Defending Against Denial of Service Attacks

Version 1.3
Released: October 31, 2012

Securosis, L.L.C. 515 E. Carefree Highway Suite #766 Phoenix, AZ 85085 **T** 602-412-3051
info@securosis.com www.securosis.com

Author's Note

The content in this report was developed independently of any sponsors. It is based on material originally posted on the Securosis blog, but has been enhanced, reviewed, and professionally edited.

Special thanks to Chris Pepper for editing and content support.

Licensed by Corero Network Security

Corero Network Security (CNS:LN), an organization's First Line of Defense, is an international network security company and the leading provider of Distributed Denial of Service (DDoS) defense and Next Generation Intrusion Prevention Systems (NGIPS) solutions. As the First Line of Defense, Corero's products and services stop DDoS attacks, protect IT infrastructure and eliminate downtime. Customers include enterprises, service providers and government organizations worldwide. Corero's appliance-based solutions are dynamic and automatically respond to evolving cyber attacks - known and unknown - allowing existing IT infrastructure such as firewalls to perform their intended purposes. Corero's products are transparent, highly scalable and feature the lowest latency and highest reliability in the industry. Corero is headquartered in Hudson, Massachusetts with offices around the world. For more information, visit http://www.corero.com/.

Contributors

The following individuals contributed significantly to this report through comments on the Securosis blog and follow-on review and conversations (in no particular order):

Dan G
Robert Hansen
rybolov
Paul
Zhou Youan
Rakesh Shah

Copyright

Table of Contents

Introduction

For years security folks have grumbled about the role compliance has assumed in driving investment and resource allocation in security. It has become all about mandates and regulatory oversight driving a focus on protection, ostensibly to prevent data breaches. We have spent years in the proverbial wilderness, focused entirely on the "C" (Confidentiality) and "I" (Integrity) aspects of the CIA triad, largely neglecting "A" (Availability). Given how many breaches we still see every week, this approach hasn't worked out too well.

We have spent years in the proverbial wilderness focused entirely on the "C" (Confidentiality) and "I" (Integrity) aspects of the CIA triad, largely neglecting "A" (Availability).

Regulators pretty much only care whether data leaks out. They don't care about the availability of systems – data can't leak if the system is down, right? Without a clear compliance-driven mandate to address availability (due to security exposure), many customers haven't done and won't do anything to address availability. Of course attackers know this, so they have adapted their tactics to fill the vacuum created by compliance spending. They increasingly leverage availability-impacting attacks to both cause downtime (costing site owners money) and mask other kinds of attacks. These availability-impacting attacks are better known as Denial of Service (DoS) attacks.

For most security professionals DoS attacks aren't new. It may be hard to remember back over a decade ago, but in the heyday of the Internet bubble we saw many old-fashioned Distributed DoS (DDoS) attacks targeting high profile web properties (think Yahoo and E*Trade, back in the day), with attackers like Mafiaboy doing the damage more for notoriety than to cause real economic damage. Over the past decade attackers have reoriented toward financially motivated attacks, which means increasingly application-centric attacks designed to evade detection and exfiltrate lucrative data.

Obviously knocking down a target interferes with looting it, so this attack vector was deemphasized as the focus shifted to financial fraud. But DDoS never really went away – it merely became a supplementary extortion tactic. Attackers communicate with a company and promise to knock down their site unless they receive a ransom. It's a simple shakedown move, and many targets are simply unable to survive a significant outage — they pay up rather than fight. We don't hear about many of these attacks – nobody wants to publicize their vulnerability to shakedowns because it invites more shakedowns.

But that is all changing now. It's like Back to the Future a bit – the rise of hacktivism has brought the Denial of Service back into a prominent position in the nightmares of security folks. Facilitated by the availability of open source tools such as the Low Orbit Ion Cannon (LOIC) and numerous bot networks to launch attacks, a DoS renaissance is underway – which means availability once again needs to become a major factor in security architecture and control design.

We focus on forward-looking research at Securosis. So we have started poking around, talking to practitioners about their DoS defense plans, and we have discovered a clear knowledge gap around the Denial of Service attacks in use today and the defenses needed to maintain availability. There is an all too common belief that the defenses that protect against run of the mill network and application attacks will stand up to a DoS. That's just not the case, so this paper will provide detail on the attacks in use today, suggest realistic defensive architectures and tactics, and explain the basic process required to have a chance of defending your organization against a DoS attack. Let's start with the major kinds of attacks.

Flooding the Pipes versus Exhausting the Servers

We will dig into specific attack tactics in much more depth later in this paper, but to understand Denial of Service we need to start with a clear distinction between *network-based* and *application-based* attacks. Both have the same objective — to impair availability – but they go about it in fundamentally different ways.

Network-based attacks overwhelm the network equipment and/or totally consume network capacity by throwing everything including the kitchen sink at a site — this interferes with legitimate traffic reaching the site. This *volumetric* type of attack is what most folks consider Denial of Service, and it realistically requires blasting away from many devices, so current attacks are called Distributed Denial of Service (DDoS). If your adversary has enough firepower it is very hard to defend against these attacks, and you will quickly be reminded that though bandwidth may be plentiful, it certainly isn't free.

Application-based attacks are different – they target weaknesses in web application components to consume all the resources of a web, application, or database server to effectively disable it. These attacks can target either vulnerabilities or 'features' of an application stack to overwhelm servers and prevent legitimate traffic from accessing web pages or completing transactions.

The beginning of a network-based attack is fairly obvious. But application-based DoS attacks are less obvious – you are unlikely to discover the attack until servers inexplicably start falling over – so they require more sophisticated defenses.

Both network-based and application-based DoS attacks have the same objective — to impair availability – but they go about it in fundamentally different ways.

Adversary Analysis

A new tactic increasingly leveraged by security practitioners is adversary analysis. It's not enough to just understand attacks and build defenses based on them – there is simply too much attack surface, and too many attack vectors. Security success depends on your ability to prioritize your efforts, as we hammered home in the Vulnerability Management Evolution paper. This involves making strategic bets about who is most likely to attack you and what tactics they tend to use. This will enable you to build control sets with an initial focus on what's likely to happen. Of course you will be wrong – attackers evolve tactics over time – but in the universe of things you can do, this approach helps narrow your options to something (mostly) manageable.

