Principia College

DDoS Attacks, Mitigations, and Known Vulnerabilities

By
Zachary Matthiesen

Spring 2018
Computer Networking 333
John Broere

Abstract

The writer of this paper is interested in learning about the various ways to thwart DDoS attacks and DoS attacks, and to understand where these attacks originated. In order to do so, the author conducted research to determine a historical timeline for DDoS vulnerabilities and attacks. Also researched: mitigations developed by researchers, and possible future developments in the field pertaining to new vulnerabilities. Digging deeper, several questions arose: What are the vulnerabilities associated with DDoS attacks? What mitigations exist for these vulnerabilities? Are these mitigations still relevant in today's landscape? How effective are these mitigations against real and theoretical attacks? Ultimately, mitigations exist for most of the vulnerabilities exposed by research, and even older mitigations remain relevant in today's networking environment, especially considering the current cyber warfare climate. As to the effectiveness of each mitigation, there were few definitive standalone solutions, and most vulnerabilities could only be addressed via scanning, filtering, monitoring and even artificial intelligence to drop illegitimate packets and ignore spoofed IP addresses.

Background

Research on DoS attacks coalesced around this definition by Loukas and Öke: "A Denial of Service attack (DoS) is any intended attempt to prevent legitimate users from reaching a specific network resource" (1). More complex and specific than a DoS attack is a DDoS attack: "A distributed denial of service attack (DDoS) is a DoS attack where attack action is distributed among several entities which act in coordination against the target" (Costa Gondim et. al. 4). Botnets (the distributed part of a DDoS attack) consist of computers, usually loaded with some type of malware that allows the attacker to issue a flood of requests against a server, that overwhelm the server's resources and cause a DoS event (Saravanan et. al. 511). NICs are network interface cards, and they handle network traffic packet switching before the traffic finds its way to the server (Palmieri et. al. 1627). NICs are sensitive to spikes in network traffic, and this sensitivity can have a knock-on effect on CPU performance server side (Palmieri et. al. 1627 - 1628). Additionally, digital signatures, as defined by Satzinger et. al. are important to understand when referring to user sessions and authentication: "[A] Digital Signature [is] a technique in which a document is encrypted using a private key to verify who wrote the document [or request]” (605). On the other hand, a Digital Certificate is a public key exchange that used by the client to start a secure session with a server (Satzinger et. al. 605). Hop Count Filtering is the process of verifying the source of a packet by measuring TTL (time to live, aka the time a packet takes to reach its destination from host to server) and comparing it to the minimum time it would theoretically take a packet from that source address to reach the server (based on the number of “hops” from network node to network node that the packet takes to make the journey), thus verifying if the packet is coming from a spoofed IP address (Loukas and Öke 8). CAPTCHAs are “Completely Automated Public Turing Test to Tell Computers and Humans Apart” used to prevent a bot net from overwhelming the heart of server’s activities (9). However, CAPTCHAs are vulnerable to attacks by AI enabled bots and have a negative effect on the visually impaired (9). The basic principle behind the verified human vs. machine traffic is sound, however (9). Aggregate based Congestion Control (ACC) is based on current wisdom, which seeks to drop packets contained within a generous threshold of the presumed attack, and thus may unfairly drop legitimate packets (10).

The research on DoS attacks centered on three principle methods: real world testing, mathematical models and historical case studies. Palmieri et. al. used the strategy of real world testing on a small rig to draw conclusions about the effects of energy focused DDoS attacks on large data centers’ power consumptions (1629). The testing for each energy-oriented attack consisted of a small, hyper-threaded, LAN (Local Area Network) connected, PC and laptops connected to it on a LAN to simulate attacks (Palmieri et. al. 1629). Saravanan et. al. Used a similar method for testing mitigations: one botnet, with a flooding attack, a secure system log and fewer live bots than total users (516). They used the scalability of their rig to infer a possible double-bad result after accounting for above average HVAC (heating, ventilation, and air conditioning) costs and server specific infrastructure damaged by an energy focused DDoS (Palmieri et. al. 1629). Both Saravanan et. al. and Loukas and Öke conducted research and testing based on historical data. Saravanan et. al. used mathematical modeling of data sets from the 1998 FIFA World Cup, and Loukas and Öke used generalized historical data to generate mathematical models from DoS based attacks (518; 1). In contrast, Costa Gondim et. al. utilized a “simulated network”, based on a unique vulnerability challenge exercise, to determine what the best general mitigation practice would be for a network under DDoS attack (21). Overall, each researching team examined historical and statistical data from models and prototypes to make sound, scientific conclusions about DDoS or basic DoS.

