Edge/Core Update Propagation/ Churn vs. Performance Preliminary Results

Avi Freedman

Akamai Technologies

Topics

- Project History & Goals
- Data sets & Infrastructure
- Traffic Density/Churn Correlation
- Future Work
- Credits

BGP Movie Titles

Return of BGP

RIP Strikes Back

Revenge of the BGP

Scary BGP; Chasing BGP

The Wrath of BGP

Fatal Announcements

Fried Green BGP

Silent Route Strikes Back

Being BGP

“Shall We Announce”

Good Flap Hunting

A Route’s Life

O Prefix Where Art Thou

The Death of a Prefix

The Unreachables

BGP Inc.; BGP Wars

Crouching Announcement, Hidden Withdrawals

A Few Good Announcements

Grumpy old BGP

The Dead Prefix Society

4 Announcements & A Withdrawal

7007: A BGP Oddysey

Much Ado About Flapping

Sense & Reachability

BGPless in Seattle

While You Were Announcing

I Know What You Announced Last Session

BGP Movie Titles (Top 14)

Fatal Announcements

“Shall We Announce”

Good Flap Hunting

O Prefix Where Art Thou

The Death of a Prefix

The Unreachables

Crouching Announcement, Hidden Withdrawals

A Few Good Announcements

The Dead Prefix Society

4 Announcements & A Withdrawal

Much Ado About Flapping

Sense & Reachability

While You Were Announcing

I Know What You Announced Last Session

The BGP Song

Yesterday
All the withdrawals seemed so far away
I thought my prefixes were here to stay
Oh, I believe in Yesterday.

Suddenly
It's not half the table it used to be
There's a black hole hanging over me
Oh, I believe in Yesterday.

Why they had to flap, announce and draw away?
They sent something bad, now I long for yesterday.

Yesterday
Routing was such an easy game to play
Now my packets all hide away
Oh, I believe in Yesterday

Project History & Goals

Project History & Goals

- Akamai has an interest in knowing what data semantics correlate with 'bad user performance' for various protocols.
- Akamai's network folks have an interest in understanding the Internet better. Many are frustrated network engineers who had no data. Now we have too much.
- Historically, the project has been one person in spare time. It is still 1-2 people in spare time.

Data Sets & Infrastructure

Data Sets: Active UDP/HTTP

- "That problem we weren't having yesterday, is it better?"
- Developed for mapping the 'net, SLA verification from us to customers and from us to providers.
- Catches (even at coarse grain) surprising #s of CEF bugs and partial unreachabilities that providers don't know about.

Data Sets: Active UDP/HTTP

- Active UDP and HTTP (1k-ish object) transfer every 3 minutes between a matrix of 20x50 of the 'public' Akamai regions (public == available to send traffic to any prefix; region == a location in a network).
- Active UDP and HTTP (1k-ish object) transfer every 30 minutes between 30 'public' Akamai regions and 1200 'private' Akamai regions (private == restricted prefix candidate set for Akamai serving).
- Caveat: NOT raw TCP performance; involves Akamai web servers.

Data Sets: BGP

- Akamai has BGP sessions for data collection with 350 ASNs, over half (and growing) full feeds, and 250+ non-core providers (fastest-growing segment). Many of the non-core providers (roughly ½) do no non-transit peering at all.
- Currently using home-grown software for reflecting updates (NOTIFY is tricky) and logging. Zebra also, but there are many issues with it. Now using MRT format.
- BGP used to determine 'acceptable' prefixes for 'regions', and to look at performance and structure.

Data Sets: Billing Logs

- 5-15 billion http transactions/day. If not complete and correct, Akamai can't get revenue (SNMP on switch ports does nothing for us…).
- Interesting data includes interrupted transfers and throughput, and traffic density per IP/prefix/AS/time.

Data Sets: Traffic Density

- Billing logs give us 'access traffic' over time per prefix or /24.
- Only an accurate proxy for where the Akamai customer HTTP and streaming traffic is going, but informal surveys indicate that it is a good proxy for 'eyeball density'.
- External sources include cache logs and flow data, but not enough to give a complete picture of 'server/service density'.