So let's coarsely group the kinds of adversaries who use DoS attacks.

- **Protection Racketeers:** These criminals use a DoS threat to demand ransom money. Attackers hold a site hostage by threatening to knock it down, and sometimes follow through. They get paid. They move on to the next target. The only thing missing is the GoodFellas theme music.

- **Hacktivists:** DoS has become a favored approach of hacktivists seeking to make a statement and shine the spotlight on their cause, whatever it is. Hacktivists care less about the target than their cause. The target is usually collateral damage, though they are happy to hit two birds with one stone by attacking an organization that opposes their cause when they can. You can't negotiate with these folks, and public discourse is one of their goals.

- **"CyberWar:"** Although we don't really like the term "cyberwar," since no one has been killed by browsing online (yet), the reality is you'll likely see online attacks as a precursor to warplanes, ships, bombing, and ground forces. By knocking out power grids, defense radar, major media, and other critical technology infrastructure, the impact of an attack can be magnified.

- **Exfiltrators:** These folks use DoS to divert attention from the real attack – stealing data they can monetize. This could be an intellectual property theft or a financial attack – for example, stealing credit cards. Either way, they figure that if they blow up your front door you will be too distracted to notice your TV scooting out through the garage. They are usually right.

- **Competitors:** They say all's fair in love and business. Some folks take that one a bit too far, and actively knock down competitor sites for an advantage. Maybe it's during the holiday season. Maybe it happens if the competitor resists an acquisition or merger advance. It could be locking folks out from bidding on an auction. Your competition may also screen scrape your online store to make sure they beat your pricing, which causes a flood of traffic on a very regular and predictable basis. You may also have the competitor try to ruin your hard-earned (and expensive) search rankings. Regardless of the reason, don't assume the attacker is a nameless, faceless crime syndicate in a remote nation. It could be the dude down the street trying to get any advantage he can – legalities be damned.

- **Success:** Sometimes a promotion or outside activity creates an innocent DoS situation. We see this all the time, like when a new iPhone is available for order and the carrier's network can't handle the traffic. Once the carrier figures how to deliver the traffic, pre-ordering applications fail under the load. It's friendly fire. Some say this is a good problem to have, but it is still a problem. Customers can't complete transactions, which causes dissatisfaction and grumpiness. So part of network and application architecture has to be handling peak traffic situations – catastrophic success – which can look an awful lot like a DoS attack.

Elasticity Enables Economic Denial of Service

The cloud may scale up to meet the computing demand, but your credit card can't expand to meet the financial demand.

The adversaries we described above tend to utilize network-based and application-based attacks. But cloud computing complicates things by creating a new kind of DoS attack. For those of you not familiar with cloud computing, one of its key aspects is elasticity – the cloud infrastructure can be expanded and contracted as needed, as dictated by business conditions. During peak traffic the system expands – typically automatically, based on utilization levels – to meet demand by provisioning new instances, adding storage, etc. When the peak passes, the system can deprovision devices and return to normal size. Sounds great, right?

Elasticity can be awesome. But another essential characteristics of cloud computing is measured service: you pay for what you use. So if an adversary figures out how to consume your resources (via network or application based DoS), forcing the provisioning of additional capacity, they can run up your credit with the could provider to the limit which takes you offline. The cloud may scale up to meet the computing demand, but your credit can't expand to meet the *financial* demand. This attack is becoming known as an "Economic Denial of Service" – the technology works fine, but the target may run out of money to keep their systems up and running in the cloud.

Understanding Attacks and Designing Defenses

With the sordid history of Denial of Service attacks, and the classes of adversaries who leverage DoS attacks, we have set the stage for the rest of this paper. We will start by going much deeper into the tactics behind these attacks. Understanding the attacks leads into a discussion of the different tactics and techniques for defending against each attack type. Finally we will wrap up by expanding your existing incident response process to handle DoS attacks.

The Attacks

Denial of Service (DoS) attacks have raised the importance of protecting availability as an aspect of security context, rather than (as previously) only addressing confidentiality and integrity. Attackers use DoS in many different ways, including extortion threats, obfuscation (to hide data exfiltration), hacktivism (to draw attention to a particular cause), and even friendly fire (when a promotion goes a little too well).

Now let's look at the types of DoS attacks you may face – attackers have many arrows in their quivers, and use them all depending on their objectives and targets.

Flooding the Pipes

The first kind of Denial of Service attack is really a blunt force object. It's basically about trying to oversaturate the bandwidth and computing resources of network (and increasingly server) devices to impair resource availability. These attacks aren't very sophisticated, but as evidenced by the ongoing popularity of volume-based attacks, they are effective. The tactics have been in use since before the Internet bubble, leveraging largely the same approach, but they have gotten easier with bots to do the heavy lifting. Of course this kind of blasting must be done somewhat carefully to maintain the usefulness of the bot, so bot masters have developed sophisticated approaches to ensure their bots avoid ISPs penalty boxes. So we see limited bursts of traffic from each bot and IP address spoofing to make it harder to track down where the traffic is coming from, but even short bursts from 100,000+ bots can flood most pipes.

A volumetric attack is really a blunt force object. It's basically about trying to oversaturate the bandwidth and computing resources of network (and increasingly server) devices to impair resource availability.

Quite a few specific techniques have been developed for volumetric attacks, but most look like some kind of flood. Floods target specific protocols (SYN, ICMP, UDP, etc.), and work by sending requests to a target using the chosen protocol, but not acknowledging the response. This quickly exhausts the ability of these devices to maintain state and causes the target to fail. Attackers need to stay ahead of Moore's Law, because targets' ability to handle floods has improved with their processing power. So network-based attacks may include encrypted traffic, forcing the target to devote additional computational resources to decrypting massive amounts of SSL traffic. Given the resource-intensive nature of encryption, this type of

attack can melt firewalls, IPS, and WAF devices unless they are configured specifically for large-scale SSL termination (which they are typically not). We also see some malformed protocol attacks but these aren't as effective nowadays because even unsophisticated network security perimeter devices drop bad packets at wire speed.

