21 Years of Distributed Denial-of-Service: Current State of Affairs

Eric Osterweil and Angelos Stavrou, George Mason University
Lixia Zhang, University of California, Los Angeles

The Internet's features and capacity have evolved, but is the nature of its security noticeably better? We examine the fundamental nature of distributed denial-of-service (DDoS) and the state of the union of our defenses in today's DDoS wars.

In 1999 (21 years ago), malware called Trin00[1] compromised a set of computers and then took down a network at the University of Minnesota. This event marked the birth of volumetric distributed denial-of-service (DDoS) attacks from robot networks (botnets). While earlier attacks exist in anecdotes and recollections, this documented case sets a lower bound on the date of birth: 21 years. The features and capacity of the Internet have evolved a lot since then, but is its security disposition demonstrably better? Trin00 used hundreds (possibly thousands) of compromised machines (bots), but today conventional botnet sizes have been seen in the millions. In relative terms, Trin00 may not seem like such a large botnet. However, this underscores that historical attack sizes are relative, and raw numbers alone do not tell the tale. Moore's law and bandwidth increases makes comparing attack volumes (bits per second) from the past to today (or tomorrow) apples-to-oranges comparisons. Consider that gigabit attacks in 2000 were considered staggering, but only because they rivaled the capacity of the infrastructure of the time. An unfortunate state of affairs is that it has always been easier to gain attack capacity than defensive capacity.[2] DDoS is an asymmetric threat with an impedance mismatch between attackers and defenders.

The gap between adversaries' barriers to attack and the price to defend has always been large, but it is growing, and the status quo does not paint a pretty picture for the future

Digital Object Identifier 10.1109/MC.2020.2983711
Date of current version: 1 July 2020

EDITOR JAMES BRET MICHAEL
IEEE Senior Member; bmichael@nps.edu

of Internet service security. In this article series, we want to sound an alarm and issue a call to action; we must discover the fundamental enablers of DDoS, and we must use these to craft efficient defenses. We feel it is time to reexamine the principles that underlie the problem space. In this two-part article, we begin by examining the fundamental nature of DDoS: reasons why our networks are susceptible, the anatomy and nature of today's DDoS attacks, and the state of the union of our defenses in today's DDoS wars. In our article's second part (in the August issue of *Computer*), we explore remediations and the evolution needed to systematically enhance the Internet and address the principles that enable DDoS.

WE ARE VICTIMS OF OUR OWN SUCCESS

The Internet has blossomed with complex and diverse network applications and services that bind our social lives, implement complex tasks, and facilitate communications, all while streamlining end users' experiences. Critical to this success has been protocol layering and abstraction (where protocols encapsulate and obscure their state from each other). Network applications sit above the transport layer, which sits above the network layer. Indeed, layering has been a central tenet of the Internet's evolvable architectures. However, it also has hidden vulnerabilities that many DDoS attacks now capitalize on. As defenders against DDoS attacks, our fundamental challenge is the onus of scrubbing attack traffic away from legitimate traffic, using deep packet inspection (DPI). The distinction of which traffic is part of an attack is often only visible at the application layer (above both the network and transport layers). For example, what network-layer information should management tools use to determine which Domain Name System (DNS) queries are real and which are participating in a reflection attack? Which Network Time Protocol (NTP) command is legitimate and which is part of an attack? Must it fall to the application programmer and operators to build custom logic to differentiate whether a `memcached`[3] query is from a genuine application or part of an attack? Or which HTTP client is trying to keep a needed connection alive and which is starving the server for resources?

Compounding the opacity across layers, network traffic is now often encrypted. A recent operational report of large-scale measurements stated that the Secure Sockets Layer (SSL) "is [sic] majority of traffic in [North America] by February 2019."[4] The necessary computational complexity, volume of traffic, and growing use of encryption often render common operational network tools ineffective in defending against attacks. When application payloads are embedded (that is, encrypted), they require multiple layers of computationally expensive decoding while exposing sensitive material. For example, performing DPI on an HTTPS flow requires decryption of the flow. Further, that also requires escrow of the end site's Transport Layer Security (TLS) private key (to terminate and inspect the embedded flow). Internet protocol layering and encryption have severely complicated scrubbing at the network layer. In short, this is high cost and low return, and it is time to investigate the fundamentals of this problem space.

