A Case for Networking as a Service

Tom Anderson
University of Washington

Joint work with:
Arvind Krishnamurthy, Sylvia Ratnasamy, and Scott Shenker

Financial support from: NSF, Cisco, and Google

The Internet Has Issues

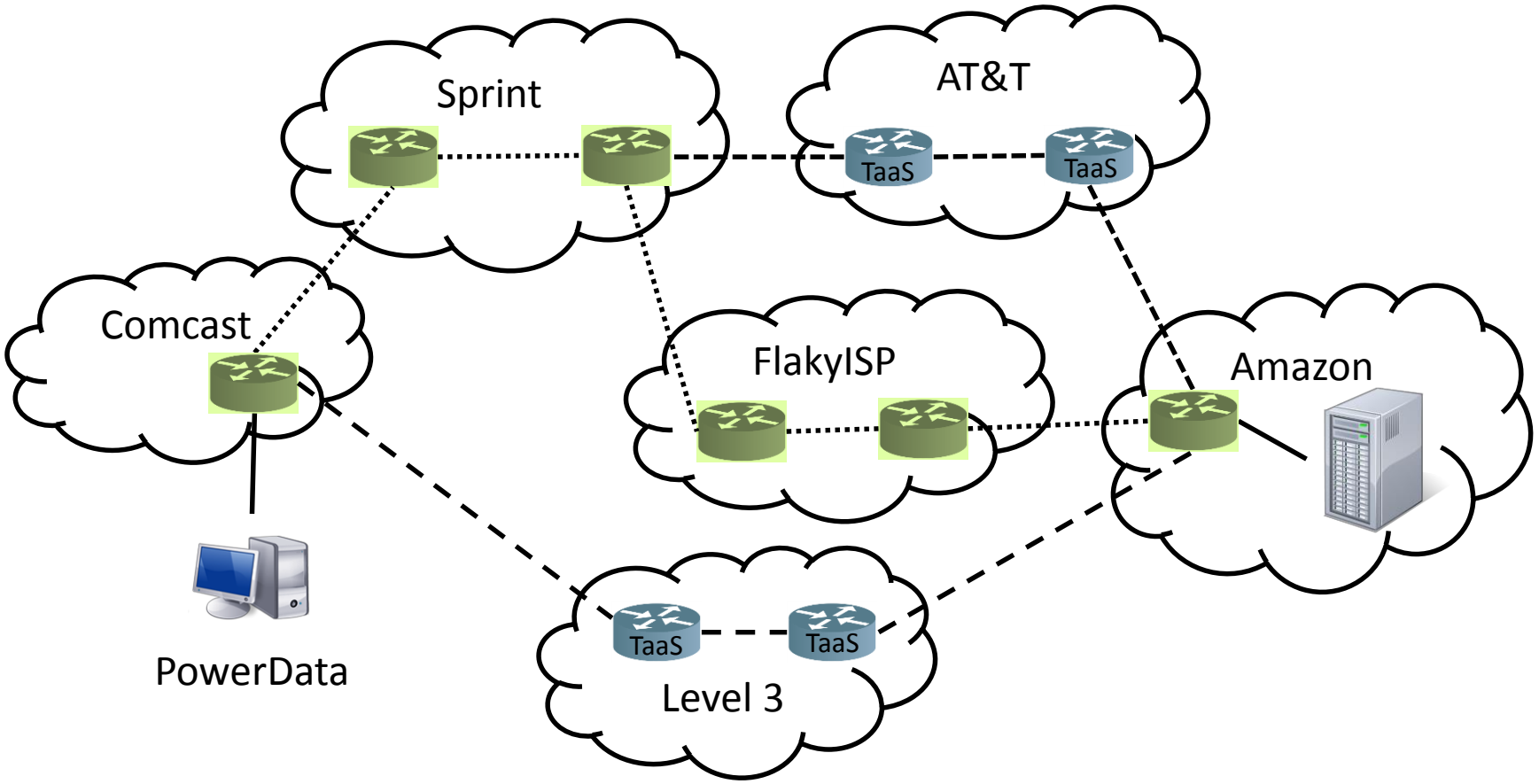
Avoidable outages and poor performance due to:

- Pathological routing policies
- Route convergence delays
- Misconfigured ISPs
- Prefix hijacking
- Malicious route injection
- Router software and firmware bugs
- Distributed denial of service

Known technical solutions to all of these issues

- Very little progress at implementing solutions

Local Problem => Global Outage



Networking as a Service

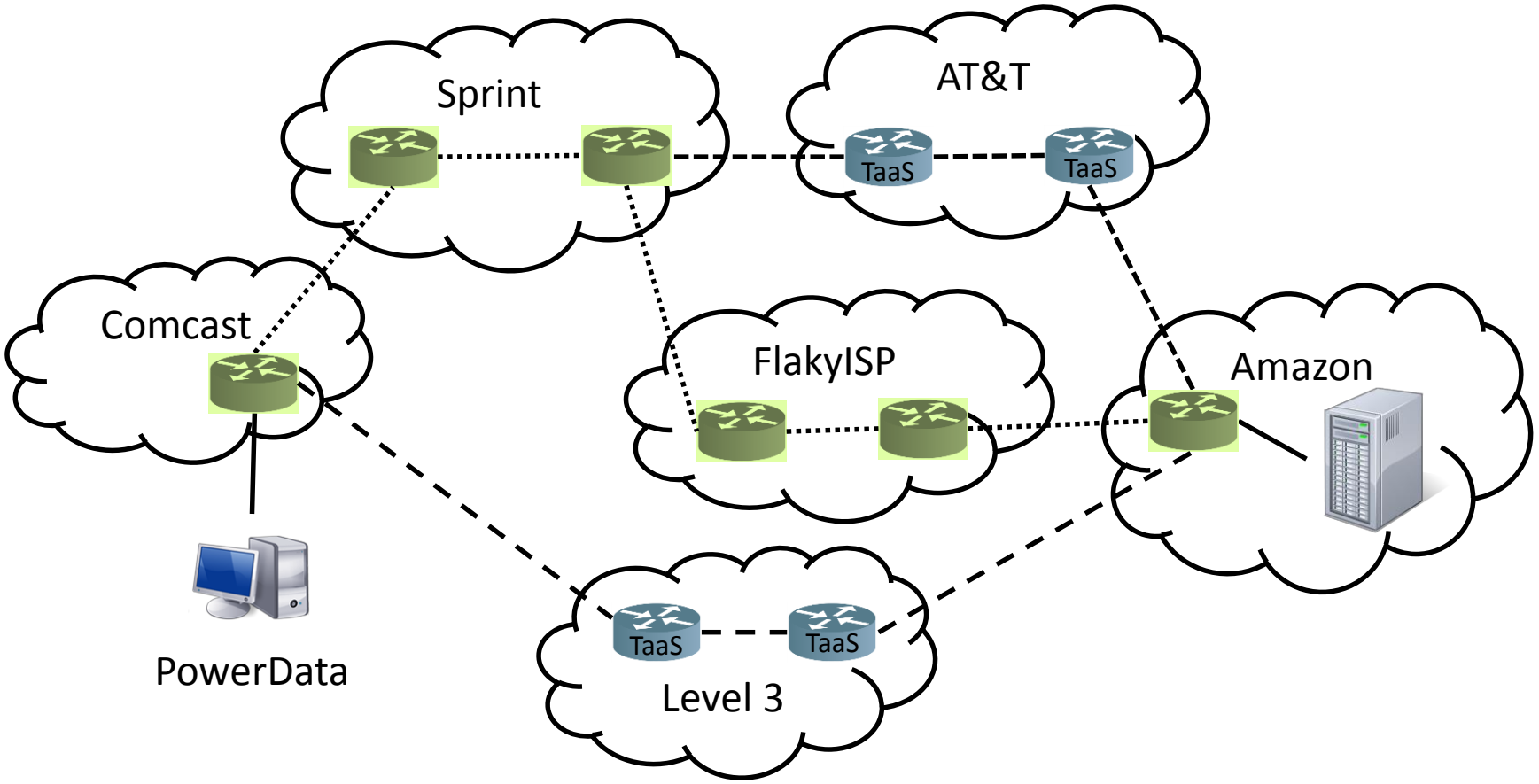
Data centers today offer computational and storage services to remote customers

- Accessible over the Internet

NaaS: Network operators offer networking services to remote customers

- Transit, packet swizzling, and packet processing
- ISPs only promise what they can *directly* provide
- Potential for much better security, reliability, worst case performance, incremental adoption than today's Internet

Local Problem => Local Outage



"A good network is one that I never have to think about" – Greg Minshall

Hot Interconnects, 1994:
A Case for Networks of Workstations

- Build scalable services out of commodity PC's connected by a scalable, switched network
  - Prediction: cost/performance benefits of volume manufacturing would beat custom hardware
  - [In 1995: first web search engine, Altavista, was built on DEC's largest shared memory multiprocessor]
  - Motivated new distributed system designs
- Low-latency/low overhead message passing
  - An early SDN: redesign of the network stack
  - Goal was 10 microsecond RTT for apps

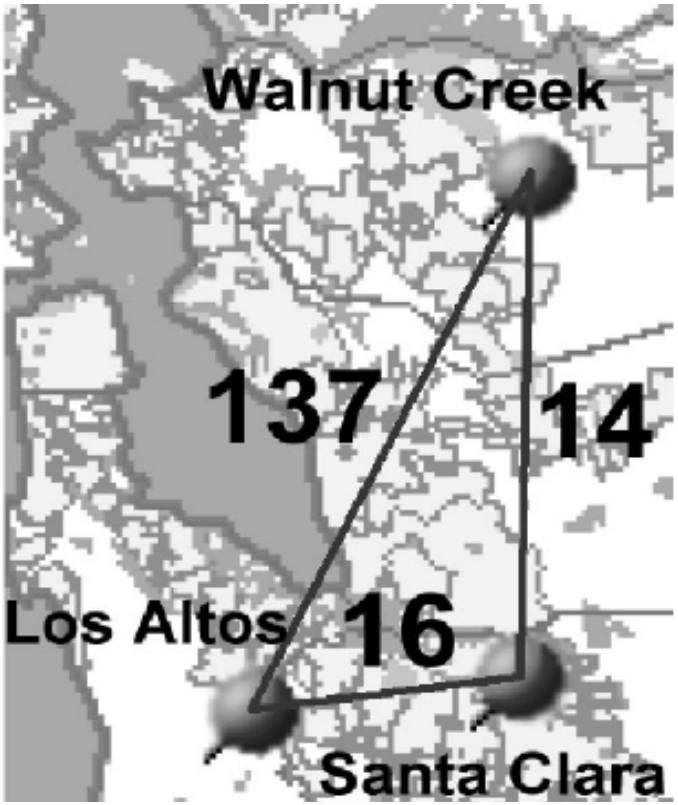
Hot Interconnects, 1998
A Case for Detour Routing

