Anycast vs. DDoS: Evaluating Nov. 30

Giovane C. M. Moura[1], Ricardo de O. Schmidt[2],
John Heidemann[3], Wouter B. de Vries[2],
Moritz Muller[1], Lan Wei[3], Christian Hesselman[1]

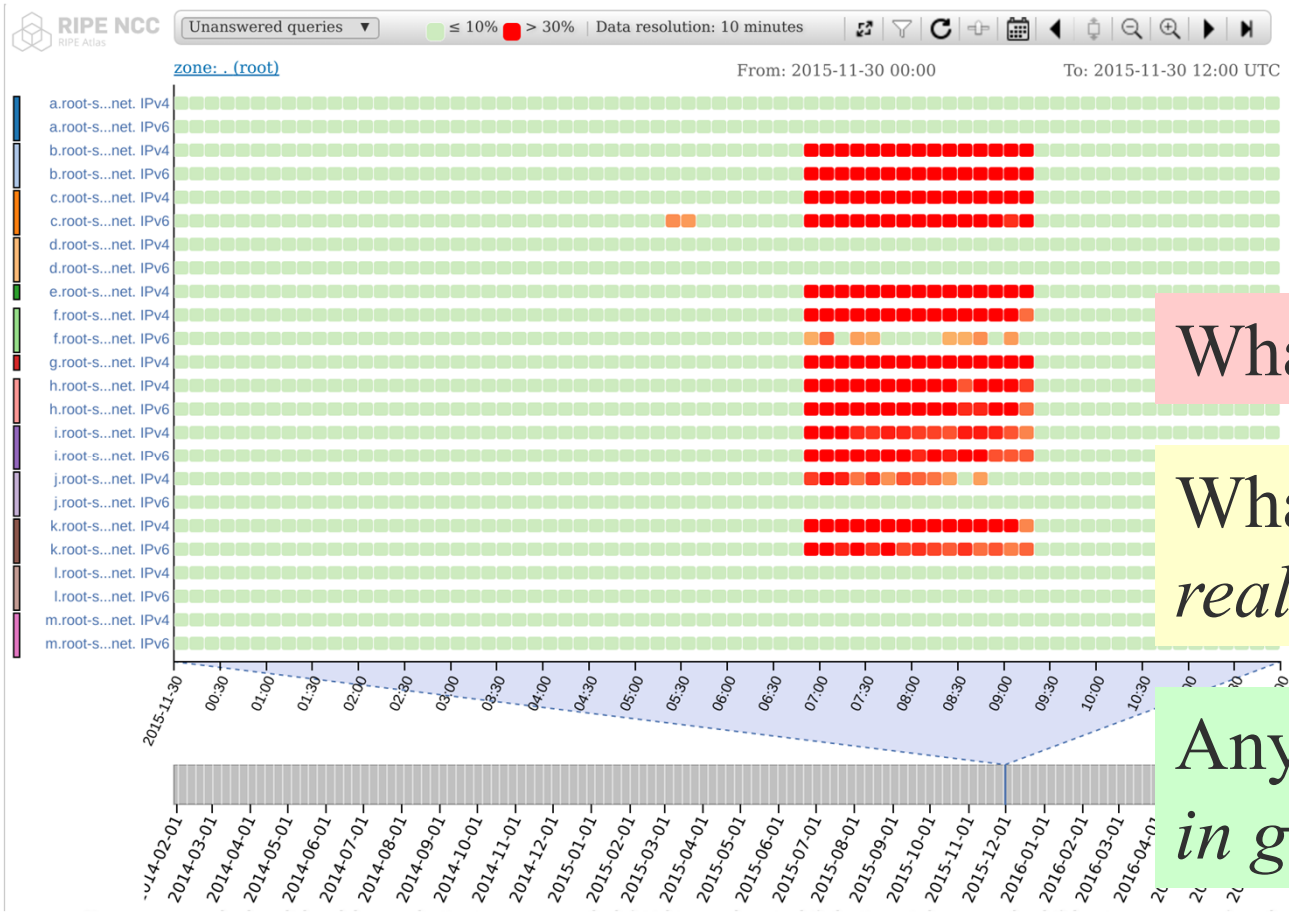
[1]SIDN Labs [2]University of Twente [3]USC/ISI

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

Based on a technical paper "Anycast vs. DDoS: Evaluating the November 2015 Root DNS Event", to appear at ACM IMC 2016.

USC Viterbi School of Engineering Information Sciences Institute
UNIVERSITY OF TWENTE.
SIDN LABS
ant.isi.edu

A Bad Day at the Root…



data: RIPE DNSmon
red: >30% loss
(some sites ~99% loss!)

What happened?

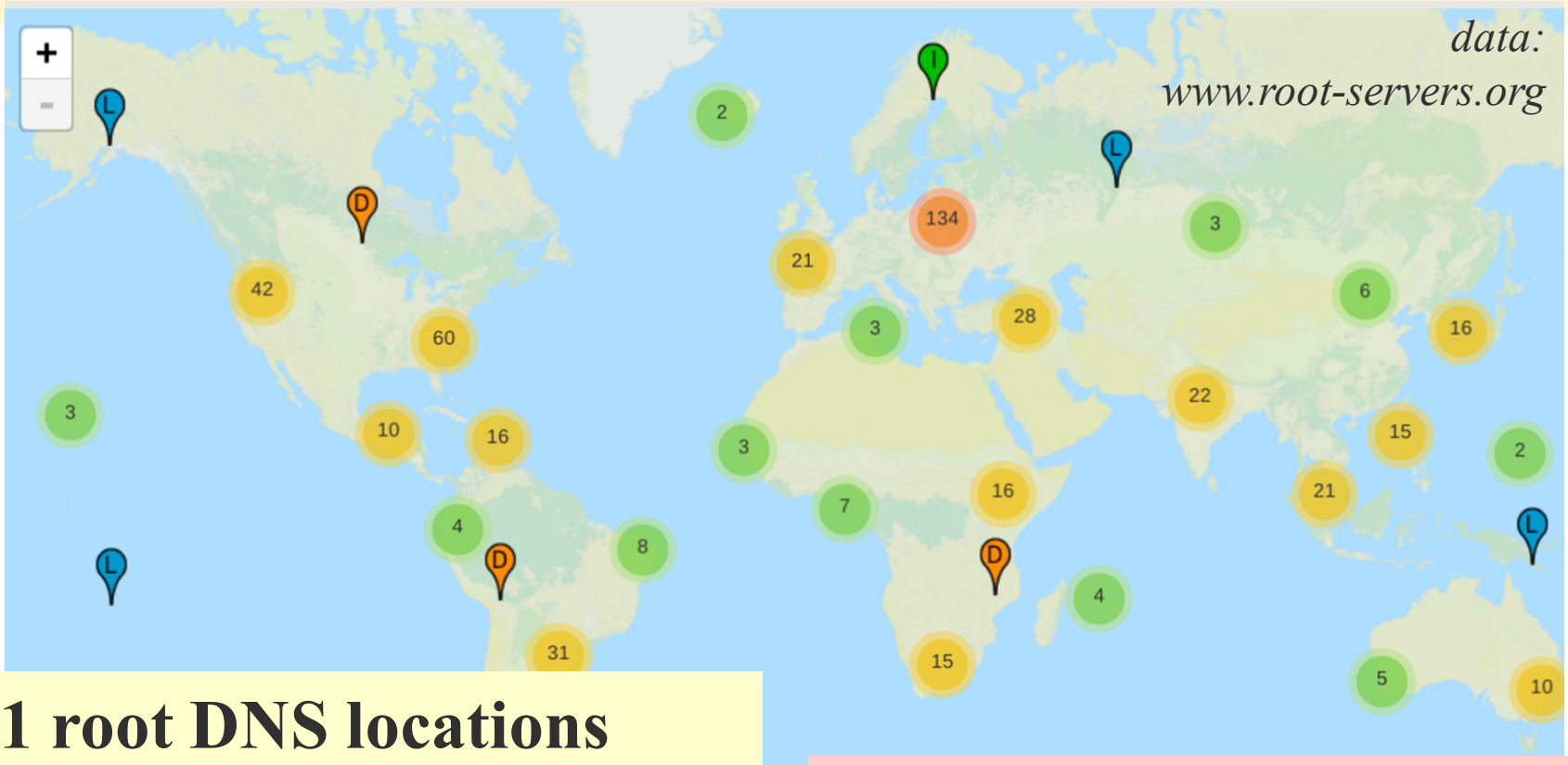
What does "red" *really* mean?