Data Sets: TCP Statistics

- Throughput, retransmits, timers for 5-15 billion HTTP transactions/day, but lives 'at the edge'. Strategy so far has been edge filtering for 'interesting' pathologies. A small percentage are sampled and pulled back, and 100% of some patterns are pulled back.
- Budget doesn't allow room for building another infrastructure for complete collection, and edge filtering makes 'anomaly' detection more tricky than just 'is it good or bad' detection.
- Still, Akamai's richest data set next to traceroute data, and future plans call for mining it. Random mapping makes it even better.
- Does NOT include streaming unless via HTTP.

Current Project: Traffic Density

- Akamai billing log-generated traffic density recorded every few days, per global prefix.
- Will eventually be broken down by hour and /24. Planning larger storage.

Performance/ Churn Correlation

Background

- Question: How do BGP churn and performance correlate?
- Monitoring the BGP infrastructure and doing active and passive measurements; no BGP or performance fault injection was performed. “No prefixes were harmed in the making of this study”
- Just looking at # of withdrawals and updates per prefix.

Thresholds

- 'Bad performer'/'Congestion' thresholds:
  - Active measurement:
    - complete failure, or
    - any UDP loss, or
    - > 150ms/5000 miles UDP, or
    - 1 packet lost on TCP;> 1 second for 10kbytes
  - Passive measurement:
    - session failure at higher than normal rate, or
    - throughput < 1/10th normal rate, or
    - retransmitted segments > 10%

Thresholds

- BGP Thresholds:
  - Enumeration of > 50% of the routes from a direct peer table invalidates a BGP session for the duration +/- 10 minutes – no remote-session-reset watch
  - Prefix must have more than 2 updates and/or withdrawals per rolling 5-minute window, and must have > 2 updates in at least 10 different-AS feeds

Limitations

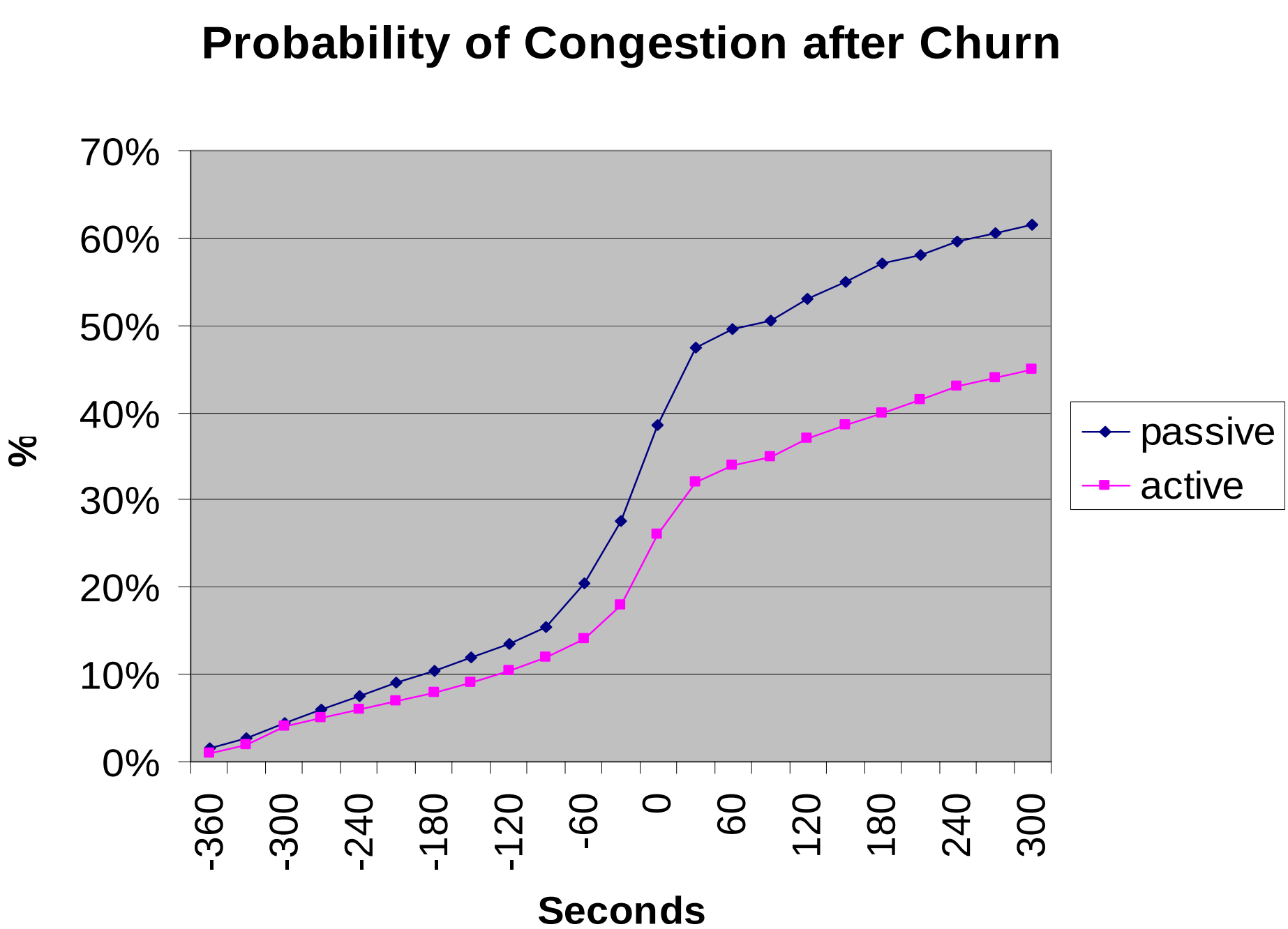
- Major concerns:
  - No churn classification (announce vs. withdraw vs. excess announce or withdraw vs. non-affecting attribute change)
  - Definition of performance
  - No sophisticated session reset elimination
- Not looked at:
  - By edge vs. core
  - By prefix length
  - Vary performance sensitivity
  - By geography
  - Varying # of announcements for “churn” def.

