A Measurement Study of BGP Misconfiguration

Ratul Mahajan, David Wetherall,
and Tom Anderson

University of Washington

Motivation

- Routing protocols are robust against failures
  - Meaning "fail-stop" link and node failures
- But what about when nodes just don't behave?
  - Misconfigurations, implementation bugs, malicious attacks
- We need to understand this to make availability guarantees
  - Many colorful anecdotes, few systematic studies
- BGP is rich ground for a study of misconfigurations
  - Thousands of ISPs, many implementations, complex to configure

This talk

- Peek at an in-progress BGP measurement study based on the RouteViews server
  - Public 2 hourly routing table snapshots from ~50 different ISPs
- Our goals:
  - Identify the common types of misconfigurations
  - Determine how frequently they occur
  - Assess their impact on the Internet as a whole
- Current focus is the analysis of origin changes (hijacks) and partial connectivity

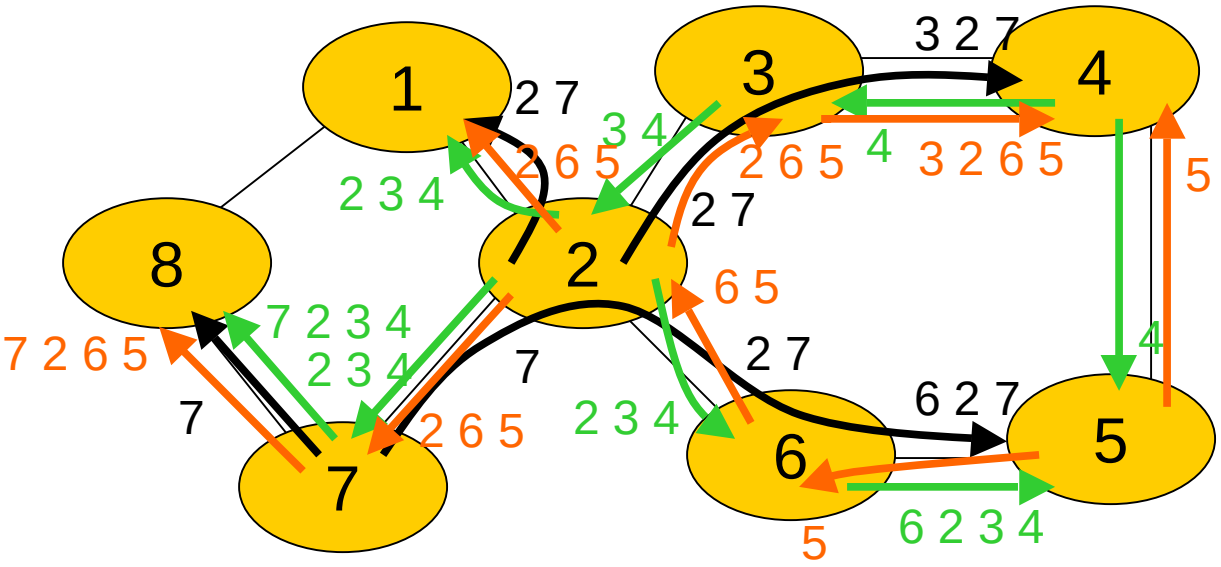
Methodology

- Define a model of acceptable BGP usage
  - Deviations from the model are “misconfigurations”
- Measure the occurrence of misconfigurations
  - Use heuristics to attribute to the likely causes
- Measure the impact of misconfigurations
  - On other, well-defined, quantities of interest
- Validate against actual ISP experiences
  - Via an email survey

BGP in a nutshell

- BGP is the routing protocol used in the Internet core, which is a graph of Autonomous Systems (ASes) or ISPs
- Each AS announces paths to other ASes that it can use to reach given prefixes (block of IP addresses)
- Announcements are aggregated where possible, e.g, one for many customers, rather than one per customer
- Imagine paths growing from origins subject to policies (transit versus peering); packets follow reverse direction

BGP in a nutshell (2)



- 2 provides transit for 7; 7 reaches and is reached via 2
- 4 and 5 peer; they exchange their customer traffic

Why we need a usage model

- BGP is defined by local operational practices, not global standards

- A contrived example: botched pre-pending
- Pre-pending by an AS is a hack used to make paths less attractive to others. Not considered to be a loop.
  - e.g., AS1 **AS77** AS4 → AS1 **AS77 AS77 AS77** AS4
- What if AS77 announces AS1 **AS77 AS66 AS77** AS4?
- Is this a mistake, or a hack for enforcing policy?