A GLIMPSE AT THE ANATOMY OF DDOS

DDoS comes in a variety of flavors. Some are called "low-and-slow," which starve servers of resources and can be hard to detect. Some are volumetric, which send overwhelming volumes of traffic that congest network links and overload servers and are hard to stop even though they are detectable by nature. In this article, we focus our discussions on these two types.

Our fundamental challenge is the onus of scrubbing attack traffic away from legitimate traffic.

Volumetric DDoS today

Today, DDoS threats are asymmetric: it is virtually free for attackers to acquire massive network capacities for their DDoS attacks, and they frequently use multiple techniques, tactics, and procedures (TTPs) at the same time. By contrast, detecting and mitigating DDoS attack traffic (for example, "packet love") requires investment in expensive infrastructure and network bandwidth (capacity). The largest recorded DDoS attacks have used source address spoofing (such as sending packets with deliberately falsified addresses) as part of their TTPs. One form of spoofing attack amplifies its volume by bouncing (or "reflecting") small queries off of Internet services to elicit larger ("amplified") responses, which are then reflected to spoofed addresses (in other words, victims). These are called *reflective amplification attacks* or just *reflector attacks*. For example, in 2016, the first publicity around a terabit attack came from an assault on a hosting provider called OVH,[5] and it reached this volume by using spoofed addresses in a reflector attack. A larger attack on Dyn[6] surpassed this volume, again in 2016, using source address spoofing. In short, the largest DDoS attacks seen today depend on address spoofing as part of their TTPs, though they have not always leveraged an amplification factor.

The increasingly relative ease of acquiring large volume attack sources has elevated the appeal of volumetric DDoS attacks to adversaries. Traffic may be DNS queries, Simple Network Management Protocol queries, NTP queries, `memcached` queries,[3] or others. Some attacks also use spoofed Transmission Control Protocol (TCP) control traffic carrying large data payload or use the TCP session itself as an amplification vector by orchestrating torrents of reset packets or data payloads via the TCP PSH option.

Server-side resource exhaustion attacks

While many headlines and defenses focus on the size of DDoS attacks, there continue to be many attacks above the transport layer. Common examples of such attacks leverage protocol aspects in the SSL, TLS, or even at the HTTP/S layer. These nonvolumetric attacks can also be crippling without a DDoS defense system and can bring Internet services down with far fewer resources.

Perhaps the earliest known resource exhaustion attacks were those that abused the TCP itself, SYN flood attacks. These DoS attacks have been used in the wild since at least 1996,[7] though they were not always distributed. Attacks like these were initially intended to exhaust servers' resources and were neither volumetric nor stealthy (low and slow).

One of the early examples of low-and-slow attacks was Slowloris,[8] where a relatively small number of stateful HTTP queries would hold connections open on webservers and thereby exhaust their ability to answer other (legitimate) clients. Other exhaustion attacks exploit TLS's cryptographic key negotiation.[9] In these types of attacks, the raw numbers of attacking clients and traffic are not as spectacular as volumetric DDoS, but perhaps more troubling is the fact that their detection and remediation more clearly requires additional state information above the network layer.

MITIGATION: STATE OF THE UNION, TODAY

Mitigation providers often do distribute their services, often called *scrubbing centers*, around the world and across the Internet's topology. However, with attack sources sometimes numbering in the millions, scrubbing centers each inevitably need to mitigate attack traffic from growing numbers of well-provisioned botnets. Scrubbing uses DPI, thereby adding computation overhead to the network/transit overhead. This frames the fundamental impedance mismatch: distributed attacks versus relatively centralized mitigations. As just an illustration, we present three examples that the state of the Internet can be categorized and evaluated: architectural, volumetric, and economic.

> The increasingly relative ease of acquiring large volume attack sources has elevated the appeal of volumetric DDoS attacks to adversaries.

