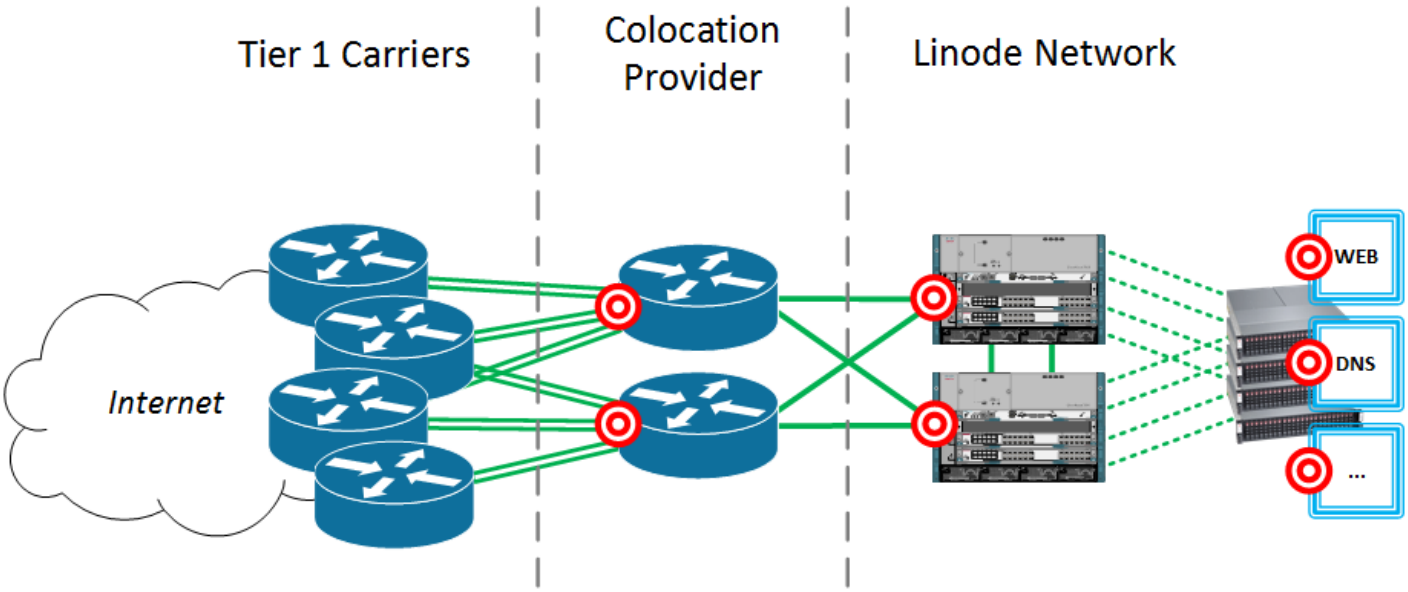
The Twelve Days of Crisis – A Retrospective on Linode's Holiday DDoS Attacks

January 29, 2016 4:30 pm

Over the twelve days between December 25th and January 5th, Linode saw more than a hundred denial-of-service attacks against every major part of our infrastructure, some severely disrupting service for hundreds of thousands of Linode customers.

I'd like to follow up on my earlier update by providing some more insight into how we were attacked and what we're doing to stop it from happening again.



Pictured above is an overview of the different infrastructure points that were attacked. Essentially, the attacker moved up our stack in roughly this order:

- Layer 7 ("400 Bad Request") attacks toward our public-facing websites
- Volumetric attacks toward our websites, authoritative nameservers, and other public services
- Volumetric attacks toward Linode network infrastructure
- Volumetric attacks toward our colocation provider's network infrastructure

Most of the attacks were simple volumetric attacks. A *volumetric attack* is the most common type of *distributed denial-of-service* (DDoS) attack in which a cannon of garbage traffic is directed toward an IP address, wiping the intended victim off the Internet. It's the virtual equivalent to intentionally causing a traffic-jam using a fleet of rental cars, and the pervasiveness of these types of attacks has caused hundreds of billions of dollars in economic loss globally.

Typically, Linode sees several dozen volumetric attacks aimed toward our customers each day. However, these attacks almost never affect the wider Linode network because of a tool we use to protect ourselves called *remote-triggered blackholing*. When an IP address is "blackholed," the Internet collectively agrees to drop all traffic destined to that IP address, preventing both good and

bad traffic from reaching it. For content networks like Linode, which have hundreds of thousands of IPs, blackholing is a blunt but crucial weapon in our arsenal, giving us the ability to 'cut off a finger to save the hand' – that is, to sacrifice the customer who is being attacked in order to keep the others online.