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
  - 2016: 3 recent big attacks (2015-11-30, 2015-12-01, 2016-06-25)

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?***

Our Work: Study Nov. 30 Event

approach and goals

- gather public info about Nov. 30 event
- study it *carefully*
  - blank
- identify design choices
- generalize for anycast
  - blank
- 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

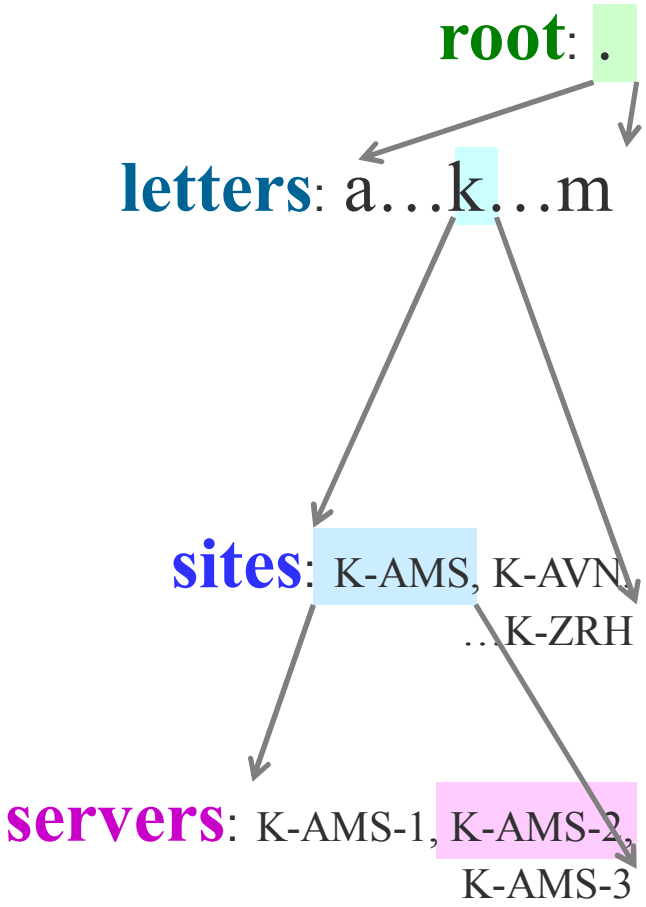
prior work for *all*, but in *private*

goals:

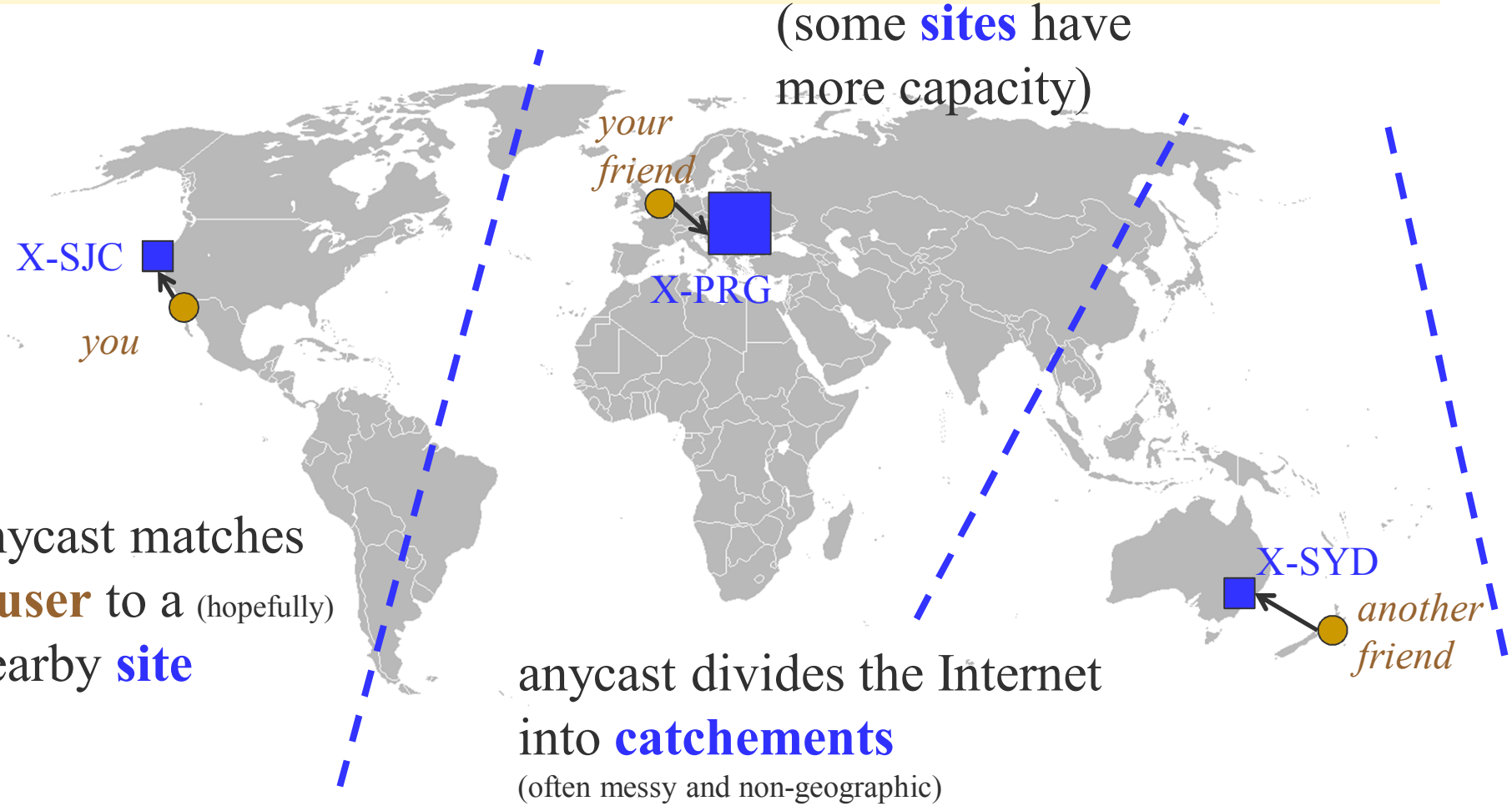
- public discussion => greater transparency
- expectation setting
- possible future defenses

