Lecture 20: Routing again

COMP 332, Spring 2023

Victoria Manfredi

WESLEYAN UNIVERSITY

Acknowledgements: materials adapted from Computer Networking: A Top Down Approach 7th edition: ©1996-2016, J.F Kurose and K.W. Ross, All Rights Reserved as well as from slides by Abraham Matta at Boston University, and some material from Computer Networks by Tannenbaum and Wetherall.

Today

1. Announcements
   - Homework 7 written due Wednesday, April 19 at 11:59p
   - Homework 7 coding due Wednesday, April 26 at 11:59p
   - Homework 8 due Wednesday, April 3 at 11:59p (no coding)
   - Homework 9 due Wednesday, May 10 at 11:59p (no written)

1. ICMP and bit-wise operations

2. Learning-based routing

3. Internet routing
   - intra-AS routing
   - inter-AS routing

INTERNET CONTROL MESSAGE PROTOCOL

OVERVIEW

Internet Control Message Protocol (ICMP)

Used by hosts & routers to communicate network-level information

- error reporting
  - unreachable host, network, port, protocol
- echo request/reply
  - used by ping)
- network-layer above IP
  - ICMP msgs carried in IP pkts

ICMP message

- type, code plus first 8 bytes of IP pkt causing error

Type	Code	Description
0	0	echo reply (ping)
3	0	dest. network unreachable
3	1	dest host unreachable
3	2	dest protocol unreachable
3	3	dest port unreachable
3	6	dest network unknown
3	7	dest host unknown
4	0	source quench (congestion control - not used)
8	0	echo request (ping)
9	0	route advertisement
10	0	router discovery
11	0	TTL expired
12	0	bad IP header

Traceroute and ICMP

Source sends series of segments or packets to destination

- first set has TTL =1
- second set has TTL=2, etc.
- unlikely port number

When *n*th set arrives to nth router

- router discards and sends source ICMP message (type 11, code 0)
- ICMP message includes name of router & IP address

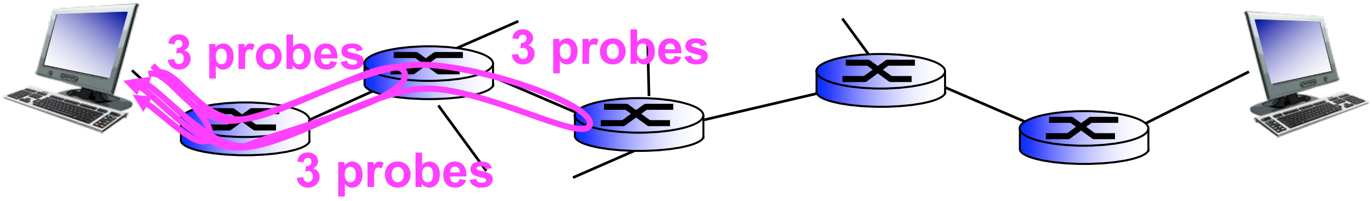
When ICMP msg arrives

- source records RTTs

Stopping criteria

TCP segment or UDP datagram eventually arrives at dst host

- dst returns ICMP "port unreachable" message
- source stops



Q: why can traceroute work with segments, datagrams, or packets?

ICMP traceroute

We're generating an ICMP echo request

Intermediate routers

- respond with ICMP TTL expired

Final destination

- responds with ICMP echo reply

NETWORK PROGRAMMING

BIT-WISE OPERATIONS IN PYTHON

Bit-wise operations on variables

x << y

- returns x with bits shifted to left by y places
    - new bits on right-hand-side are zeros
    - same as multiplying x by 2^y

x >> y

- returns x with bits shifted to right by y places
    - same as dividing x by 2^y

x & y

- does a bitwise and
    - each bit of output is 1 if corresponding bit of x AND of y is 1, otherwise 0

~ x

- returns complement of x
    - number you get by switching each 1 for 0 and each 0 for 1

What to use for?

