Mapping the Great Void

Smarter scanning for IPv6

Richard Barnes, Rick Altmann, Daniel Kerr
BBN Technologies

Agenda

- Challenges for mapping the IPv6 Internet
- Some approaches to smarter scanning
  - CIDR++
  - Registry information
  - Addressing heuristics
- Empirical results

Background: IPv6 is big

IPv6 address space is big

- How do you select the networks you trace to?
  - Ark IPv4: Each /24 covered by a BGP prefix
  - Ark IPv6: One per prefix advertised in BGP
- Supposing we view a /48 as functionally similar to a /24...
  - IPv4: 12,577,420 /24s advertised ($\sim 2^{23.6}$)
  - IPv6: 3,523,931,041 /48s advertised ($\sim 2^{31.7}$)
    - ... and that's with the current level of IPv6 deployment
- And really, /48s get subdivided too

http://www.caida.org/workshops/isma/1102/slides/aims1102_yhyun_ark.pdf
RouteViews RIB from WIDE collector, 2011/12/22

General Approach: Adaptive Probing

- Learn from previous rounds of probes to predict where you should probe next
- In the IPv4 context, focus has been on reducing impact of comprehensive measurement traffic
  - DoubleTree / Interface Set Cover algorithms find minimal set of paths to cover all interfaces
- In IPv6, focus is more on discovering the most subnets / interfaces in a feasible number of measurements
  - Some algorithms don't scale to IPv6 (e.g., subnet-centric)

http://rbeverly.net/research/papers/direct-imc10.pdf
RouteViews RIB from WIDE collector, 2011/12/22

Smarter Scanning

Going beyond BGP

- To tell two networks apart in measurements, we need to trace to a target in each of them
- Finding networks via pure random scanning within BGP-announced prefixes doesn't scale
- Start with BGP, add more information
  - Small amounts of randomness
  - Registration information (WHOIS)
  - Information gathered in earlier scans

Testing Methodology

- 5 nodes from commercial VPS services
- ICMP Paris traceroutes to selected targets
- Metric: Discovered addresses (no alias resolution)

Baseline: BGP

Technique	Traceroute Targets / Monitor	Monitors	Total Measurements	Discovered Interface Addresses	Gain Rate (New Hops Per Trace)
BGP	8380	5	41900	16986	0.405

BGP+4

- Some networks do a little bit of subdivision of an advertised prefix, but maybe not much
- Take each prefix from BGP
- Compute 16 subnets you can get by adding 4 random bits
  - Random scanning, but bounded increase in work (16x)

BGP+4

Technique	Traceroute Targets / Monitor	Monitors	Total Measureme nts	Discovered Interface Addresses	Gain Rate (New Hops Per Trace)
BGP	8380	5	41900	16986	0.405
BGP+4	73407	5	367035	20434	0.056

BGP ⋂ WHOIS + Rand48

- People sometimes register WHOIS information at a higher level of granularity than they advertise in BGP
- Download bulk WHOIS information and build a list of prefixes from inet6num objects
- Find routable WHOIS prefixes, covered by prefixes advertised in BGP
- If a given BGP prefix has no more specifics in WHOIS, sample five random /48s

BGP ∩ WHOIS + Rand48

Prefix		Network	BGP	Gain
2a02:f8:7:1a::/64	IT	AISA-NET-1	/32	32
2a01:4f8:141:22::/64	DE	FORMER-03-GMBH	/32	32
2406:4800::/64	SG	DOCOMOinterTouch-HQ-V6	/40	24
2405:2000:ff10::/56	IN	CHN-CXR-TATAC	/32	24
2607:f6f0:100::/56	US	EQUINIX-EDMA-V6-CORP-01	/40	16
2001:42c8:ffd0:100::/56	ZA	CAPETOWN-KLT-TATA	/32	24

BGP ⋂ WHOIS + Rand48

Technique	Traceroute Targets / Monitor	Monitors	Total Measurements	Discovered Interface Addresses	Gain Rate (New Hops Per Trace)
BGP	8380	5	41900	16986	0.405
BGP+4	73407	5	367035	20434	0.056
BGP ⋂ WHOIS + Rand48	90817	4	363268	40074	0.110

Sequence Completion

- As we do traceroutes, we get addresses back in the source addresses of responses
- Sometimes these addresses hint at the use of addressing schemes
- Look for runs within each hex digit, then complete sequences

```
2001:db8:1:47c8::797f
2001:db8:1:47c9::47db
2001:db8:1:47cb::8a03
2001:db8:1:47cd::4d33
2001:db8:1:47cf::b221
```