The attackers have compounded the severity of the attack using tactics like DNS amplification and reflection. In a reflection attack at the same time they target your site with a variety of floods, they are also spoofing your IP address and sending traffic to thousands of other sites. Those sites reply to your IP address to establish the session, and further increase the traffic hitting your site. Likewise, the attackers can also send a simple DNS request (using your spoofed IP address) to an open DNS resolver, which results in upwards of 50:1 amplification of the traffic directed to you via the response. Again, these aren't very sophisticated techniques, but can be very effective in magnifying the impact of the network-based attack.

These volume-based attacks are climbing the stack as well, targeting web servers by actually completing connection requests and then making simple `GET` requests and resetting the connection, over and over again, with approximately the same impact as a volumetric attack – overconsumption of resources effectively knocking down servers. These attacks may also include a large payload to further consume bandwidth. The now famous Low Orbit Ion Cannon (LOIC), a favorite tool of the hacktivist crowd, has undertaken a similar evolution, first targeting network resources and now proceeding to target web servers as well. It gets even better – these attacks can be magnified to increase their impact by simultaneously spoofing the target's IP address and requesting sessions from thousands of other sites, which then bury the target in a deluge of misdirected replies, further consuming bandwidth and resources.

Fortunately defending against these network-based tactics isn't overly complicated, as we will discuss later in this paper, but without a sufficiently large network device at the perimeter to block these attacks or an upstream service provider/traffic scrubber to dump offending traffic, devices tend to fall over in short order.

Overwhelming the Applications

But attackers don't only attack the network – they increasingly attack the applications as well, following other attackers up the stack. Your typical *n*-tier web application will have some Internet-facing device (usually a web server) to present content, an application server to handle application logic, and then a database to store the data. Attackers can target all tiers of the stack to impact application availability. So let's dig into each layer to see how these attacks work.

The termination point is usually the first target in application DoS attacks. They started with simple `GET` floods as described above but quickly evolved to additional attack vectors. The best known application DoS attack is probably RSnake's Slowloris, which consumes web server resources by sending partial HTTP requests, effectively opening connections and leaving sessions open by sending additional headers at regular intervals. This approach is far more efficient than the `GET` flood, requiring only hundreds of requests at regular intervals rather than thousands continuously, and only requires one device to knock down a large site. These application attacks have evolved over time and now send complete HTTP requests to evade IDS and WAF device rules which drop incomplete HTTP requests, but they tamper with payloads to confuse real

applications and consume resources. As defenders learn the attack vectors and deploy defenses, attackers evolve their attacks. The cycle continues.

Web server based attacks can also target weaknesses in the web server platform. For example the Apache Killer attack sends a malformed HTTP range request to take advantage of an Apache vulnerability. The Apache folks quickly patched the code to address this issue, but it clearly shows attackers targeting weaknesses in the underlying application stack to knock servers over. And of course unpatched Apache servers at many organizations are still vulnerable today — which represents a key issue for defending against these attacks. Even if application and database server patches are available to address the attack, if the patch doesn't happen you are still exposed.

> Traditional network security devices are little more than speed bumps to an application-centric attack, so developers need to ensure their applications can deal with attacks.

Attackers can also target legitimate application functionality. One such attack targets the search capability within a web site. If an attacker scripts a series of overly broad searches, the application can waste a large amount of time querying the database and presenting results. Likewise, attackers can game shopping carts by opening many shopping sessions, adding thousands of items to each cart, constantly refreshing the carts, and then abandoning them. Unless the application is architected to handle these malicious use cases efficiently, such attacks generally succeed. For most sites, failing to return search results or track shopping carts is a complete failure with severe business ramifications — DoS success.

The advantage of application attacks is their ability to evade many of network-centric anti-DoS techniques. The concept of evasion is well known and involves the attacker profiling an organization's defenses to plan attacks around the preventative controls. Many traditional network security devices (IPS and firewalls) are little more than speed bumps to an application-centric attack, so developers need to ensure their applications can deal with attacks, either through a purpose-built device such as a web application firewall (WAF with proper anti-DoS protections) or via tight coordination with the security team.

Targeting Alternative Sites

Some organizations don't keep their disaster recovery (DR) sites as current or protected as the main site. So if an adversary takes out a main site and forces a move to the DR location, that could open up a new set of exposures that escape ongoing security testing and assessment, or even just reduce functionality. Savvy attackers perform reconnaissance not just on main sites, but on pretty much any address space owned by, controlled by, or linked to their targets.

Impacting the Wallet

The so-called EDoS (economic denial of service) attack is a focused attempt to increase the cost of the target's technology infrastructure. In a network attack, bad guys take advantage of excessive bandwidth charges. So even if a target successfully defends against an attack, bandwidth costs to absorb or mitigate a multi-gigabyte attack could cost as much or more than an outage. Similarly, if a target leverages public cloud infrastructure to auto-provision new instances as utilization thresholds are met, an application attack can have a substantial financial cost without threatening availability. One potential endgame for cloud-based attacks is reaching the target's credit limit with their cloud provider, triggering an outage when the provider caps usage. There is always more than one way to impact a target.

The so-called EDoS (economic denial of service) attack involves a focused attempt to increase the cost of the target's technology infrastructure.

Attackers can (and do) mix and match a variety of different DoS attacks to achieve their goal of adversely impacting availability of an application or service.

Defense, Part 1: The Network

The previous section discussed both network-based and application-targeting Denial of Service (DoS) attacks. Given the radically different techniques between these types, it's only logical that we use different defense strategies for each. But aspects of both network-based and application-targeting DoS attacks are often combined for maximum effect. Most anti-DoS products and services to consider defend against both. This section focuses on defending against network-based volumetric attacks.