A brief overview of the history of DoS attacks answers the questions: did DoS attacks peak in the 2000s? If DDoS largely disappeared from the public eye (Patrikakis et al. 2004), then where is it headed in the future? Loukas and Öke agree with Patrikakis et. al. and further claim that: "Since 2004, DoS incidents have been deliberately not widely publicized, as the scene has shifted to the sensitive Feld of economic crime and DoS incidents harm the victims' reputation in the eyes of the increasingly security-aware public" (4). In summary, DDoS may have disappeared from the headlines, but the attacks have not disappeared from the face of the earth and may even increase in frequency as networks increase in complexity.

Security issues can become a finger pointing exercise, but the aggressiveness of the adversarial hacker can render a lack of action fatal. A quality threat mitigation must deliver an effective solution in a short amount of time, with minimal cost, and be compatible with the philosophy of the design inherent to the product. In conclusion, according to multiple sources, the effectiveness of these attacks peaked in the 2000s, but they may experience a resurgence in the current era.

The Foundations of DDoS Vulnerabilities

DoS, as previously mentioned stands for Denial of Service attack. A distributed DoS (DDoS) attack is: "a huge number of remotely controlled machines can be used as the origin of multiple simultaneous attacks against a single target or a whole organization" (Palmieri et. al. 1625). There exist two basic types of DoS attacks: volumetric attacks vs. protocol attacks (Costa Gondim et. al. 3). Protocol attacks, in this author's opinion, peaked before web 2.0 standards launched and HTTP 5 became an accepted standard. However, this does not preclude them from resurging if vulnerabilities appear in a future revision of network protocol. Volumetric attacks are likely to remain relevant because the nature of networks as an interconnected system of nodes are vulnerable to traffic-spike based attacks.

Costa Gondim et. al. uses a broad definition of attacker and attack, as is necessary for consideration of DoS style attacks (3). Botnet and zombie-net style attacks are just a two of the more common attacks, and an attacker can use a myriad of viruses, trojans and scam ware to control host computers in a Botnet or zombie-net formation against a particular network infrastructure (Costa Gondim et. al. 5). These control structures can launch attacks against vulnerable servers because each infected computer sends requests without alerting the server to the location or identity of the attack coordinator (Costa Gondim et. al. 3). When an attacker uses many computers to distribute the task of attacking a server, it is said to be distributed and volumetric, thus the name DDoS (Costa Gondim et. al. 3).

The research conducted by Saravanan et. al. provides two examples of DDoS attacks: TCP SYN flooding and HTTP flooding (510). SYN flood attacks as not specifically the problem with energy related attacks, but rather “Continuous Transaction” attempts over HTTP, and can be more acute if they contain SSL and HTTPs service requests, because these are more CPU intensive (Palmieri et. al. 1626). TCP SYN flooding seemed to be a 1990’s phenomenon, based on the research done by Loukas and Öke (2). However, it is possible that TCP SYN flooding continues to plague low security data centers. As with other types of infrastructure, the internet often does not progress in a uniform manner towards greater security, and software security engineers still chatter about the lack of protection from SQL injection attacks on surprisingly large websites.