Fundamentals of the state of the union

Our reliance on DPI for detection and remediation has resulted in increasing dependence on keeping our defenses in large computation/network capacity data centers. For example, with reflector attacks leveraging application-level semantics (for example, NTP's monlist and memcached's GET) and the increased use of TLS, terminating and interpreting traffic has necessitated backhauling traffic to DPI in scrubbing centers. This has framed a fundamental asymmetry: large volumes of attack traffic from more sources with increasingly better provisioned networks versus fewer and centralized remediation. This asymmetry is further exacerbated by the increased complexity of web applications and use of encryption. Our mitigation techniques are predicated on matching mitigation bandwidth to ever-growing aggregate distributed attack volumes, and we need a different/more distributed solution. For example, in 2015, the Defense Advanced Research Projects Agency announced a call for "Extreme DDoS Defense" that included a solicitation to "[disperse] cyber assets (physically and/or logically)."[10]

Some techniques to disperse network-based remediation focus on using network-layer routing, like Border Gateway Protocol's (BGP's) FlowSpec, remotely triggered black-holing, and others. However, without the necessary application-level expressiveness, this can unfortunately lead to collateral damage to well-behaving (nonattack) sources that happen to be on the same network (such as in the same BGP prefix) as attackers.

Another network-layer defense, called Internet Protocol (IP) anycast, uses BGP routing to replicate services. Anycast allows operators to position services near clients and provides redundancy. However, Internet Architecture Board RFC 7094[11] describes some known limitations: "IP control packets from a DNS client may initially be routed to one anycast instance, but subsequent IP packets may be delivered to a different anycast instance." Recent work[12] examined the DNS anycast root server system while under sustained DDoS and concluded that there is a "need to understand anycast design for critical infrastructure, paving the way for future study in alternative policies that may improve resilience."

Volumetric state of the union

The volumetric state of the union—volumes of attack traffic versus carriers' and providers' provisioned capacities—paints a similarly disconcerting picture. Service providers (SPs) buy transit in gigabits per second (Gbit/s) links in multiple locations from multiple carriers. Internet exchange points and carrier capacity are also often offered in Gbit/s. Large carriers' global aggregate capacity may approach, and in some cases achieve, terabits per second (Tbit/s). However, this does not mean any given ingress point to a carrier's network is itself a Tbit/s link. Generally, aggregate capacity in Tbit/s is a summation of router/regional capacities (Gbit/s). However, the aggregate attack traffic of the largest DDoS attacks is already over 1 Tbit/s. In an aggregate view, a recent observation from operational measurements quotes that "attacks [are] growing in size faster than network growth."[3]

Unfortunately, often the aggregate capacity is not near attack sources, and it can be topologically very far from attack sources. While the volume of observed DDoS attacks has already crippled critical infrastructure, the potential sizes of attacks is far worse than anything that we have seen to date. "The Internet's capacity attenuates the total throw weight a DDoS attack can generate; the farther a target is from components of a network, the less traffic that will make it across any congested links between the target and the attack source."[13] In other words, this can result in service degradation and outages to other Internet services whose traffic shares congested routing infrastructure as they become collateral damage. This was also noted during the Spamhaus/Cloudflare DDoS of 2013.[14] When attack sources are topologically far from mitigation, their traffic is backhauled across transit and peering infrastructure to scrubbing centers causing terabits of attack traffic to potentially be routed to gigabit scrubbing centers. Even in the case of high-capacity scrubbing centers, the centralized nature of the mitigation enables attack traffic to permeate network links far from the sources of attack.