First the obvious: you cannot just throw bandwidth at the problem. Your adversaries likely have an unbounded number of bots at their disposal and are getting smarter at using shared virtual servers and cloud instances to magnify the amount at their disposal. So you can't just hunker down and ride it out. They likely have a bigger cannon than you can handle. You need to figure out how to deal with a massive amount of traffic, and separate good traffic from bad while maintaining availability. Find a way to dump bad traffic before it hoses you somehow, without throwing out the baby (legitimate application traffic) with the bathwater.

First the obvious: you cannot just throw bandwidth at the problem. Your adversaries likely have an unbounded number of bots at their disposal and are getting smarter at using shared virtual servers and cloud instances to magnify the bandwidth at their disposal.

We need to be clear about the volumes we are talking about. Recent attacks have blasted upwards of 80-100gbps of *sustained* network traffic at targets. These are not just peak data rates. Unless you run a peering point or some other network-based service, you probably don't have (or can't afford) that kind of bandwidth. Keep in mind that even if you have big enough pipes, the weak link may be the network security devices connected to them. Successful DoS attacks frequently target network security devices to overwhelm their session management capabilities. Your huge expensive IPS might be able to handle 80gbps of traffic in ideal circumstances, but fall over due to session table overflow.

Before you just call up your favorite anti-DoS service provider, ISP, or content delivery network (CDN) and ask them to scrub your traffic, understand that approach is no silver bullet either. It's not like you can just flip a switch and have all your traffic instantly go through a scrubbing center. Redirecting traffic incurs latency, assuming you can even communicate with the scrubbing center (remember, your pipes are overwhelmed with attack traffic). With a little simple recon, attackers can understand enough about your defenses and mitigations to choose a mix of network and application attacks for maximum effect.

No, we won't *only* talk about more problems, but it's important to keep everything in context. *Security is not a problem you can ever solve – you need to figure out how much loss you can accept.* And consciously choose the trade-offs. If a few hours of downtime is fine, then you can take action to ensure you are back up within that timeframe. If no downtime is acceptable you need a different (and far more expensive) approach. There are no right answers – just a series of trade-offs to manage to the availability requirements of your business, within the constraints of your funding and available expertise.

There are no right answers – just a series of trade-offs to manage to the availability requirements of your business, within the constraints of your funding and available expertise.

Handling network-based attacks involves mixing and matching a number of different architectural constructs, involving both customer premise devices and network-based service offerings, which must cooperate to blunt an attack.

Customer Premise-based Devices

The first category of defenses is based around a device on the customer premises. These appliances are purpose-built to deal with DoS attacks. Before you turn your nose up at the idea of installing another box to solve such a specific problem, take a look at your perimeter. There is a reason you already have so many different devices. The existing devices already in your perimeter aren't particularly well suited to dealing with DoS attacks. Your IPS, firewall, and load balancers aren't designed to manage extreme quantities of sessions, nor are they particularly adept at dealing with obfuscated attack traffic which looks legitimate. In fact, we're seeing these devices being targeted more frequently to capitalize on their inability to deal with large volume attacks. Nor can other devices integrate with network providers (to automatically change network routes, as we will discuss later) – and they don't include built-in DoS mitigation rules, dashboards, or forensics, designed specifically to provide the information you need to ensure availability under duress.

An anti-DoS device needs to integrate a number of existing capabilities such as IPS, network behavioral analysis, WAF, and SSL termination, combining them with highly scalable session management.

A new category of DoS mitigation devices has emerged to deal with these attacks. They include both optimized IPS-like rules to prevent floods and other network anomalies, and simple web application firewall capabilities which we will discuss later in this paper. Additionally, we see a number of anti-DoS features such as session scalability, combined with embedded IP reputation capabilities, to discard traffic from known bots without performing full inspection. To understand IP reputation's role let's recall how email connection management devices

enabled anti-spam gateways to scale up to handle spam floods. It is computationally expensive to fully inspect every inbound email, so immediately dumping messages from known bad senders focuses inspection on email that might be legitimate, and keeps mail flowing. The same concept applies here. Keep the latency inherent in checking a cloud-based reputation database in mind — you will want the device to aggressively cache known bad IPs to avoid a lengthy cloud lookup for every incoming session.

These devices should be as close to the edge of the perimeter as possible, to get rid of the maximum amount of traffic before attacks impact anything else. It can be a little counter-intuitive to put yet another box inline within the network perimeter, but given that traditional network security devices do not provide anti-DoS mitigation, you may not have another choice. Some devices can be deployed out-of-band as well, to monitor network traffic and alert on attacks, but we believe that's of limited value since the point of these devices is to maintain availability. Telling you that your network is down doesn't really help the situation. And of course you will want a high-availability deployment – an outage due to a failed security device is likely to be even more embarrassing than simply succumbing to a DoS.

But anti-DoS devices include their own limitations. First and foremost is the simple fact that if your pipes are overwhelmed and the traffic can't reach your network, a device on your premises is irrelevant. Additionally, SSL attacks are increasing in frequency. It's cheap for an army of bots to use SSL to encrypt all their attack traffic, but expensive for a network security device to terminate all SSL sessions and check all their payloads for attacks. That kind of computational cost arbitrage can put defenders in a world of hurt. Even load balancers, which are designed to terminate high SSL volumes, can face challenges from SSL DoS attacks due to session management limitations.

So an anti-DoS device needs to integrate a number of existing capabilities such as IPS, network behavioral analysis, WAF, and SSL termination, combining them with highly scalable session management to cope with the onslaught. And all that is still not enough – you will always be limited by the amount of bandwidth coming into your site. Which brings us to network services as a compliment to premise-based devices.