- use to pack ip_version and ip header length into 8 bits

https://wiki.python.org/moin/BitwiseOperators
https://www.tutorialspoint.com/python3/bitwise_operators_example.htm

Control Plane

LINK STATE VS. DISTANCE VECTOR ROUTING

Message complexity

n nodes
E links

Link state

- O(nE) messages sent
  - every node floods its link state message out over every link in network to reach every node
- smaller messages
  - message size depends on the number of neighbors a node has
  - any link change requires a broadcast

Distance vector

- # of messages depends on convergence time which varies
  - nodes only exchange messages between neighbors
- larger routing update messages
  - message size is proportional to the number of nodes in the network
  - if link changes don't affect shortest path, no message exchange

Speed of convergence

n nodes
E links

Link state

- $\Sigma_{i=1}^{n-1} i = n(n+1)/2 = O(n^2)$
  - search through n-1 nodes to find min, recompute routes
  - search through n-2 nodes to find min, recompute routes
  - …
- converges quickly but may have oscillations
  - route computation is centralized
  - a node stores a complete view of the network

Distance vector

- slow to converge and convergence time varies
  - route computation is distributed
- may be routing loops, count-to-infinity problem

What happens if router malfunctions?

n nodes
E links

Link state

- node can advertise incorrect link cost
- each node computes only its own table

Distance vector

- DV node can advertise incorrect path cost
- each node's DV used by others: errors propagate through network

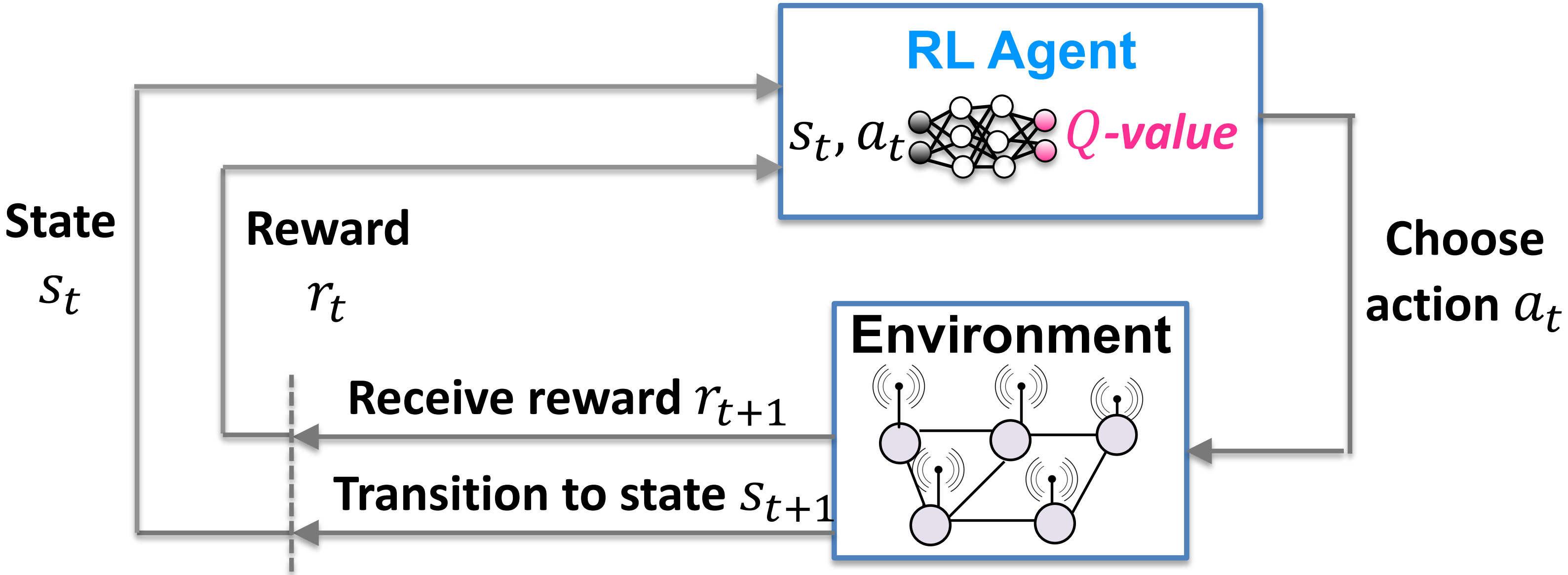
Both have strengths and weaknesses.
One or the other is used in almost every network