A model of BGP usage

- Private identifiers are not be leaked in public
- The origin AS owns the address space it announces
- The advertised AS path matches the forwarding path
- Announcements are aggregated where possible
- AS paths obey policy constraints
- Providers are connected to the entire Internet

- Deviations are defined to be “misconfigurations”

Impacts of misconfiguration

- Alteration of selected paths
  - Not what you preferred
- Increased routing load
  - More routing announcements to process
- Loss of connectivity
  - No paths at some/all locations that reach a prefix

- The last is most serious and visible to users
- The two deviations we focus on can affect connectivity

Measuring routes with incorrect origins

- Are there easy ways to detect misconfigured origins?
  - Multiple origins for a prefix; increasingly common practice
  - Internet Routing Registries (IRRs); found to be inaccurate

- We observe that origins tend to change on human timescales, except for failures and misconfigurations
  - We analyze changes in the RouteViews BGP snapshots
  - We divide them by duration (short vs. long-lived)
  - Then we attribute probable causes to changes
  - Finally we assess their impact on reachability

IRRs: do they detect incorrect origins?

	Total Prefixes	Registered Origins	Consistent Origin(s)	Inconsistent Origin (s)
Single Origin AS	115228	101952	70458 (69%)	31494 (31%)
Multiple Origin AS's	1720	1523	293 (19%)	1230 (81%)

Causes of origin changes

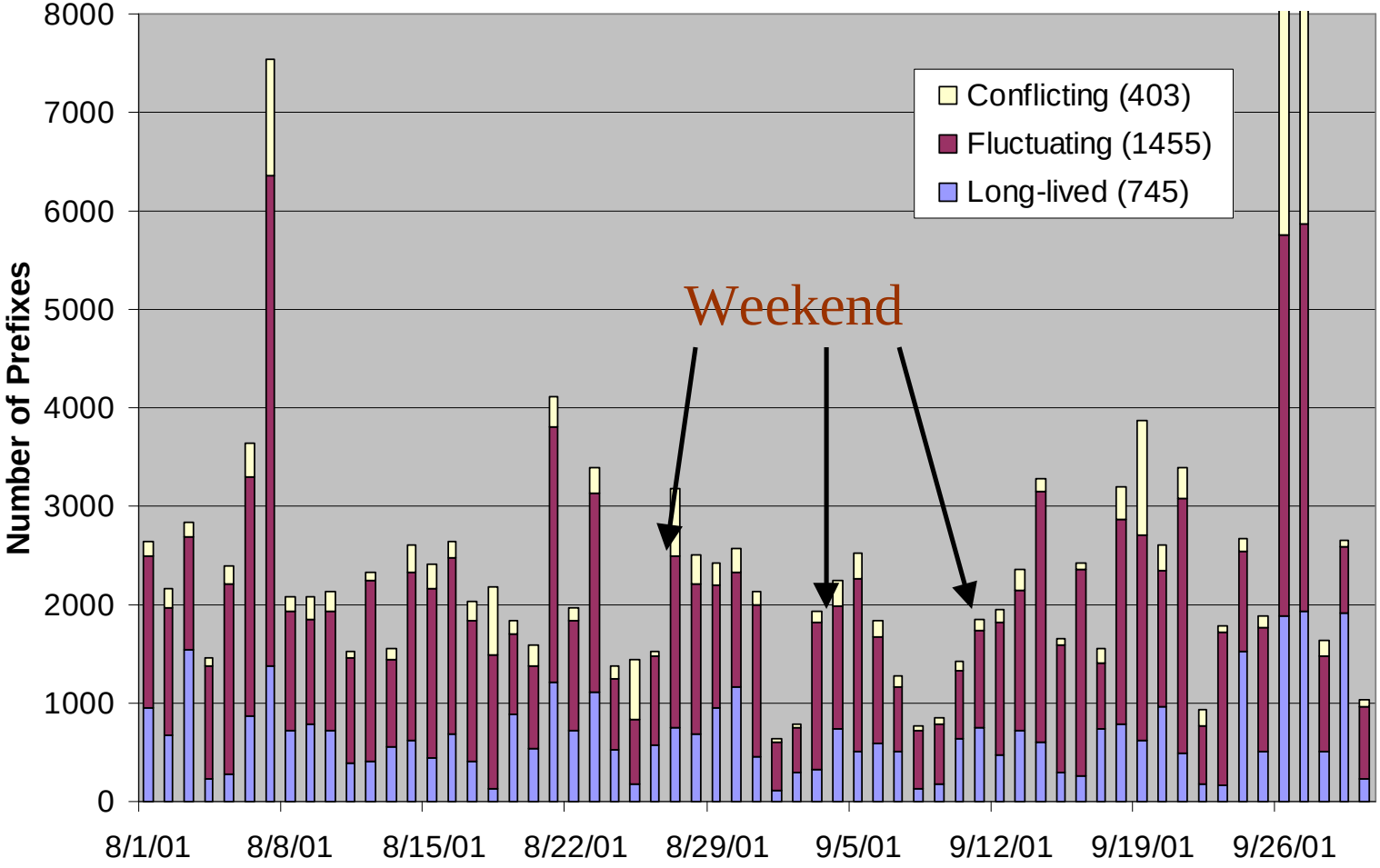
Long-lived	Fluctuating	Conflicting
More Specific Added	Self Deaggregation	AS-Path Stripping
More Specific Deleted	Failures (unreachable)	Strip Deaggregation
Origin Added	Backups	Extra Last Hop
Origin Deleted		Foreign Deaggregation
Origin Changed		Other
New Address Space		
Address Space Deleted		

