Anycast vs. DDoS: Evaluating Nov. 30

Giovane C. M. Moura¹, Ricardo de O. Schmidt²,
John Heidemann³, Wouter B. de Vries²,
Moritz Muller¹, Lan Wei³, Christian Hesselman¹

¹SIDN Labs  ²University of Twente  ³USC/ISI

DNS-OARC    Dallas, Texas, USA    2016-10-16

Copyright © 2016 by John Heidemann
A Bad Day at the Root…

Data: RIPE DNSmon

Red: >30% loss
(some sites ~99% loss!)

What happened?

What does “red” really mean?

Anycast vs. DDoS in general?
DDoS: Bad and Getting Worse

• big and **getting bigger**
  – 2012: first 100Gb/s [Arbor12a]
  – 2016: 100Gb/s common; 540Gb/s seen; 1Tb/s possible

• easy and **getting easier**
  – 2012: several 1000+-node botnets
  – 2016: DDoS-as-a-service (booters): few Gb/s @ US$1

• frequent and **getting frequent-er**
  – 2002: the October 30 DNS root event
How Well Does Anycast Defend?

561 root DNS locations for 13 services (in 2016-01) large capex and opex

is 561 too few? too many? what happens under stress?

data: www.root-servers.org
Our Work: Study Nov. 30 Event

approach and goals

• gather public info about Nov. 30 event
• study it carefully
• identify design choices
• generalize for anycast
• suggest future defenses

non-approach and non-goals

• no inside information
• not bashing operators
• not just intentional, but also emergent policies
• not only about DNS and roots
• not help attackers
Contributions

• public evaluation of anycast under stress
• public articulation of design options
• evaluation of collateral damage

prior work for all, but in private

goals:
• public discussion => greater transparency
• expectation setting
• possible future defenses
Parts of Root DNS’ Anycast

- one root “.”
  - Q: .com’s NS? A: 192.5.6.30
- provided by 13 letters
  - 12 operators, 13 deployments
  - each different
  - each thoughtful
  - each constrained (peering, funding, etc.)
- 11 use IP anycast sites
  - 5 to 144 anycast sites for each anycast letter
  - (1 uses primary/secondary, 1 is single site)
- sites may have multiple servers
Anycast in Good Times

(some **sites** have more capacity)

anycast matches a **user** to a (hopefully) nearby **site**

anycast divides the Internet into **catchements** (often messy and non-geographic)

---

**X-SJC**

you

**your friend**

**X-PRG**

**another friend**

**X-SYD**
Anycast Under Stress

You, your friend X-SJC, and another friend X-SYD are under an attack from many attackers. The attackers are targeting your site X-SJC and another site X-SYD.

Too many attackers overwhelm your site, and your queries get lost. A similar size attack may be absorbed at a bigger site. Other attackers catchments also isolate sites from attackers.
1. **nothing**: X-SJC is **degraded absorber**, protecting X-SYD’s users 😞😊😊

**Anycast Reactions to Stress**

(do nothing?)
Anycast Reactions to Stress
(withdraw some routes?)

1. nothing: X-SJC is degraded absorber, protecting X-SYD’s users

2. withdraw routes from X-SJC; may shift attackers to big site
Anycast Reactions to Stress
(withdraw other routes?)

1. nothing: X-SJC is degraded absorber, protecting X-SYD’s users
2. withdraw routes from X-SJC; may shift attackers to big site
3. withdraw wrong routes from X-SJC; may shift attackers to other site

3. **withdraw** wrong routes from X-SJC; may shift attackers to other site
Best Reaction to Stress?
You Don’t Know

don’t know:
number of attackers
location of attackers
affects of routing change

don’t fully control
routing and catchments

1. nothing: X-SJC is degraded absorber, protecting X-SYD’s users
2. withdraw routes from X-SJC; may shift attackers to big site
3. withdraw wrong routes from X-SJC; may shift attackers to other site

don’t know:
number of attackers
location of attackers
affects of routing change
hard to make informed choices
What Actually Happens?