→

```
2001:db8:1:47c7::/48
2001:db8:1:47c8::/48
2001:db8:1:47c9::/48
2001:db8:1:47ca::/48
2001:db8:1:47cb::/48
2001:db8:1:47cc::/48
2001:db8:1:47cd::/48
2001:db8:1:47ce::/48
2001:db8:1:47cf::/48
2001:db8:1:47d0::/48
```

Sequence Completion

BGP
2a01:198::/32

BGP ∩ WHOIS
SIXXS-DEDUS01
2a01:198:200::/40

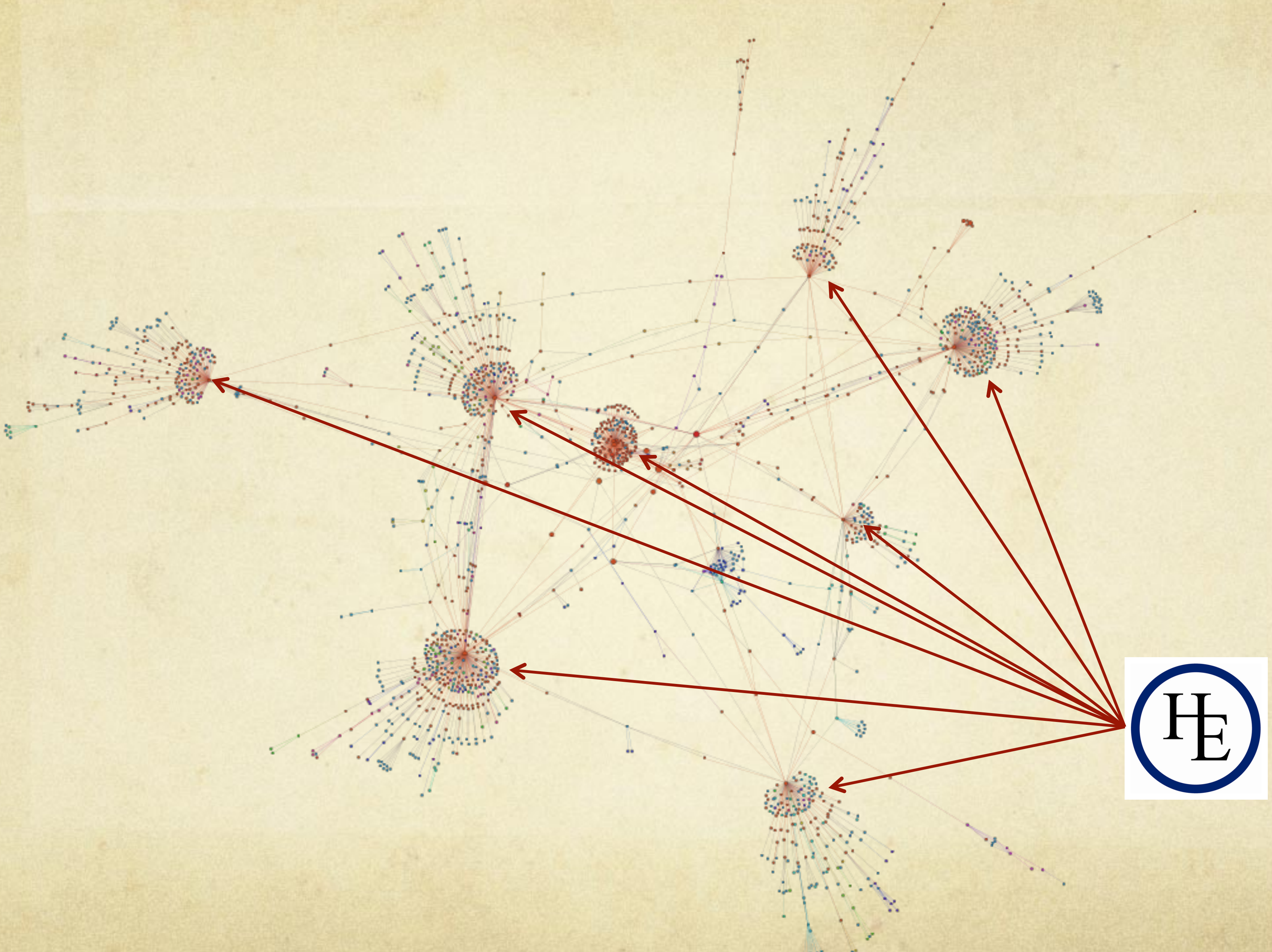
```
2a01:198:200:000::/52
2a01:198:200:100::/52
2a01:198:200:200::/52
2a01:198:200:300::/52
2a01:198:200:400::/52
2a01:198:200:500::/52
2a01:198:200:600::/52
2a01:198:200:700::/52
2a01:198:200:800::/52
2a01:198:200:900::/52
2a01:198:200:a00::/52
```

Scanning within the /40...
Completing the sequence...