Proxies & CDN

The first service option most organizations consider is a Content Delivery Network (CDN). These services enhance web site performance by strategically caching content. Depending on the nature of your site, a CDN might be able to dramatically reduce your ingress network traffic – if they can cache much of your static content. They also offer some security capabilities, especially for dealing with DoS attacks. The CDN acts as a proxy for your web site, so the provider can protect your site by using its own massive bandwidth to cope with DoS attacks for you. They have significant global networks, so even a fairly large volumetric attack shouldn't look much different than a busy customer day – say a software company patching an operating system for a hundred million customers. Their scale enables them to cope with much larger traffic onslaughts than your much smaller pipes. Another advantage of a CDN is its ability to obscure

The CDN acts as a proxy for your web site, so the provider can protect your site by using its own massive bandwidth to cope with DoS attacks for you.

the real IP addresses of your site, making it more difficult for attackers to target your servers. CDNs can handle SSL termination if you allow them to store your private keys.

What's the downside? Protecting each site individually. If one site is not running through the CDN, attackers can find it through simple reconnaissance and blast the site directly. Even for sites running through the CDN, if attackers can find your controlling IPs they can target them directly, bypassing the CDN. Attackers can also randomize web page and image requests, forcing cache misses within the CDN and floods of requests for what "dynamic content" directly from your servers. Obviously you want the CDN to be smart enough to detect and blunt these attacks before they melt your pipes and servers.

Your site becomes inextricably linked to the CDN in this model. Thus, if the CDN has an SSL problem (it's been known to happen), that becomes a trust issue for you. Moreover, if the CDN has any downtime or undue latency, then that means downtime and latency for you. So picking any service provider for a web site is a critical decision.

Also be wary of excessive bandwidth costs. At the low end of the market, CDNs charge a flat fee and just eat the bandwidth costs if a small site is attacked. But enterprise deals are a bit more involved, charging for both bandwidth and protection. A DoS attack can explode bandwidth costs, causing an "Economic DoS", and perhaps shutting down the site when the maximum threshold (by contract or credit card limit) is reached. When setting up contracts be sure to get some kind of protection from excessive bandwidth charges in case of attack.

Anti-DoS Service Providers

CDN limitations drive some organizations to consider more focused network-based anti-DoS service providers. These folks run *scrubbing centers* – big data centers with lots of anti-DoS mitigation gear to process inbound floods and keep sites available. You basically flip a switch to send your traffic through the scrubbing center once you detect an attack. The switch usually controls BGP routing, so as soon as DNS updates and the network converges, the scrubbing center handles all inbound traffic. On the backend you receive legitimate traffic through a direct connection – typically GRE tunnels to leverage the Internet, or a dedicated network link from the scrubbing center. Keep in mind that there is disruption while redirection is activated and deactivated, and running through the scrubbing center increases latency.

Scrubbing centers are big data centers with lots of anti-DoS mitigation gear to process inbound floods and keep sites available. You basically flip a switch to send your traffic through the scrubbing center once you detect an attack.

But what does a scrubbing center actually do? The same type of analysis as a premise-based device. They manage sessions, drop traffic based on network telemetry and IP reputation, block application-oriented attacks, and otherwise keep your site up and available. Most scrubbing centers have substantial anti-DoS equipment investments, amortized across all their customers.

You pay for what you need when you need it, rather than over-provisioning your network and buying a bunch of anti-DoS equipment for whenever you are actually attacked.

Getting back to our email security analogy, think of an anti-DoS service provider like a cloud email security service. Back in the early days of spam, most organizations implemented their own email security gateways to deal with it. When the volume inbound of email overwhelmed the gateways, organizations needed to deploy more gateways to deal with the additional traffic. This made anti-spam gateways a good business, until a few service providers started selling cloud email security services to deal with the issue. You'd route your mail through their networks and only good stuff (usually less than 10% of all traffic) would actually get delivered to your email servers. Spam flood? No worries – it's the provider's problem. Obviously there are differences – most notably that email filtering is full-time, while DoS filtering is generally only on-demand during attacks.

Of course there are issues with this type of service, aside from the inevitable latency, which causes disruption while you reroute traffic to the scrubbing center. Scrubbing centers have the same SSL requirement as CDNs: termination requires access to your private key. Depending on your security tolerance, this could be a serious problem. Many large sites have tons of certificates and can-cross sign keys for the scrubbing center, but it does complicate management of the service provider.

You will also need to spell out a process for determining when to redirect traffic, which generally involves an internal workflow to detect emerging attacks, evaluation of the situation, and then a determination to move the traffic.

You will also need to spell out a process for determining when to redirect traffic. We will talk about this more when we go through the DoS defense process, but it generally involves an internal workflow to detect emerging attacks, evaluation of the situation, and then a determination to move the traffic – typically rerouting via BGP. But if your anti-DoS provider uses the same equipment as you have onsite you might be able to leverage proprietary signaling protocols to automatically shift traffic based on thresholds. As these attacks become more prevalent, the need to automate the redirection via these signaling protocols will become more important. Although there is a downside to increased automation, as some network operations folks don't enjoy letting Skynet redirect their traffic through different networks. What could possibly go wrong with that?

Selection of an anti-DoS service provider is a serious decision. We recommend a fairly formal procurement process, which enables you to understand the provider's technical underpinnings, network architecture, available bandwidth, geographic distribution, ability to handle SSL attacks, underlying anti-DoS equipment, and support for various signaling protocols. Make sure you are comfortable with the robustness of their DNS infrastructure because DNS is a popular target and critical to several defenses. Also pay close attention to process handoffs, responsiveness of their support group, and their research capabilities (to track attackers and mitigate specific attacks).

The Answer: All of the Above

Ultimately your choice of network-based DoS mitigations will involves trade-offs. It is never good to over-generalize, but most organizations will be best suited by a hybrid approach, involving both a customer premise-based appliance and a contract with a CDN or anti-DoS service provider to handle severe volumetric attacks. It is simply not cost-effective to run all your traffic through a scrubbing center constantly, and many DoS attacks target the application layer – demanding use of a customer premise device anyway.

In terms of service provider defense, many organizations can and should start with a CDN. The CDN may be more attractive initially for its performance benefits, with anti-DoS and WAF capabilities as nice extras. Until you are attacked – at which point, depending on the nature of the attack, the CDN may save your proverbial bacon. If you are battling sophisticated attackers, or have a complicated and/or enterprise-class infrastructure, you are likely to look for a dedicated anti-DoS service provider. Again, this will usually be a retainer-based relationship which gives you the ability to route your traffic through the scrubbing center when necessary – paying when you are under attack and sending them traffic.