- Long-lived changes last more than one day

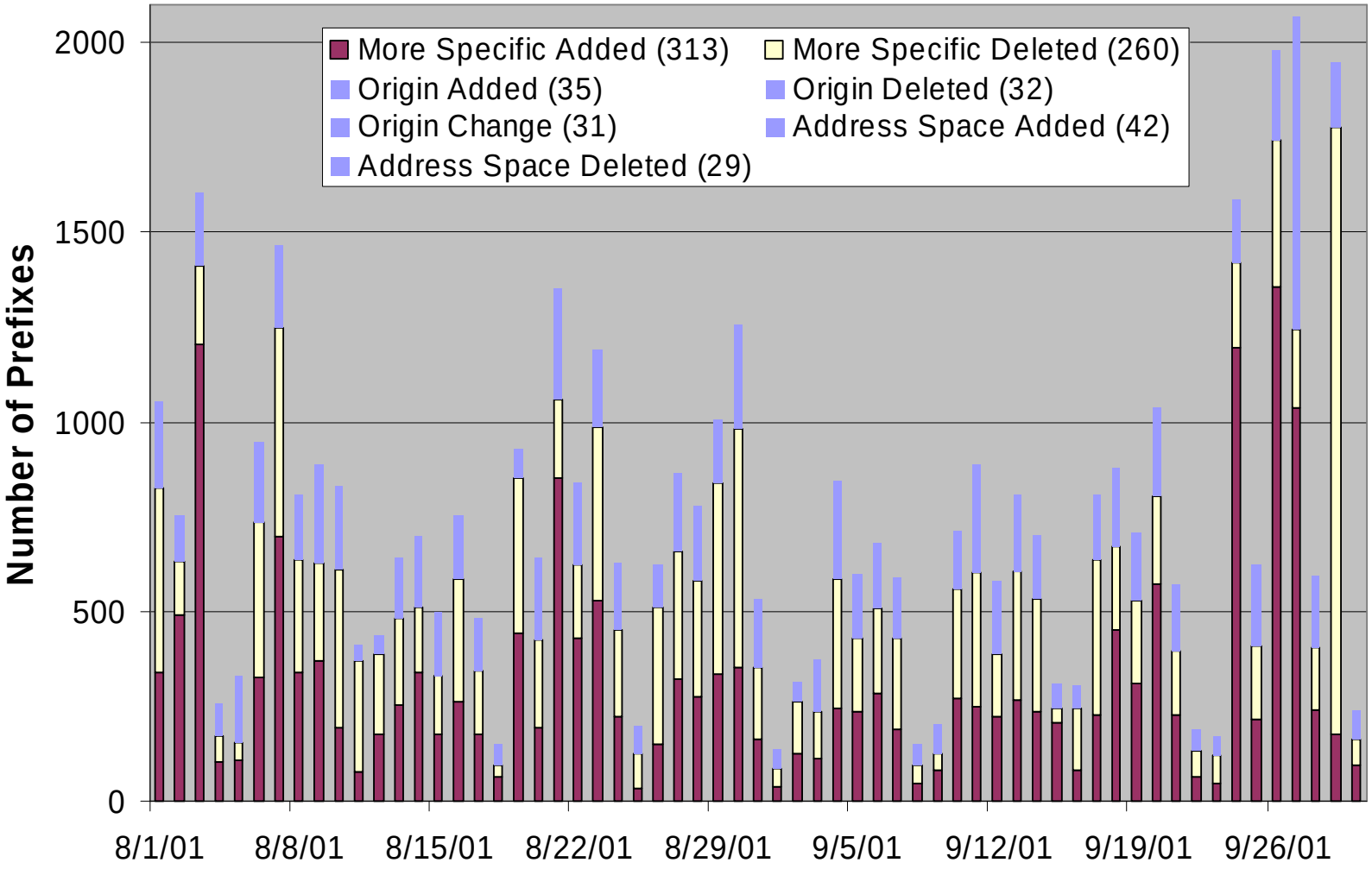
Definitions of short-lived changes

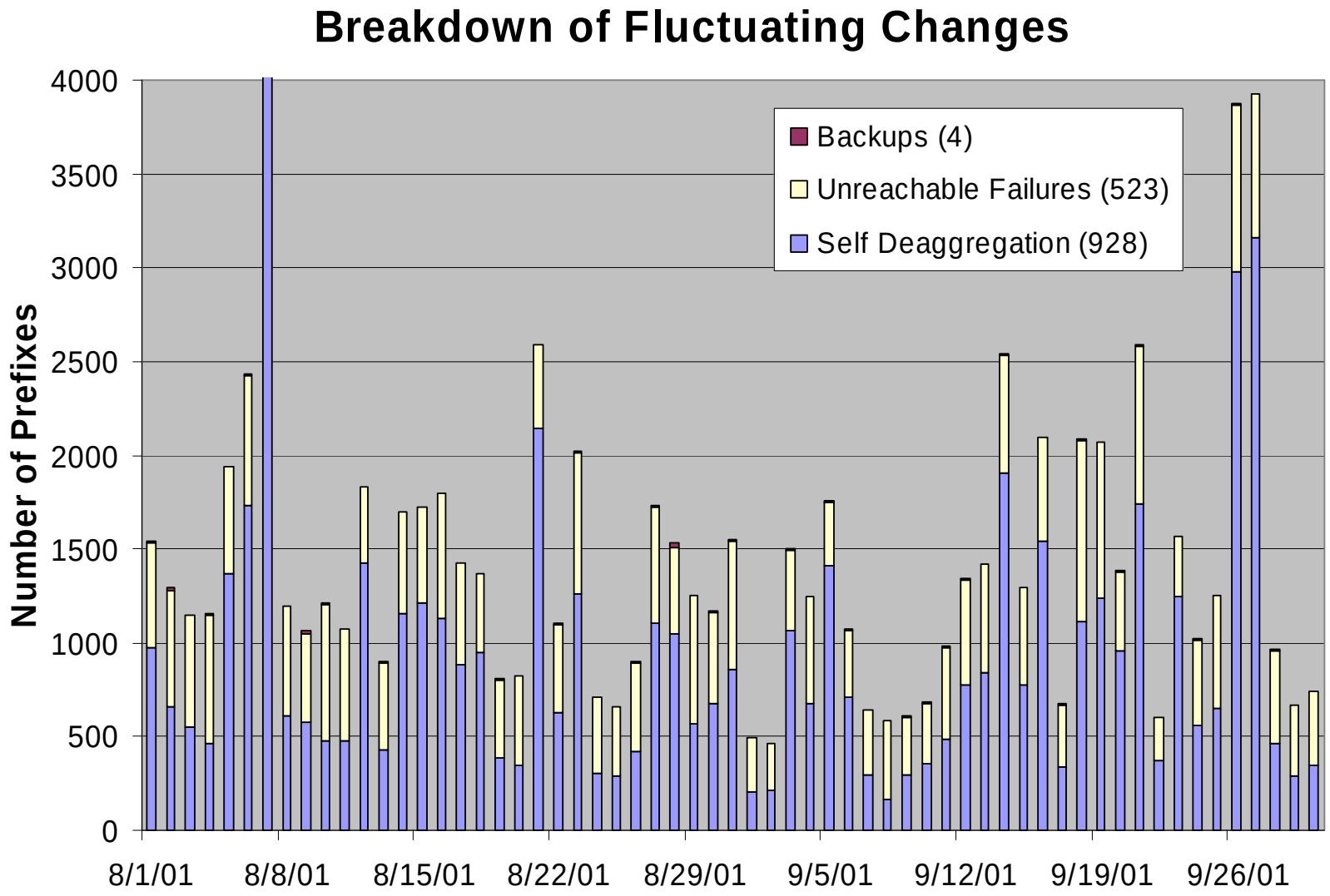
	Stable Announcements		Short-lived Announcements	
Self Deaggregation	a.b.0.0/16	X-Y-Z	a.b.c1.0/24 a.b.c2.0/24	X‘-Y‘-Z X‘-Y‘-Z
AS-Path Stripping	a.b.c.d/s	X-Y-Z	a.b.c.d/s	X‘-Y
Strip Deaggregation	a.b.0.0/16	X-Y-Z	a.b.c1.0/24 a.b.c2.0/24	X‘-Y X‘-Y
Extra Last Hop	a.b.0.0/16	X-Y-Z	a.b.c1.0/24 a.b.c2.0/24	X‘-Y‘-Z-O X‘-Y‘-Z-O
Foreign Deaggregation	a.b.0.0/16	X-Y-Z	a.b.c1.0/24 a.b.c2.0/24	X‘-Y‘-O X‘-Y‘-O