Control Plane

OTHER APPROACHES TO MAKE ROUTING DECISIONS

Reinforcement learning to make routing decisions

RL agent learns to choose actions to maximize expected future reward



Define RL agent for routing. Requires us to define *states*, *actions*, and *rewards* useful for routing

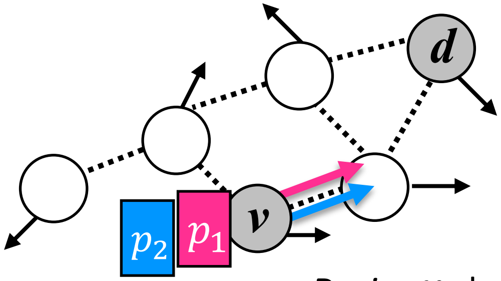
Given trained model install at routers using Software-Defined Networking

Key ideas

1. Packet-centric decisions

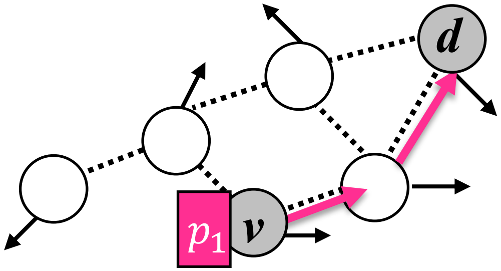
2. Relational features

Problem: Normally a device chooses a packet's next hop … but a device's state doesn't track what happens to the packet



Device v chooses next hop for each outgoing packet

Solution: Use packet agents to simplify s, a, s', r experience sequence and define reward



Packet p_1 chooses its next hop at each device

Key ideas

Problem: How to define generalizable states and actions?

Solution: Use relational features that model the relationship between devices instead of describing a specific device

1. Packet-centric decisions

2. Relational features

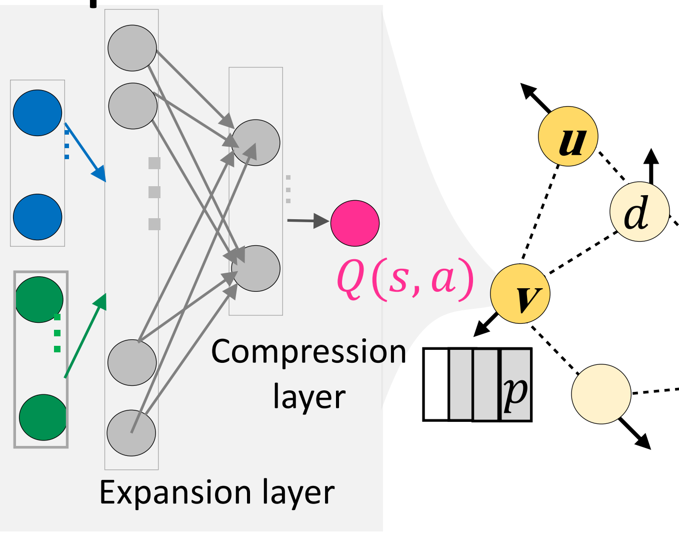
For packet p at device v with 1-hop neighbors $Nbr(v)$

- **Packet features $f_{packet}(p)$: p's TTL**
- **Device features $f_{device}(v,d)$: v's** queue length, queue length for packets to d, node degree, node density
- **Neighbor features, $f_{neighbor}(Nbr(v),p,t)$:** summarize varying # of neighbors using min, mean, max of $f_{device}(Nbr(v),p,t)$
- **Path features $f_{path}(v,d)$:** distance or delay from v to d

State features $f_s(s)$

Action features $f_a(a)$

Deep Neural Network



Packet p separately considers each action u

- **device features** $f_{device}(u,d)$,
- **neighborhood features** $f_{nbrhood}(u,d)$,
- **device features** $f_{device}(u,d)$
- **context features** $f_{context}(p,u)$ which indicate whether p has recently visited u

Internet Routing

OVERVIEW

From graph algorithms to routing protocols

Need to address Internet reality

1. Internet is network of networks
   - hierarchical structure
   - routers not all identical
     - some routers connect different networks together
   - each network admin may want to control routing in its own network