Sequence Completion

Technique	Traceroute Targets / Monitor	Monitors	Total Measurements	Discovered Interface Addresses	Gain Rate (New Hops Per Trace)
BGP	8380	5	41900	16986	0.405
BGP+4	73407	5	367035	20434	0.056
BGP ⋂ WHOIS + Rand48	90817	4	363268	40074	0.110
Sequence Completion	21279.75	4	85119	22919	0.269

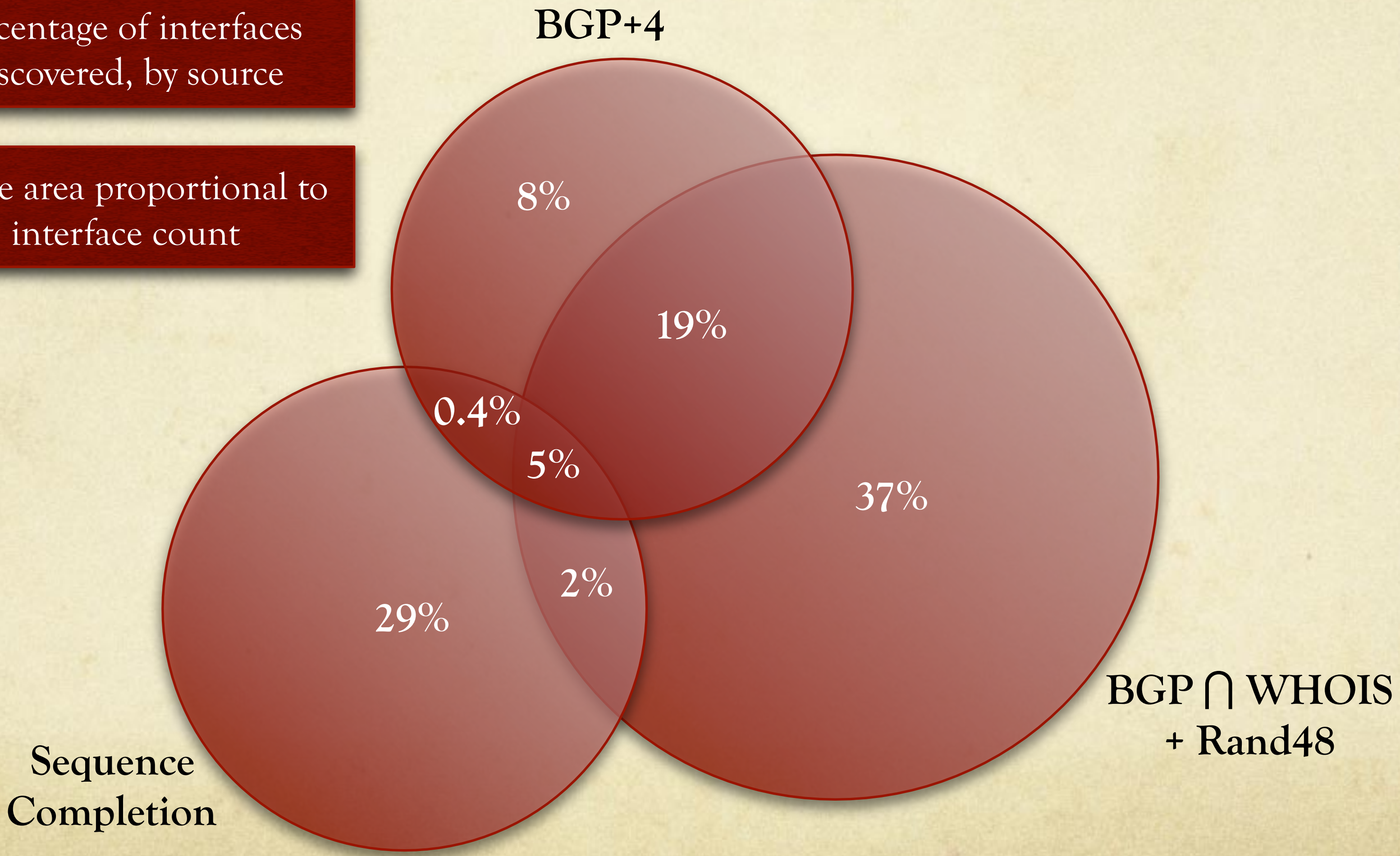
How much did we learn?