Ultimately your choice of network-based DoS mitigations will involves trade-offs. Most organizations will be best suited by a hybrid approach, involving both a customer premise-based appliance and a contract with a CDN or anti-DoS service provider to handle severe volumetric attacks.

All this assumes your sites reside within your data center. Cloud computing fundamentally alters the calculations, requiring different capabilities and architectures. If your apps reside in the cloud you don't have a customer premise where you can install devices, so you would instead consider either virtual instances, routing traffic through your site before it hits the cloud, or using a CDN for all inbound traffic. You could also architect your cloud infrastructure to provision more instances as necessary to handle traffic, but it is easy to convert a DoS attack into an economic attack as you pay to scale up in order to handle bogus traffic. There are no clear answers yet – it is still quite early in the evolution of cloud computing – but keep this in mind as your application architects keep talking about this cloud thingy.

Defense, Part 2: Applications

Whereas defending against volumetric DoS attacks requires resilient network architectures and service providers, dealing with application-targeted DoS puts the impetus for defense squarely back on your shoulders. As we mentioned earlier, overwhelming an application typically entails messing with its ability to manage session state and targeting weaknesses in the application stack. These attacks don't require massive bandwidth, bot armies, or much more than a well-crafted series of `GET` or `POST` requests.

While defending against network-based DoS involves handling brute force, application attacks require a more nuanced approach. Many of these attack tactics are based on legitimate traffic. For example, even legitimate application transactions start with a simple application request. So the challenge is to separate the good from the bad without impacting legitimate traffic in a way that that gets you in hot water with your operations folks.

What about WAFs?

Application-targeted DoS attacks look an awful lot like every other application attack. Just the end goal is different. DoS attacks try to knock the application down, whereas more traditional application attacks involve compromising either the application or the server stack as a first step to either application/data tampering or exfiltration. Most organizations work to secure their applications either by building security in via a secure SDLC (Software Development Lifecycle) or front-ending the application with a WAF (Web Application Firewall). Or, in many cases, both.

Application-targeted DoS attacks look an awful lot like every other application attack. Just the end goal is different, with DoS attacks trying to knock the application down.

So is building security in a solution to application DoS attacks? Obviously effectively managing state within a web app is good practice, and building anti-DoS protections directly into each application will help. But given the sad state of secure application development and the prevalence of truly simplistic attacks like SQLi, it's hard to envision anti-DoS capabilities becoming a key specification of web apps any time soon. Yeah, that's cynical, and we recommend that you keep DoS mitigation in mind during application security and technology planning, but it will be a while before that is widespread. A long while.

What about WAFs? Are they reasonable devices for dealing with application DoS attacks? Let's circle back to the trouble with existing WAFs: ease of evasion and the difficulty of keeping policies current. We recently did an entire paper on maximizing the value of WAF: Pragmatic WAF Management, highlighting positive polices based on what applications should do, and negative policies to detect and handle attacks.

It turns out that many successful WAF implementations start with stopping typical application attacks — something like a purpose-built IPS to protect applications. Those WAF policies can be helpful in stopping application DoS attacks too. Whether you're talking about Slowloris-type session manipulation, or application stack vulnerabilities, a WAF is well positioned to deal with those attacks.

Of course, a customer-premise-based WAF is another device that can be targeted, just like a firewall or IPS. Given the compute and state management requirements to detect and block application attacks, overwhelming a WAF device can be trivial, and most DoS attacks involve not just application-centric attack but also network-based tactics. So the WAF needs anti-DoS capabilities built in, and the architectural protections we discussed earlier when discussing network attacks should be used to protect the WAF from brute force attacks.

Many anti-DoS vendors offer the best of both worlds, helping you block network attacks (via load balancing, anti-DoS techniques, and coordination with scrubbing centers), as well as implementing both negative and positive WAF policies.

Anti-DoS Devices

Anti-DoS devices have emerged to detect volumetric attacks, drop bad traffic locally as long as possible, and then redirect excessive traffic to a scrubbing center. Another key capability of anti-DoS devices is their ability to deal with application DoS attacks. From this perspective these device have a lot in common with WAFs — focusing on protecting the applications with negative policies without the capabilities to profile applications or implement positive policies. This is just fine if you are deploying equipment specifically to deal with DoS attacks.

But you don't need to choose between a WAF and an anti-DoS device. Many anti-DoS vendors also offer full-featured WAF products. These products may offer the best of both worlds, helping you block network attacks (via load balancing, anti-DoS techniques, and coordination with scrubbing centers), as well as implement both negative and positive WAF policies within a single policy management system.

Managed WAF Services and CDN

As with network-based DoS attacks, there is no lack of service options for handling application attacks. We will compare and contrast the types. We discussed managed WAF services in Pragmatic WAF Management; they tend to focus on compliance with regulations such as PCI-DSS. These cloud WAFs typically implement

slimmed-down rule bases, focused mostly on negative policies – exactly what you need to defend against application DoS attacks.

Managed WAFs are largely offered by folks who offer Content Delivery Networks (CDN), as a value-added offering or possibly part of the core service. Obviously the less the service costs, the fewer capabilities you have to customize the rule base, which impacts the usefulness of a general-purpose WAF. But managed WAF services provide the additional bandwidth and 24/7 protection you are need to deal with attacks, and if the primary use case is DoS mitigation, a CDN or managed WAF can be a good fit.

Keep in mind that you will need to run all your application traffic through the managed WAF service, and many of the same issues crop up as with CDN. If you don't protect an application with the managed WAF it can be attacked directly. If its direct IP address can be discovered, the application can be attacked directly. And be clear on response processes, notifications, handoffs, and forensics with the service provider before things go live, so you are ready when an attack starts.