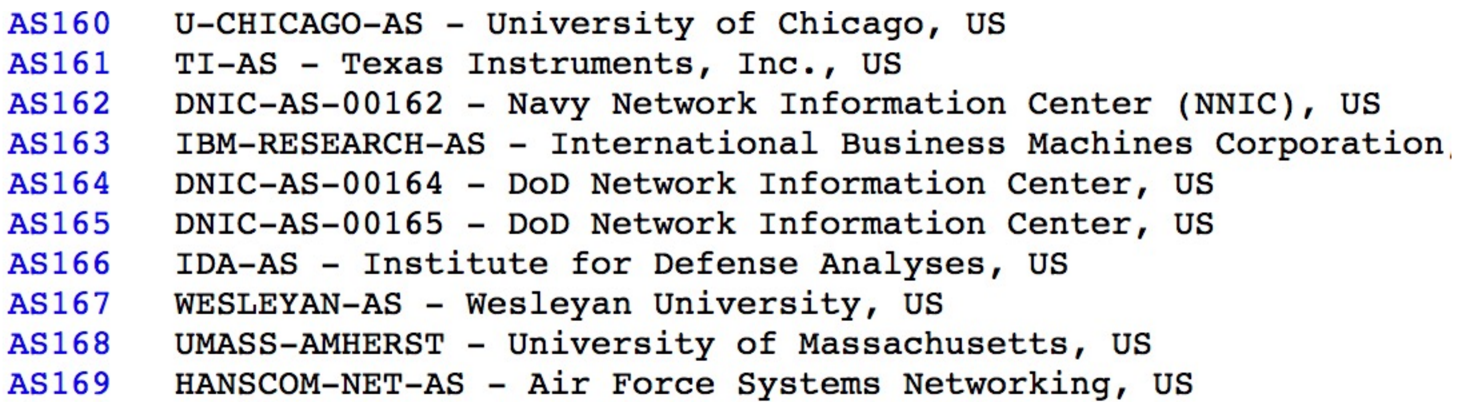
2. Scalability with billions of destinations
   - don't all fit in one routing table
   - can't exchange routing tables this big
     - would use all link capacity

Scalable routing on the Internet

Aggregate routers into regions called Autonomous Systems

Autonomous Systems (AS)

- aka domain
- network under single administrative control
  - company, university, ISP, …
- 30,000+ ASes: AT&T, IBM, Wesleyan …
- each AS has a unique 16-bit AS #
  - Wesleyan: AS167
  - BBN: used to be AS1: was first org to get AS # then L3 later acquired



AS160	U-CHICAGO-AS - University of Chicago, US
AS161	TI-AS - Texas Instruments, Inc., US
AS162	DNIC-AS-00162 - Navy Network Information Center (NNIC), US
AS163	IBM-RESEARCH-AS - International Business Machines Corporation,
AS164	DNIC-AS-00164 - DoD Network Information Center, US
AS165	DNIC-AS-00165 - DoD Network Information Center, US
AS166	IDA-AS - Institute for Defense Analyses, US
AS167	WESLEYAN-AS - Wesleyan University, US
AS168	UMASS-AMHERST - University of Massachusetts, US
AS169	HANSCOM-NET-AS - Air Force Systems Networking, US

Hierarchical routing

2-level route propagation hierarchy

1. intra AS routing protocol between routers in same AS
   - aka intra domain routing protocol
   - aka interior gateway protocol
   - each AS selects its own