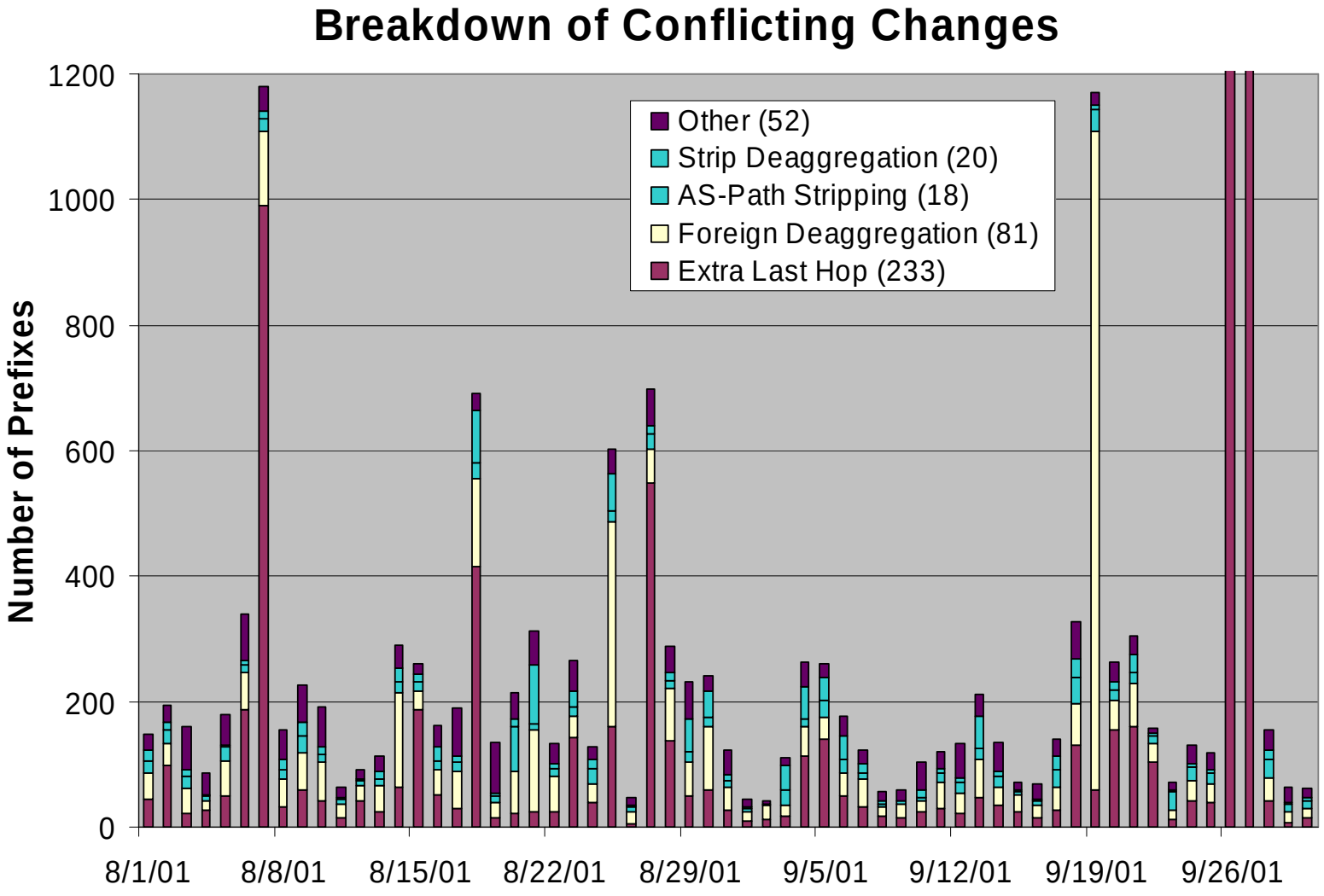
Distribution of Origin Changes



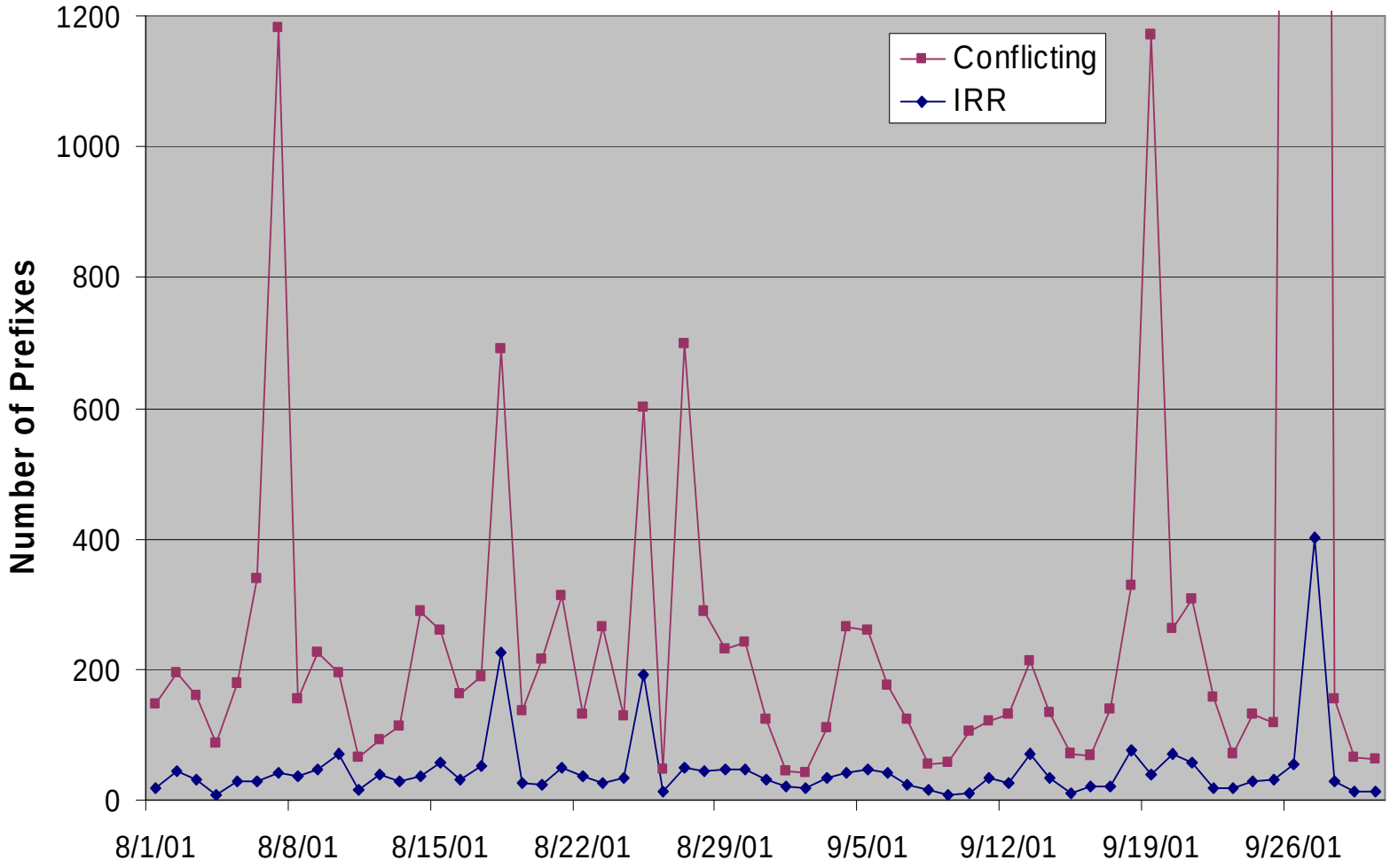
Breakdown of Long-Lived Changes




Breakdown of Fluctuating Changes
Number of Prefixes
4000
3500
3000
2500
2000
1500
1000
500
0
8/1/01
8/8/01
8/15/01
8/22/01
8/29/01
9/5/01
9/12/01
9/19/01
9/26/01
Backups (4)
Unreachable Failures (523)
Self Deaggregation (928)


Breakdown of Conflicting Changes
Number of Prefixes
1200
1000
800
600
400
200
0
8/1/01
8/8/01
8/15/01
8/22/01
8/29/01
9/5/01
9/12/01
9/19/01
9/26/01
Other (52)
Strip Deaggregation (20)
AS-Path Stripping (18)
Foreign Deaggregation (81)
Extra Last Hop (233)

IRR suggests Conficting cases contain misconfigs



Validation via an email survey

- Interesting exercise in its own right …
- 30% of emails bounce outright
- More find their way to /dev/null
  - "Your support request has been accepted by our team, a case has been opened with reference 12345 …"
- Surprise and lack of a clue