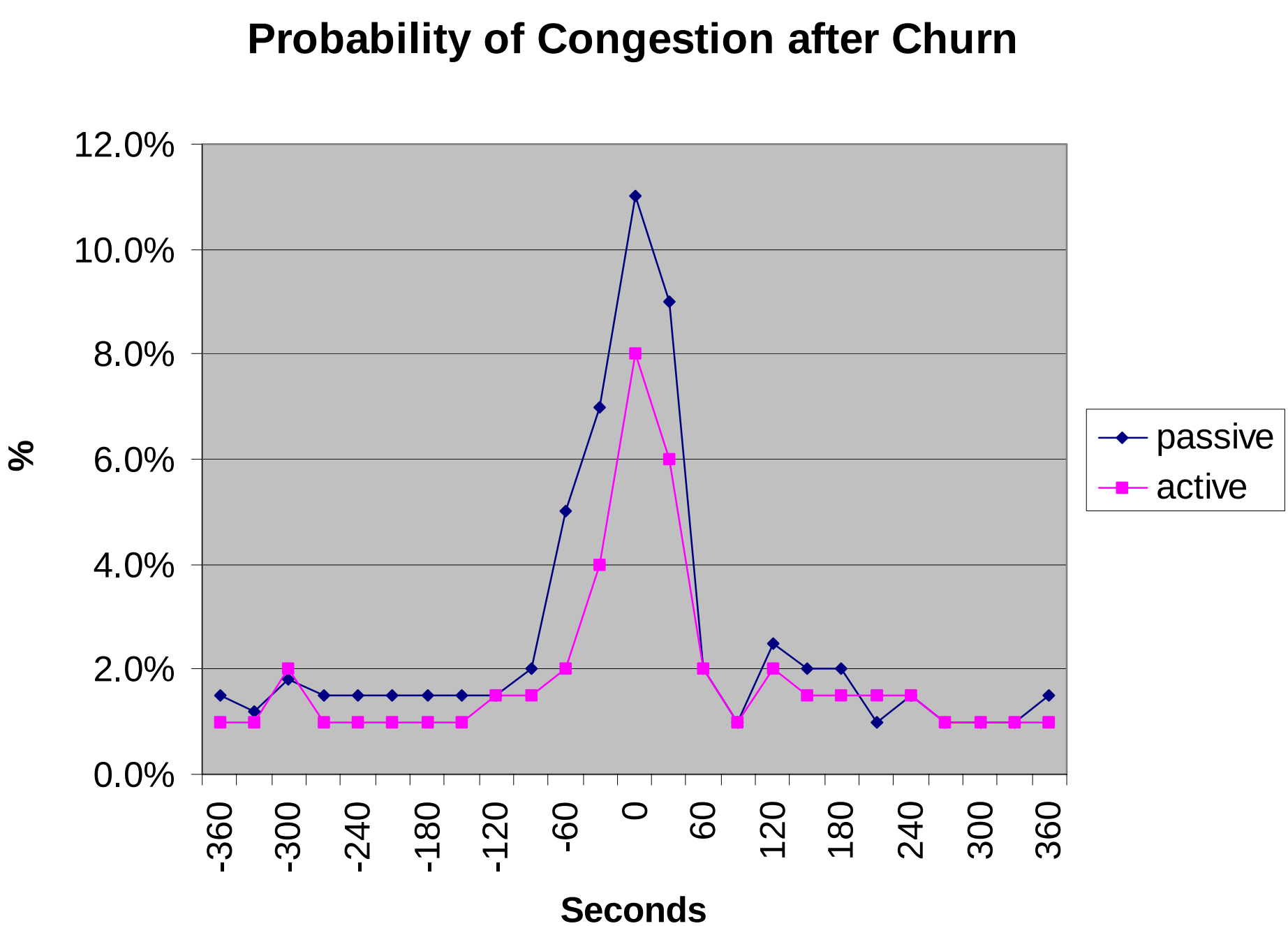
Limitations (2)