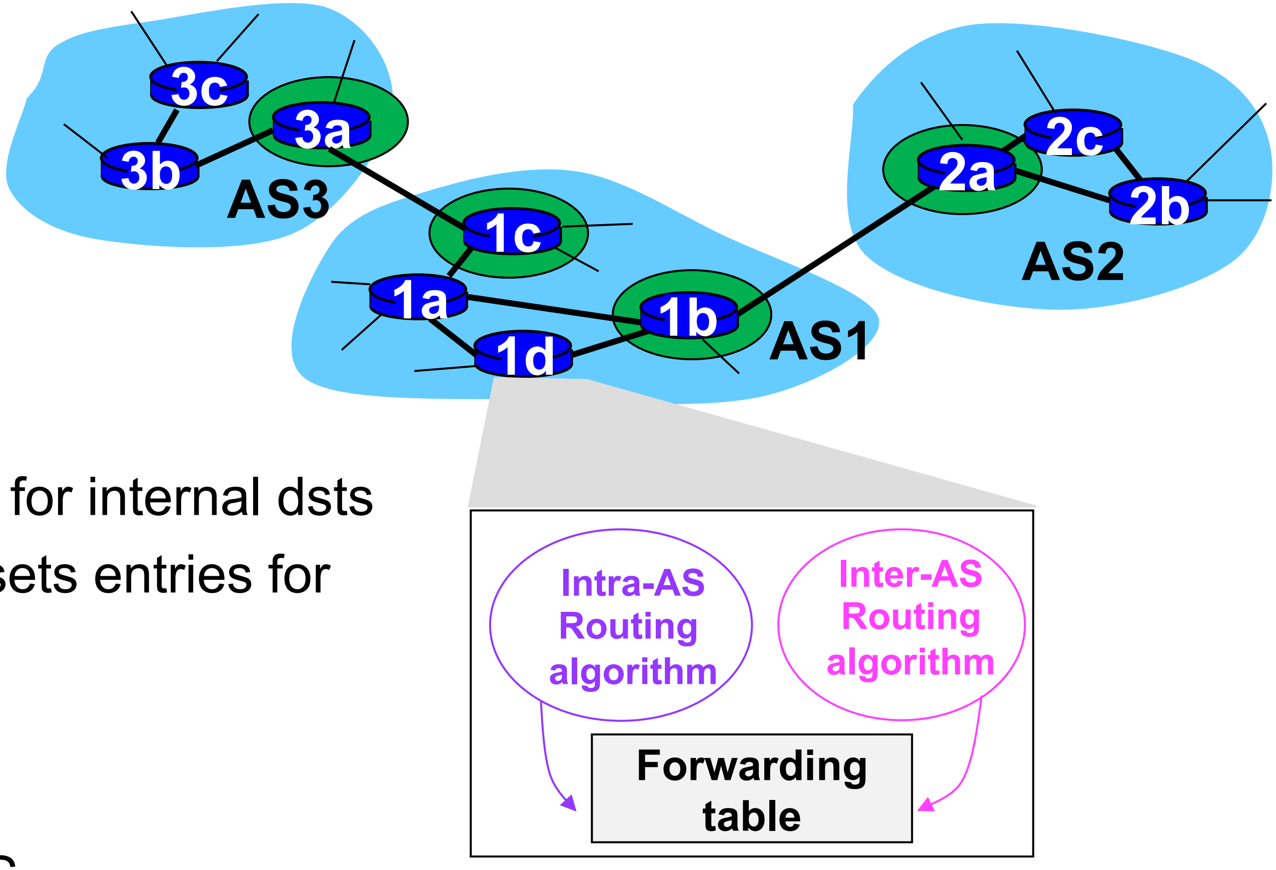
Focus is performance

2. inter AS routing protocol between gateway routers in different ASes
   - aka inter domain routing protocol
   - aka exterior gateway protocol
   - Internet-wide standard

Policy may dominate performance

Q: Can routers in different ASes run different intra AS routing protocol?

Hierarchical routing



Intra-AS Routing algorithm

Inter-AS Routing algorithm

Forwarding table

Forwarding table

- intra-AS sets entries for internal dsts
- inter-AS & intra-AS sets entries for external dsts

Gateway router

- at edge of its own AS
- direct link to router in another AS
- perform inter-AS as well as intra-AS routing
- distributes results of inter-AS routing to other routers in AS