Overlap in Discovered Interfaces

Percentage of interfaces discovered, by source

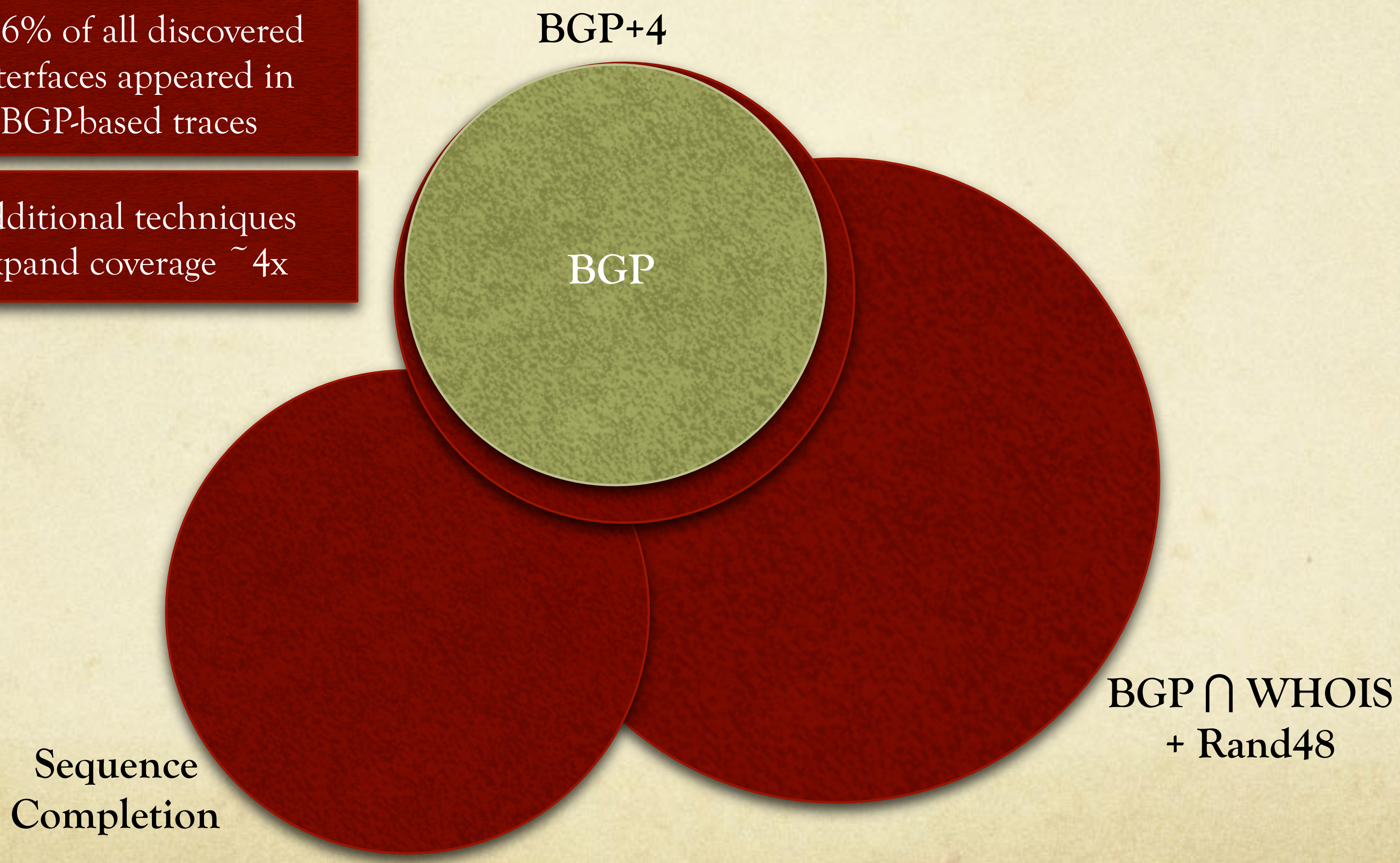
Circle area proportional to interface count