In order to set the stage for a discussion of DoS attacks, consider the history of DoS attacks from the past three decades. Loukas and Öke collected major DoS attacks through the mid 2000s and the highlights of their research creates a sturdy familiarity for DoS moving forward (1). The first few DoS attacks occurred in the 1980s, but over time the attacks became more and more complex and distributed (Loukas and Öke 1). In 1996, the first SYN flood attack occurred on commercial operations, laying the groundwork for future attacks focused, not on stealing user data or compromising privacy of users but on simply preventing the smooth operation of network systems and possibly demanding ransoms to abate the relentless traffic spikes (2). Later, in 1997, an young hacker (a “script kiddie”) attacked Undernet and ISPs (internet service providers) from several countries, using ping and SYN Floods (IRC-based DDoS attack) (2). Loukas and Öke reported that in 1998 “DALnet and other IRC networks became targets of “smurfing”, where the attacker is using ICMP echo request packets directed to IP broadcast addresses from remote locations to generate DoS attacks” (2).

By the turn of the century, DoS attacks had moved on to a global stage, with larger consequences. 2000 Mafia boy from Canada attacks e-commerce sites caused 1.7 billion in damages and caused DDoS to truly enter the mainstream (Loukas and Öke 3). 2001 Distributed Reflector DoS attack, where attackers compromise computers to generate connection requests to legit computers and spoofing the source of these packets at the target's IP address (like throwing a boomerang) (Loukas and Öke 3).

By 2001, DDoS reached peak destructiveness, and global conflicts became cyberwars using DDoS attacks. Code Red worm launches DoS attack against the White House generated 2.6 billion in damages and, port of Houston in the US had its system disabled by a DDoS attack, which was the first known cyber-attack on national infrastructure (Loukas and Öke 3). In 2002 India hacks Pakistan with a worm and DoS end effect YAHA (3). In 2002 DNS root servers attacked, in response a mitigation called Anycast which removed the treat, after a follow-up attack against DNS and ICANN failed (3). 2003 SQL slammer worm disrupts global infrastructure 60% of South Korea lacks firewall and gets smashed, and a nuclear power plant in Ohio has safety disabled for 5 hours (3). In 2004 HTTP flood and spidering attacks developed to saturate HTTP traffic and spidering which takes the Wikipedia game automated to generate recursive requests by link by link (4). In 2006, a 500,000 computer-strong botnet became available on the black-market, and malicious entities could rent the botnet to conduct DDoS attacks (4). In 2008, DoS attacks used in the war between Georgia and Russia escalated the conflict (4). DoS mitigations applied to today's infrastructure can easily defeat casual, script-kiddie type attacks but only more sophisticated mitigations (that are not available to every datacenter or server) can counter intelligent DoS attacks (4). For the moment, DoS attacks may have fallen out of the spotlight, but are by no means a forgone conclusion as a global network vulnerability.

The Present Status of DDoS as a Threat

Using anti-spoofing and filter rules at endpoints, servers mitigate volumetric DDoS attacks (Costa Gondim et. al. 4). Loukas and Öke claim a trade-off exists between broad and narrow attack mitigation from anomaly-based at the low end of accuracy to signature-based at the low end of portability (8). Unfortunately, ISPs need to communicate directly with victim servers to help filter traffic efficiently without the added overhead of proxy servers (Costa Gondim et. al. 5). To determine when a DDoS attack is occurring detection schemes look for, IP spoofing, abnormal traffic flow, classification of clients, and thresholds for service patterns (like "hot" and "cold" pages) are used (Loukas and Öke 1).

It is this writer's opinion that, in the modern day, a single user would be hard pressed to overwhelm an entire industrial scale server by themselves. To gain the upper hand against modern network solutions, an attacker must recruit significant external resources to help them launch the attack, similar to the current pattern of distributed systems, which require a distributed attack to disrupt (Saravanan et. al. 511). Botnets are the most common application level attack in the modern day (511). The first DDoS attacks were SYN floods using this method, where the attacker sends SYN messages to the server, but the attacker chooses to be unresponsive to the ACK message the server may send back, so the server has many half open connections, and eventually becomes swamped (Loukas and Öke 1). Also used at the time was a "fraggle" (in reference to the popular television show *Fraggle Rock* created by Jim Henson, similar to "smurfing") attack using UDP packets instead of TCP packets to cause the server to hang waiting for traffic to clear up, instead of generating half-open sockets (Loukas and Öke 2).