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**



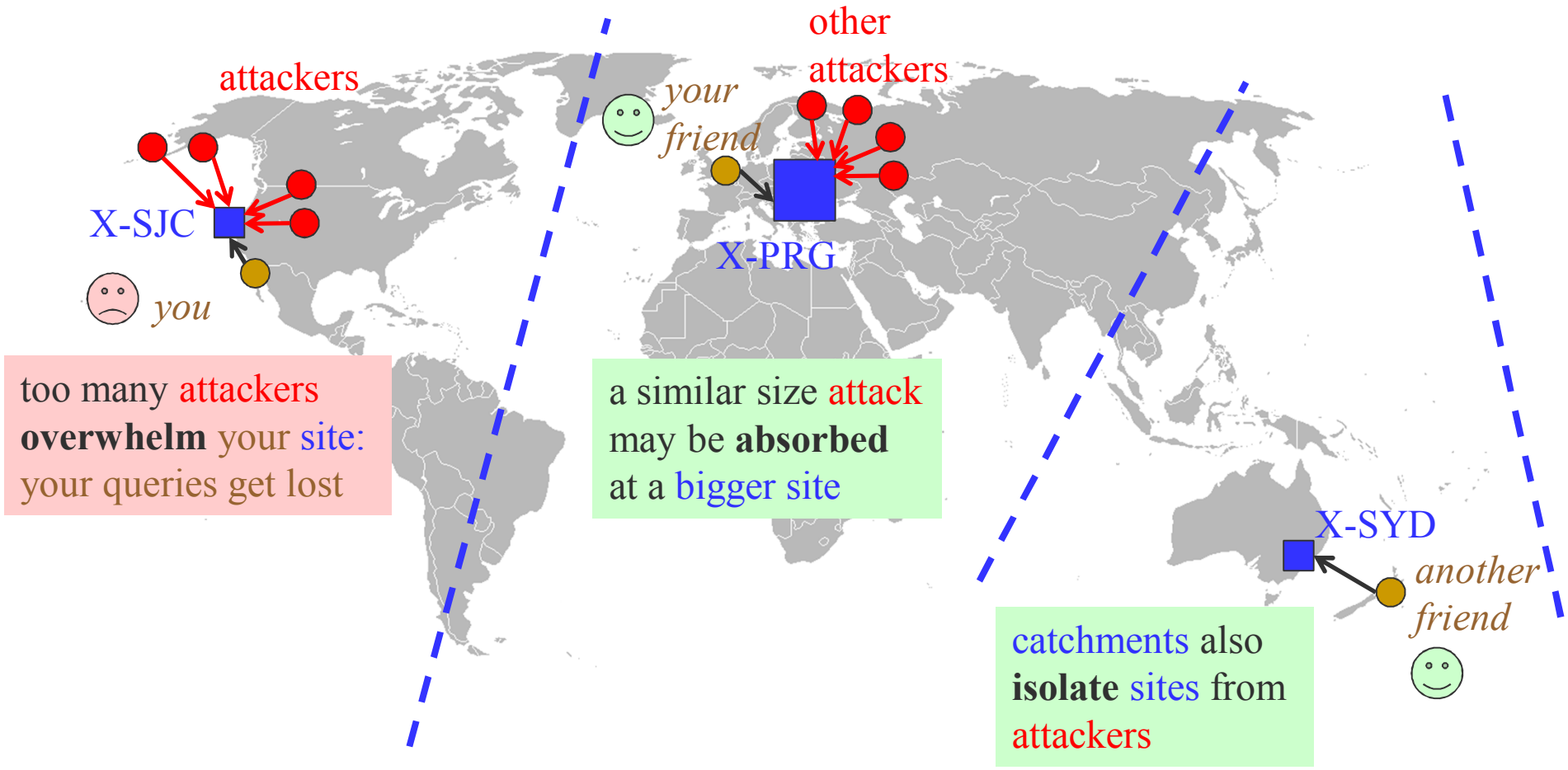
Anycast in Good Times



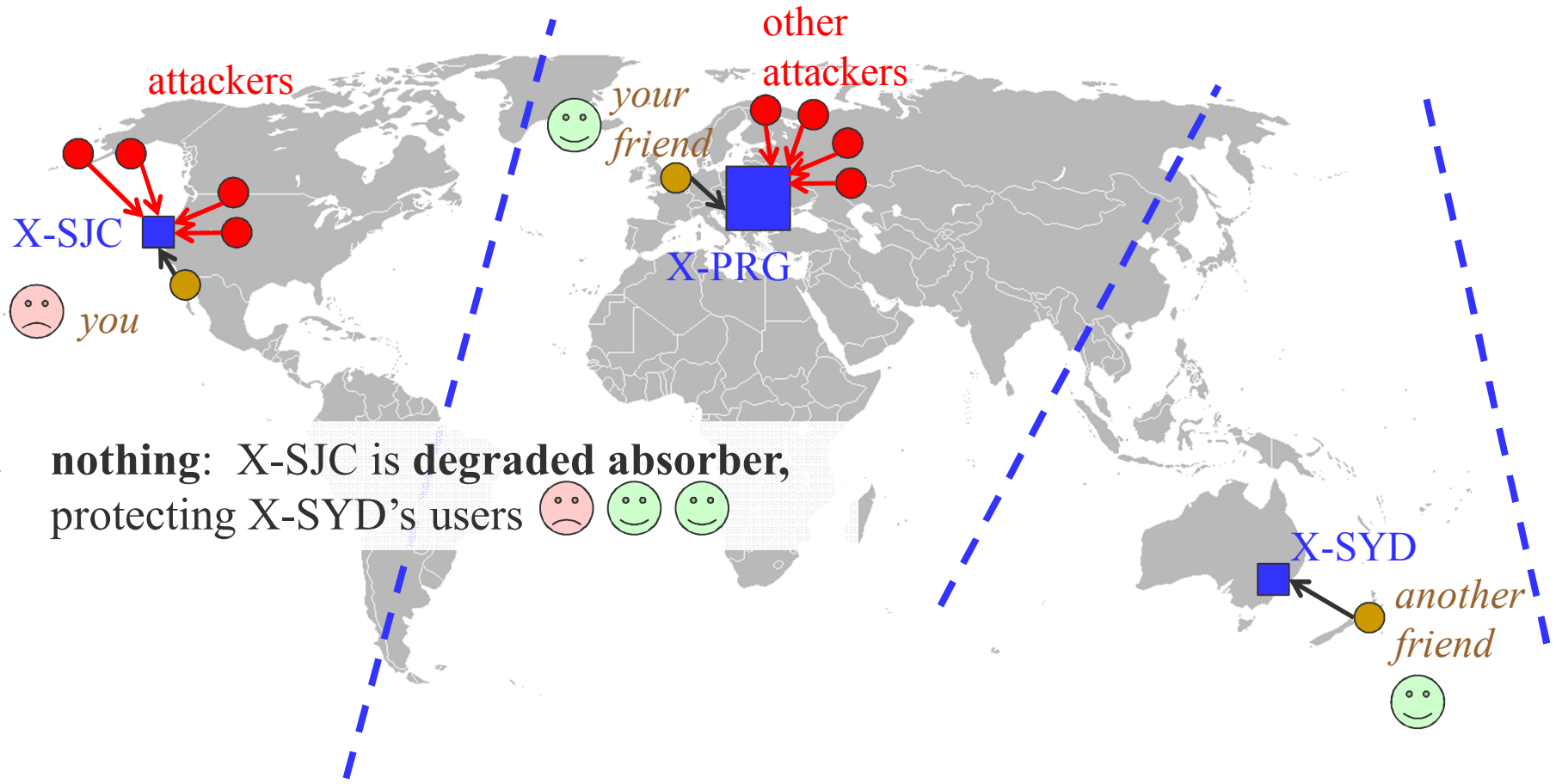
anycast matches a **user** to a (hopefully) nearby **site**

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

Anycast Under Stress

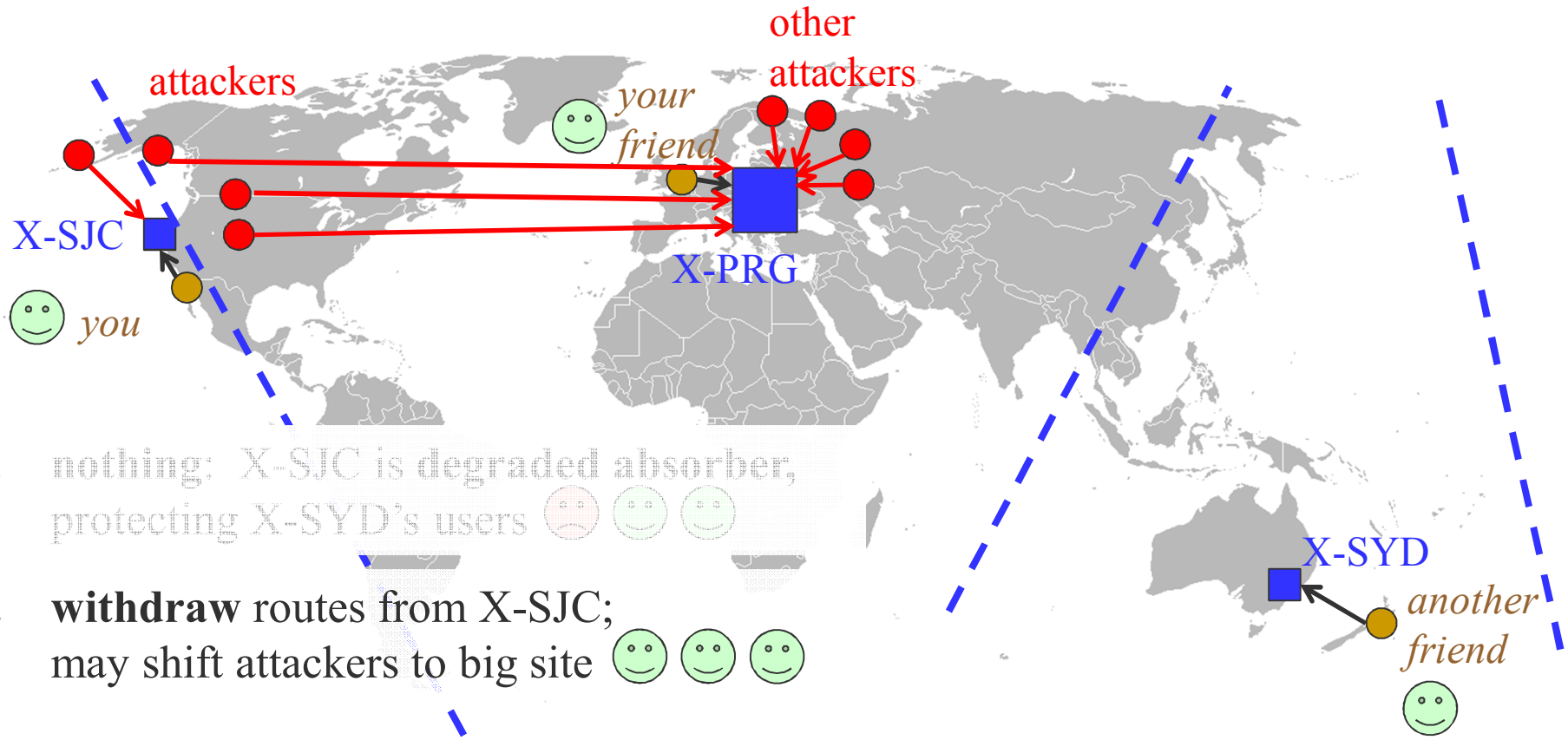


Anycast Reactions to Stress
(do nothing?)