Routes in the Internet do not obey the triangle inequality

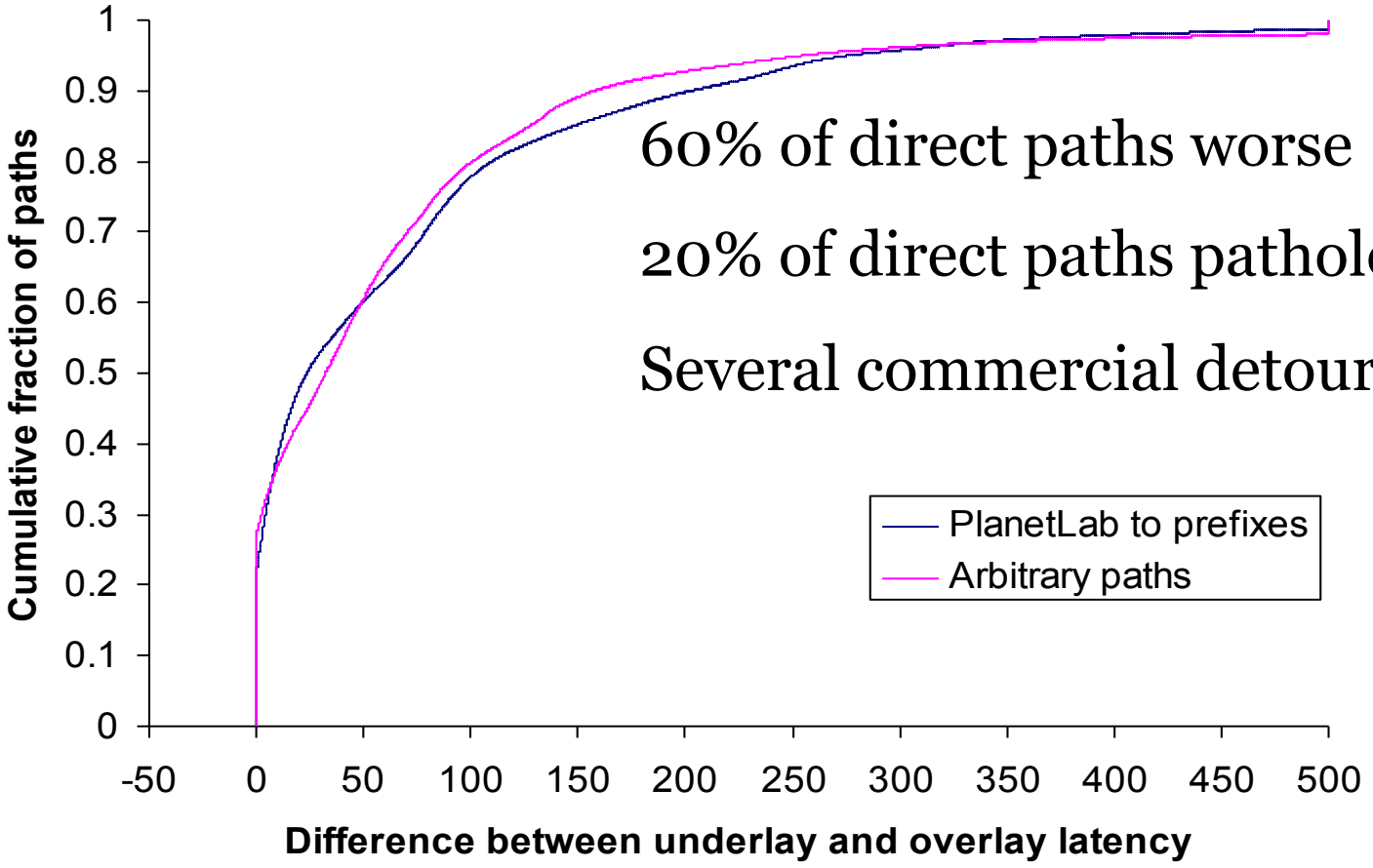
- 40% of all Internet routes
- 10% pathological
- Similar issues with route availability

Fix via overlay routing?

Embarrass ISPs into improving their routing?