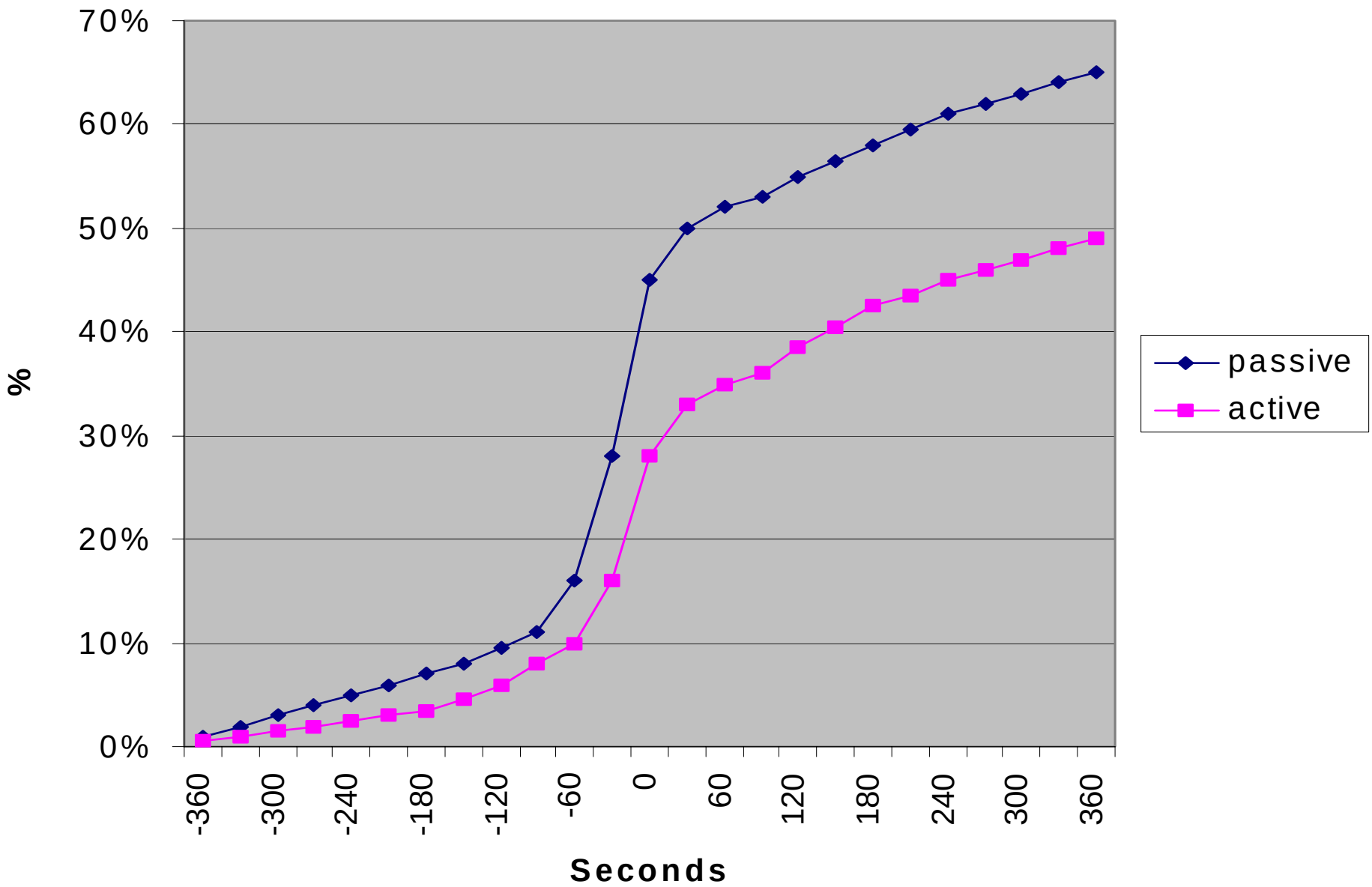
- No beacons
- Could be under-counting: BGP was live; windows were live; code could be missing cases.
- Not looked at: % churn for good performers