Anti-DoS Service Providers

We discussed handling volumetric attacks with scrubbing centers. Can a scrubbing center detect and block an application DoS attack? Of course they have racks of anti-DoS gear and sophisticated SOC infrastructures to detect and defeat attacks. But that doesn't make this kind of service well suited to application DoS mitigation. Application DoS is not a brute force attack. It works by gaming the innards of an application stack or the application code itself. By the time you realize the application is under attack, the servers are likely to be down, and switching traffic to a scrubbing center won't change that.

If you run all your traffic through a scrubbing center (or mega-CDN), it should detect and block many of the attacks before the bad traffic ever gets to your network. But that's not economically feasible for most organizations, so scrubbing centers are not the primary mitigation for application DoS attacks.

The Answer? Both.

As we have seen with recent DoS attacks on financial institutions, attackers are now attacking all levels of the stack. The web infrastructure is blasted with 60gbps of traffic, while simultaneously being hit by application attacks, typically using either SSL overhead or application weaknesses to knock applications out — even if the traffic flood can be handled. So you cannot really choose between tactics for the two types of DoS attacks.

You need a mitigation architecture to handle ***both*** volumetric and application DoS — combining on-premise solutions (or managed WAF/CDN) with scrubbing centers for full coverage against this class of attacks. Of course, if you need to phase in defenses, you can start with a CDN-based option for performance benefits with basic DoS protection. As your adversaries become more sophisticated and your risk profile changes, implementing on-premise defenses and contracting with a scrubbing center are next logical steps.

The Process

As we have mentioned throughout this paper, a strong underlying process is your best defense against a Denial of Service (DoS) attack. Tactics change as attack volumes increase, but if you don't know what to do when your site goes down it will be out for a while.

The good news is that the DoS Defense process is a close relative of your general incident response process. We have already published a ton of research on the topic, so check out both our Incident Response Fundamentals series and our React Faster and Better paper. If your incident handling process isn't where it needs to be, you should start there.

The good news is that the DoS Defense process is a close relative of your general incident response process. Building off the IR process, think about what you need to do as a set of activities before, during, and after the attack.

Building off the IR process, think about what you need to do as a set of activities before, during, and after the attack:

- **Before:** Before an attack, spend time figuring out the triggers for an attack and ensuring you perform persistent monitoring to ensure you have both sufficient warning and enough information to identify the root cause of an attack. This must happen before the attack because you only get one chance to collect that data — while things are happening. In Before the Attack we defined a three-step process for these activities: define, discover/baseline, and monitor.
- **During:** How can you contain the damage as quickly as possible? By identifying the root cause accurately and remediating effectively. This requires identifying the attack (Trigger and Escalate), identifying and mobilizing the response team (Size up), and then containing the damage in the heat of battle. During the Attack summarizes these steps.
- **After:** Once the attack has been contained, focus shifts to restoring normal operations (Mop up) and making sure it doesn't happen again (Investigation and Analysis). This involves a forensics process and some introspection as described in After the Attack.

But there are key differences when dealing with DoS, so let's amend the process a bit. We have already discussed preparation, in terms of controls and architectures to maintain availability in the face of DoS attacks. That may involve network-based approaches, focusing on the application layer, or more realistically both.

Before we jump into *during* the attack, we need to mention the importance of practice. You practice your disaster recovery plan, right? You should practice your incident response plan, and even a subset of that practice for DoS attacks. The worst time to discover gaping holes in your process is when your site is melting under a volumetric attack. That doesn't mean you need to blast yourself with 80gbps of traffic either. Practice handoffs with the service provider, tuning the anti-DoS gear, and ensuring everyone knows their roles and accountability for the real thing.

You should practice your incident response plan, and even a subset of that practice for DoS attacks. The worst time to discover the gaping holes in your process is when the site is melting under a volumetric attack.

Trigger and Escalate

There are a number of ways you might detect a DoS attack in progress. You could see increasing bandwidth volumes or a spike in DNS traffic. Perhaps your applications get a bit flaky and fall down, or you see server performance issues. You might get lucky and receive an alert from your CDN — you did set the CDN to alert on anomalous volumes, right? More likely you'll just lose your site. Increasingly these attacks tend to come out of nowhere in a synchronized series of activities targeting your network, DNS, and applications. We are big fans of setting thresholds and monitoring everything, but DoS is a bit different — you may not see it coming despite your best efforts.

Size up

Now your site and/or servers are down, and all hell is likely breaking loose. So you need to notify the powers that be, assemble the team, and establish responsibilities and accountabilities. You will also have guys starting to dig into the attack. They will need to identify root cause, attack vectors, and adversaries, and figure out the best way to get the site back up. You need to figure out what you're dealing with before you have any chance of mitigating an attack.

Restore

There is considerable variation in what comes next. It depends on what network and application mitigations are already in place. Optimally your contracted CDN and/or anti-DoS service provider already has a team working on the problem. If it's an application attack, perhaps some tuning of your anti-DoS appliance can block the attacks. But hope isn't a strategy, so you need plan B, which usually entails redirecting your traffic to a scrubbing center.

The biggest decision you will face is when to actually redirect the traffic. If the site is totally down it's easy. But if it's an application performance issue caused by an application or network attack you will need more information – particularly an idea of whether or not redirection will even help. In many cases it will, since the service provider will then see the traffic and they likely have more expertise and can more effectively diagnose the issue, but there will be a lag as the network converges after a change.

Finally, there is the issue of targeted organizations without scrubbing center contracts. In that case your best bet is to cold call an anti-DoS provider and hope they can help. These folks are in the business of fighting DoS, so they will likely be able to help, but do you want to take a chance on that? We don't, so we suggest you at least have a conversation with an anti-DoS provider before you are attacked – if only to understand their process and how they can help when you are under attack. Talking to a service provider doesn't mean you need to contract for their service. It means you know who to call and what to do under fire.

We suggest you at least have a conversation with an anti-DoS provider before you are attacked – if only to understand their process and how they can help when you are under attack.

Mop up

You have weathered the storm and your sites are operating normally again. In terms of mopping up, you will shunt traffic back from the scrubbing center and perhaps loosen up anti-DoS/WAF rules. You will keep monitoring for signs of trouble, and probably hope to grab a couple days of sleep to catch up.