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

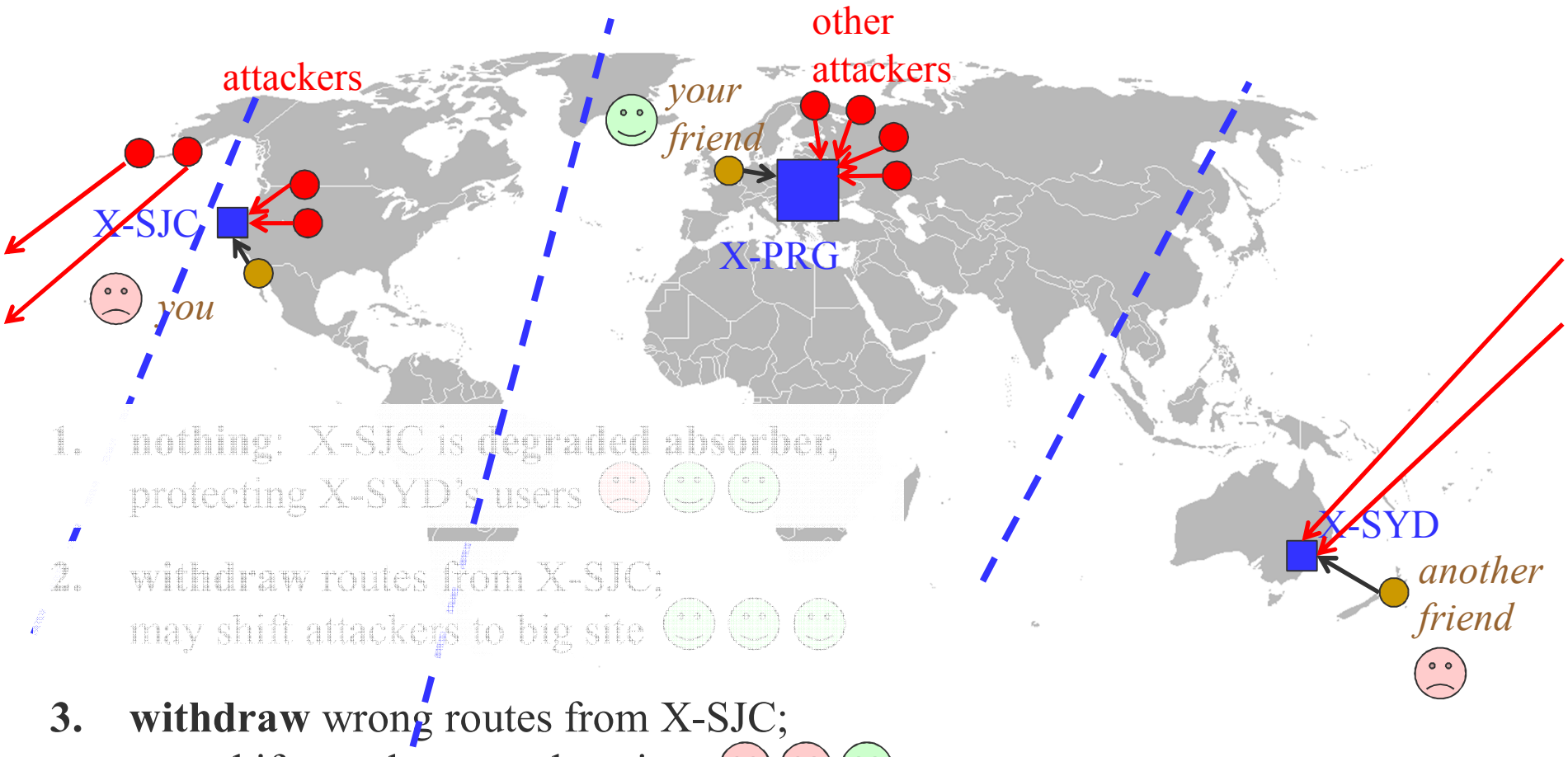
Anycast Reactions to Stress
(withdraw some routes?)



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

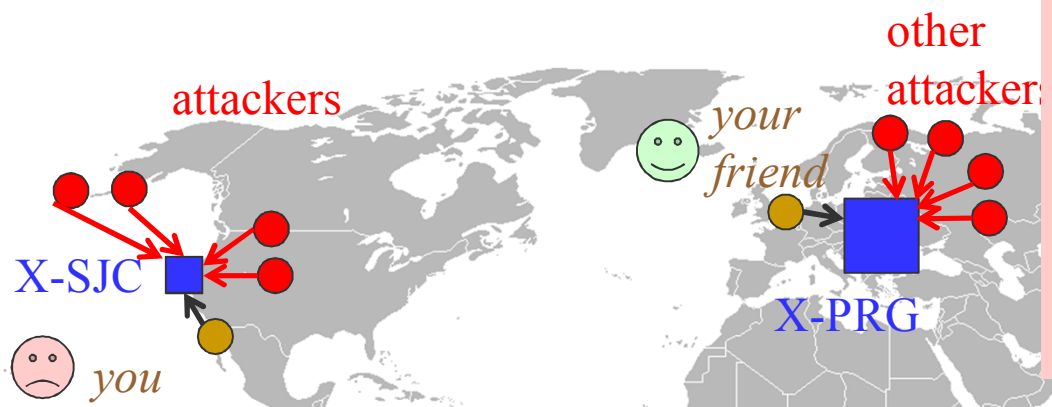
2. **withdraw** routes from X-SJC; may shift attackers to big site

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

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 al
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

-SYD
another
friend

hard to make
informed choices

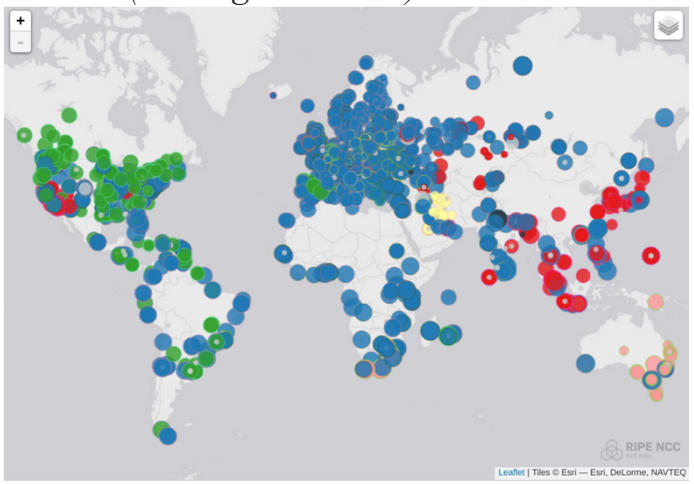
What Actually Happens?

- 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**

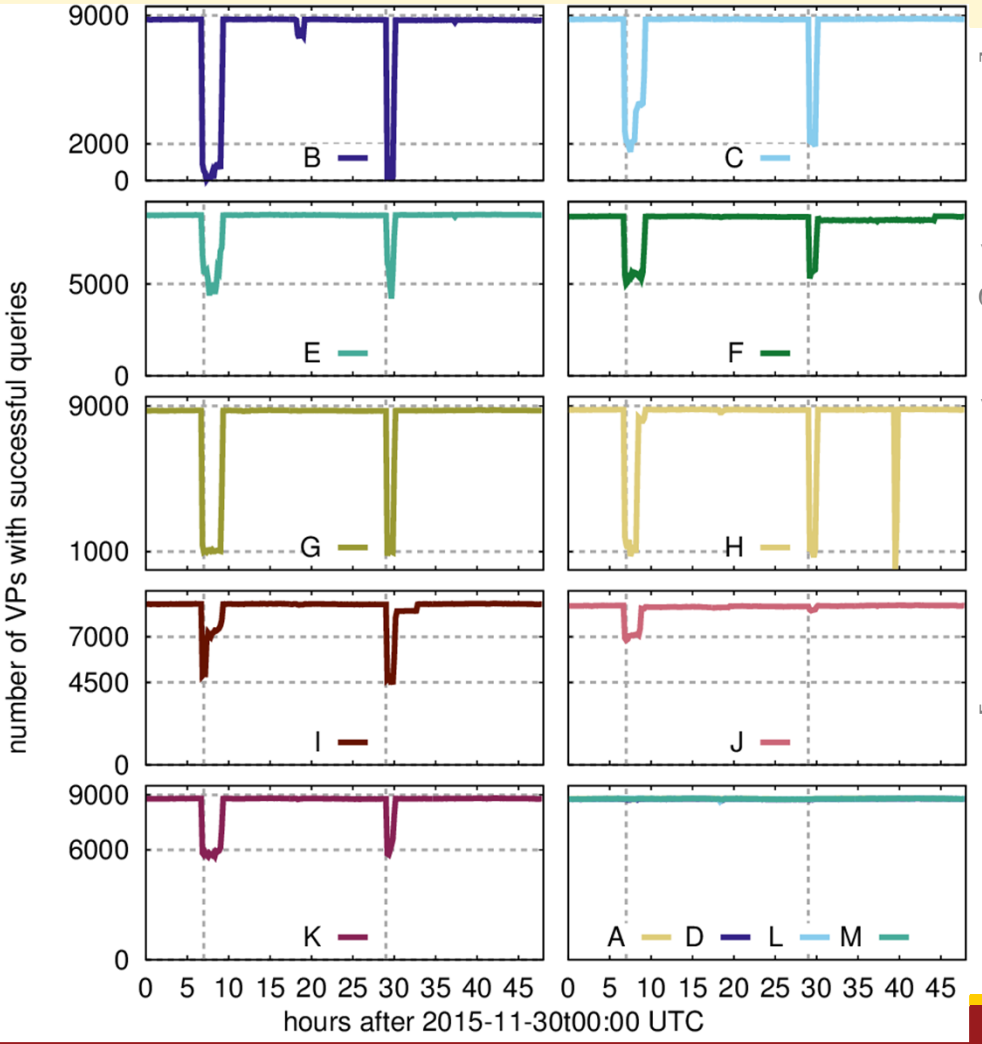
data:
A-Root had full view (Verisign presentation);
RSSAC-002 reports

How About the Letters?