Data Sets

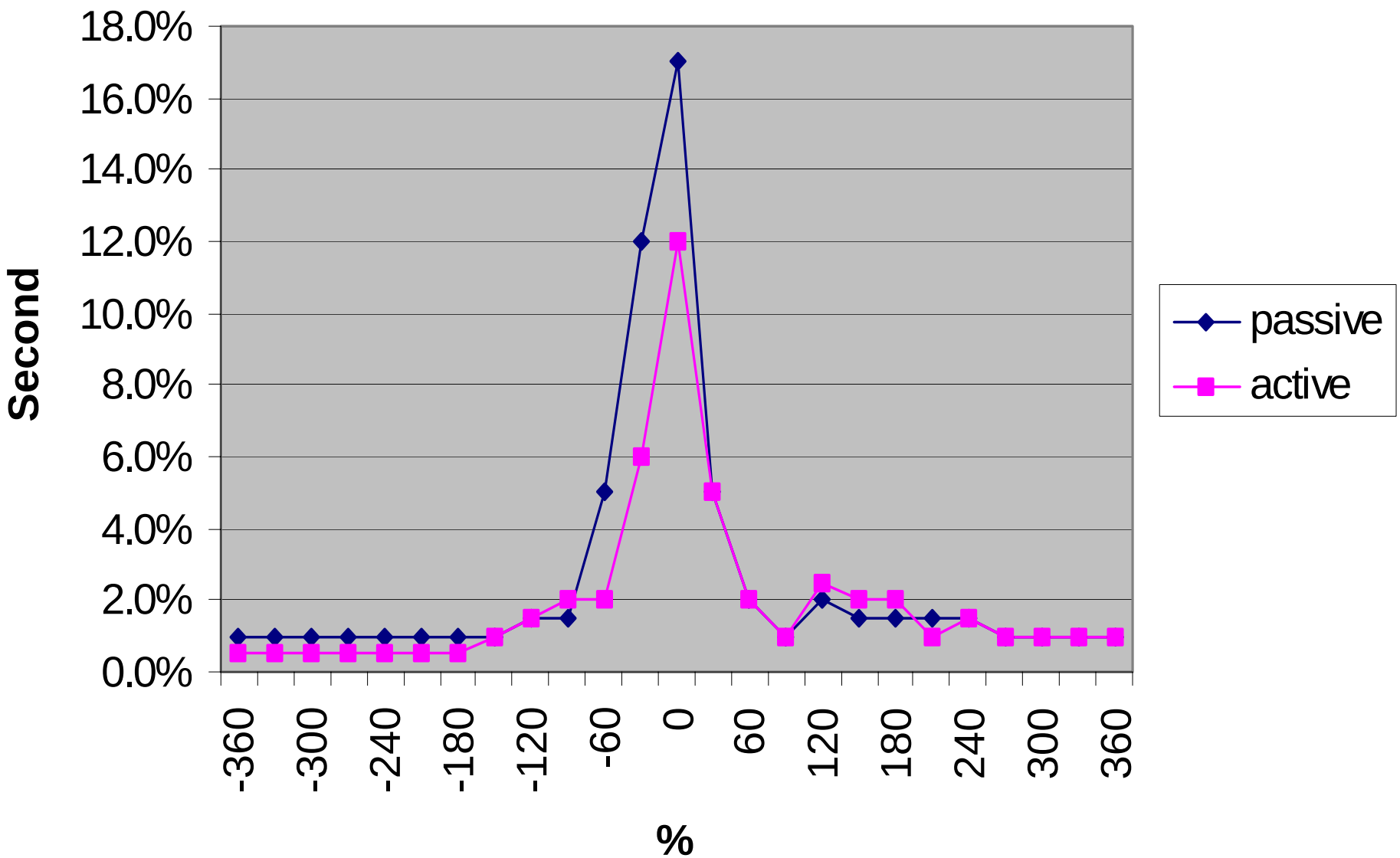
- 30 days of BGP feeds (live) - August:
  - 45 Core, 45 Edge, with 5 duplicate Core ASs and 6 duplicate Edge ASs
- 30 days of UDP/HTTP active measurements for both the 20:50 core tests – 23,903 congestion matches
- 30 days of TCP statistics logs for 2 machines in each of 5 regions (10 machines) – 94,820 congestion matches