  - "Thanks for alerting us … I am a bit surprised …"
  - "Ratul, … can you help us?", "No idea really …"
  - "I believe research has shown routes appear and disappear every day"
- Defensiveness
  - "Yes, we leaked … but took pre-emptive action right away …"
  - "The information you are requesting is covered by NDA …'
- Hard information and encouragement
  - "You caught us. This is what happened …"
  - "I enjoyed your NANOG talk …"

Validation results

Cause	Total	Replies	Misconfig	Connect?	False +ve
extra-last-hop	111	38	31 (82%)	7 (18%)	7 (18%)
as-path-strip	760	730	723 (99%)	2 (0%)	7 (1%)
self-deagg	1222	243	180 (73%)	42 (17%)	63 (26%)
other	91	36	24 (67%)	12 (33%)	12 (33%)
strip-deagg	150	85	82 (96%)	5 (6%)	3 (4%)
foreign-deagg	188	45	41 (91%)	18 (40%)	4 (10%)
all	2522	1177	1081 (92%)	86 (7%)	96 (8%)

- Caveat: these stats are for prefixes, not incidents.

Causes of origin changes

Real misconfigurations:

- Buggy ACLs/route-maps
- Relying on upstream
- Forgot auto-summary
- Redistribution
- Over-aggregating
- Hijacking
- Old routers …

False positives:

- Just testing
- Failures
- Temp. load balancing
- Migration
- Re-numbering

Speculation

- Complexity of configuration is a root cause of error
  - Scope for greater "type-checking"
- Operational practices are diverse
  - Makes systematic identification of errors difficult
- Authoritative databases will be inaccurate
  - Use for automatic blocks is problematic
- ISPs depend on one another to a significant degree
  - "I thought you'd handle that"
- Connectivity can persist despite many misconfigs
  - Route leaks, redistribution, de-aggregation, …

Also: Measuring partial connectivity

- Advertised address space is not reachable from all places in the Internet!

- Causes:
  - Convergence delays
  - route flap damping
  - policy (filtering on prefix length, or commercial relationships)

- Failures do not lead to partial connectivity
- We can distinguish the above causes by timescale