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

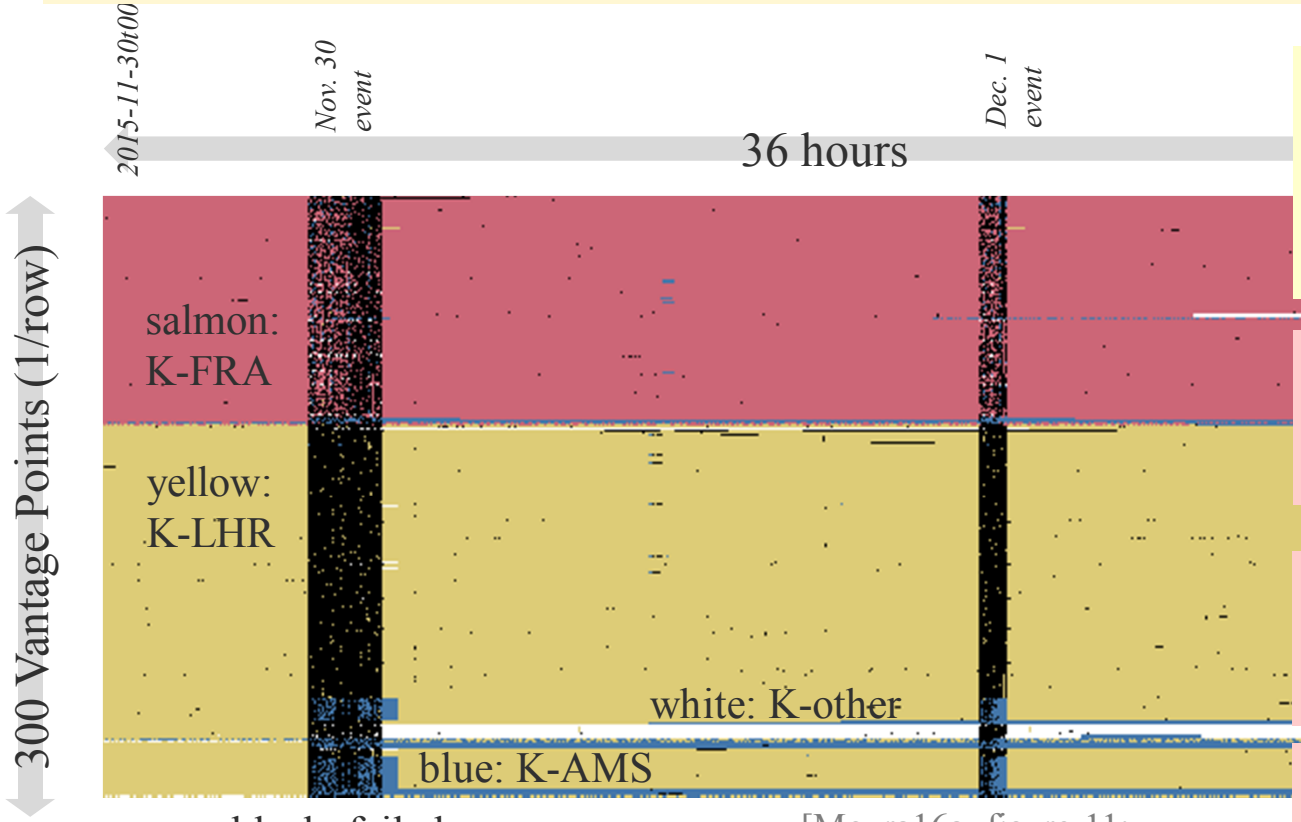
most suffered:
a bit (E, F, I, J, K)
or a lot (B, C, G, H)

but does "x%"
measure what
users actually see?



[Moura16a, figure 3; data: RIPE Atlas]

View from Atlas Vantage Points



[Moura16a, figure 11; data: RIPE Atlas]

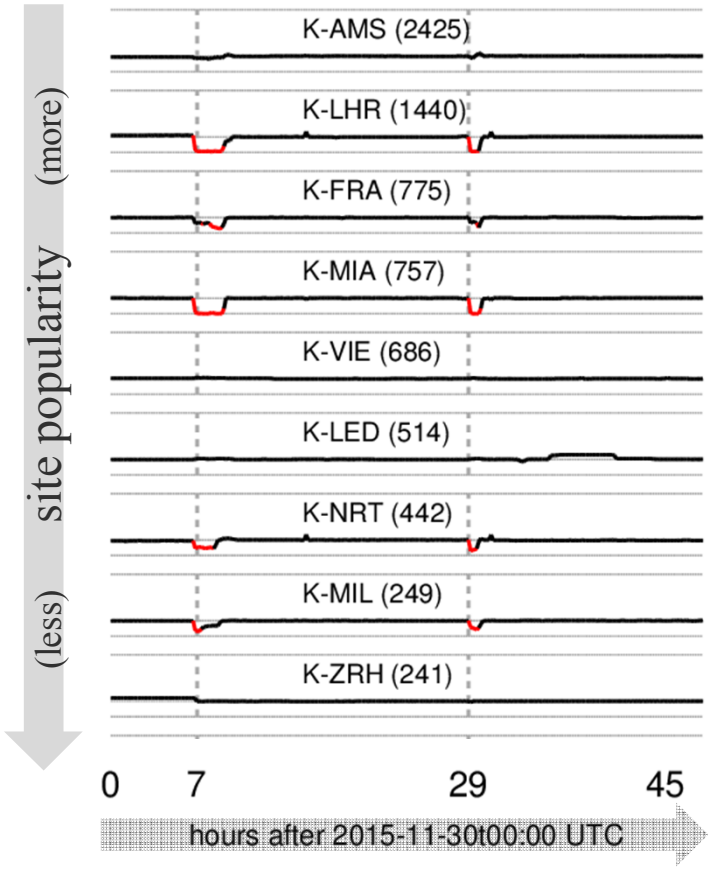
K overall: ~30% loss (not bad)