Probability of Congestion after Churn
70%
60%
50%
40%
30%
20%
10%
0%
%
-360
-300
-240
-180
-120
-60
0
60
120
180
240
300
Seconds
passive
active


Probability of Congestion after Churn
12.0%
10.0%
8.0%
6.0%
4.0%
2.0%
0.0%
%
-360
-300
-240
-180
-120
-60
0
60
120
180
240
300
360
Seconds
passive
active

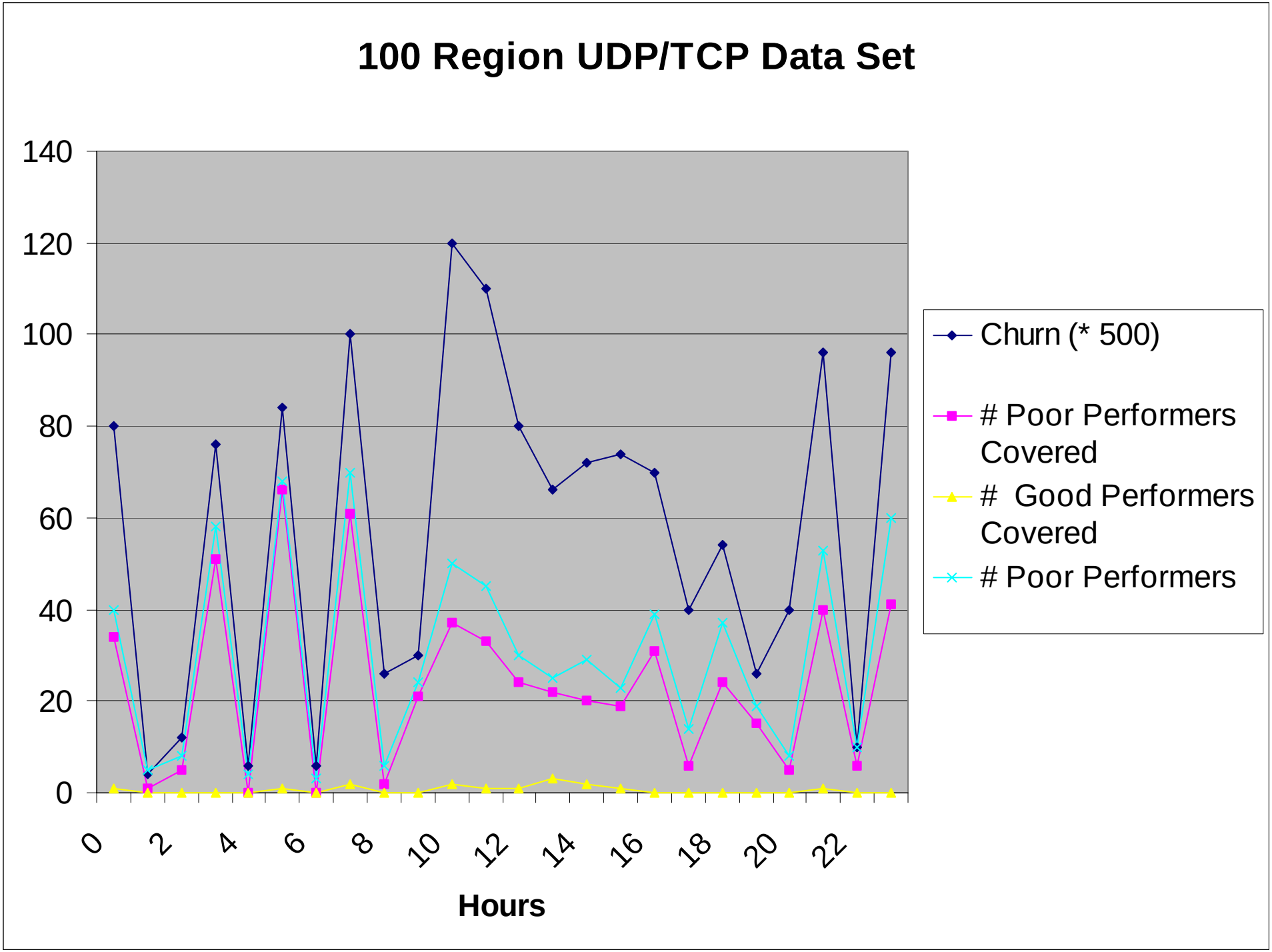

Probability of Churn after Congestion
70%
60%
50%
40%
30%
20%
10%
0%
%
-360
-300
-240
-180
-120
-60
0
60
120
180
240
300
360
Seconds
passive
active