Overlap in Discovered Interfaces

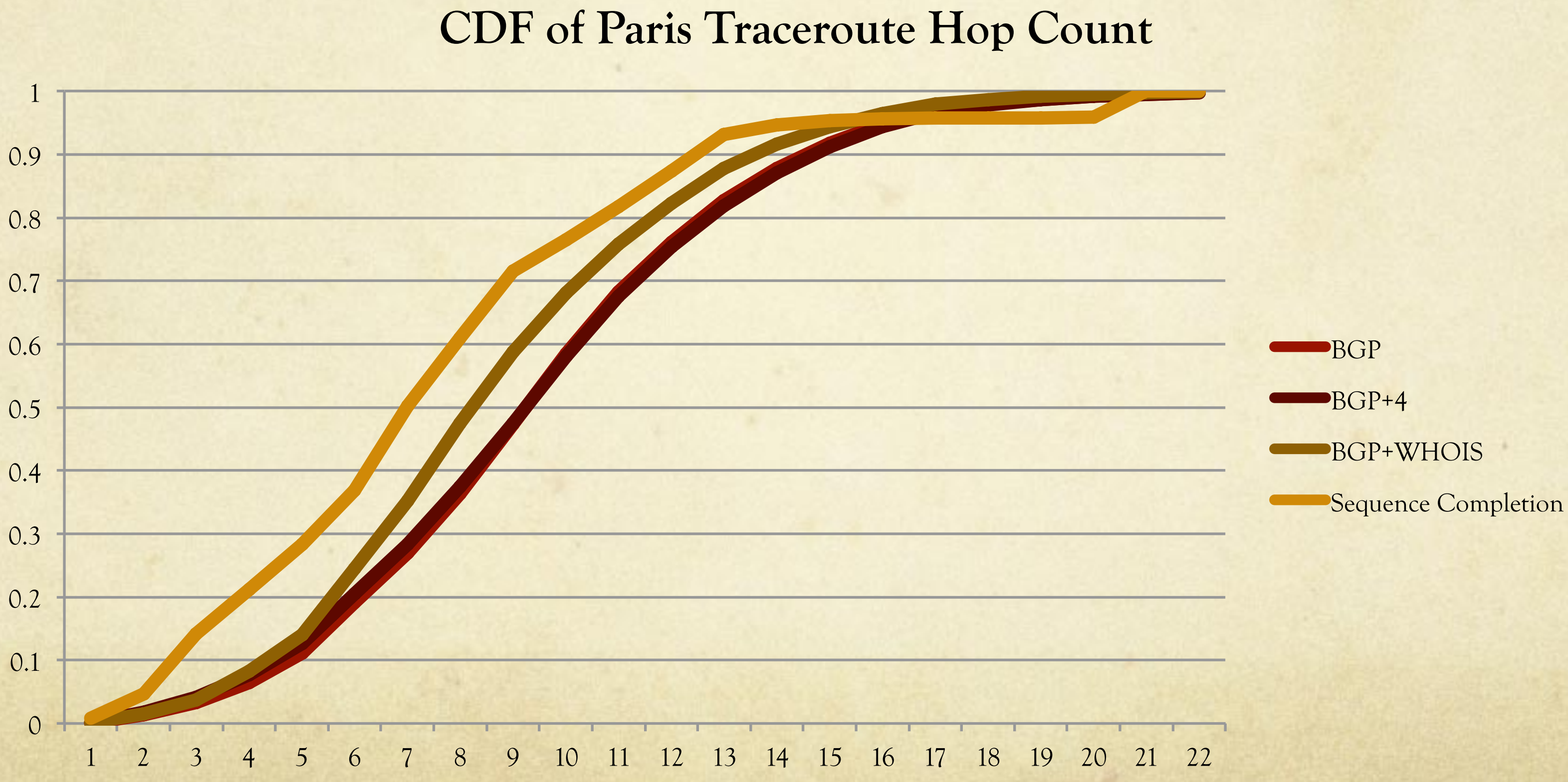
26.6% of all discovered interfaces appeared in BGP-based traces

Additional techniques expand coverage ~4x



Broader or Deeper?