AR-DDoS attack uses an additional layer to overwhelm the target, as opposed to the conventional DDoS two-layer attack (Costa Gondim et. al. 5). This Third layer is the reflector layer, and simply repeats the bot net's attack over multiple channels, more effectively overwhelming the victim because they need fewer bots when more reflectors are available (5). The order of a AR-DDoS attack goes as follows: "… the attack master sends control information to intermediate layer hosts in order to coordinate the attack. Then, those hosts send probes crafted with the victim's address as source address to reflectors. When a reflector receives one of those probes, it sends a properly amplified reply..." (5). The application layer is a powerful place to launch DDoS attacks because they fly under the radar for some filtering schemes, and real users can see their packets dropped when traffic spikes from an attack (Saravanan et. al. 510).

No single mitigation can thwart an application level attack or detect a flash crowd DDoS, so servers implement approaches combining several mitigation techniques (Saravanan et. al. 512). Three primary methods exist to determine if a user is legitimate: if the user visits "hot pages" (web pages popular with human users: think of the front page of a newspaper as a "hot page" vs the "cold page" of the obituary section, which fewer people read on a daily basis), frequently create sessions, and have random request flows (511). In order to best ascertain which clients are legitimate, servers can separate the web pages into "hot" and "cold" pages by measuring the volume of requests to each page on a website on a normal (non-DDoS attacked) day (514). The clients that visit mainly “hot” pages are then determined to be legitimate, and clients with a “colder” page set than a weighted mean the server classifies as suspect (514).

Despite the vulnerabilities, there are some reasons why AR-DDoS may be easier to mitigate on IoT networks. IoT networks often have a topology with indirect connections to most nodes, and “Network Address Translation” acts similarly to an access filter on the IoT (Costa Gondim et. al. 5). Thus, IoT networks can resist plain DDoS attacks by decentralizing traffic and avoiding IP spoofing (5). However, these resilient properties are not enough, in this author’s opinion to preclude the existence of significant attack vectors against IoT networks. Additionally, Costa Gondim et. al. counterpoises the previous argument of resilience with a comment that “AR-DDoS is chosen among other forms of DDoS for its relatively low complexity and attacker effort, large availability of potential reflectors, and high efficiency” (9). This implies an AR-DDoS attack’s ability to negate many of the structural resiliencies of IoT with its ease of attack and built in design flaws.

Amplified Reflection DDoS are the most common attacks since 2013 (Costa Gondim et. al. 2). Interestingly, given the time of publication on the article as 2015, Palmieri et. al. described DoS attacks as “…an ever-increasing menace for corporate and government organizations…” (1624). Costa Gondim et. al. cites the academic discussion on IoT centered around the growing threat the IoT environment poses as a security risk, because of the bare-bones nature of the networking framework underlying the technology (3).

While AR-DDoS has become more common, coercive parsing can also drain CPU availability quickly by exploiting XML’s verbose structures and tags, and SOAP’s embedded security blocks in the message body that parse with deep complexity (Palmieri et. al. 1626). These requests are difficult to filter, because they do not require such massive numbers as a normal DDoS attack so the ramp up may be indistinguishable from normal high level of traffic (1626).

Server farms and data centers do not experience perfectly consistent energy draw, because of random “flash crowd” events and DoS attacks can cause the energy usage to shoot above threshold as previously discussed (Palmieri et. al. 1624). An additional vulnerability, currently in a state of partial mitigation, is a flash crowd DDoS or AR-DDoS attack A flash crowd occurs when the server experiences extremely high legitimate traffic suddenly, as with a popular web page going viral (Saravanan et. al. 510). Cloud computing led to a more distributed computing environment generally, but this caused larger and larger data center construction, and thus greater stress on the power grid, especially during flash crowd attacks, which can cause outages (Palmieri et. al. 1621).