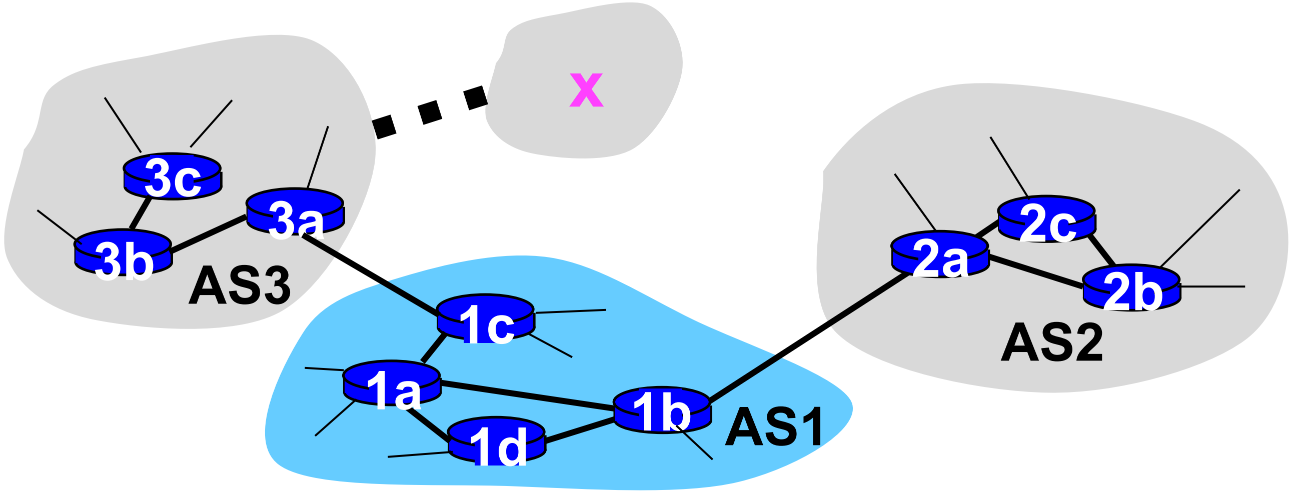
Example: set forwarding table in router 1d

AS1 learns (from inter-AS protocol)

- subnet x is reachable via AS3 (gateway 1c) but not via AS2

Router 1d learns (from intra-AS protocol)

- that its interface y is on least cost path to 1c.
- installs forwarding table entry (x,y)



Q: What if multiple ASes can be used to reach x?

Internet ROUTING

INTRA-AS ROUTING

Most common intra-AS routing protocols

RIP

- Routing Information Protocol
- distance vector protocol

(E)IGRP

- (Enhanced) Interior Gateway Routing Protocol
- Cisco proprietary for decades, until 2016
- distance vector protocol

IS-IS

- Intermediate System to Intermediate System
- link state protocol

OSPF

- Open Shortest Path First
- link state protocol

Open Shortest Path First (OSPF)

Open

- i.e., publicly available

Link-state algorithm

1. Each router floods its link state to all other routers in AS
   - msgs carried directly over IP, authentication possible
   - supports unicast (1src –1dst) and multicast (1src - multiple dst)
2. Each router builds topology map
3. Route computation using Dijkstra’s
   - can have multiple paths with same cost
     - traffic can go over different paths
   - can have different costs per link depending on type of service
     - e.g., satellite link cost: low for best effort, high for real time

Internet ROUTING

INTER-AS ROUTING

Inter-AS routing

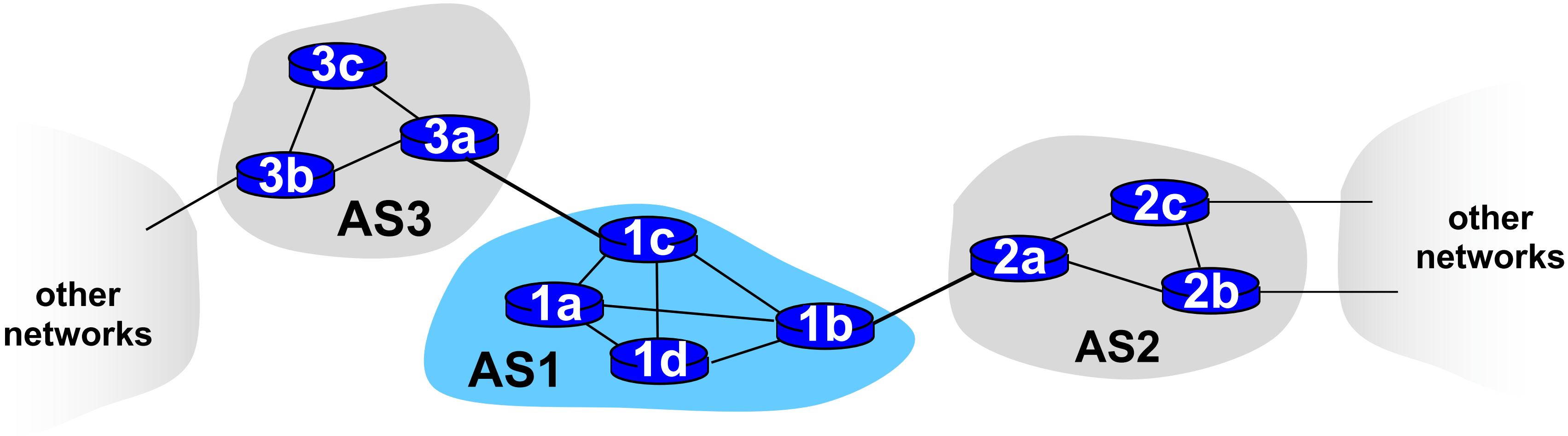
Router in AS1 receives pkt destined outside of AS1