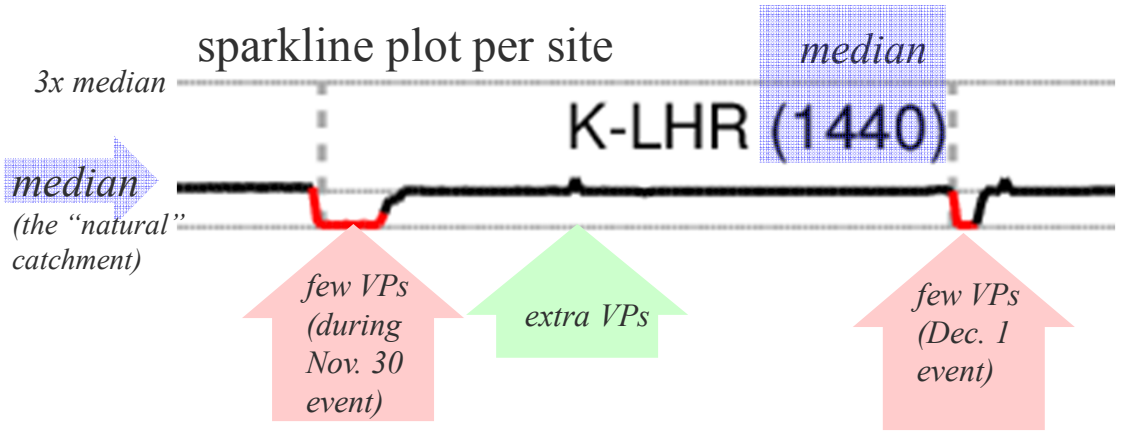
but *these 300 VPs*: 70-90% loss to K

=> loss is uneven; some users very sad

=> "30% loss" may imply all VPs lose; doesn't show *uneven distribution*

Reachability at K's Sites

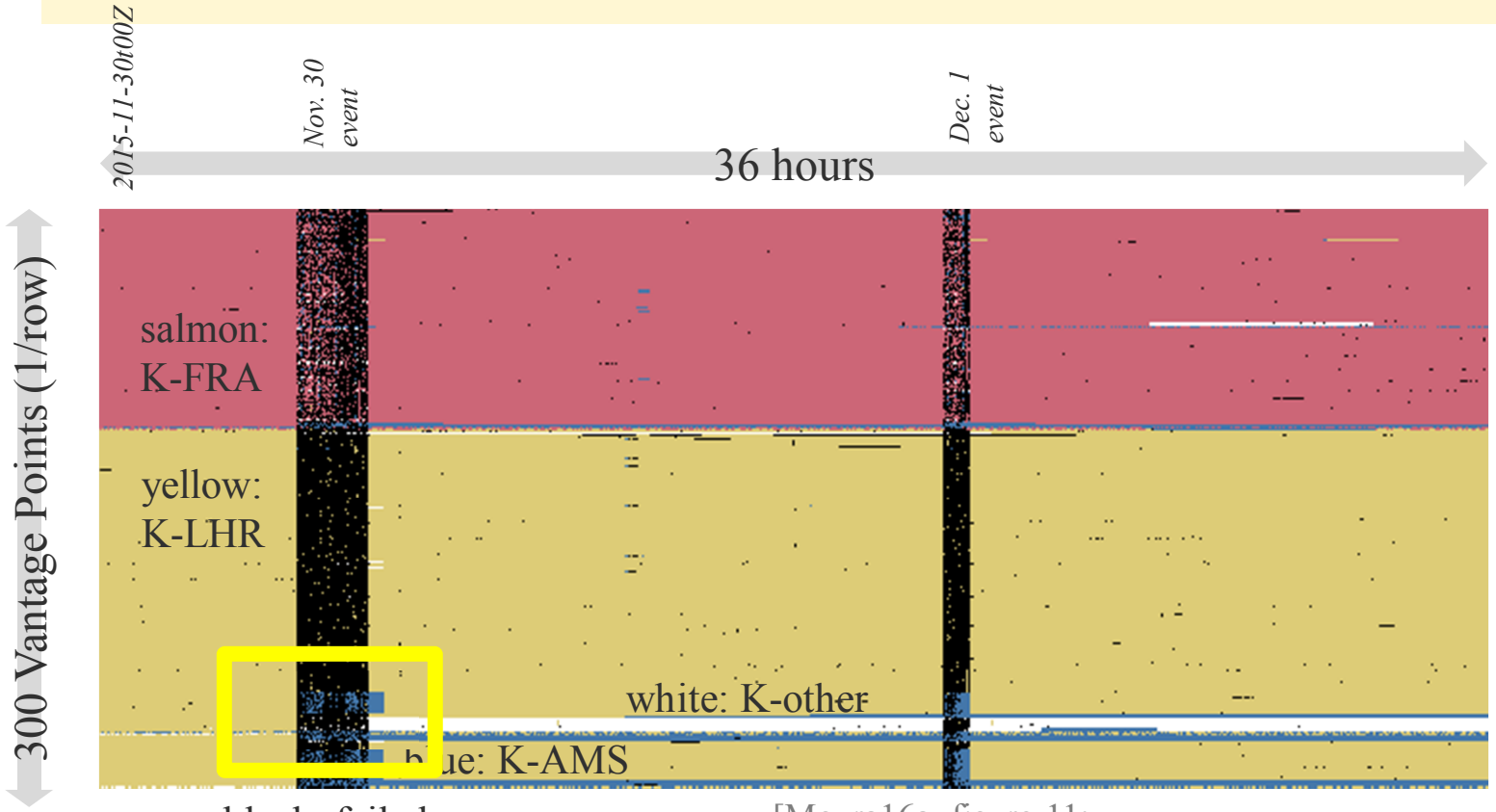




sites see fewer VPs, but why?

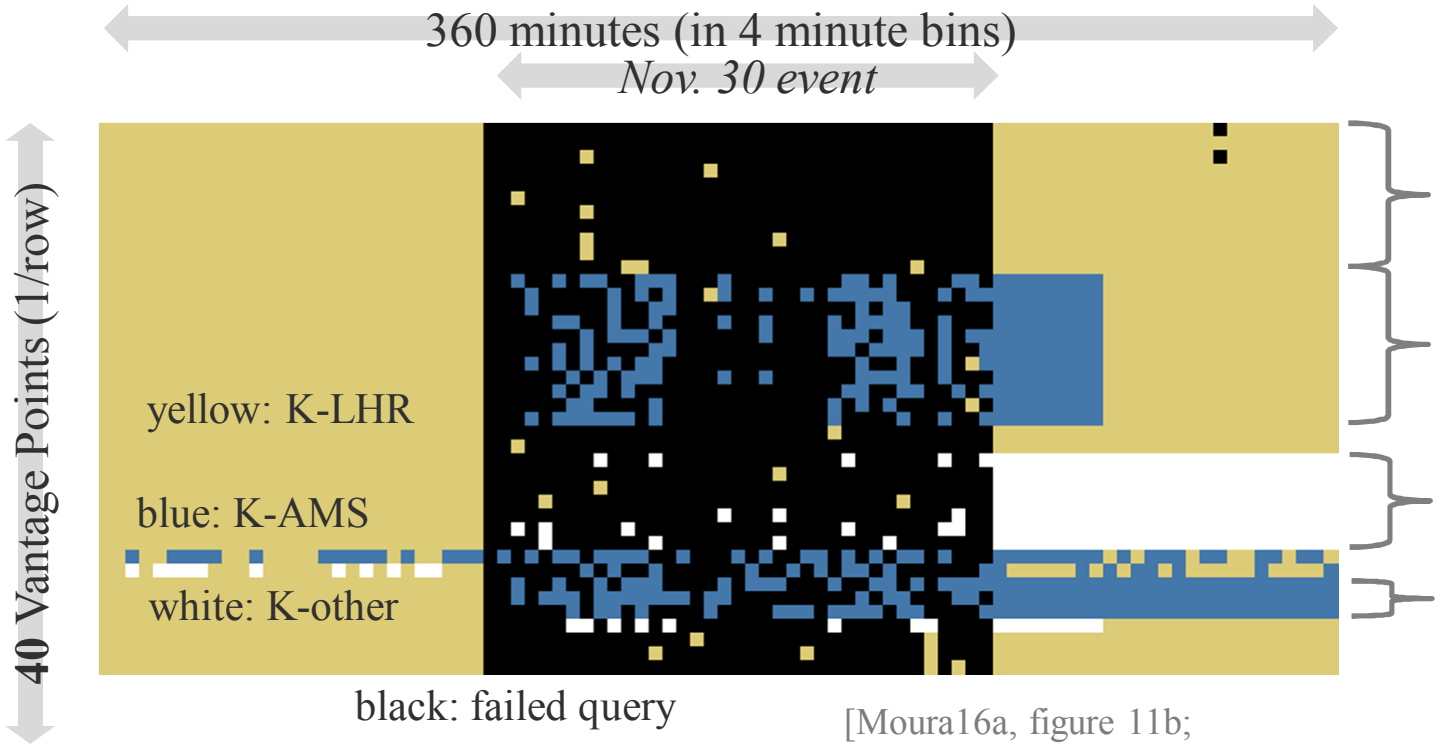
- query loss? site absorbs attack, but sad customers
- route change? who? why? where?

Site *Flips* from Routing Changes



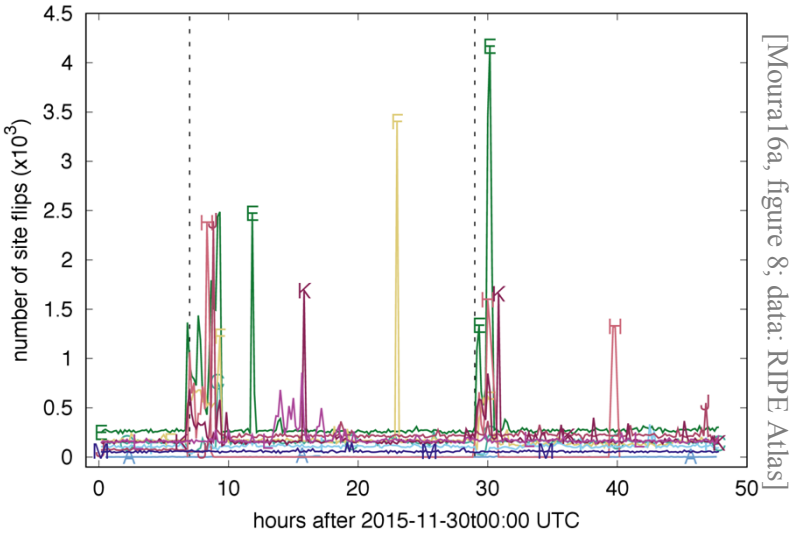
[Moura16a, figure 11; data: RIPE Atlas]