The Future of DDoS and DoS Style Attacks

Recently, new re-inventions of the internet have gained popularity (and have entered into common knowledge) including the internet of things (IoT) and the so-called “cloud” where computing power distributed by centralized servers in large data centers take care of the heavy-duty processing (Costa Gondim et. al. 1). However, the IoT is susceptible to cyber-attack because the lightweight standards developed for IoT do not strongly mitigate AR-DDoS attacks, (1). As with most new innovations, the IoT probably needs more time to develop robust standards of network communication. But, for the time being, manufacturers and end-users must cooperate and develop effective patches and mitigations. Costa Gondim et. al. used case studies to explore AR-DDoS and non-amplified DDoS using the IoT as a vector for the attack because IoT networks are vulnerable (6). Phases of the test used to determine threat significance in IoT space Task Specification, Tactical Definition, Initial Scanning, Target Selection, execution, Reporting, Analysis and Reporting (9).

On top of protecting against AR-DDoS, server infrastructure faces a slew of more subtle attacks. The crack in the modern armor of a network solution is flash crowds because they cannot detect the difference between request flows (Saravanan et. al. 512). One such subtle attack occurs in the context of a flash crowd, when the traffic patterns of a flash crowd parrot back at the server in enough volume to overwhelm the system (511). Source IP address distribution and traffic speed set application level DDoS apart from legitimate flash crowds, so filters correctly configured with IP address tables and frequency counters may stop malicious traffic from reaching the server (511). IP address distribution is limited when a DDoS attach occurs, because of the botnet’s closer correlation to one area (515). Mitigations based on IP address distribution trace their origin to mitigations from 2009 where detecting an anomaly in traffic volume or signature composition could prompt a response from the server to drop suspicious packets (Loukas and Öke 4). The character of request flows using the Hellinger distance, referred pages and a client analysis. Even in a flash crowd, the Hellinger distance is significant among legitimate user request flows, and small amount botnet request flows (Saravanan et. al. 513). One more helpful mitigation is to perform a trace-back on the suspicious traffic to find out from where it is coming (Loukas and Öke 5).

The first mitigation measure used against DoS attacks was a traffic threshold connection reset, but its early versions did not prevent attacks with randomized IP addresses at a high traffic flow (Loukas and Öke 2). Additionally, the classic method of using a secure system log to do the pre-processing work for determining user behavior (hot pages) and returning users provides an effective mitigation for DoS vulnerabilities (516). To determine the threshold usages of each webpage, the mean traffic to each page is used, in conjunction with the frequency of the traffic (516). Generally speaking, the preparation phase of the application level DDoS mitigation consists of keeping tabs of general user frequency to each page using averaging, and distances calculations are made at zero hour when a crowd flash occurs (517).

In the real world, energy focused DDoS attacks could be launched between rival corporations or governments, causing profit margins to dissipate and infrastructure to fail (Palmieri et. al. 1639). A proxy server is the best way to detect all factors for pre-sorting malicious packets to prevent server overload (Saravanan et. al. 515). Proxies (more than a few) loaded with the filtering data from the previous step, send legitimate requests to the server in turn (518). However, the more proxies do not scale well compared with servers: “Too small a number may lead to request drops due to the overwhelming number of incoming requests, and higher numbers of proxies may lead to processing overhead and poor detection rate” (518). New attacks are targeting data center electrical bills, and aim to cause outages and downtime (Palmieri et. al. 1621). Packets must be determined to be legitimate or illegitimate, and treated accordingly with either a strait drop or a redirection to continue vetting if the system has the capacity (Loukas and Öke 4).

Power consumption in data centers primarily originates from CPUs (25 - 55%) and HVAC (Heating, Ventilation and Air-conditioning) systems (up to 38%) (Palmieri et. al. 1622). Electrical power companies charge a flat rate to data centers, as long as they do not exceed a particular threshold of usage (as measured in kilowatt or megawatt hours) (1623 - 1624). Because of the costs associated with electrical usage, it is important to dispatch packets of malicious intent as far downstream from the server as possible before they limit a server's CPUs' capacity and force them to do simple -- but high volume -- packet filtering (1624).

Mass storage memory devices can experience failure when repeated DoS style attacks use specialized requests to drain the read/write endurance on memory cells in SSDs and the cache on hybrid magnetic drives (Palmieri et. al. 1627). Hash collisions from the algorithms in uses from network applications and scripting can "…cause worst-case behavior in hash table usage" and thus degrade the read/write endurance of drives (Palmieri et. al. 1626).