Detour Routing Today

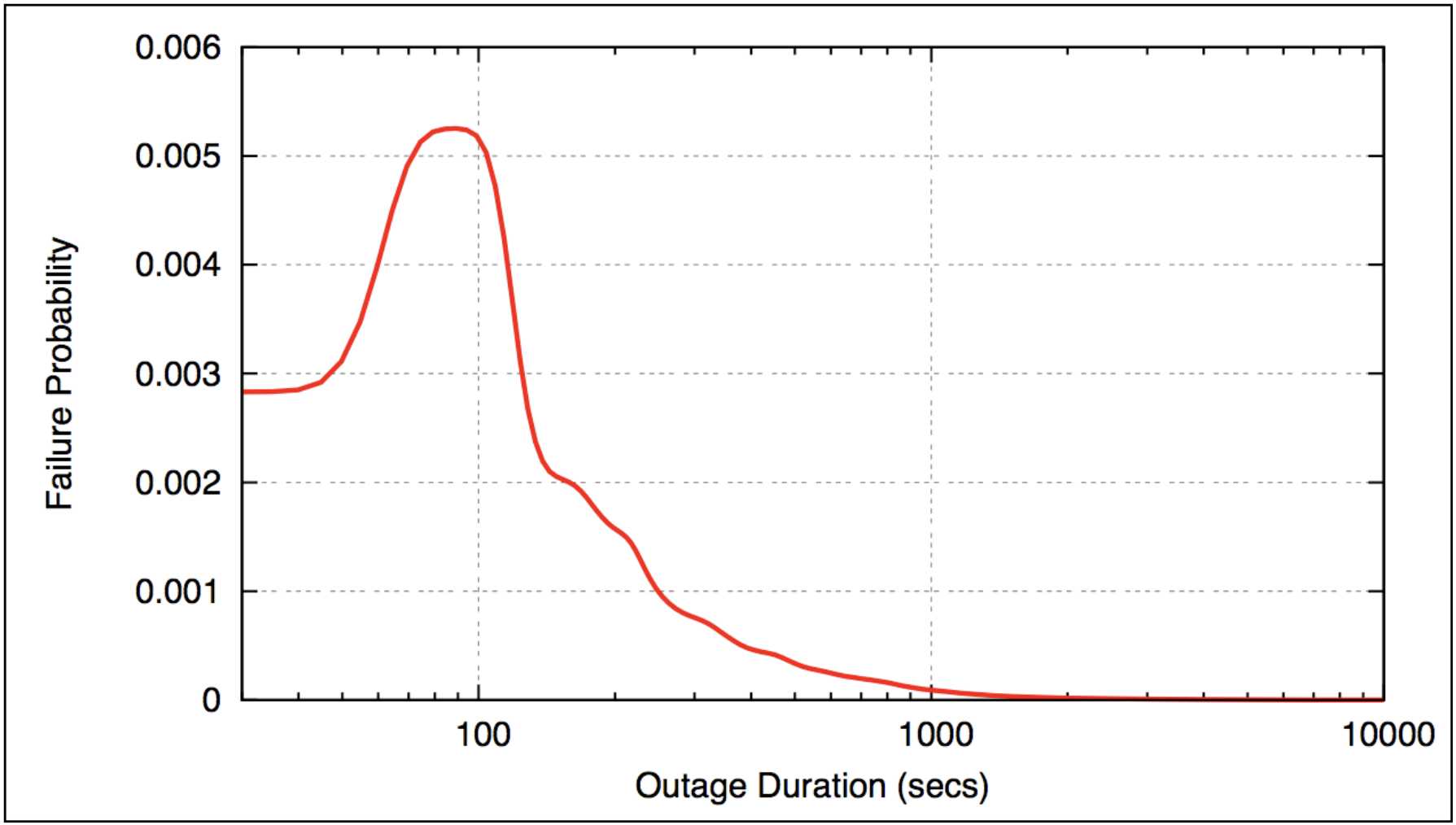


60% of direct paths worse

20% of direct paths pathological

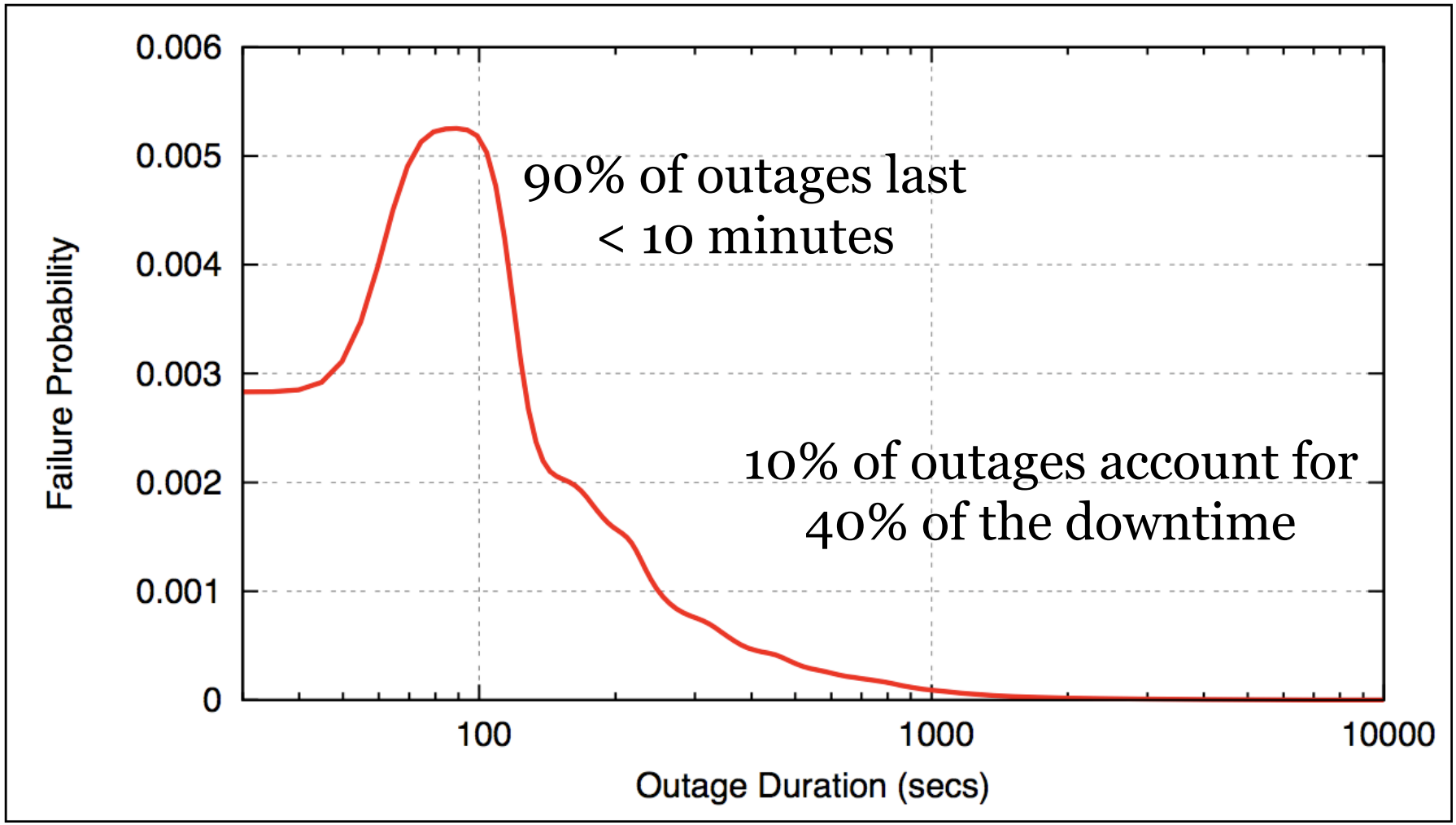
Several commercial detour services

Characterizing Internet Outages

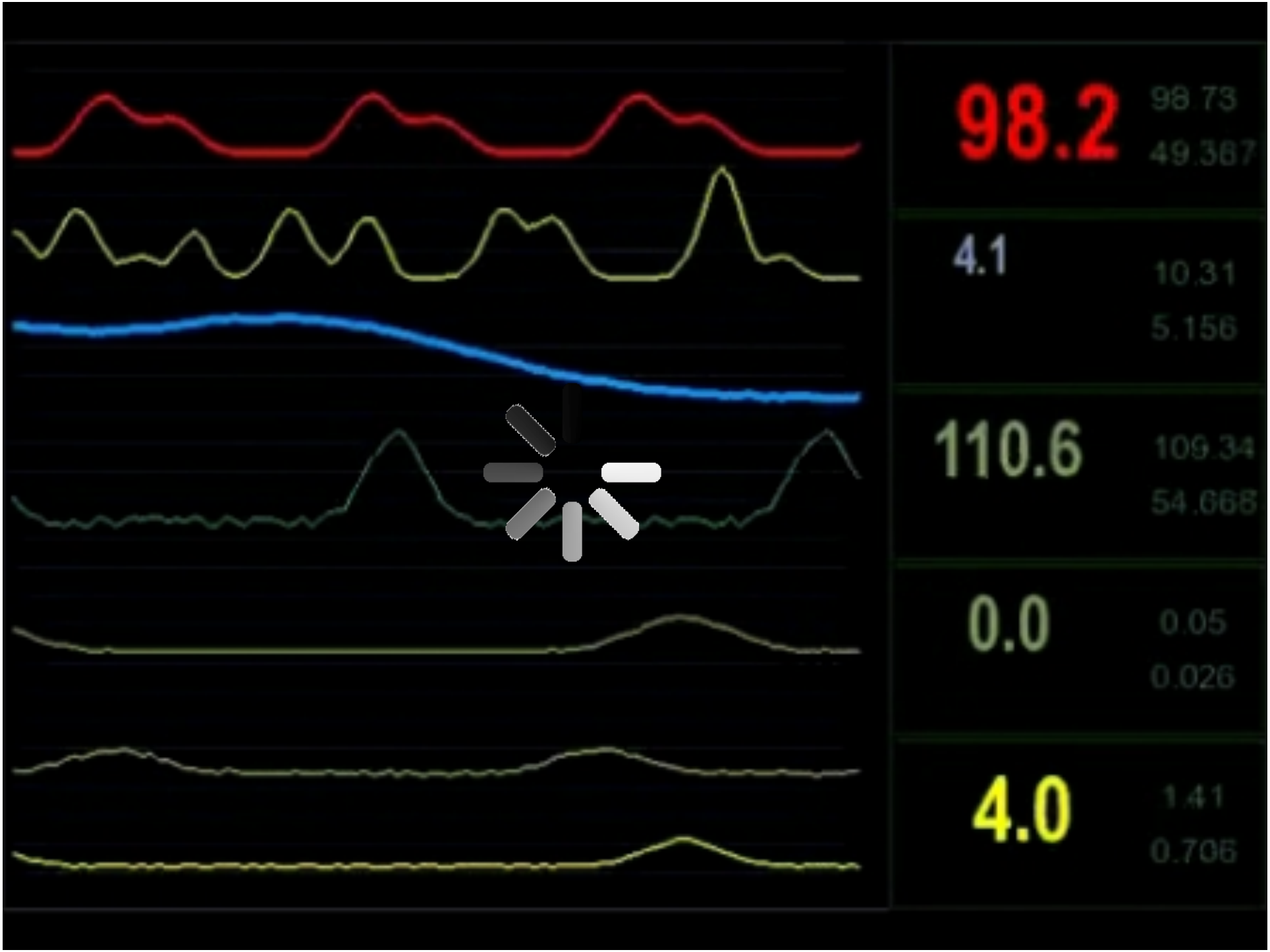


Two month study: more than 2M outages, most partial

Characterizing Internet Outages



Two month study: more than 2M outages, most partial


98.2
98.73
49.387
4.1
10.31
5.156
110.6
109.34
54.668
0.0
0.05
0.026
4.0
1.41
0.706

Roadmap

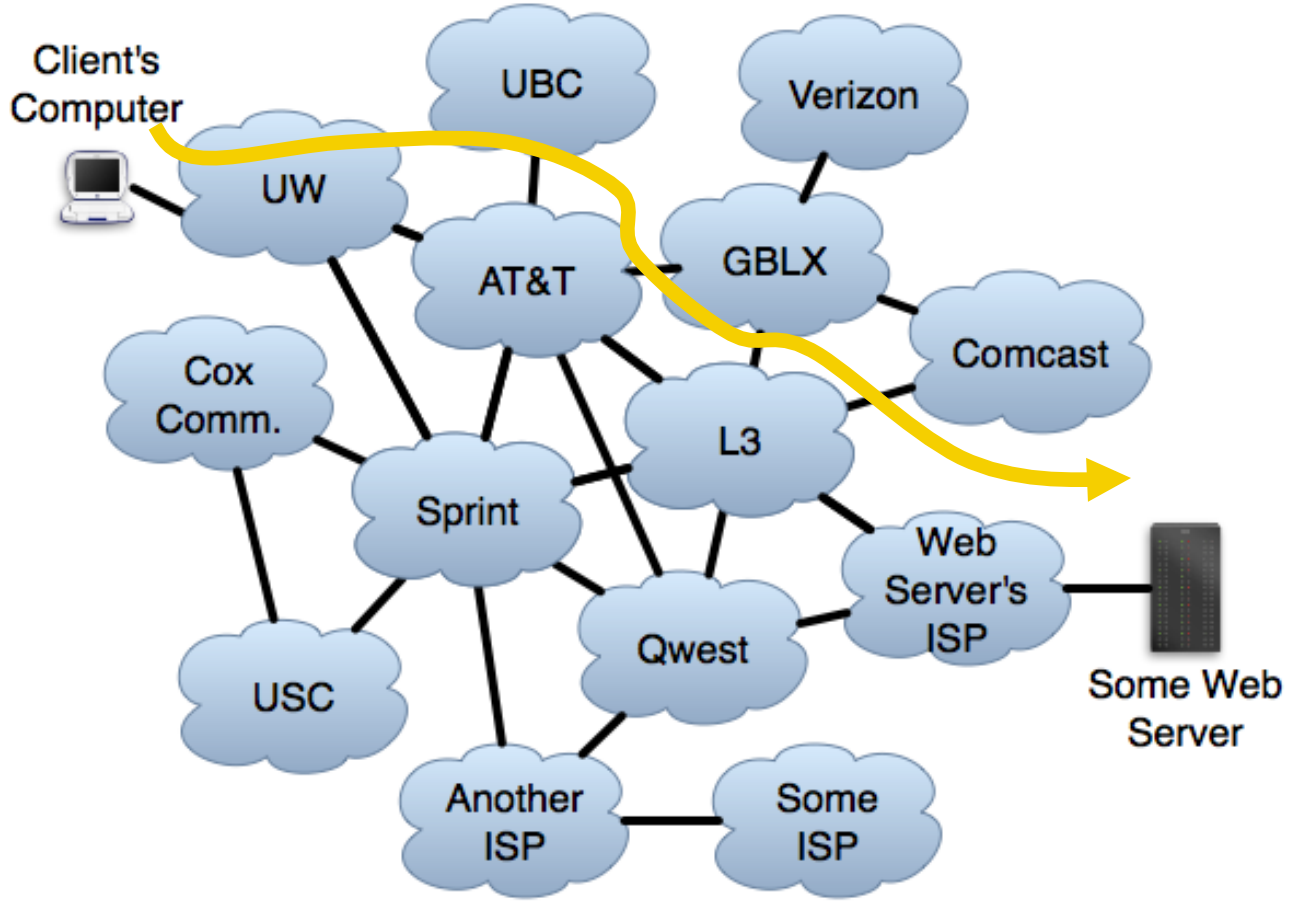
Brief primer on Internet routing

A catalog of Internet vulnerabilities

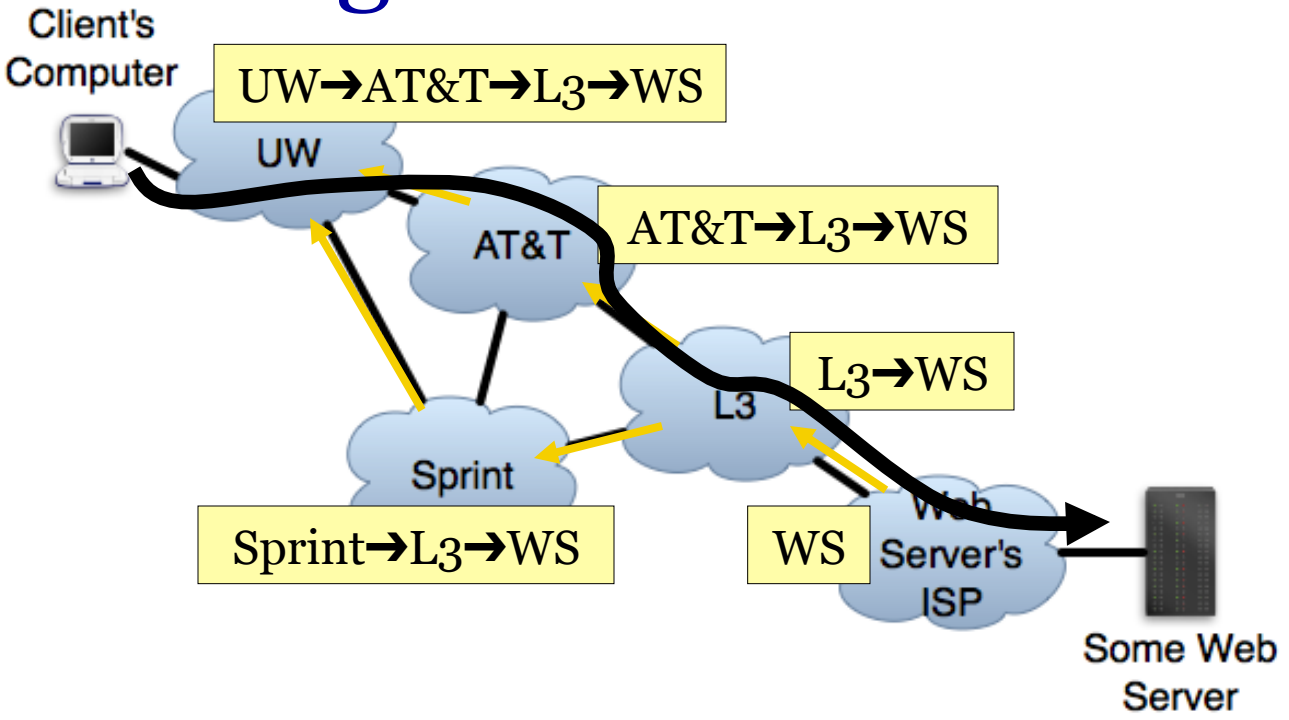
- Performance anomalies
- Delayed route convergence
- Outages due to misconfigurations
- Prefix and route hijacking
- Distributed denial of service
- …

A Case for Networking as a Service

Federation of Autonomous Networks



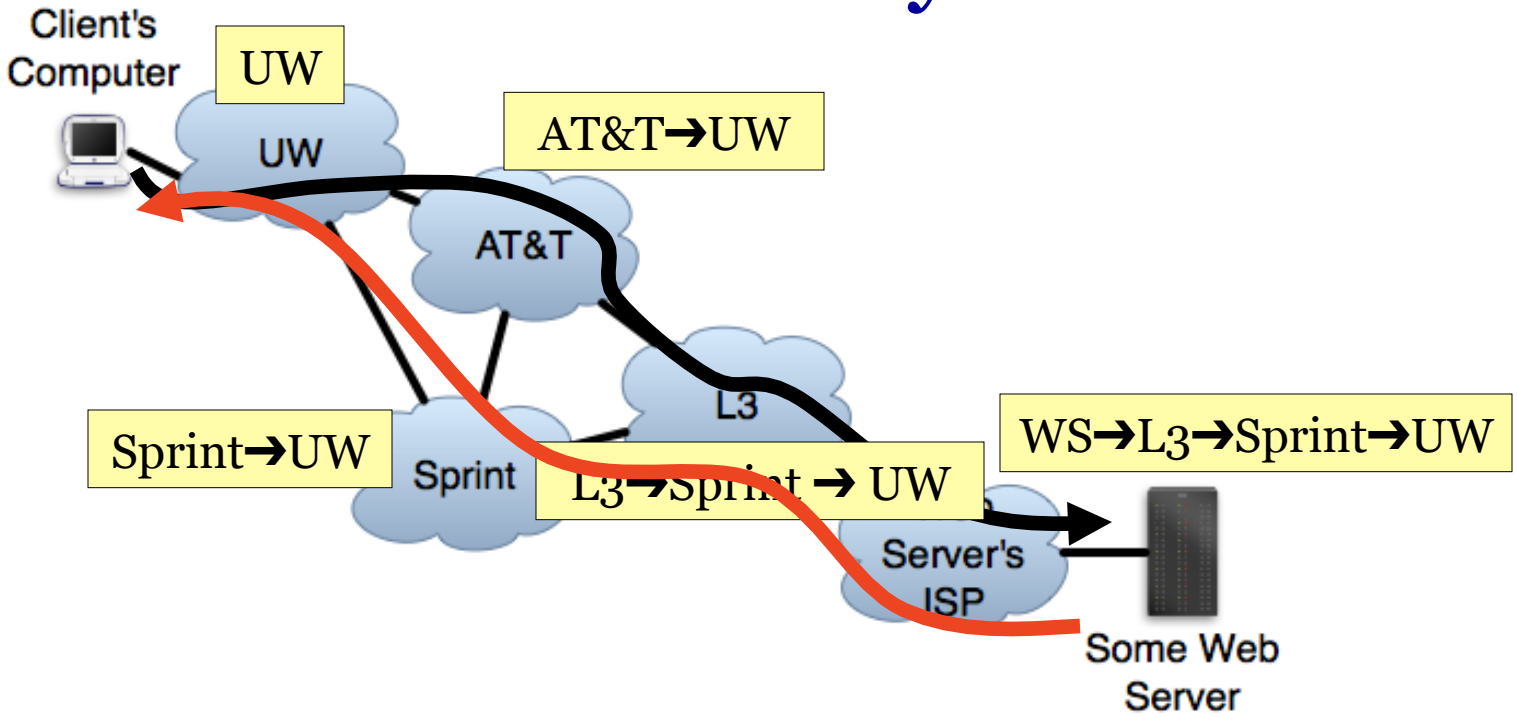
Establishing Inter-Network Routes



Border Gateway Protocol (BGP)