Probability of Churn after Congestion

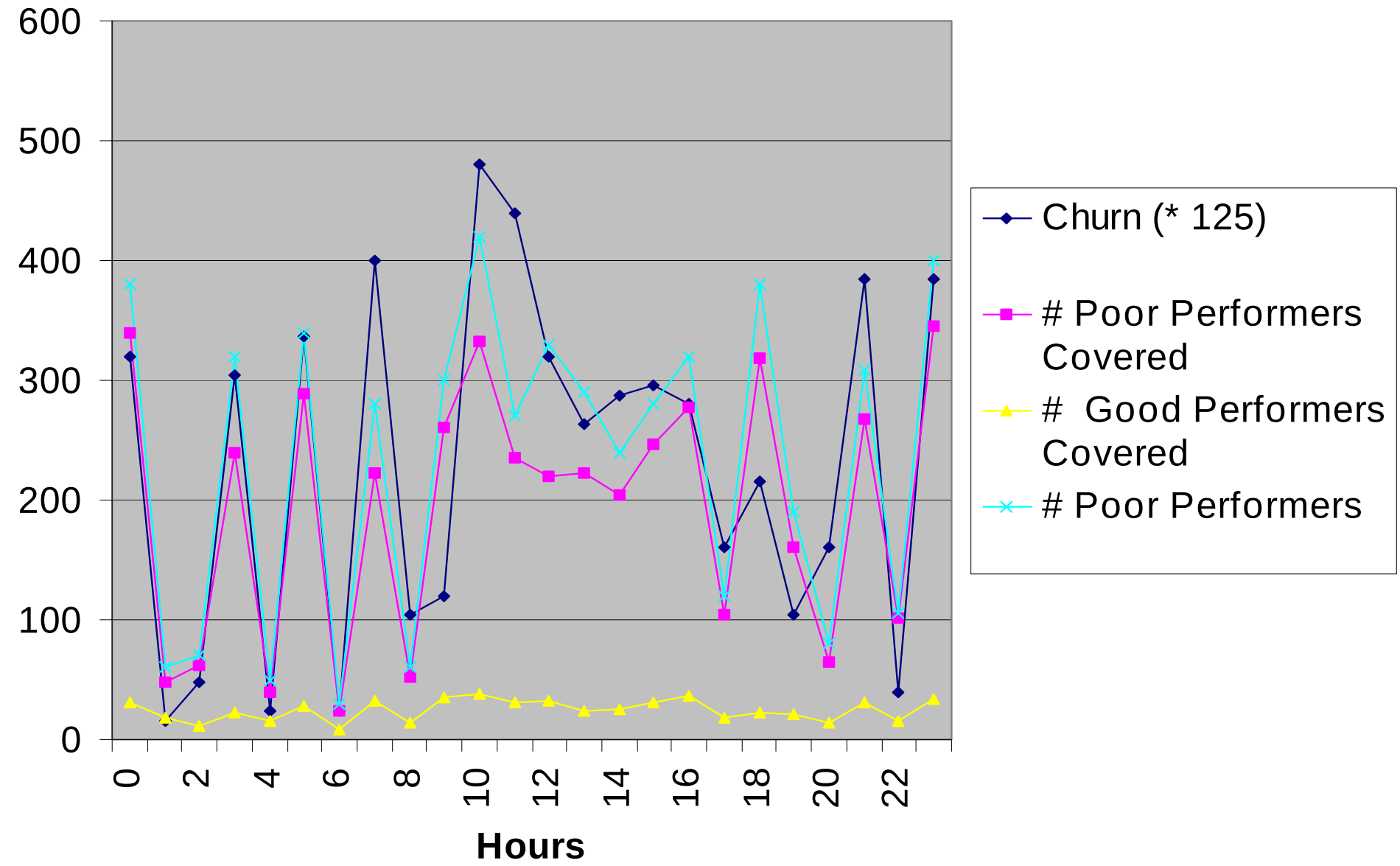


Data from ISMA 01

- Since ISMA - stopped using active measurement set of 1200 regions measured infrequently (very noisy).
- ISMA data was one day, active only


100 Region UDP/TCP Data Set
140
120
100
80
60
40
20
0
0
2
4
6
8
10
12
14
16
18
20
22
Hours
Churn (* 500)
Poor Performers Covered
Good Performers Covered
Poor Performers

1200 Region UDP/TDP Data Set



Analysis

- It appears that there is a real correlation (no surprise, Labovitz/Ahuja found packet loss:fault correlation, but always good to check common wisdom) between BGP churn and performance as seen across the Internet.
- We encourage others to look at this, as we may not have time this year to do more analysis.

Future Work

- Classifying by type of churn (patterns such as insertion/deletion of prefix, or withdrawals vs. updates, etc.)
- Classifying by complete failures vs. high latency completed transactions; using less or more sensitive parameters for ‘poor’ performers.

Misc: Future Work

Misc: Topics of Interest (1)

- AS taxonomy (peering, transit, partial peering)
- BGP advertisement/withdrawal activity classification (patterns of withdrawal/updates, looping)
- Active vs. passive measurements and 'performance'
- BGP churn of differing taxonomy vs. 'performance'
- Path vs. routing
- # of as links found or possible to estimate
- 'Shape of table and churn' over time; Inter-AS topologies; vs. traffic load.

Misc: Q's/Topics (NOT original)

- How does intra-AS BGP churn differ from inter-AS BGP churn? Differing correlations with performance?
- Is BGP growing at 'the edge'?
- Other chronic looping in BGP?
  - Designing a system to find other patterns (healthy/unhealthy) in BGP?
- Does user performance differ by protocol for same-time-window and same endpoint communications?
- Possible to build a Cisco CEF ('Customer Enragement Feature') confusion detector from active and/or passive measurements?

Credits due:

Akamai netarch team

CAIDA/ISMA presentations

Leiden '00 discussions/presentations

Various NANOG presentations

Particularly, Ahuja, Labovitz, Griffin, Gao, kc, Broido

Thanks.
Questions?

avi@akamai.com
avi@freedman.net