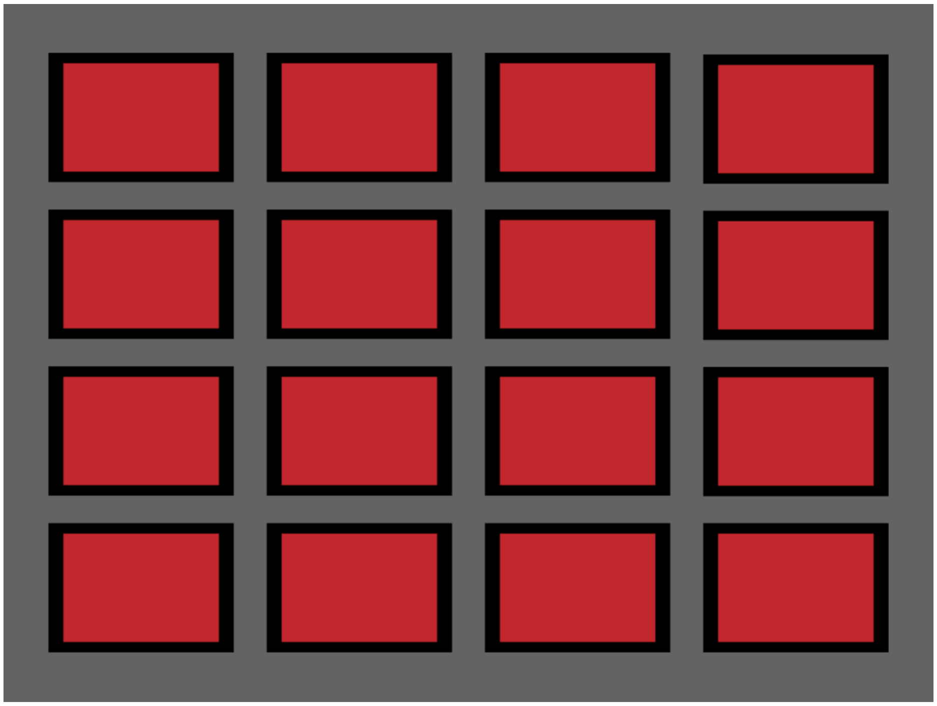
LILY HAY NEWMAN SECURITY 07.28.17 12:05 PM

HOW NETFLIX DDOS'D ITSELF TO HELP PROTECT THE ENTIRE INTERNET



HOTLITTLEPOTATO

IN JUNE 2016, Netflix security engineer Scott Behrens ran a massive infrastructure test on the streaming system in front of dozens of coworkers. In the process, he brought the site down. But instead of panic or embarrassment, it was a moment of celebration. Behrens, working with cloud security engineer Jeremy Heffner and others, had successfully shown that Netflix was in fact vulnerable to an unorthodox type of distributed denial of service attack. And proving it worked was the first step toward preventing it in the future—not just for Netflix but for the entire internet.

junk traffic requests, overwhelming the system to either crash it completely or burden it until it can′t function normally. Those would have a hard time impacting Netflix, though; the service is already built to handle more than 35TB per second of data during peak hours, and has a network of Open Connect devices that localizes most of its traffic anyway. Aiming a botnet at Netflix would be like shoveling dirt into Carlsbad Caverns.

But Behrens conceived of a different type of DDoS, one that turned Netflix′s application programming interface against itself. Netflix′s API acts as a sort of gateway to a complex array of middle and backend application services—all the stuff that happens under the hood. Behrens realized that an attacker could send a very small number of resource-intensive, carefully chosen requests designed to trigger more and more requests, cascading deep into the system. In this way, an attacker could easily and cheaply cause significant resource burden, and even take Netflix down.

RELATED STORIES

BRIAN BARRETT
Netflix′s Grand, Daring, Maybe Crazy Plan to Conquer the World

CADE METZ
The Counterintuitive Tech Behind Netflix′s Worldwide Launch

BRIAN BARRETT
Netflix Isn′t Made for the US Anymore—It′s for the Whole World

″It was pretty cool. We were actually able to really test this in the environment that our customers would have been impacted in, as opposed to simulating or hypothesizing that it was an issue without actually proving it,″ says Behrens, who presented his findings at the DefCon security conference in Las Vegas on Friday. ″Maybe we send one request to the API, but it results in 10,000 requests on the inside of the network, meaning we can

Behrens tested his attack at what Netflix calls a "Chaos Kong," a time when Netflix engineers reroute customers away from a certain region of production servers so they can have a real-world sandbox in which to experiment. The process also helps ensure that Netflix can continue to provide service to its customers even if one of its regions goes down or experiences problems; during a Chaos Kong all user traffic gets rerouted from a particular region, ideally without customers noticing.

Application DDoS attacks like the one Behrens devised are rare, but not completely unheard of. A recent Akamai State of the Internet report notes that they account for less than 1 percent of all DDoS attacks. But Behrens says that Netflix's application security team works to stay two steps ahead of attackers, so even such a small percentage merited closer examination. Especially given that the attack takes fewer resources than the more common standard version—meaning it could spike in popularity.

The type of assault Behrens envisioned wouldn't translate effortlessly to an attack on any company. Only those that use an "API gateway" microservices architecture—the iceberg approach, where the internet-connected interface is the small portal to a huge array of services underneath—like Netflix would be so vulnerable to it. But many companies do use this type of setup. And if attackers started working to expand this type of attack, they could probably find ways to apply the concept of high-cost, low-volume request attacks to other architectures.

"If attackers could potentially pull off the same objective with a lot less requests, it's lower cost for them," Behrens says. "As a security researcher I'm always looking for ways to increase the cost for adversaries and attackers. We really wanted to position ourself in such a way that we could give folks the tools and the frameworks to find this in their own applications, so they can build in those remediations before that number [of these

To improve protections against these types of attacks, Behrens suggests more robust monitoring of middle-tier and backend service traffic and behavior, so operators have more insight into what's going on deep in their systems and can spot problems early, before they spiral into a mess of junk requests. Most companies—including Netflix, until Behrens pulled off his attack—don't bother keeping track of traffic that far down the stack. Behrens also advocates for tools that can help us understand behavior patterns, and distinguish legitimate customer requests from malicious traffic so that the system can automatically work to prioritize real requests.

On Friday, Netflix also released two open-source tools, called Repulsive Grizzly and Cloudy Kraken, to help developers do their own small-scale testing once they identify potential vulnerabilities to this type of attack. These tools aren't production-grade solutions in themselves, but do represent a first step toward making testing options more available for this type of weakness.

"The combination of those things has really raised the bar for causing this sort of issue against the product," Behrens says. "A lot of the mitigations that I discuss definitely did hold true, but we have to be humble and realize that there's always going to be something that might pop up. It's a cat and mouse game, so we just continue to try to find ways to make our testing more sophisticated and then build in stronger remediations."

The evolution of attacker strategies never ends, but if companies adopt Netflix's suggestions for protecting against this type of application DDoS, it represents one opportunity for everyone to keep ahead of the danger.

RELATED VIDEO