Investigate and Analyze

Once you are well rested, don't fall into the trap of turning the page and moving on. There are tons of lessons to be learned. What worked? What didn't? Who needs to be added to the team? Who just got in the way? The post-mortem needs to identify the good, the bad, and the ugly. Some folks may end up with bruised egos. Tell them to get over it. The sacred cows must be slain if you don't want to relive the nightmare soon.

More important, dig into the attack. What controls would have damped its impact? Would running all your traffic through a CDN help? Did the network redirection work effectively? Did you get the proper level of support from the service provider? The more questions you ask the better.

Then update your process as needed. Implement new controls if necessary. Swap out your service provider if they didn't get it done. If you aren't learning from every attack you are missing opportunities to improve response next time. And you know there will be a next time – there always is.

Summary

Years of compliance-centric security have forced security controls, funding, and expertise to address data leakage and protection issues — rather than focusing on availability impacting attacks. No good deed goes unpunished, so attackers have adapted their techniques to add Denial of Service (DoS) attacks to their repertoires. Not just to impact availability and cause loss, but also to hide their more important activity: exfiltration of sensitive data.

Today's DoS attacks use far more sophisticated tactics than the packet blasting escapades of yesteryear. They involve both network-based and application-based techniques to confuse defenders and exploit weaknesses in their targets' defenses. So organizations need a multi-faceted approach to defending against DoS, which likely involves deploying both anti-DoS equipment on-site and contracting with a service provider (either a scrubbing center or a content delivery network) to handle excessive traffic. DoS mitigations do not stand in isolation, rather on-premise devices and services are co-dependent to provide adequate protection.

There are trade-offs with all DoS defense options. Selecting an optimal mix of defensive tactics requires some adversary analysis and a fine assessment of just how much downtime is survivable. If a few hours of downtime can be survived, defensive tactics can be much different than in situations where no downtime is ever acceptable – which demands more expenditure and implementation of much more sophisticated defenses.

But the truism that "security is a process, not a product or service" holds for DoS attacks. The process for DoS mitigation is an extension of your existing incident response process — and requires the same level of preparation, practice, and post-mortem analysis to address shortcomings that can only be identified in the rear-view mirror. We discussed process last, but it's the first thing you need to tackle.

If you have any questions on this topic, or want to discuss your situation specifically, feel free to send us a note at info@securosis.com or ask via the Securosis Nexus (http://nexus.securosis.com/).

About the Analyst

Mike Rothman, Analyst/President

Mike's bold perspectives and irreverent style are invaluable as companies determine effective strategies to grapple with the dynamic security threatscape. Mike specializes in the sexy aspects of security — such as protecting networks and endpoints, security management, and compliance. Mike is one of the most sought-after speakers and commentators in the security business, and brings a deep background in information security. After 20 years in and around security, he's one of the guys who "knows where the bodies are buried" in the space.

Starting his career as a programmer and networking consultant, Mike joined META Group in 1993 and spearheaded META's initial foray into information security research. Mike left META in 1998 to found SHYM Technology, a pioneer in the PKI software market, and then held executive roles at CipherTrust and TruSecure. After getting fed up with vendor life, Mike started Security Incite in 2006 to provide a voice of reason in an over-hyped yet underwhelming security industry. After taking a short detour as Senior VP, Strategy at eIQnetworks to chase shiny objects in security and compliance management, Mike joined Securosis with a rejuvenated cynicism about the state of security and what it takes to survive as a security professional.

Mike published The Pragmatic CSO <http://www.pragmaticcso.com/> in 2007 to introduce technically oriented security professionals to the nuances of what is required to be a senior security professional. He also possesses a very expensive engineering degree in Operations Research and Industrial Engineering from Cornell University. His folks are overjoyed that he uses literally zero percent of his education on a daily basis. He can be reached at mrothman (at) securosis (dot) com.

About Securosis

Securosis, LLC is an independent research and analysis firm dedicated to thought leadership, objectivity, and transparency. Our analysts have all held executive level positions and are dedicated to providing high-value, pragmatic advisory services.

Our services include:

- **The Securosis Nexus**: The Securosis Nexus is an online environment to help you get your job done better and faster. It provides pragmatic research on security topics that tells you exactly what you need to know, backed with industry-leading expert advice to answer your questions. The Nexus was designed to be fast and easy to use, and to get you the information you need as quickly as possible. Access it at <https://nexus.securosis.com/>.

- **Primary research publishing**: We currently release the vast majority of our research for free through our blog, and archive it in our Research Library. Most of these research documents can be sponsored for distribution on an annual basis. All published materials and presentations meet our strict objectivity requirements and conform to our Totally Transparent Research policy.

- **Research products and strategic advisory services for end users**: Securosis will be introducing a line of research products and inquiry-based subscription services designed to assist end user organizations in accelerating project and program success. Additional advisory projects are also available, including product selection assistance, technology and architecture strategy, education, security management evaluations, and risk assessment.

- **Retainer services for vendors**: Although we will accept briefings from anyone, some vendors opt for a tighter, ongoing relationship. We offer a number of flexible retainer packages. Services available as part of a retainer package include market and product analysis and strategy, technology guidance, product evaluation, and merger and acquisition assessment. Even with paid clients, we maintain our strict objectivity and confidentiality requirements. More information on our retainer services (PDF) is available.

- **External speaking and editorial**: Securosis analysts frequently speak at industry events, give online presentations, and write and/or speak for a variety of publications and media.

- **Other expert services**: Securosis analysts are available for other services as well, including Strategic Advisory Days, Strategy Consulting engagements, and Investor Services. These tend to be customized to meet a client's particular requirements.

Our clients range from stealth startups to some of the best known technology vendors and end users. Clients include large financial institutions, institutional investors, mid-sized enterprises, and major security vendors.

Additionally, Securosis partners with security testing labs to provide unique product evaluations that combine in-depth technical analysis with high-level product, architecture, and market analysis. For more information about Securosis, visit our website: <http://securosis.com/>.