- router forwards pkt to gateway router, but which one?

AS1 must learn which dsts reachable through neighbor ASes

- propagate this reachability info to all routers in AS1

⇒ job of inter-AS routing!



Border Gateway Protocol (BGP)

Defacto inter-domain routing protocol

- allows subnet to advertise its existence to rest of Internet
- path vector protocol

BGP provides way to find good routes to other networks

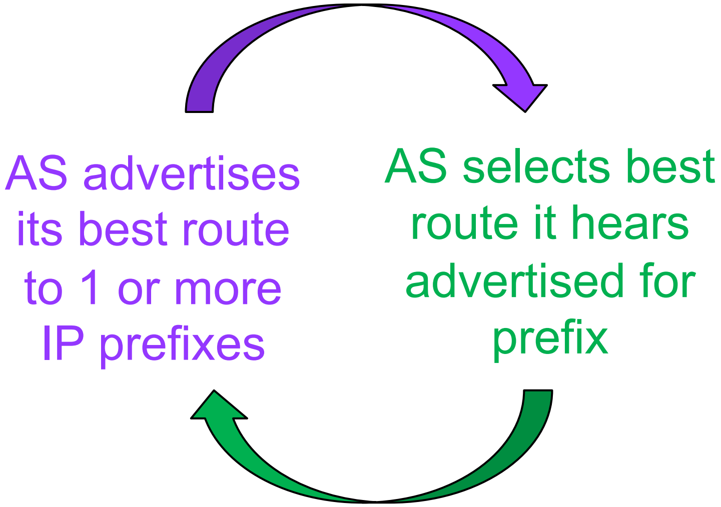
- based on reachability info and policy

Q: why must all ASes use same inter-AS protocol

How BGP works

Similarities with distance vector

- per dst route info advertised
- no global sharing of network topology
- iterative distributed convergence



AS advertises its best route to 1 or more IP prefixes

AS selects best route it hears advertised for prefix

Differences from distance vector

- selects best route based on policy not min cost
- path vector routing
  - advertises entire path for each dst rather than cost
    - allows policies based on full path
    - avoids loop: if your AS is in path then discard
  - selective route advertisements
    - choose not to advertise route to dst for policy reasons
    - aggregate routes for scalability: e.g., a.b.*.* and a.c.*.* become a.*.*.*

Policy-shaped route selection

Political, economic, security considerations

Shaped by business relationships between ASes

- AS1 is customer of AS2 (AS 1 pays AS2)
- AS1 is provider of AS 2
- AS1 is peer of AS 2 (peers don’t pay each other to exchange traffic)

E.g.,

- don’t want to carry commercial traffic on university network
- traffic to apple shouldn’t transit through google
- pentagon traffic shouldn’t transit through Iraq

Why BGP is so complicated!

Why different intra- vs. inter-AS routing?