Blackholing fails as an effective mitigator under one obvious but important circumstance: when the IP that's being targeted – say, some critical piece of infrastructure – can't go offline without taking others down with it. Examples that usually come to mind are "servers of servers," like API endpoints or DNS servers, that make up the foundation of other infrastructure. While many of the attacks were against our "servers of servers," the hardest ones for us to mitigate turned out to be the attacks pointed directly toward ours and our colocation providers' network infrastructure.

Secondary Addresses

The attacks leveled against our network infrastructure were relatively straightforward, but mitigating them was not. As an artifact of history, we segment customers into individual /24 subnets, meaning that our routers must have a "secondary" IP address inside each of these subnets for customers to use as their network gateways. As time has gone by, our routers have amassed hundreds of these secondary addresses, each a potential target for attack.

Of course, this was not the first time that our routers have been attacked directly. Typically, special measures are taken to send blackhole advertisements to our upstreams without blackholing in our core, stopping the attack while allowing customer traffic to pass as usual. However, we were unprepared for the scenario where someone rapidly and unpredictably attacked many dozens of different secondary IPs on our routers. This was for a couple of reasons. First, mitigating attacks on network gear required manual intervention by network engineers which was slow and error-prone. Second, our upstream providers were only able to accept a limited number of blackhole advertisements in order to limit the potential for damage in case of error.

After several days of playing cat-and-mouse games with the attacker, we were able to work with our colocation providers to either blackhole all of our secondary addresses, or to instead drop the traffic at the edges of their transit providers' networks where blackholing wasn't possible.

Cross-Connects

The attacks targeting our colocation providers were just as straightforward, but even harder to mitigate. Once our routers were no longer able to be attacked directly, our colocation partners and their transit providers became the next logical target – specifically, their cross-connects. A *cross-connect* can generally be thought of as the physical link between any two routers on the Internet. Each side of this physical link needs an IP address so that the two routers can communicate with each other, and it was those IP addresses that were targeted.

As was the case with our own infrastructure, this method of attack was not novel in and of itself. What made this method so effective was the rapidity and unpredictability of the attacks. In many of our datacenters, dozens of different IPs within the upstream networks were attacked, requiring a level of focus and coordination between our colocation partners and their transit providers which was difficult to maintain. Our longest outage by far – over 30 hours in Atlanta – can be directly attributed to frequent breakdowns in communication between Linode staff and people who were sometimes four-degrees removed from us.

We were eventually able to completely close this attack vector after some stubborn transit providers finally acknowledged that their infrastructure was under attack and successfully put measures in place to stop the attacks.

Lessons Learned

On a personal level, we're embarrassed that something like this could have happened, and we've learned some hard lessons from the experience.

Lesson one: don't depend on middlemen

In hindsight, we believe the longer outages could have been avoided if we had not been relying on our colocation partners for IP transit. There are two specific reasons for this:

First, in several instances we were led to believe that our colocation providers simply had more IP transit capacity than they actually did. Several times, the amount of attack traffic directed toward Linode was so large that our colocation providers had no choice but to temporarily de-peer with the Linode network until the attacks ended.

Second, successfully mitigating some of the more nuanced attacks required the direct involvement of senior network engineers from different Tier 1 providers. At 4am on a holiday weekend, our colocation partners became an extra, unnecessary barrier between ourselves and the people who could fix our problems.

Lesson two: absorb larger attacks

Linode's capacity management strategy for IP transit has been simple: when our peak daily utilization starts approaching 50% of our overall capacity, then it's time to get more links.

This strategy is standard for carrier networks, but we now understand that it is inadequate for content networks like ours. To put some real numbers on this, our smaller datacenter networks have a total IP transit capacity of 40Gbps. This may seem like a lot of capacity to many of you, but in the context of an 80Gbps DDoS that can't be blackholed, having only 20Gbps worth of headroom leaves us with crippling packet loss for the duration of the attack.

Lesson three: let customers know what's happening

It's important that we acknowledge when we fail, and our lack of detailed communication during the early days of the attack was a big failure.