- studying Nov. 30
- we see withdrawals and degraded absorbers
- some clients lose service
- results vary
  – by anycast deployment
Data About Nov. 30

- RIPE Atlas
  - ~9000 vantage points (RIPE Atlas probes)
  - try every letter every 4 minutes
    - except A-root, at this time, was every 30 minutes
    - CHAOS query identifies server and implies site
    - targets letters, not Root DNS (cannot switch letter)
  - global, but heavily biased to Europe
  - we map server->site
    - map will be public dataset
- RSSAC-002 reports
  - self-reports from letters
  - not guaranteed when under stress
- BGPmon routing
  - control plane

6996 RIPE Atlas VPs on 2015-11-30 (looking at K-Root)
Summary of the Events

- two events
  - 2015-11-30T06:50 for 2h40m
  - 2015-12-01T05:10 for 1h
- affected 10 of 13 letters
- about 5M q/s or 3.5Gb/s per affected letter
  - aggregate: 34Gb/s
- real DNS queries, common query names, from spoofed source IPs
- implications:
  - some letters had high loss
  - overall, though DNS worked fine
    - clients retried other letters (as designed)
  - but want to do better

data:
A-Root had full view
(Verisign presentation);
RSSAC-002 reports
How About the Letters?

some did great:
D, L, M: not attacked
A: no visible loss

most suffered:
a bit (E, F, I, J, K)
or a lot (B, C, G, H)

but does “x%” measure what
users actually see?
View from Atlas Vantage Points

K overall: ~30% loss (not bad)

but these 300 VPs: 70-90% loss to K

=> loss is uneven; some users very sad

=> “30% loss” may imply all VPs lose; doesn’t show uneven distribution
Reachability at K’s Sites

sites see fewer VPs, but why?
- query loss? site absorbs attack, but sad customers
Site Flips from Routing Changes

[Moura16a, figure 11; data: RIPE Atlas]
Site Flips from Routing Changes

360 minutes (in 4 minute bins)

Nov. 30 event

stay at K-LHR;
(sad during event)

flip to K-AMS;
(less) sad during event;
back to K-LHR after

flip to K-other
and stay there

flip to K-AMS

[Moura16a, figure 11b; data: RIPE Atlas]
flips common during events for most letters

flips seen in BGP
Flips Across Letters: E and K

to evaluate flips over two days: compare minimum and maximum catchment (measured in VPs/site)

normalize to median VPs (the natural catchment), to correct for uneven Atlas locations

sites acquiring VPs (during event?)

(red sites: <20 VPs; not enough to provide meaningful results)

sites shedding VPs

[Moura16a, figure 5; data: RIPE Atlas]
Flips: Implications

- Some ISPs are “sticky” and won’t flip
  - Will suffer if their site is overloaded
- Some ISPs will flip
  - But new site may not be much better
- Result depends on many factors
  - Actions taken by root operator
  - Routing choices by operator and peer
    - And perhaps peer’s peers, depending on congestion location
  - Implementation choices
    - DNS, routing
Anycast Under Stress: What Should Happen?

- consider a service
  - 3 sites: s1, s2, S3
  - s1 and s2: 1Gb/s
  - S3: 10Gb/s
- with clients
  - 4 clients: c0 to c3
- the attack
  - A0 and A1
  - each: 0.49, 0.99, 4.9, or 6Gb/s
- what is the optimal, ideal defense?
  - assume static attackers
  - defender knows attack strengths
  - defender controls routing
- metric: *Happiness* $H$: number of clients served
Anycast Under Stress: What Should Happen?