- Internet's interdomain routing protocol
- Network chooses path based on its own opaque policy
- Forward your preferred path to neighbors

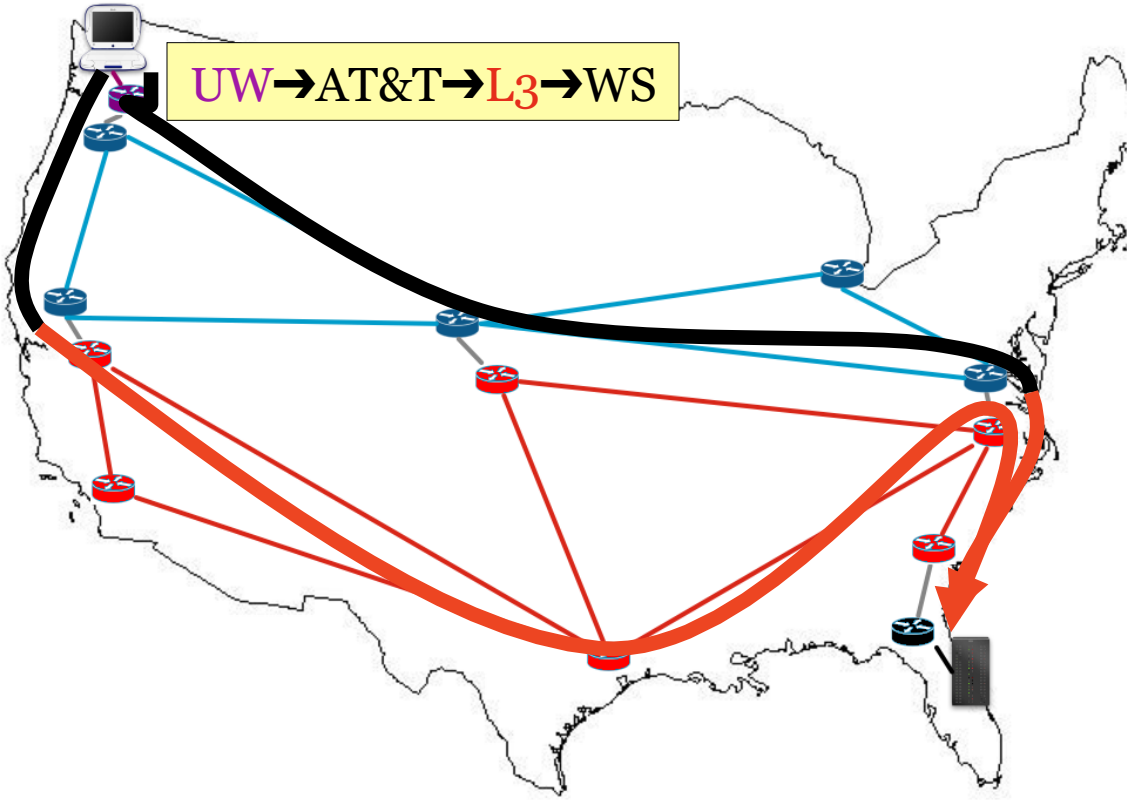
BGP Paths Can Be Asymmetric



Asymmetric paths are a consequence of policy

- Available paths depend on policy at other networks
- Network chooses path based on its own opaque policy ($$)
- Allowing policy-based decisions leads to asymmetry

From Interdomain Path to Router-Level



Each ISP decides how to route across its network and where to hand traffic to next ISP

End-to-end depends on interdomain + intradomain

- Performance and availability stem from these decisions

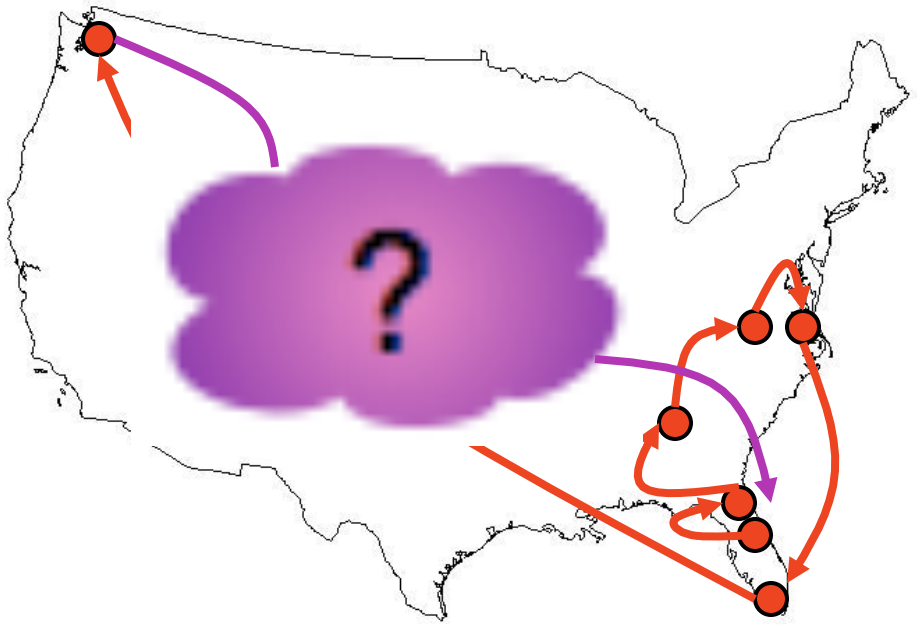
Example of an Inflated Path

150ms round-trip time Orlando to Seattle, 2-3x expected

- E.g., Content provider detects poor client performance

(Current practice) Issue traceroute, check if indirect

Hop no.	DNS name / IP address
1	132.170.3.1
2	198.32.155.89
3	**JAX-FL**. . . net.flrnet.org
4	**ATLANTA**ix.cox.com
5	**ASH**. . . as.cox.net
6	core2. . . **WDC**.pnap.net
7	cr1.**WDC**. . . internap.net
8	cr2-cr1.**WDC**. . . internap.net
9	cr1.**MIA**. . . internap.net
10	cr1.**SEA**. . . internap.net



Indirectness: FL→DC→FL

But only explains half of latency inflation

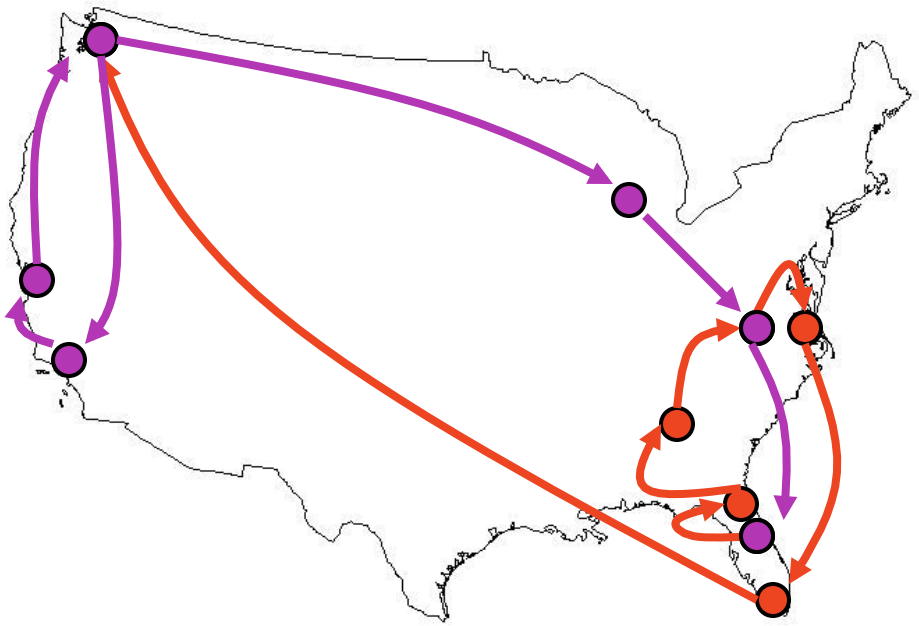
Example of an Inflated Path

(Current practice) Issue traceroute, check if indirect

- Does not fully explain inflated latency

(Our tool) Use reverse traceroute to check reverse path

Hop no.	DNS name / IP address
1	cr1.**SEA**...internap.net
2	cr1.**SEA**...internap.net
3	internap...**LSANCA**01.transitrail.net
4	te4...**LSANCA**01.transitrail.net
5	te4...**PLALCA**01.transitrail.net
6	te4...**STTLWA**01.transitrail.net
7	te4...**CHCGIL**01.transitrail.net
8	te2...**ASBNVA**01.transitrail.net
9	132.170.3.1
10	planetlab2.eecs.**UCF.EDU**



Indirectness: WA →LA→WA

Bad reverse path causes inflated round-trip delay

What Then?