Site *Flips* from Routing Changes

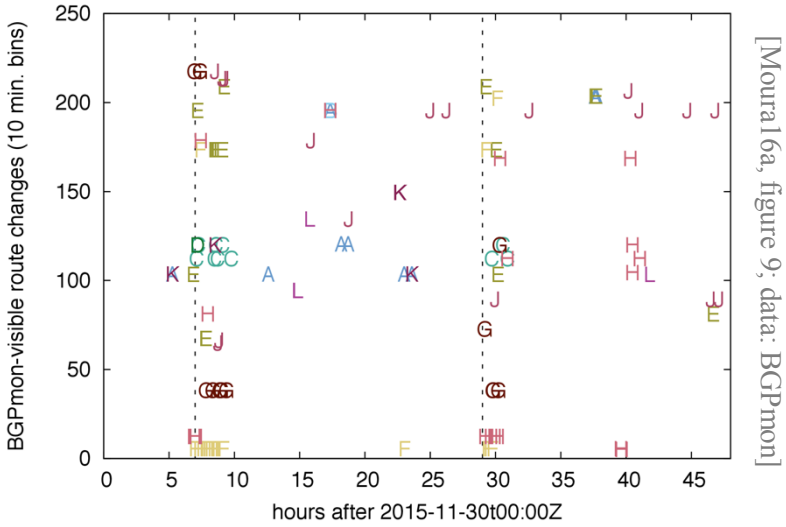


[Moura16a, figure 11b; data: RIPE Atlas]

Confirming Flips in BGP



flips common during events for most letters

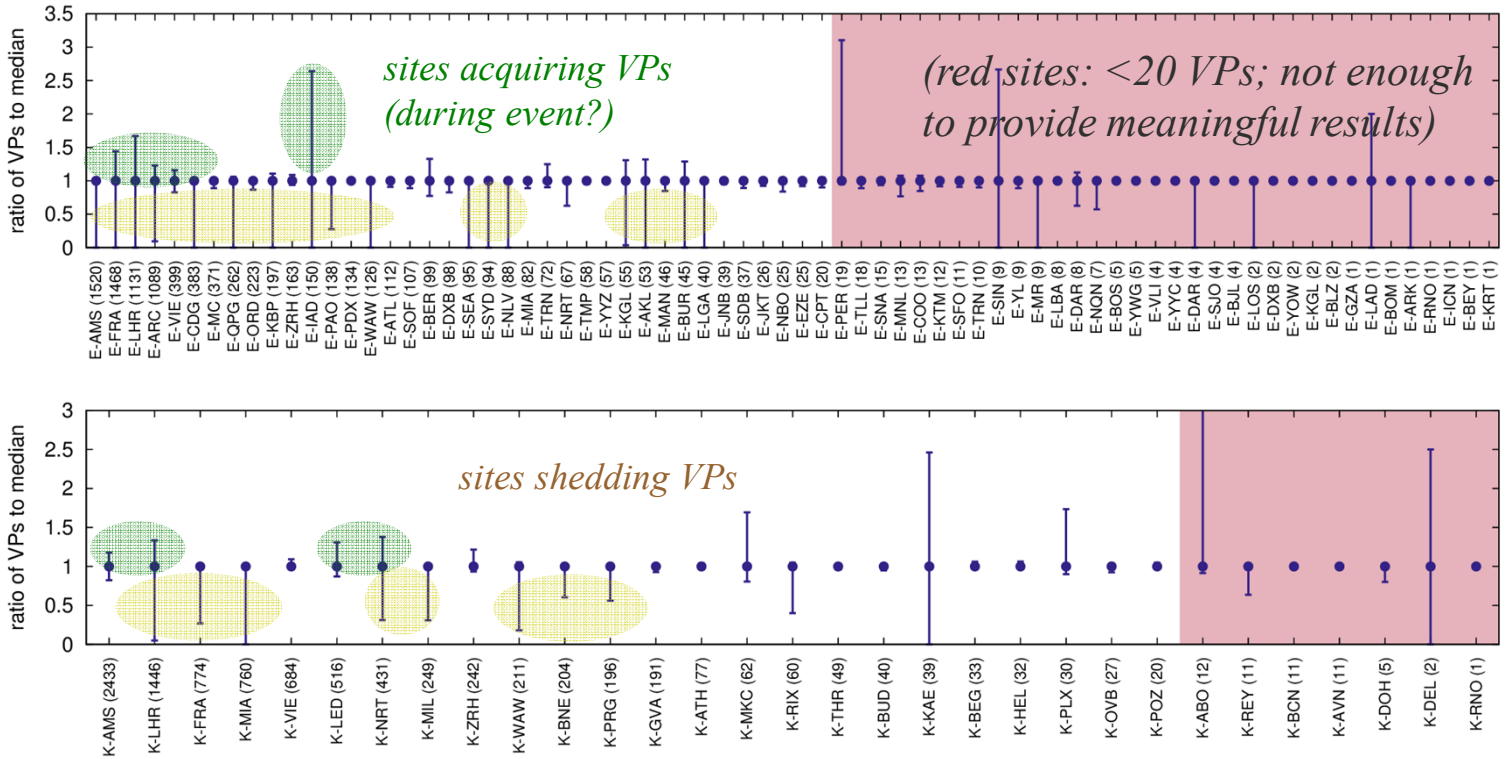


flips seen in BGP

Flips Across Letters: E and K

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

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

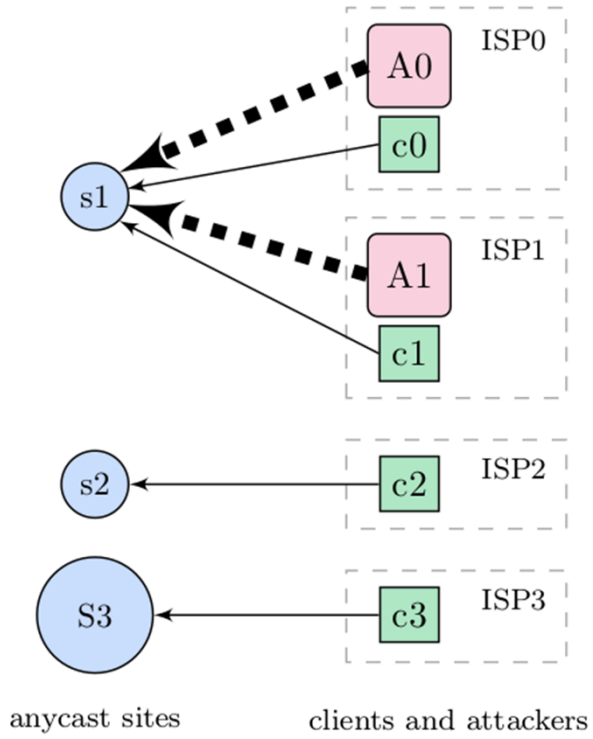


[Moura16a, figure 5; data: RIPE Atlas]

Flips: Implications

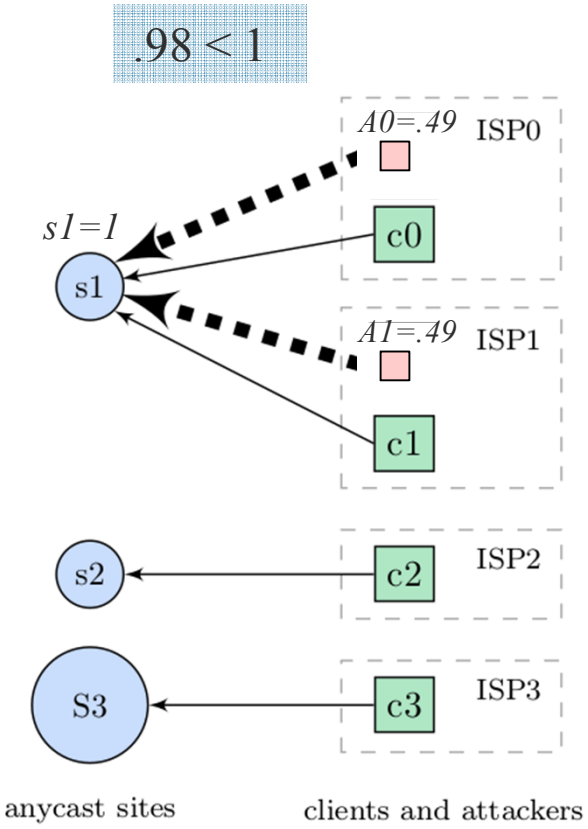
- some ISPs are “sticky” and won’t flip
  - will suffer if their site is overloaded
- some ISPs will flip
  - but new site may not be much better
- result depends on many factors
  - actions taken by root operator
  - routing choices by operator *and peer*
    - and perhaps *peer’s peers*, depending 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?