1. \(A_0 + A_1 < s_1\): do nothing; \(H=4\)
2. \(A_0 < s_1\) and \(A_0 + A_1 > s_2\): shed load; \(H=4\)
   - vs. \(H=2\) if do nothing
3. \(A_0 > s_1\) and \(A_0 + A_1 < s_3\):
   keep only big site; \(H=4\)
   - vs. \(H=2\) if nothing
4. \(A_0 + A_1 > S_3\): do nothing (\(s_1\) is degraded absorber); \(H=2\)

⇒ with today’s uncertainty: “do nothing” looks good
⇒ future goal: what is needed (measurement and control) to do better?
Anycast Under Stress: What Should Happen?

1. $A_0 + A_1 < s_1$: do nothing; $H=4$
2. $A_0 < s_1$ and $A_0 + A_1 > s_2$: shed load; $H=4$
   - vs. $H=2$ if do nothing
3. $A_0 > s_1$ and $A_0 + A_1 < s_3$: keep only big site; $H=4$
   - vs. $H=2$ if nothing
4. $A_0 + A_1 > S_3$: do nothing ($s_1$ is degraded absorber); $H=2$

⇒ with today’s uncertainty: “do nothing” looks good
⇒ future goal: what is needed (measurement and control) to do better?
Anycast Under Stress: What Should Happen?

1. $A_0 + A_1 < s_1$: do nothing; $H=4$
2. $A_0 < s_1$ and $A_0 + A_1 > s_2$: shed load; $H=4$
   vs. $H=2$ if do nothing
3. $A_0 > s_1$ and $A_0 + A_1 < s_3$: keep only big site; $H=4$
   vs. $H=2$ if nothing
4. $A_0 + A_1 > S_3$: do nothing ($s_1$ is degraded absorber); $H=2$

$\Rightarrow$ with today’s uncertainty: “do nothing” looks good
$\Rightarrow$ future goal: what is needed (measurement and control) to do better?
Anycast Under Stress: What Should Happen?

1. $A_0 + A_1 < s_1$: do nothing; $H=4$
2. $A_0 < s_1$ and $A_0 + A_1 > s_2$: shed load; $H=4$
   \hspace{1cm} \text{vs. } H=2 \text{ if do nothing}
3. $A_0 > s_1$ and $A_0 + A_1 < s_3$: keep only big site; $H=4$
   \hspace{1cm} \text{vs. } H=2 \text{ if nothing}
4. $A_0 + A_1 > S_3$: do nothing (s1 is degraded absorber); $H=2$

⇒ with today’s uncertainty: “do nothing” looks good
⇒ future goal: what is needed (measurement and control) to do better?
During An Event: 
Active Routing Changes or Not?

- no active routing changes
  - should expect partial loss in future attacks
    - inevitable: non-uniform attacker and defender capacity
  - overloaded catchments will suffer during attack
  - need to pre-deploy excess capacity
  - *operators understand and are doing these; but what about user expectations?*

- active routing changes
  - important when aggregate attack and defense capacity is similar
    - if one exceeds the other, no need to bother
  - requires *much* better measurement and route control
    - seems like a research problem; AFAIK no tools today
  - important to reduce client losses at smaller sites
  - *seems necessary to get to 0% loss*
Aside: Collateral Damage

• can an event hurt non-targets?
• yes! ...a risk of shared datacenters

D-FRA and D-SYD: less traffic (even though D was not directly attacked)

In other attacks, B-Root’s ISP saw loss to other customers

.NL-FRA and .NL-AMS: \textit{no} traffic
Recommendations

• current approach reasonable
  – build out capacity in advance
  – no active re-routing during attack
  – should expect some loss during each attack
• need true diversity to avoid collateral damage
• longer-term
  – need research to improve measurement and control
  – active control can improve loss during some attacks
• how many sites needed?
  – there is a *lot* of capacity already
  – many small sites seem to increase partial outages
Conclusions

- Anycast under stress is complicated
  - Some users will see persistent loss
  - “x% loss” is not a complete picture

- Options:
  - Pre-deploy + no change during
    is a reasonable choice today
  - To avoid loss, will need to do more

- More info:
  - Paper: http://www.isi.edu/~johnh/PAPERS/Moura16b
  - Data: https://ant.isi.edu/datasets/anycast/