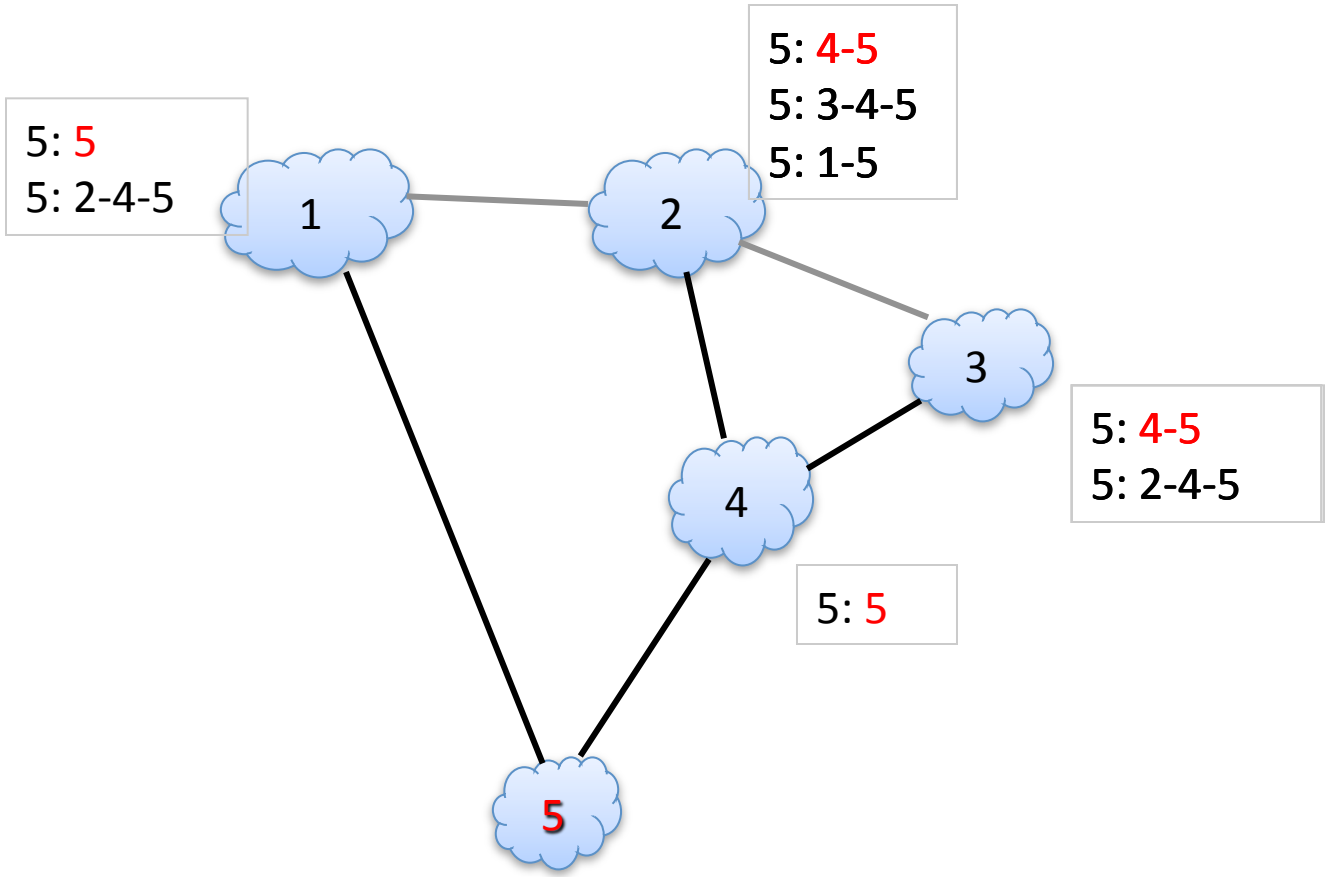
With forward and reverse traceroute, we (now) have tools that can diagnose root cause of pathologically poor paths

- Root cause is typically in a remote ISP, with no direct commercial relationship with source or destination
- Your IT group can send them email, but no recourse if the problem isn't fixed
- Even if fixed, path can change back without notice

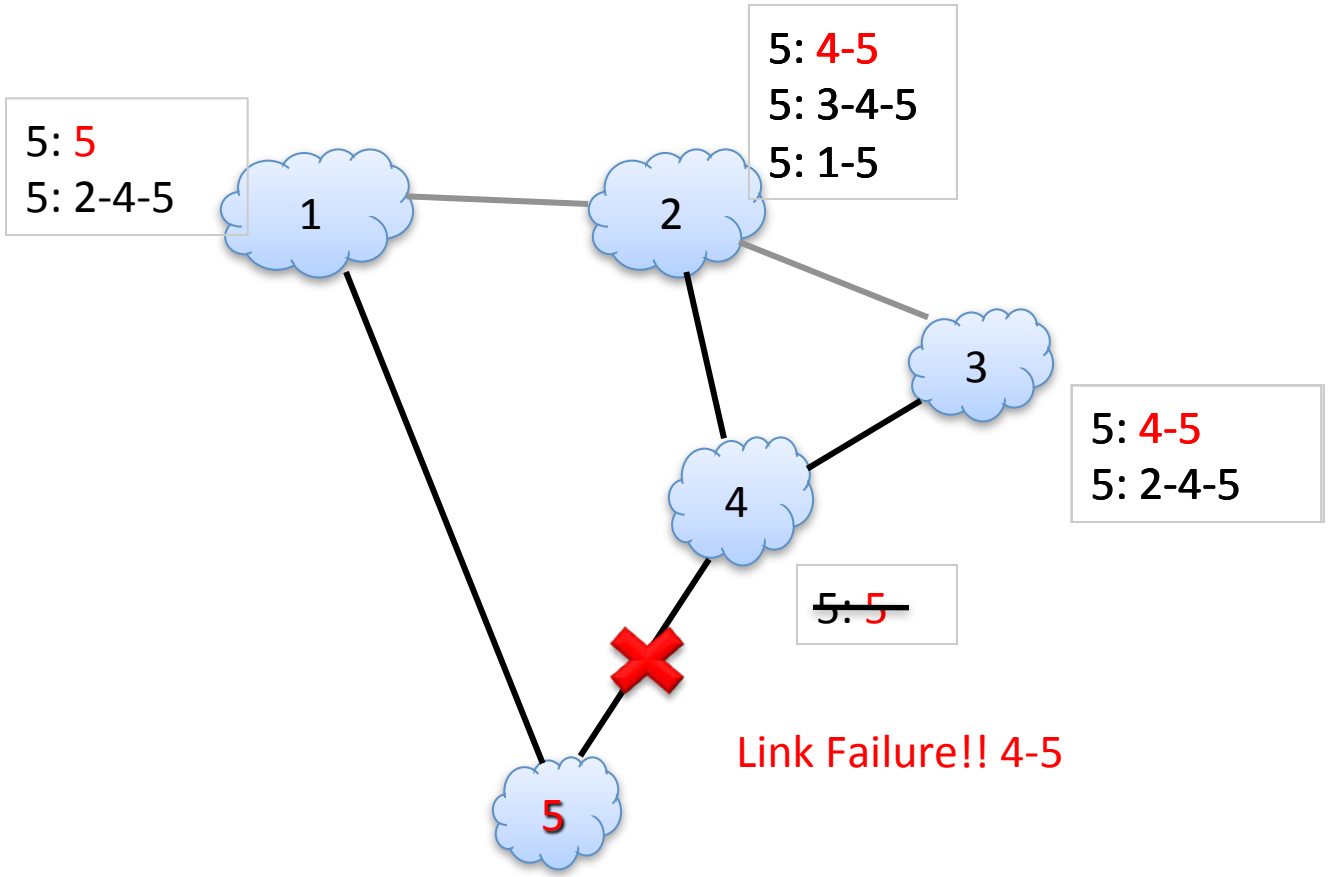
Border Gateway Protocol

- Key idea: *opaque policy routing* under local control
  - Preferred routes visible to neighbors
  - Underlying policies are not visible
- Mechanism:
  - ASes send their most preferred path (to each IP prefix) to neighboring ASes
  - If an AS receives a new path, *start using it right away*
  - Forward the path to neighbors, with a *minimum inter-message interval*
    - essential to prevent exponential message blowup
  - Path eventually propagates in this fashion to all AS's