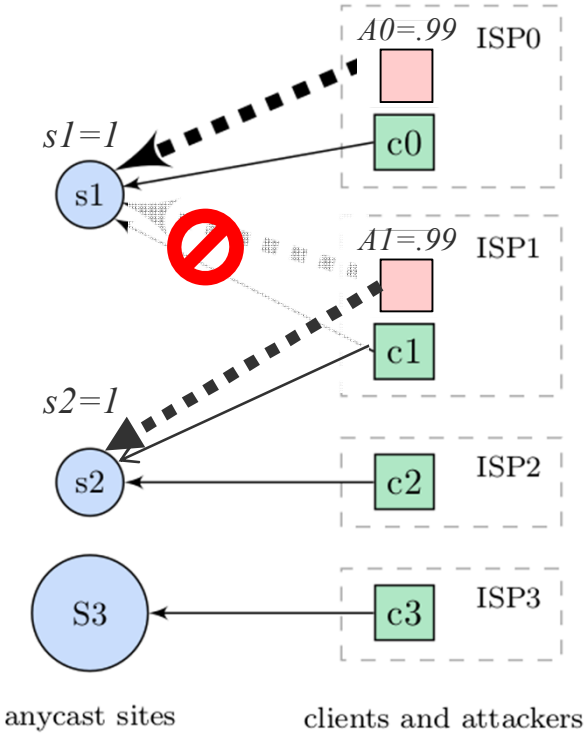
1. A0+A1 < s1: **do nothing; H=4**
2. A0 < s1 and A0+A1 > s2: shed load; H=4
   - vs. H=2 if do nothing
3. A0 > s1 and A0+A1 < s3: keep only big site; H=4
   - vs. H=2 if nothing
4. A0+A1 > S3: do nothing (s1 is degraded absorber); H=2

⇒ with today's uncertainty: "do nothing" looks good

⇒ future goal: what is needed (measurement and control) to do better?

Anycast Under Stress: What *Should* Happen?

.99 < 1 and 1.98 > 1



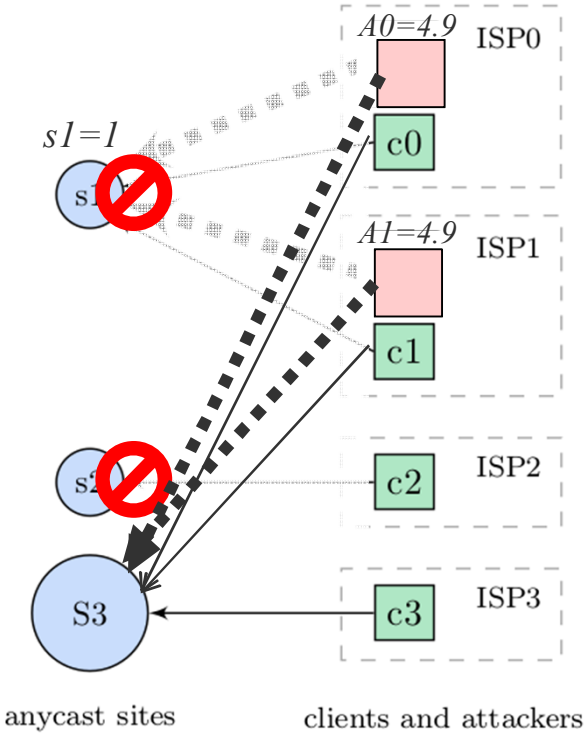
1. A0+A1 < s1: do nothing; H=4
2. A0 < s1 and A0+A1 > s2: **shed load; H=4**
   - vs. H=2 if do nothing
3. A0 > s1 and A0+A1 < s3: keep only big site; H=4
   - vs. H=2 if nothing
4. A0+A1 > S3: do nothing (s1 is degraded absorber); H=2

⇒ with today's uncertainty: "do nothing" looks good

⇒ future goal: what is needed (measurement and control) to do better?

Anycast Under Stress: What *Should* Happen?

4.9 > 1 and 9.8 < 10

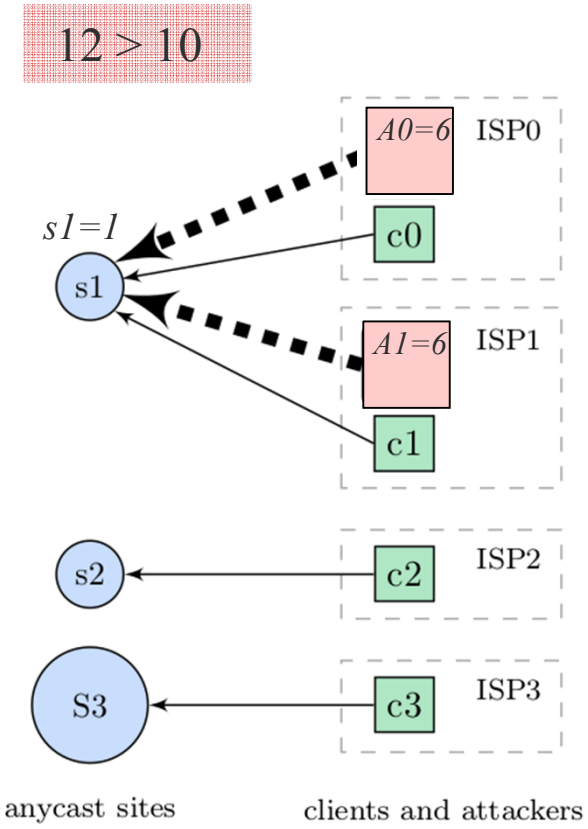


1. A0+A1 < s1: do nothing; H=4
2. A0 < s1 and A0+A1 > s2: shed load; H=4
   - vs. H=2 if do nothing
3. A0 > s1 and A0+A1 < s3: **keep only big site; H=4**
   - vs. H=2 if nothing
4. A0+A1 > S3: do nothing (s1 is degraded absorber); H=2

⇒ with today's uncertainty: "do nothing" looks good

⇒ future goal: what is needed (measurement and control) to do better?

Anycast Under Stress: What *Should* Happen?



1. A0+A1 < s1: do nothing; H=4
2. A0 < s1 and A0+A1 > s2: shed load; H=4
   - vs. H=2 if do nothing
3. A0 > s1 and A0+A1 < s3: keep only big site; H=4
   - vs. H=2 if nothing
4. A0+A1 > S3: **do nothing** (s1 is degraded absorber); **H=2**

⇒ with today's uncertainty: "do nothing" looks good

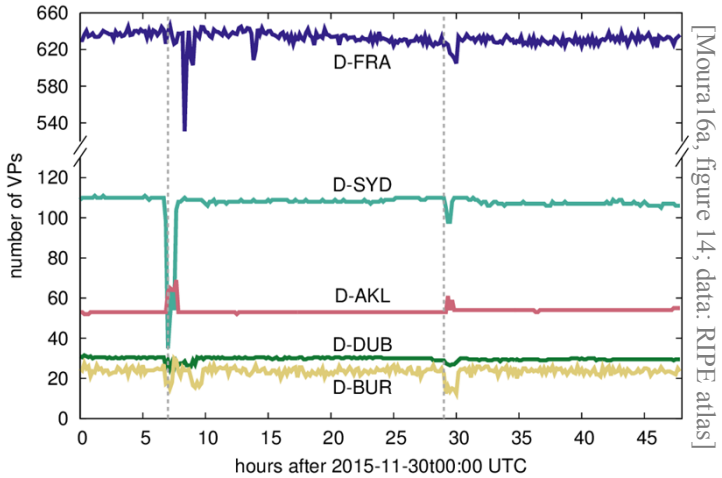
⇒ future goal: what is needed (measurement and control) to do better?

During An Event: Active Routing Changes or Not?

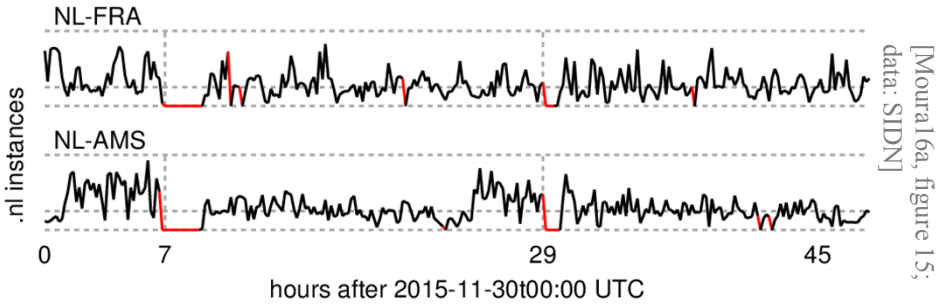
- no active routing changes
  - should expect partial loss in future attacks
    - inevitable: non-uniform attacker and defender capacity
  - overloaded catchments will suffer during attack
  - need to pre-deploy excess capacity
  - *operators understand and are doing these; but what about user expectations?*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



.NL-FRA and .NL-AMS: *no* traffic

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

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 complete picture
- options:
  - pre-deploy + no change during is 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/