To prevent a DoS attack, filters examine the IP addresses of incoming packets, the frequency of the traffic, and then use of baseline readings on the network to establish normal thresholds of usage (Loukas and Öke 5). Because bots tend to request both hot and cold pages, but human users tend to access mostly hot pages, especially during a flash crowd, the filtering mechanisms on the server side must be sensitive to the behavior of each client in a session (Saravanan et. al. 514). Precedence to regular users should be the top priority in this situation, as usual, regular users care the most about access to their favorite content on a server (Saravanan et. al. 515). Secure Overlay Services (SOS) route only the authenticated users' traffic to the server and drop everything else (Loukas and Öke 10).

The victim of a flash crowd can usually determine ahead of time which pages are hot, and program that into the server (Saravanan et. al. 514). Client Loyalty, QoS honors (if a client has made an agreement to provide credentials in return for a particular quality of service) and IP time of appearance because most attacks take time to "ramp up" so the bot net takes time to maneuver into peak traffic flow, allowing for a brief window of observation to occur before the flood becomes overwhelming to the server (Loukas and Öke 5). Triggering the anti-virus systems and anti spam services on the server can also drain power efficiency, and actual viruses and worms on the user side can subtly prod the server to increase power consumption (Palmieri et. al. 1627 - 1628).

Conclusion

Drawing together the research available on DDoS, the future appears to be energy focused DDoS, as discussed by Palmieri et. al. and IoT vulnerabilities, as described by Costa Gondim et. al. Some of the components to these attacks can be seen in the progression of DoS and DDoS over time, as outlined by Loukas and Öke. The study of energy usage during DDoS attacks concluded that a spike of up to 1360 kWh would occur (Palmieri et. al. 1636). This implies the destructive potential of such attacks is great, but not on par with the multi-billion-dollar disasters of the early 2000s (Loukas and Öke 1).

What were the vulnerabilities associated with DDoS attacks? The IoT showed vulnerabilities in and large server farms showed vulnerabilities because of their energy usage, and Loukas and Öke described vulnerabilities with servers historically, including overlapping vulnerabilities due to flawed web services.

What mitigations now exist for these vulnerabilities? Most of the sources surveyed indicate that packet filtering and trace-back to the source of the packets involved in suspicious activity are the most reliable mitigations for normal DDoS attacks, with variations on the method for IoT and energy specific needs. It is likely that the mitigations proposed by Palmieri et. al. and Costa Gondim et. al will continue to be relevant in the modern landscape because they conducted research within the last 5 years, and they used careful data analysis to support their findings.

Works Cited

Costa Gondim, João, et al. “A Methodological Approach for Assessing Amplified Reflection Distributed Denial of Service on the Internet of Things.” *Sensors*, vol. 16, no. 11, Nov. 2016, p. 1855. *CrossRef*, doi:10.3390/s16111855.

Loukas, Georgios, and Gülay Öke. “Protection against Denial of Service Attacks: A Survey.” *The Computer Journal*, vol. 0, no. 0, May 2009, p. 19, doi:0.1093/comjnl/bxh000.

Palmieri, Francesco, et al. “Energy-Oriented Denial of Service Attacks: An Emerging Menace for Large Cloud Infrastructures.” *The Journal of Supercomputing*, vol. 71, no. 5, May 2015, pp. 1620–41. *CrossRef*, doi:10.1007/s11227-014-1242-6.

Patrikakis C, Masikos M, Zouraraki O. 2004. Distributed Denial of Service Attacks - The Internet Protocol Journal. The Internet Protocol Journal. 7:13–36.

Saravanan, Renukadevi, et al. “Behavior-Based Detection of Application Layer Distributed Denial of Service Attacks during Flash Events.” *TURKISH JOURNAL OF ELECTRICAL ENGINEERING & COMPUTER SCIENCES*, vol. 24, 2016, pp. 510–23. *CrossRef*, doi:10.3906/elk-1308-188.

Satzinger J, Jackson R, Burd S. 2009. Systems Analysis & Design in a Changing World. Fifth. Boston: Cenage Learning.