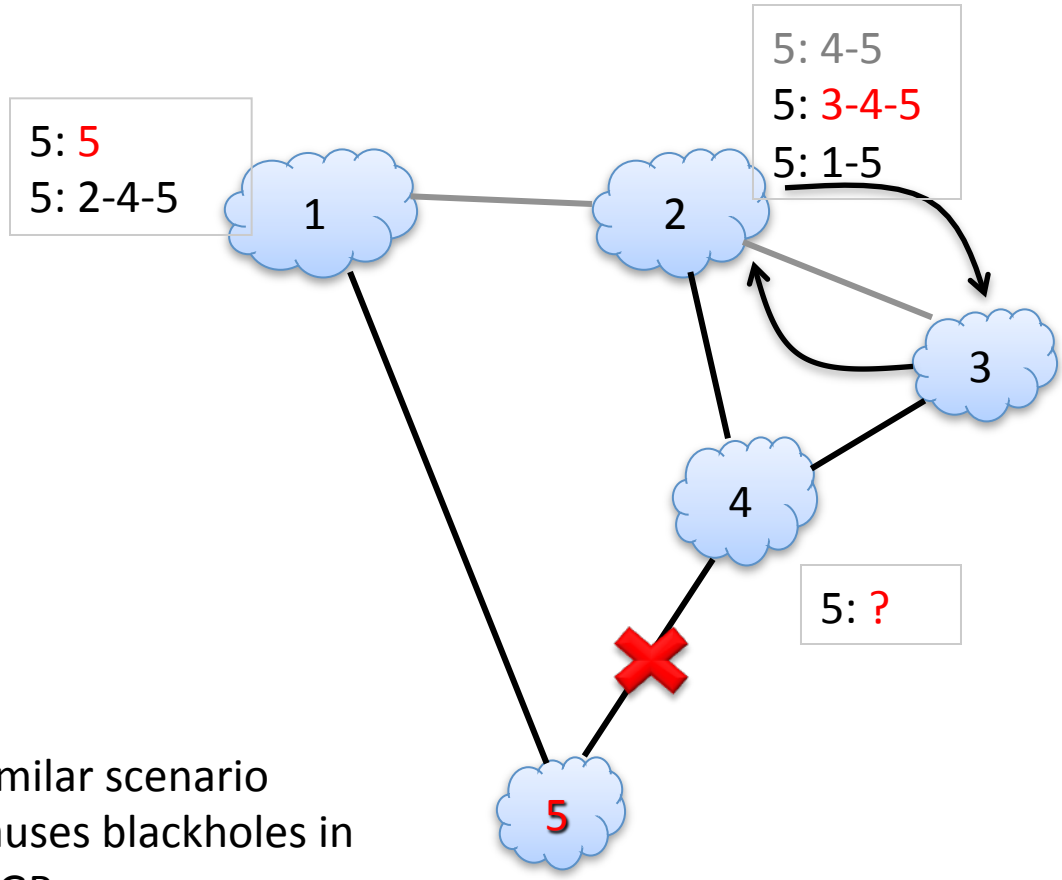
Failures Cause Loops in BGP



Failures Cause Loops in BGP



Failures Cause Loops in BGP

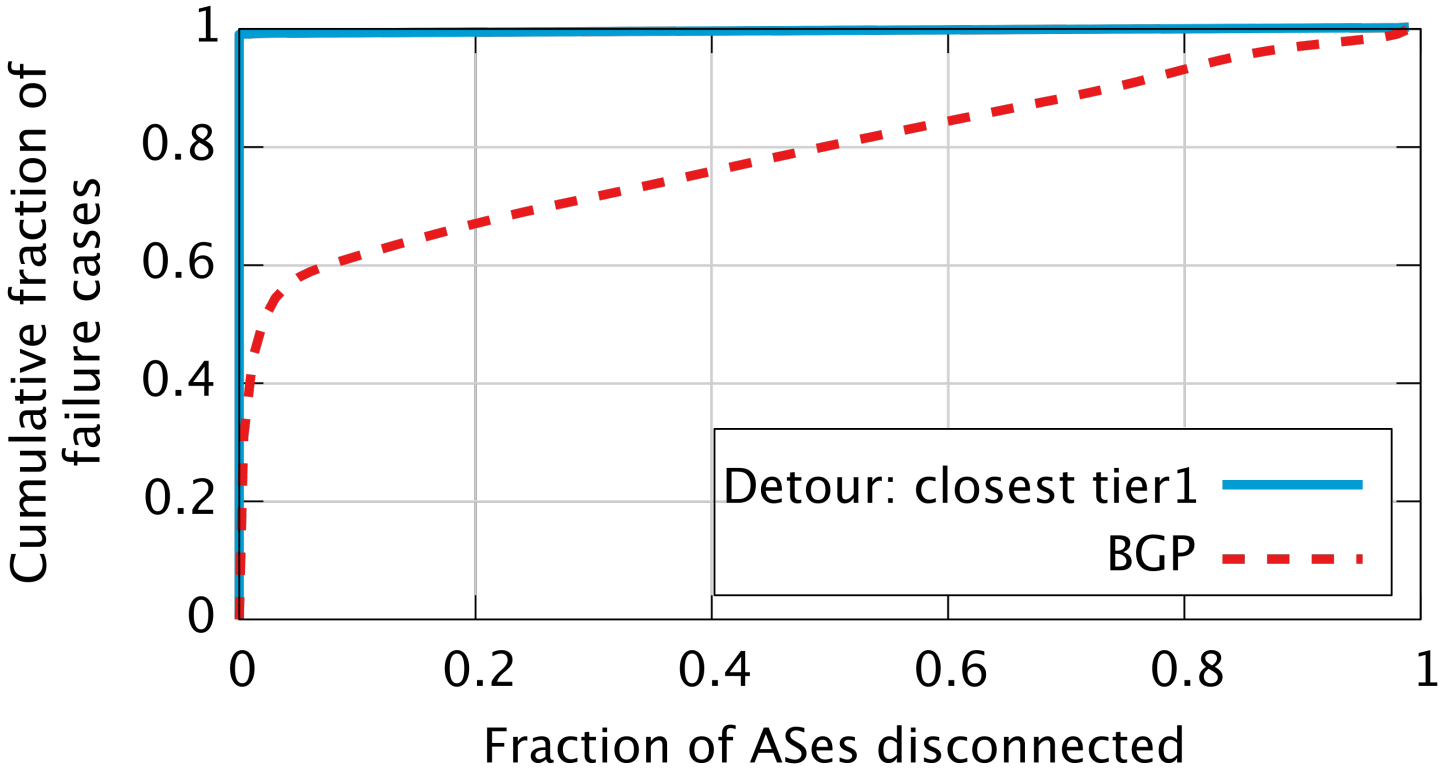


AS2 and AS3 now switch to next best path

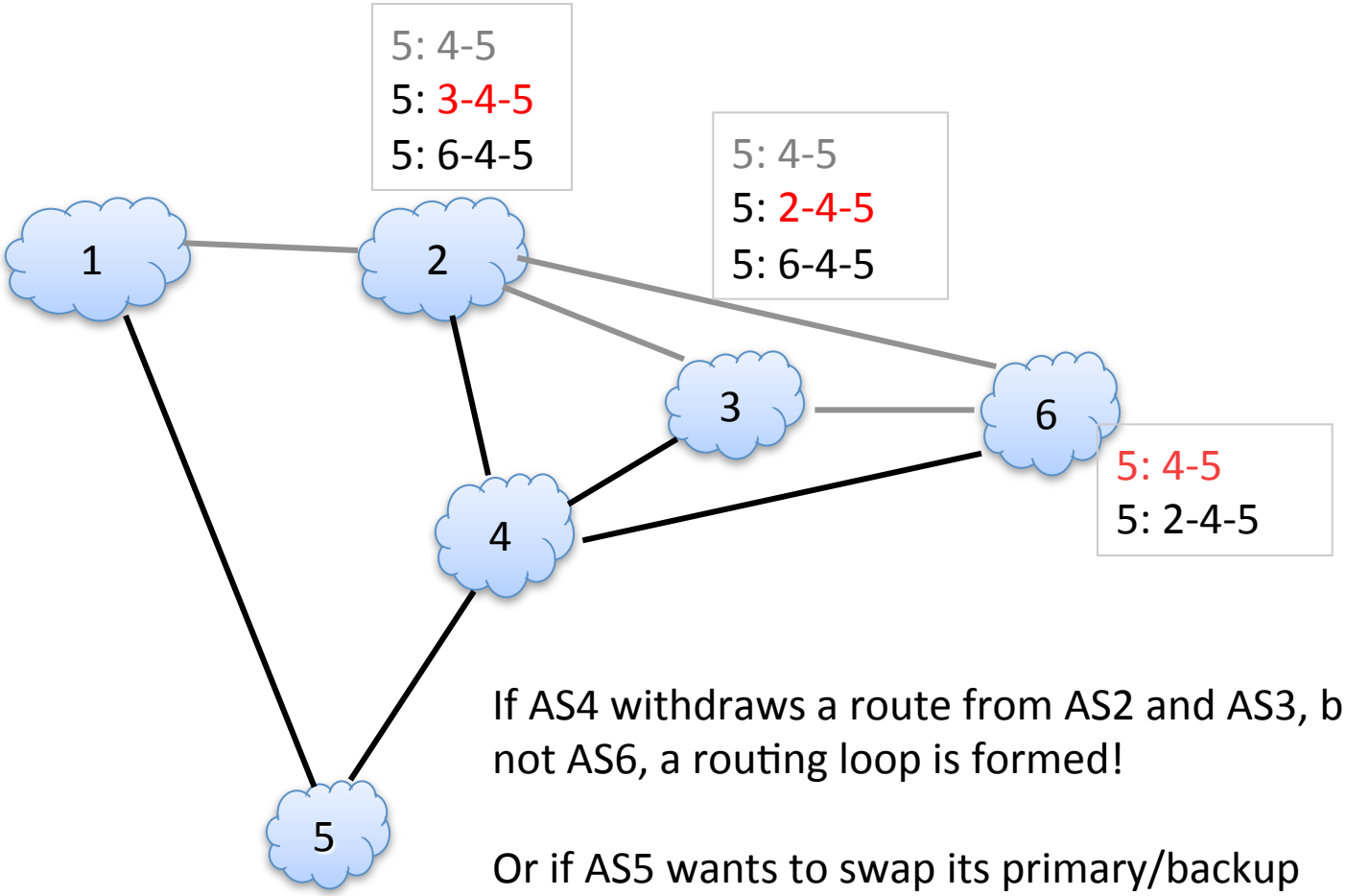
A routing loop is formed between AS2 and AS3!

Similar scenario causes blackholes in iBGP

Availability After Failure



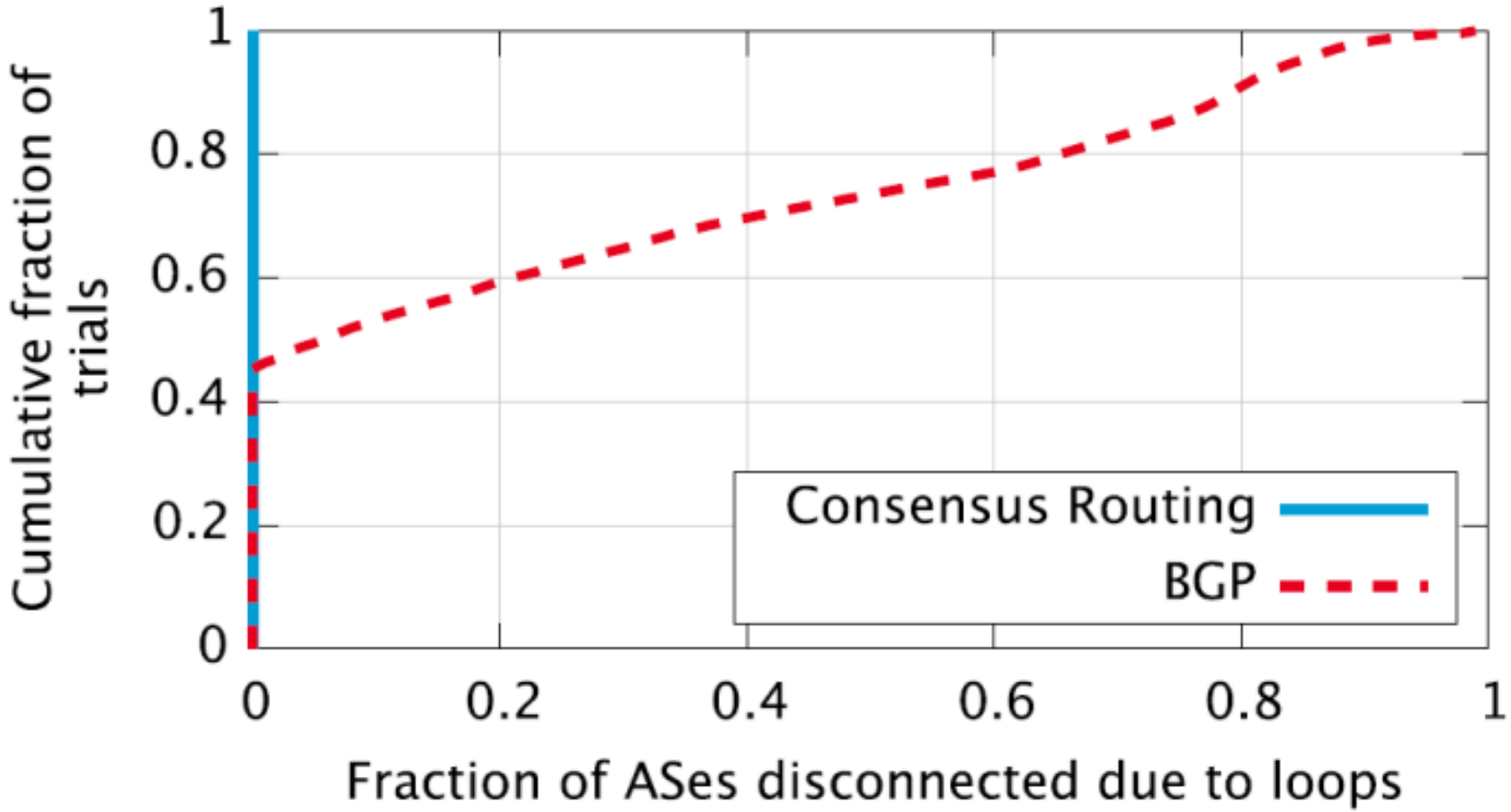
Policy Changes Cause Loops in BGP



If AS4 withdraws a route from AS2 and AS3, but not AS6, a routing loop is formed!

Or if AS5 wants to swap its primary/backup provider from 4 -> 1, or 1->4, a loop is formed

BGP loops, prefix engineering



Delayed Route Convergence

Fixable at the protocol level (consensus routing)

- Interdomain routes switch synchronously across all ISPs, only once all ISPs have learned of new route
- In meantime, use backup detour path through tier 1

Requires global agreement or regulatory mandate

- Delayed routing convergence was fixed *within* ISPs twenty years ago

Operators Struggle to Locate Failures

"Traffic attempting to pass through Level3's network in the Washington, DC area is getting lost in the abyss. Here's a trace from Verizon residential to Level3."

Outages mailing list, December 2010

Mailing List User 1

1 Home router
2 Verizon in Baltimore
3 Verizon in Philly
4 Alter.net in DC
5 Level3 in DC
6 * * *
7 * * *

Mailing List User 2

1 Home router
2 Verizon in DC
3 Alter.net in DC
4 Level3 in DC
5 Level3 in Chicago
6 Level3 in Denver
7 * * *
8 * * *

Reasons for Long-Lasting Outages

Repaired over slow, human timescales

Caused by routers advertising paths that do not work

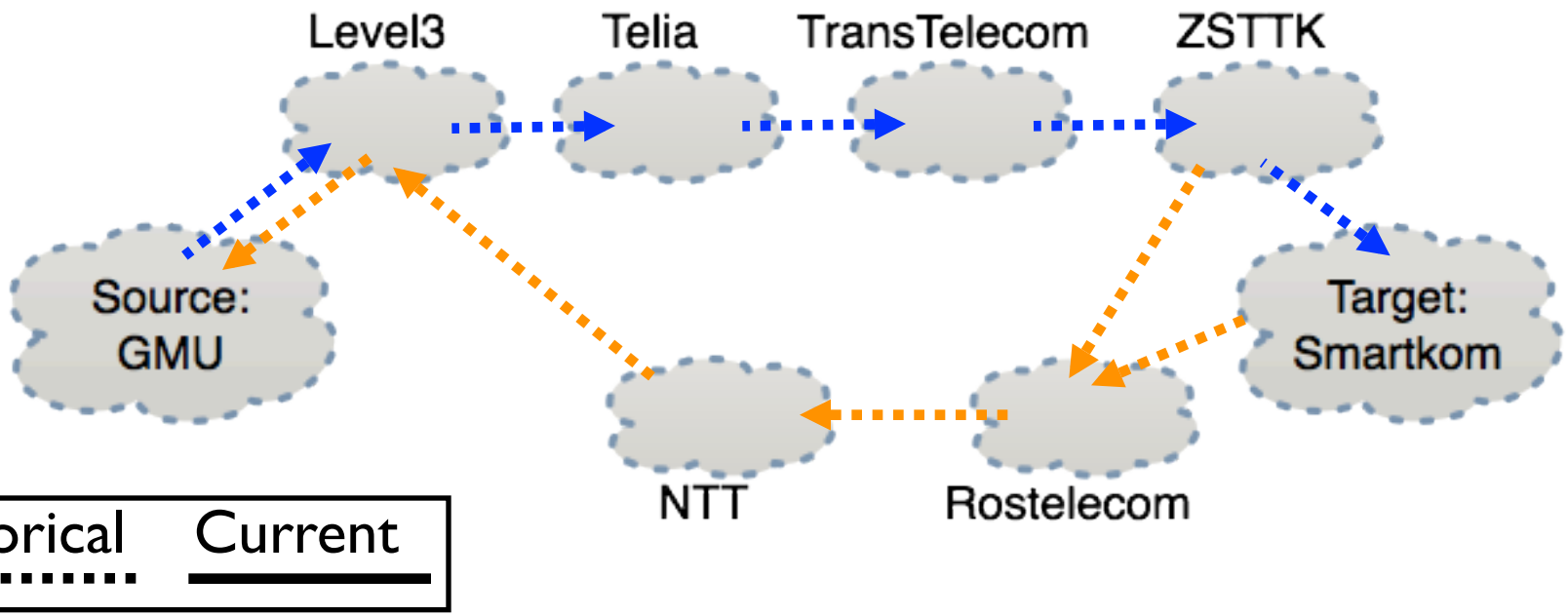
- Corrupted memory on line card causes black hole
- Bad cross-layer interactions causes failed MPLS tunnel
- Misconfigured backup paths triggered by router outage