Partial connectivity analysis

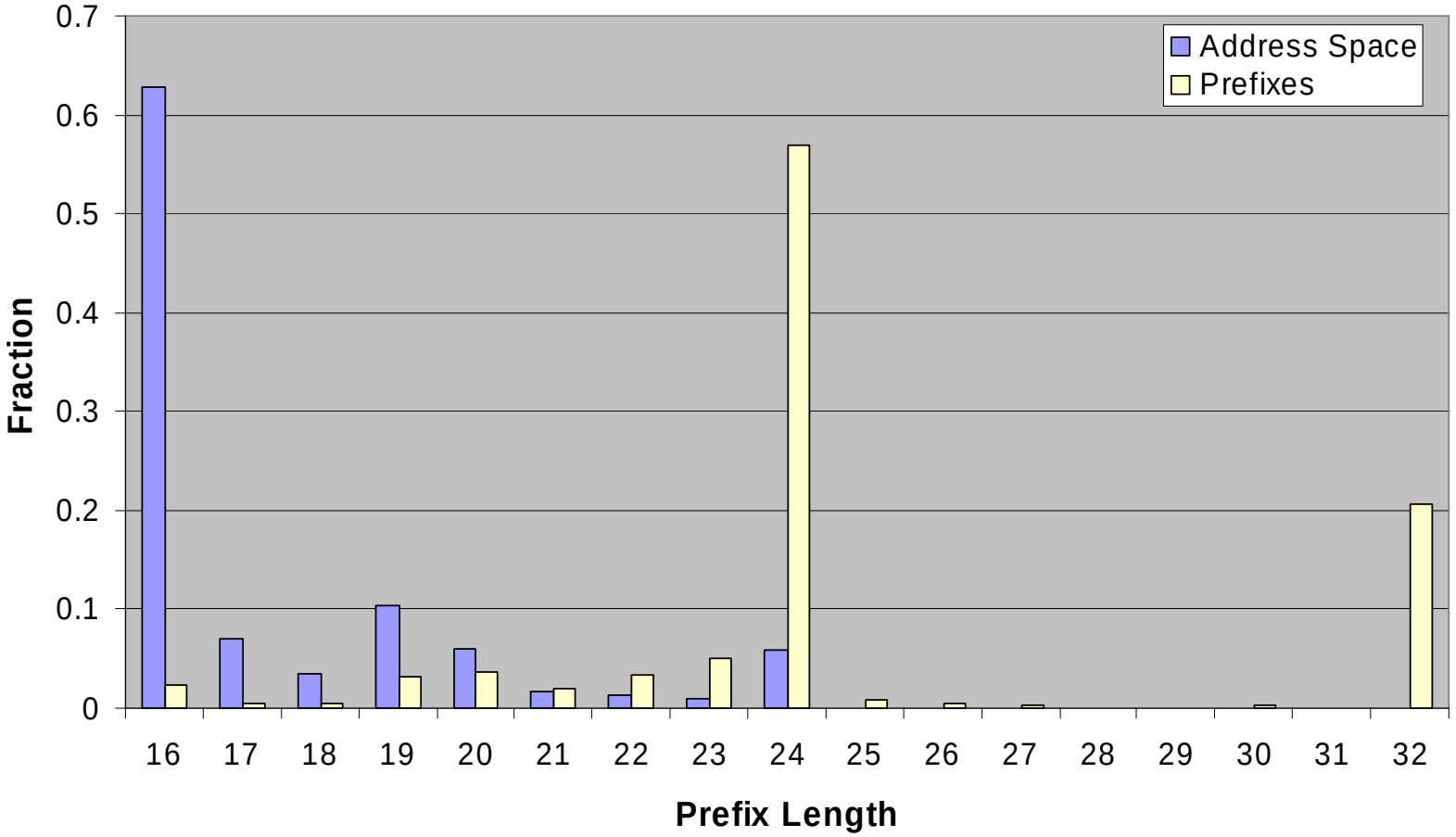
- Identify partially connected address space (!= prefix) from the BGP table
- Consult BGP snapshots 15 minutes before and after to identify partial connectivity due to convergence delays
- Correlate against partial connectivity across days to differentiate between route flap damping and filtering based partial connectivity
- Verify using public looking glasses to guard against restrictive export policies and default pointing

Partial connectivity: results

- Express as percentage of advertised address space.

- Convergence: 0.005-0.02%
- Route flap damping: 0.1-0.8%
- Filtering: 0.7%

Prefix Length Distribution of Partially Connected Address Space



Tentative conclusions

- There is considerable churn in prefix origins
  - More than 2% of the prefixes are affected every day
  - 1/3 to 1/2 of this churn is due to misconfigurations
- The causes of misconfigurations are diverse
- Connectivity is surprisingly robust
  - ~ 3 in 4 incidents do not cause reachability to be lost
- The address space is not fully connected
  - ~1% persistently partially connected at any time

- Many thanks to the ISP community for its support
- Feedback: http://www.cs.washington.edu/homes/ratul/bgp/