The largest DDoS attacks that we have seen are already larger than the provisioned capacity of many of the large providers' and carriers' capacities. In 2016, the U.S. Department of Homeland Security started a program called DDoS Defense, whose starting position was that "one day" DDoS could swell to 1 Tbit/s.[15] By 2017, the largest DDoS attacks had already reached that, and in 2018 DDoS attacks quantifiably exceeded that, as shown in Figure 1.[16] Recent work has estimated that the Internet-wide capacity to launch volumetric reflective amplification DDoS attacks is "two orders of magnitude larger than the Dyn attack."[17]

Economic state of the union

Using SPs' outlays to protect against DDoS also paints a grim picture. In 2000, DDoS attacks on Yahoo, eBay, and several other major Internet services led the news and raised alarms. Now, almost 21 years later, protection rackets exist in gaming spheres. Online gaming and gambling sites are frequently held hostage for ransom by DDoS threats,[18] and sometimes attacks are launched simply to gain gaming advantages. Generally, all online services today need DDoS protection, and companies expect to pay for defensive protections against inevitable DDoS attacks. The DDoS mitigation market was US$1.94 billion in 2018 and is growing. Furthermore, there has also been a DDoS-for-hire (sometimes known as a *booter*) grey-market for roughly a decade.

The motivations for launching DDoS attacks can be diverse. For example, in 2015 the hacktivism group Anonymous threatened to—and then did—launch a DDoS attack against the DNS root server system.[16] The stated goal of this attack was to disrupt all transactions on the Internet by rendering the DNS inoperable. While unsuccessful, this attack illustrates that sometimes DDoS attacks are launched to wreak Internet havoc and do not have a specific target.

> The motivations for launching DDoS attacks can be diverse.

ADDRESSING ROOT CAUSES

Internet providers and clients seek protection from DDoS attacks in advance of, during, and subsequent to them. However, there are no official authorities to enforce or remedy DDoS. There is no government mandate or Internet regulatory body that has the authority or is even in a position to offer remediation

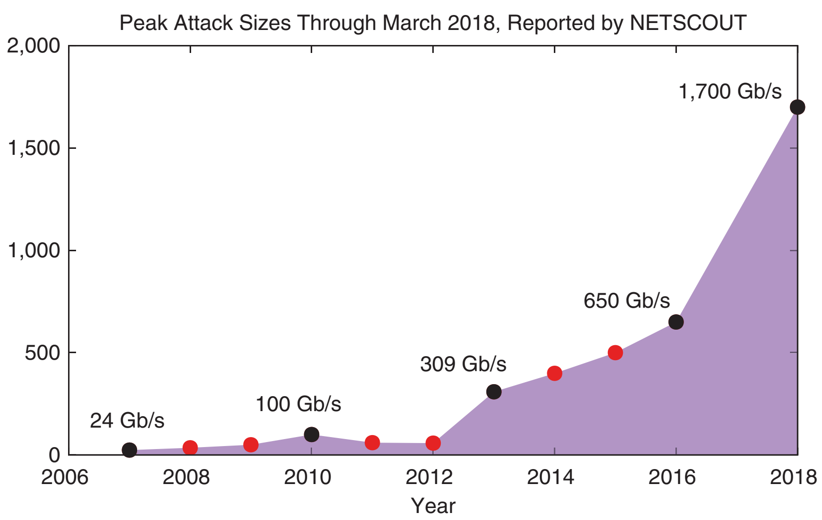

FIGURE 1. The peak attack sizes through March 2018.[16]

or help, and not everyone would want to bestow such global authority to an organization. As of today, we must pay for help, and we have privatized our defenses. This begs the unfortunate question: what would a proper remedy to DDoS attacks even look like? Considering that DDoS traffic often originates from multiple countries, may transit through separate jurisdictions, may lie about where it comes from (spoofing), and the fact that the Internet's infrastructure is operated by private corporations, is it even feasible that any entity could provide official remedies to DDoS attacks? We feel the biggest challenge, which must be addressed first, is to examine the root-cause vulnerabilities that enable DDoS attacks and then to develop mitigation techniques.