Control plane does not need to match data plane

Complicated by lack of visibility into or control over routes in other ISPs

How To Locate Failure

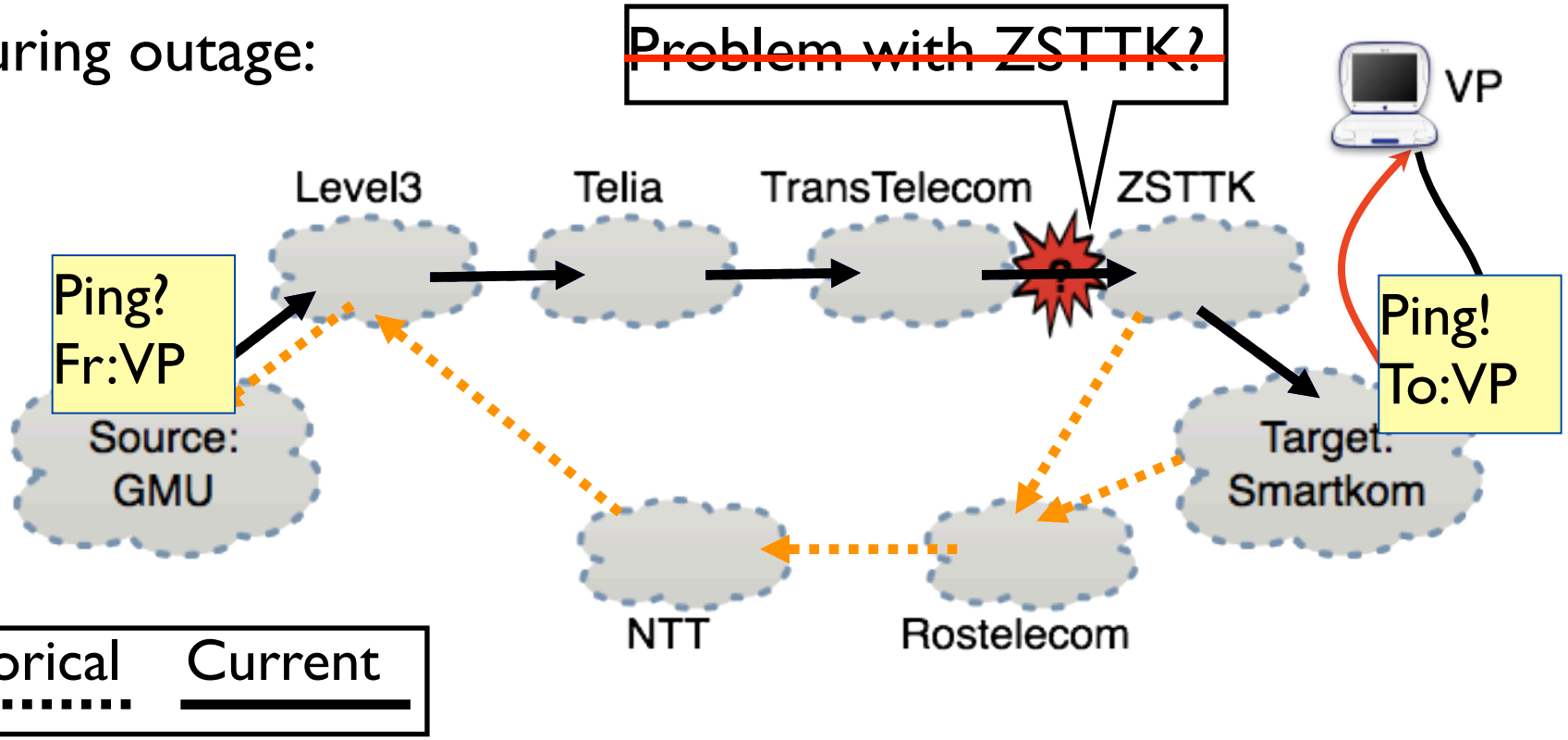
Before outage:



- Historical atlas enables reasoning about changes
- Traceroute yields only path from GMU to target
- Reverse traceroute reveals path asymmetry

How To Locate Failure

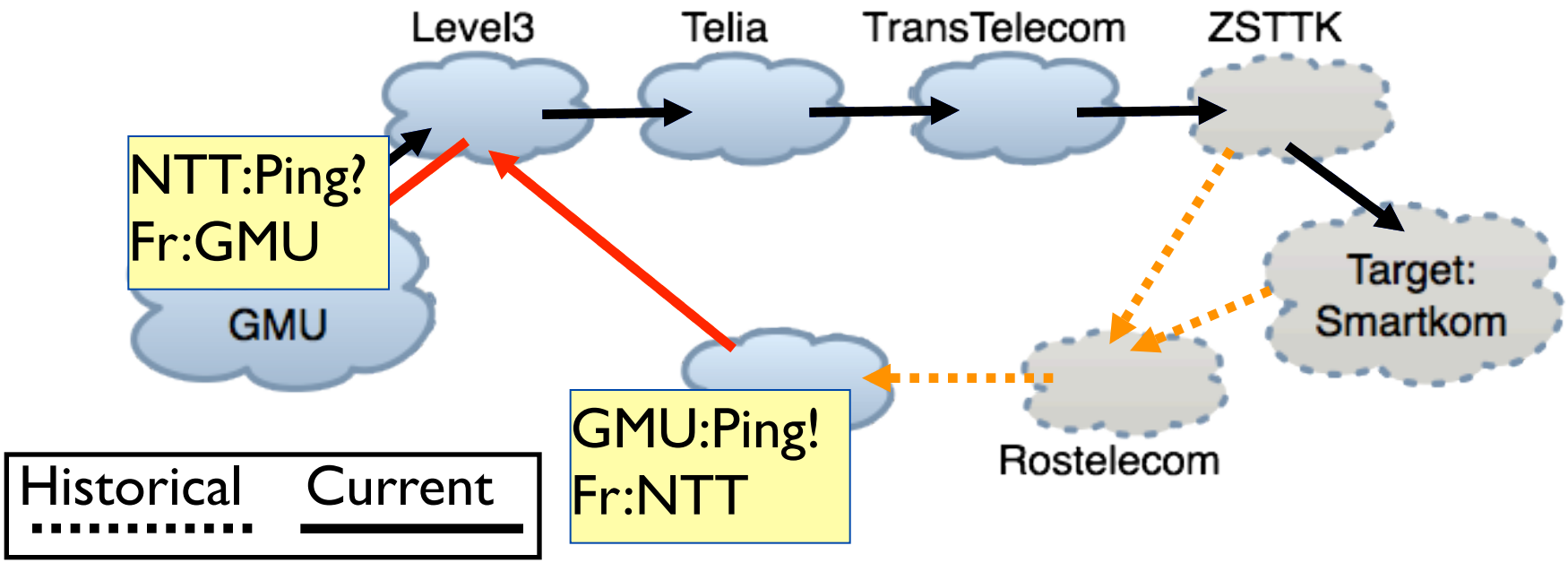
During outage:



- Forward path works

How To Locate Failure

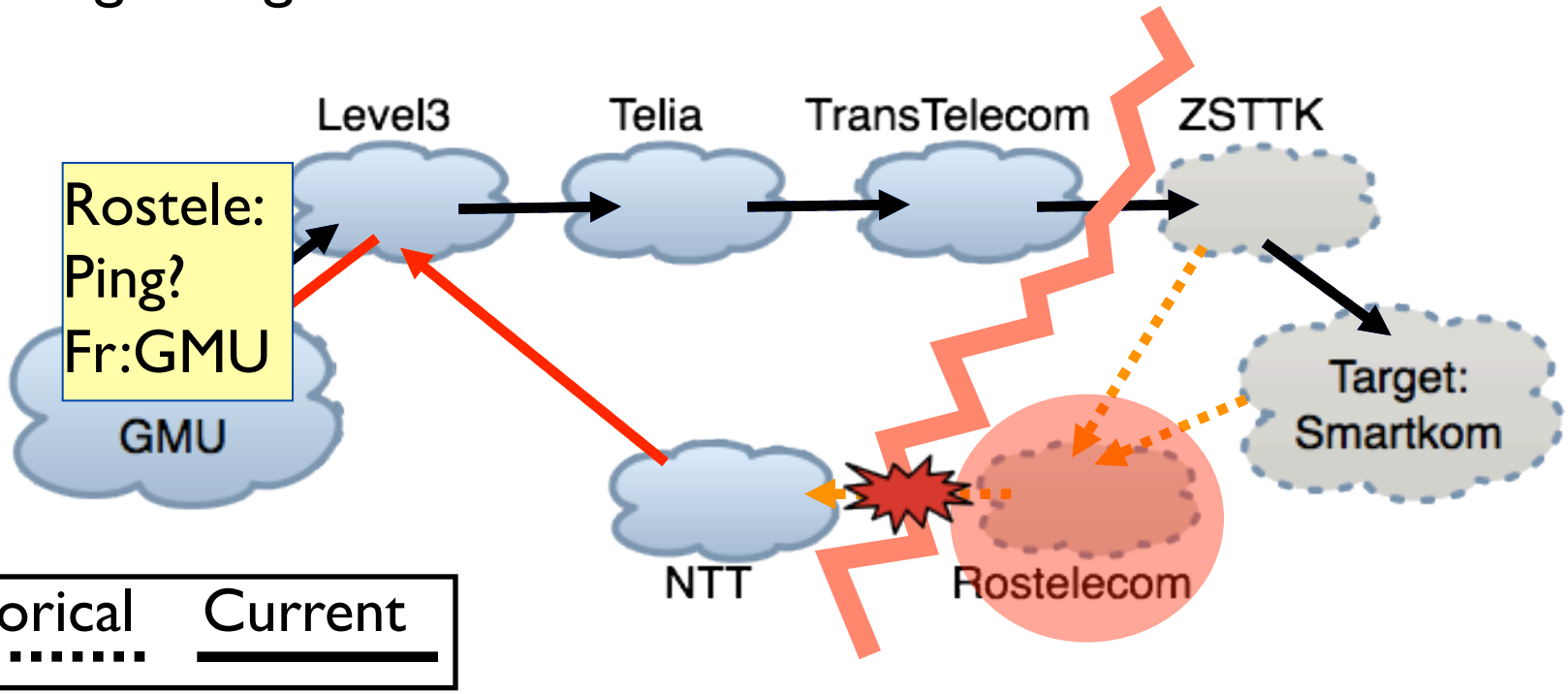
During outage:



- Forward path works

How To Locate Failure

During outage:



- Forward path works
- Rostelcom is not forwarding traffic towards GMU

What Then?