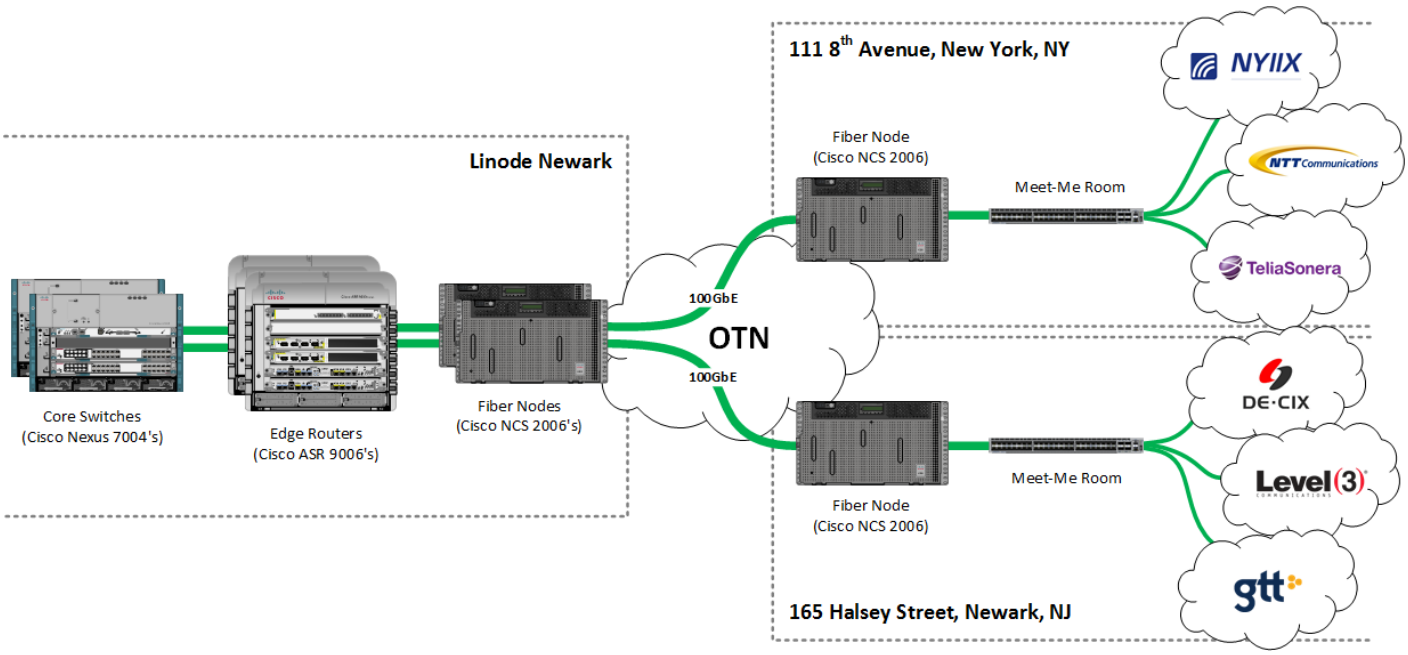
Providing detailed technical updates during a time of crisis can only be done by those with detailed knowledge of the current state of affairs. Usually, those people are also the ones who are firefighting. After things settled down and we reviewed our public communications, we came to the conclusion that our fear of wording something poorly and causing undue panic led us to speak more ambiguously than we should have in our status updates. This was wrong, and going forward, a designated technical point-person will be responsible for communicating in detail during major events like this. Additionally, our status page now allows customers to be alerted about service issues by email and SMS text messaging via the "Subscribe to Updates" link.

Our Future is Brighter Than our Past

With these lessons in mind, we'd like you to know how we are putting them into practice.

First, the easy part: we've mitigated the threat of attacks against our public-facing servers by implementing DDoS mitigation. Our nameservers are now protected by Cloudflare, and our websites are now protected by powerful commercial traffic scrubbing appliances. Additionally, we've made sure that the emergency mitigation techniques put in place during these holiday attacks have been made permanent.

By themselves, these measures put us in a place where we're confident that the types of attacks that happened over the holidays can't happen again. Still, we need to do more. So today I'm excited to announce that Linode will be overhauling our entire datacenter connectivity strategy, backhauling **200 gigabits of transit and peering capacity** from major regional points of presence into each of our locations.



Carriers shown are for example purposes only. All product names and logos are the property of their respective owners.

Here is an overview of forthcoming infrastructure improvements to our Newark datacenter, which will be the first to receive these capacity upgrades. The headliner of this architecture is the optical transport networks that we have already begun building out. These networks will provide fully diverse paths to some of the most important PoPs in the region, giving Linode access to hundreds of different carrier options and thousands of direct peering partners.

Compared to our existing architecture, the benefits of this upgrade are obvious. We will be taking control of our entire infrastructure, right up to the very edge of the Internet. This means that, rather than depending on middlemen for IP transit, we will be in direct partnership with the carriers who we depend on for service. Additionally, Linode will quintuple the amount of bandwidth available to us currently, allowing us to absorb extremely large DDoS attacks until properly mitigated. As attack

sizes grow in the future, this architecture will quickly scale to meet their demands without any major new capital investment.

Final Words

Lastly, sincere apologies are in order. As a company that hosts critical infrastructure for our customers, we are trusted with the responsibility of keeping that infrastructure online. We hope the transparency and forward-thinking in this post can regain some of that trust.

We would also like to thank you for your kind words of understanding and support. Many of us had our holidays ruined by these relentless attacks, and it's a difficult thing to try and explain to our loved ones. Support from the community has really helped.

We encourage you to post your questions or comments below.

Filed under: announcements, retrospectives, upgrades by Alex Forster