In next month's column, we will detail the approaches that are used to combat DDoS today, examining scenarios and conditions under which they perform well. We will also explore opportunities and directions for basic and applied research to address the fundamental attack vectors that DDoS attacks exploit. ■

REFERENCES

1. "CERT incident note IN-99-04," CERT Coordination Center, Pittsburgh, PA, 1999. [Online]. Available: https://web.archive.org/web/20081115163511/http://www.cert.org/incident_notes/IN-99-04.html
2. G. Huston, "Why is securing the Internet so hard?" in *Proc. Asia Pacific Regional Internet Conf. Operational Technologies*, 2019.
3. "Memcached DDoS explained," Akamai, Cambridge, MA. [Online]. Available: https://www.akamai.com/us/en/resources/our-thinking/threat-advisories/ddos-reflection-attack-memcached-udp.jsp
4. C. Labovitz, "Internet traffic 2009–2019," in *Proc. Asia Pacific Regional Internet Conf. Operational Technologies*, 2019.
5. E. Kovacs, "Hosting provider OVH hit by 1 Tbps DDoS attack," SecurityWeek, 2016. [Online]. Available: https://www.securityweek.com/hosting-provider-ovh-hit-1-tbps-ddos-attack
6. S. Mansfield-Devine, "DDoS goes mainstream: How headline-grabbing attacks could make this threat an organisation's biggest nightmare," *Netw. Secur.*, vol. 2016, no. 11, pp. 7–13, 2016. doi: 10.1016/S1353-4858(16)30104-0.
7. "CERT advisory CA-1996-21 TCP SYN flooding and IP spoofing attacks," CERT Coordination Center, Pittsburgh, PA, 1996.
8. R. Hansen, J. Kinsella, and H. Gonzalez, "Slowloris HTTP DoS," ha.ckers, 2009. [Online]. Available: https://web.archive.org/web/20150426090206/http://ha.ckers.org/slowloris
9. C. Castelluccia, E. Mykletun, and G. Tsudik, "Improving secure server performance by rebalancing SSL/TLS handshakes," in *Proc. ACM Symp. Information, Computer and Communications Security*, 2006, pp. 26–34.
10. J.M. Smith, "Extreme DDoS defense (XD3)," Defense Advanced Research Projects Agency, Arlington, VA. 2015. [Online]. Available: https://www.darpa.mil/program/extreme-ddos-defense
11. D. McPherson, D. Oran, D. Thaler, and E. Osterweil, "Architectural considerations of IP anycast," RFC-7094, 2014.
12. G. M. Moura et al., "Anycast vs. DDoS: Evaluating the November 2015 root DNS event," in *Proc. Internet Measurement Conf.*, ACM, 2016, pp. 255–270. doi: 10.1145/2987443.2987446.
13. "[state of the internet]/security: A year in review," Akamai, Cambridge, MA, vol. 4, no. 5, 2018.
14. T. Samson, "Spamhaus DDoS attack just another day for ISPs," *InfoWorld*, 2013. [Online]. Available: https://www.infoworld.com/article/2613993/spamhaus-ddos-attack-just-another-day-for-isps.html
15. "Distributed denial of service defense (DDoSD)," DHS Science and Technology Directorate, Washington, D.C., 2016. [Online]. Available: https://www.dhsgov/sites/default/files/publications/FactSheet%20DDoSD%20FINAL%20508%20OCC%20Cleared.pdf
16. C. Morales, "NETSCOUT Arbor confirms 1.7 Tbps DDoS avttack; The terabit attack era is upon us," NETSCOUT: Westford, MA, Mar. 5, 2018. [Online]. Available: https://www.netscout.com/blog/asert/netscout-arbor-confirms-17-tbps-ddos-attack-terabit-attack-era
17. E. Leverett and A. Kaplan, "Towards estimating the untapped potential: A global malicious DDoS mean capacity estimate," *J. Cyber Policy*, vol. 2, no. 2, pp. 195–208, 2017. doi: 10.1080/23738871.2017.1362020.
18. S. Mansfield-Devine, "The growth and evolution of DDoS," *Netw. Secur.*, vol. 2015, no. 10, pp. 13–20, 2015. doi: 10.1016/S1353-4858(15)30092-1.

ERIC OSTERWEIL is an assistant professor in the Department of Computer Science at George Mason University. Contact him at eoster@gmu.edu.

ANGELOS STAVROU is a professor in the Department of Computer Science at George Mason University. Contact him at astavrou@gmu.edu.

LIXIA ZHANG is the Jonathan B. Postel Professor of Computer Science at the University of California, Los Angeles. Contact her at lixia@cs.ucla.edu.