Once we know root cause of Internet outages

- Often in a remote ISP, with no direct commercial relationship with end user or enterprise
- Human intervention required
- No recourse for user if the problem isn't fixed
- Even if fixed, path can break again without notice

Possible to fix at protocol level

- Add BGP hint: tell ISPs along path to avoid using failing AS on routes to destination

The Sad Case of Prefix Hijacking

In BGP, any ISP can announce a route to an IP range, with no authentication required

- Easy to configure new customers
- Easy to misconfigure new customers

Well-known problem (from mid-90's)

- Frequent small outages
- Infrequent mass outages

Well-known solutions

- s-BGP, so-BGP, BGP secure routing extension, …
- Little benefit unless everyone adopts

The Sad Case of Route Hijacking

Any ISP can advertise a short route to a destination

- Even if they don't have a route!

Other ISPs will use route, as long as its consistent with their policy

Well-known problem, well-known solutions

- Little progress at adoption

The Sad Case of Route Control

Enterprises often want to control which paths are taken

- DoD wants to avoid sending its traffic through the PRC

Not technically difficult

Distributed Denial of Service

Internet allows anyone to send any amount of traffic to anyone

- Easy to overwhelm a target
- Vastly easier with massive botnets

Research community converged on a solution:

- Receiver permission to send
- Enforced by network

Low benefit to early adopters, no adoption

The Internet Has Issues

Avoidable outages and poor performance due to:

- Pathological routing policies
- Route convergence delays
- Misconfigured ISPs
- Prefix hijacking
- Malicious route injection
- Router software and firmware bugs
- Distributed denial of service

Known technical solutions to all of these issues

- Trustworthy network requires fixes to *all* of the above

Underlying Problem

ISP offers a service that depends on trustworthiness of every other ISP on the planet

- To coordinate application of route updates
- To not misconfigure routers
- To not hijack prefixes
- To squelch DDoS attacks
- …

NaaS: construct a network where ISP's only promise what they can directly deliver through their own network

Networking as a Service

Data centers today offer computational and storage services to remote customers

- Accessible over the Internet

NaaS: Network operators offer networking services to remote customers

- Transit, packet swizzling, and packet processing
- ISPs only promise what they can *directly* provide
- Potential for much better security, reliability, worst case performance, incremental adoption than today's Internet

Networking as a Service

ISPs sell networking services to remote customers

- Transit from entry PoP to exit PoP over ISP's network
- Packet swizzling (e.g., change destination address)
- Added value services (e.g., multicast, content-centric networking)

Enterprise (or end ISP)

- Stitches end to end paths from AS hops
- Based on advertised resources from each ISP
- Portions of path may use plain old Internet
- Authenticator in packet prevents hijacking

Why Now?

- Very high performance software packet processing
  - 10 Gbps *per core* with minimum sized packets
- Distributing topology updates on a global scale now practical
  - No longer an engineering need to do localized topology management
- ISPs have made considerable progress at improving reliability of their internal operations
  - Often, two orders of magnitude more reliable than the global Internet

Scenario

Large scale cellular telecom

- Market share driven by perceived data network performance, reliability
- Extensive use of middleboxes for managing data traffic
- 70% of data traffic exits telecom network
- Well-developed market for premium service

NaaS provides ability to manage traffic beyond the telecom boundary

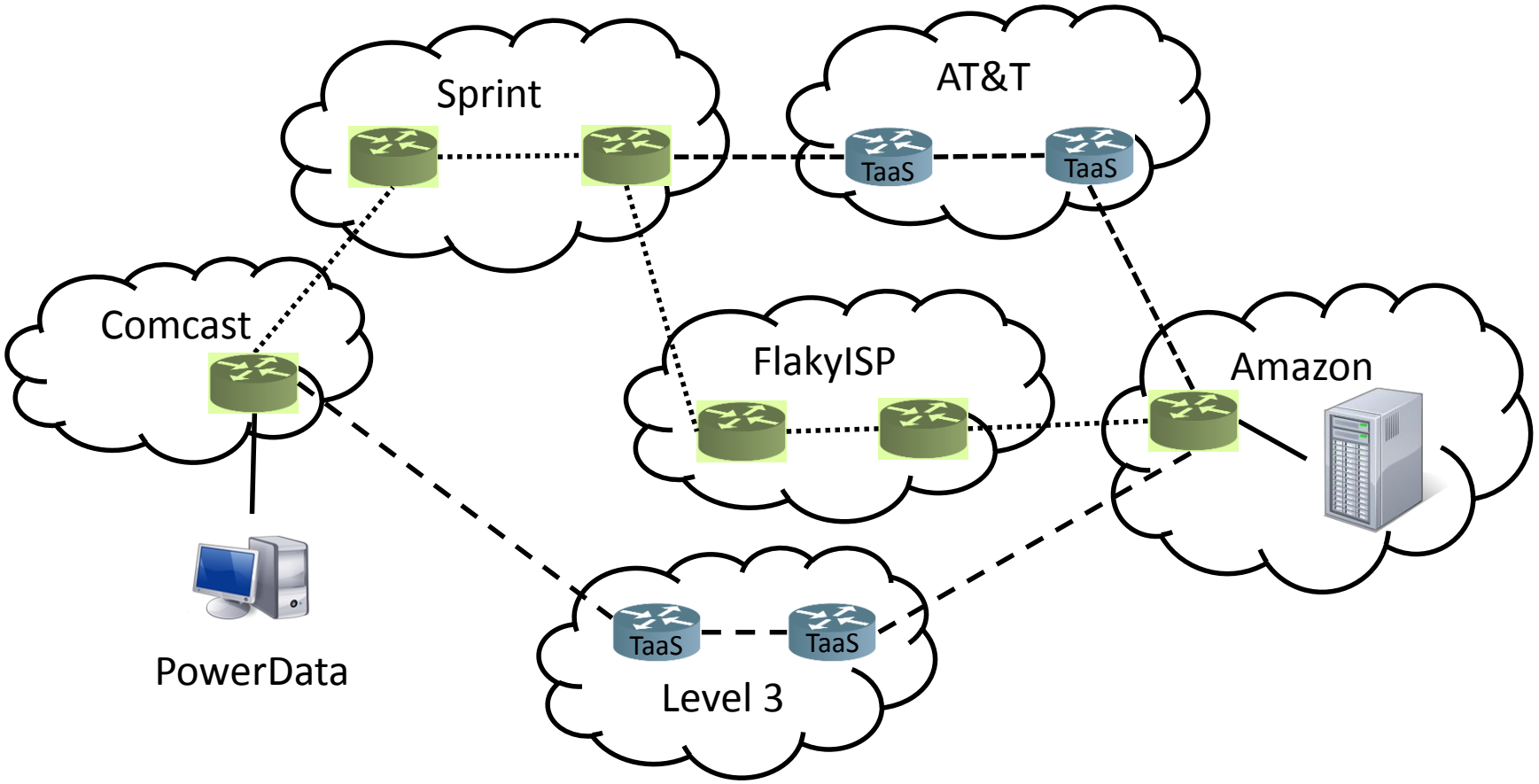
NaaS Design Principles

- Agile and reliable ISPs
  - Flexible deployment of new functionality at the edge
- Each ISP promises only what it can guarantee through its own network
  - Packet delivery, QoS from PoP to PoP
- Incentives for incremental adoption
  - Each ISP charges for its added services, without waiting for neighbors to adopt
- Security through minimal information exposure
  - Simpler protocols => smaller attack surface

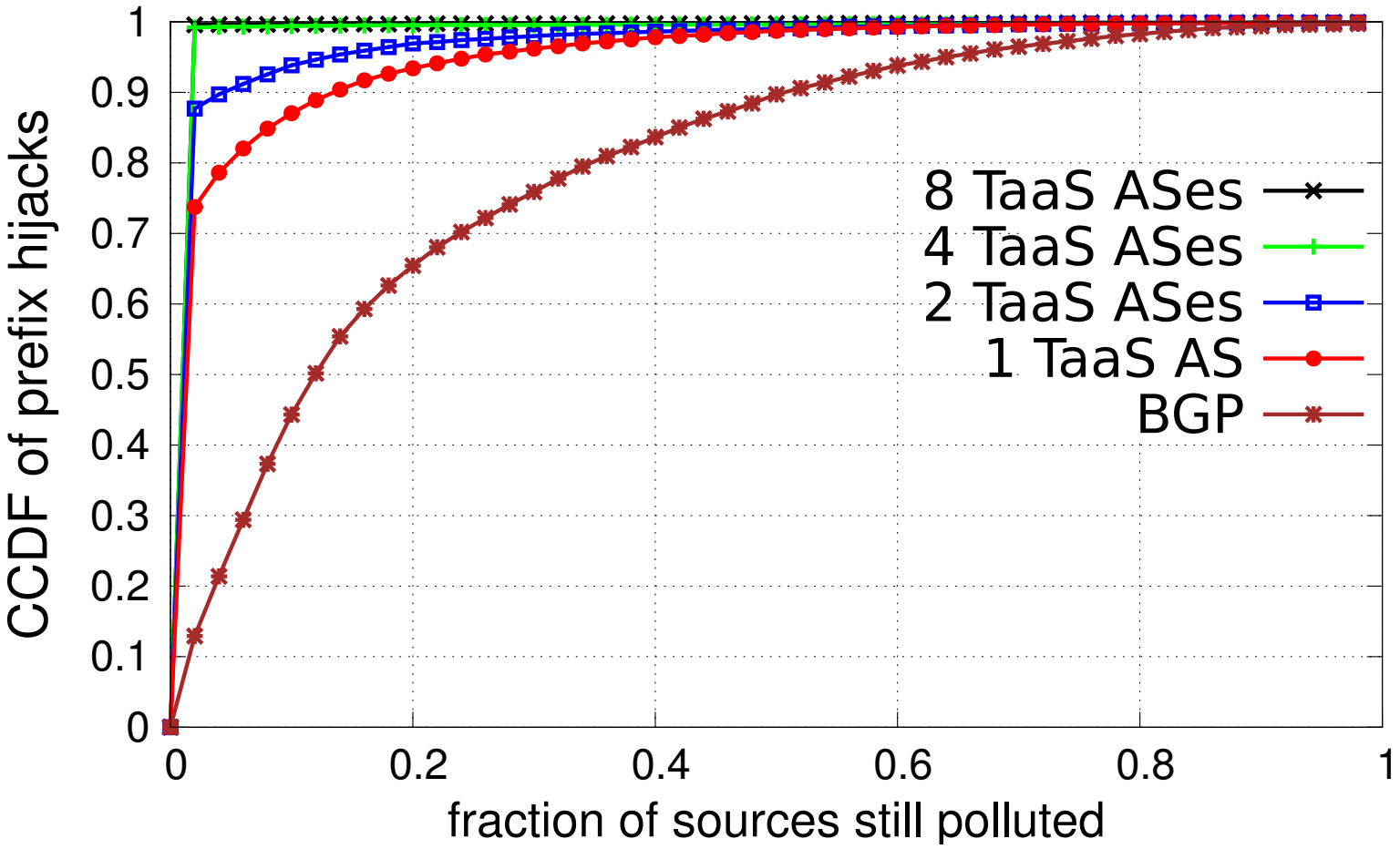
Agile and Reliable ISPs

- Software processing at the edge, hardware switching in the core
  - Software packet processing: 10 Gbps per core on modern servers (min-sized packets)
  - Total Internet traffic: 100 Tbps
  - => Need 10K cores to process *every* packet in the public Internet
- Fault tolerant control plane layer
  - Setup/teardown circuits, install filters, …

Incremental Adoption



Resilience to Prefix Hijacking



Networking as a Service

Network operators offer networking services to remote customers

- Transit, packet swizzling, and packet processing
- ISPs only promise what they can *directly* provide

Potential for much better security, reliability, worst case performance, incremental adoption