assumptions that were reasonably justified. In particular, we combined all bots and systems connected to an IXP as one node, which allowed us to approximate the amount of attack traffic a node can generate, and the number of systems impacted if a node is experiencing traffic congestion. To approximate the traffic generated by bots, we used bots data from a security website, and to approximate the number of systems connected to an IXP, we used population density and the Internet penetration survey data. To make our network test environment more realistic, we used the fiber-optics map of the US from a recent research study. Using this novel network simulation test-bed, we simulated results for two different attack scenarios to understand the traffic flow as a function of attack-bandwidth. The traffic flow visualization enabled us to find the edges (fiber-optic cables) that are more prone to congestion in case of an attack. We also used real data from downdetector.com website to compare both simulation results and found a reasonably good correlation. We provided a list of assumptions that limit our study, but we hope that the approach we have used could be used by the Internet Infrastructure companies or the Homeland Security to better understand the Internet infrastructure vulnerabilities of the US.

In future, we would like to expand the experiment to a global scale. For this, we plan to include the undersea optical fiber map available at http://www.submarinecablemap.com/ for modeling the global landscape of Internet traffic flow.

REFERENCES

Betker, A., I. Gamrath, D. Kosiankowski, C. Lange, H. Lehmann, F. Pfeuffer, F. Simon, and A. Werner. 2014. "Comprehensive topology and traffic model of a nationwide telecommunication network". *Journal of Optical Communications and Networking* 6 (11): 1038–1047.

Durairajan, R., P. Barford, J. Sommers, and W. Willinger. 2015. "InterTubes: a study of the US long-haul fiber-optic infrastructure". In *ACM SIGCOMM Computer Communication Review*, Volume 45, 565–578. ACM.

DYN INC 2016, Nov. "DYN statement on DDOS attack". http://dyn.com/blog/dyn-statement-on-10212016-ddos-attack/.

Grunewald, D., M. Lützenberger, J. Chinnow, R. Bye, K. Bsufka, and S. Albayrak. 2011. "Agent-based network security simulation". In *The 10th International Conference on Autonomous Agents and Multi-agent Systems-Volume 3*, 1325–1326. International Foundation for Autonomous Agents and Multiagent Systems.

Inofsecurity 2016, Nov. "DDOS Attacks increase inn Size and Frequency". http://www.infosecurity-magazine.com/news/ddos-attacks-increase-in-size-and.

Kandula, S., D. Katabi, M. Jacob, and A. Berger. 2005. "Botz-4-sale: Surviving organized DDoS attacks that mimic flash crowds". In *Proceedings of the 2nd conference on Symposium on Networked Systems Design & Implementation-Volume 2*, 287–300. USENIX Association.

Kong, J., M. Mirza, J. Shu, C. Yoedhana, M. Gerla, and S. Lu. 2003. "Random flow network modeling and simulations for DDoS attack mitigation". In *Communications, 2003. ICC'03. IEEE International Conference on*, Volume 1, 487–491. IEEE.

Kotenko, I. 2005. "Agent-based modeling and simulation of cyber-warfare between malefactors and security agents in Internet". In *19th European Simulation Multiconference Simulation in wider Europe*.

Kotenko, I., and A. Ulanov. 2006. "Simulation of internet DDoS attacks and defense". In *International Conference on Information Security*, 327–342. Springer.

Kotenko, I. V., and A. Ulanov. 2005. "The Software Environment for Multi-agent Simulation of Defense Mechanisms against DDoS Attacks.". In *CIMCA/IAWTIC*, 283–289.

Mohit Kumar 2016, Nov. "More Insights On Alleged DDoS Attack Against Liberia Using Mirai Botnet". http://thehackernews.com/2016/11/ddos-attack-mirai-liberia.html.

Li, M., J. Li, and W. Zhao. 2008. "Simulation study of flood attacking of DDOS". In *2008 International Conference on Internet Computing in Science and Engineering*, 286–293. IEEE.

Malwaretech 2016, Nov. "Malwaretech". https://intel.malwaretech.com.

Omer, M., R. Nilchiani., and A. Mostashari. 2009. "Measuring the resilience of the global internet infrastructure system". 156–162. IEEE.

Omer, M., R. Nilchiani, and A. Mostashari. 2009. "Measuring the resilience of the trans-oceanic telecommunication cable system". *IEEE Systems Journal* 3 (3): 295–303.

Ottis, R. 2008. "Analysis of the 2007 cyber attacks against estonia from the information warfare perspective". In *Proceedings of the 7th European Conference on Information Warfare*, 163.

Qwasmi, N., F. Ahmed, and R. Liscano. 2011. "simulation of ddos attacks on p2p networks". In *High Performance Computing and Communications (HPCC), 2011 IEEE 13th International Conference on*, 610–614. IEEE.

Stinson, E., and J. C. Mitchell. 2007. "Characterizing bots remote control behavior". In *International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment*, 89–108. Springer.

Whitehouse.gov 2016, Nov. "Mapping the Digital Divide". https://www.whitehouse.gov/sites/default/files/wh_digital_divide_issue_brief.pdf.

Zhang, M.-q., J. Xie, M. Zhang, and X.-l. Zhang. 2008. "Modeling and Simulation of DDos Attacks Using OPNET Modeler [J]". *Journal of System Simulation* 10:055.

AUTHOR BIOGRAPHIES

Sumeet Kumar is a PhD student in the Institute for Software Research at Carnegie Mellon University. He is interested in simulations, cyber security, data mining and network analysis. He received an MS degree in Software Engineering from Carnegie Mellon University in 2013, and a B.Tech in Aerospace Engineering from the Indian Institute of Technology Kanpur (IIT Kanpur) in 2007. His email address is sumeetku@cs.cmu.edu.

Kathleen M. Carley is a professor in the School of Computer Science in the department - Institute for Software Research - at Carnegie Mellon University. She is the director of the Center for Computational Analysis of Social and Organizational Systems (CASOS), a university wide interdisciplinary center that brings together network analysis, computer science, and organization science. She and her lab have developed infrastructure tools for analyzing large scale dynamic networks and various multi-agent simulation systems. Her email address is kathleen.carley@cs.cmu.edu.