- Three techniques show similar hop count distributions
- BGP+WHOIS lower mean, but greater max by 5 hops



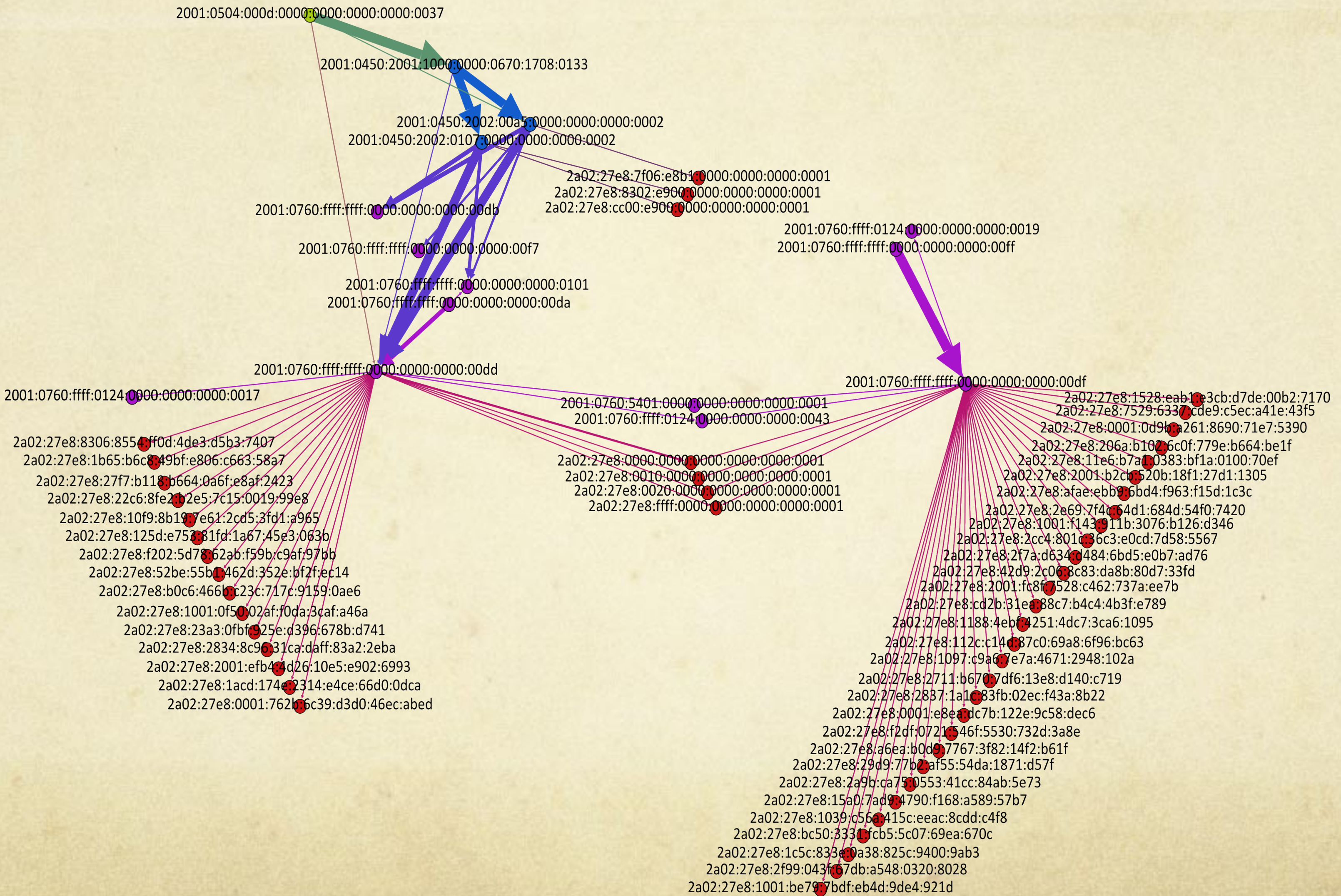
Conclusions

- CIDR prefixes derived from BGP hide a lot of topology information
- New techniques add both detail and depth relative to scanning based on BGP prefixes alone
  - "Augmented BGP": BGP+4, BGP+WHOIS
  - Inference from discovered addresses
- Each technique seems to cover different parts of the network, so combination is necessary
- Future work: Incorporate better algorithms (e.g., ISC)

Digression: Security Appliances

- There are apparently security appliances out there that respond to ICMP requests for every address in a subnet
  - Show up in measurements as highly active networks / highly connected nodes
  - May be useful for mapping out subnet boundaries
- “20% test” detects with high confidence
  - If 2 of 10 randomly chosen addresses within a network respond to pings …
  - … then there’s probably one of these devices there.

Digression: Security Appliances



Thanks!

Richard Barnes
<rbarnes@bbn.com>
Rick Altmann
Daniel Kerr