Policy

- inter-AS
  - admin wants control over how its traffic routed, who routes through its net
- intra-AS
  - single admin, so no policy decisions needed

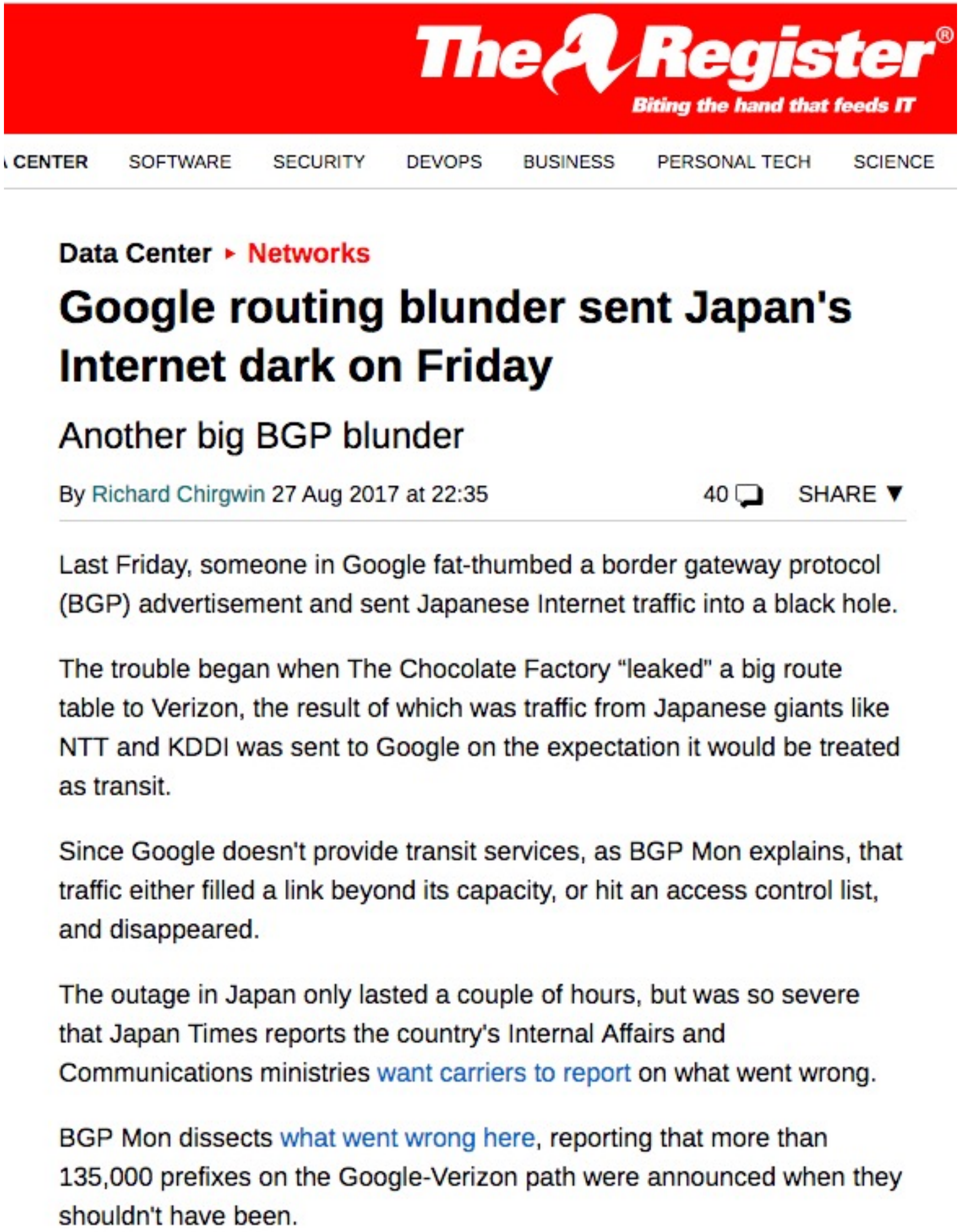
Scale

- hierarchical routing saves table size, reduced update traffic

Performance

- inter-AS
  - policy may dominate over performance
- intra-AS
  - can focus on performance

Routing blackholes



Data Center ▸ Networks

Google routing blunder sent Japan's Internet dark on Friday

Another big BGP blunder

By Richard Chirgwin 27 Aug 2017 at 22:35

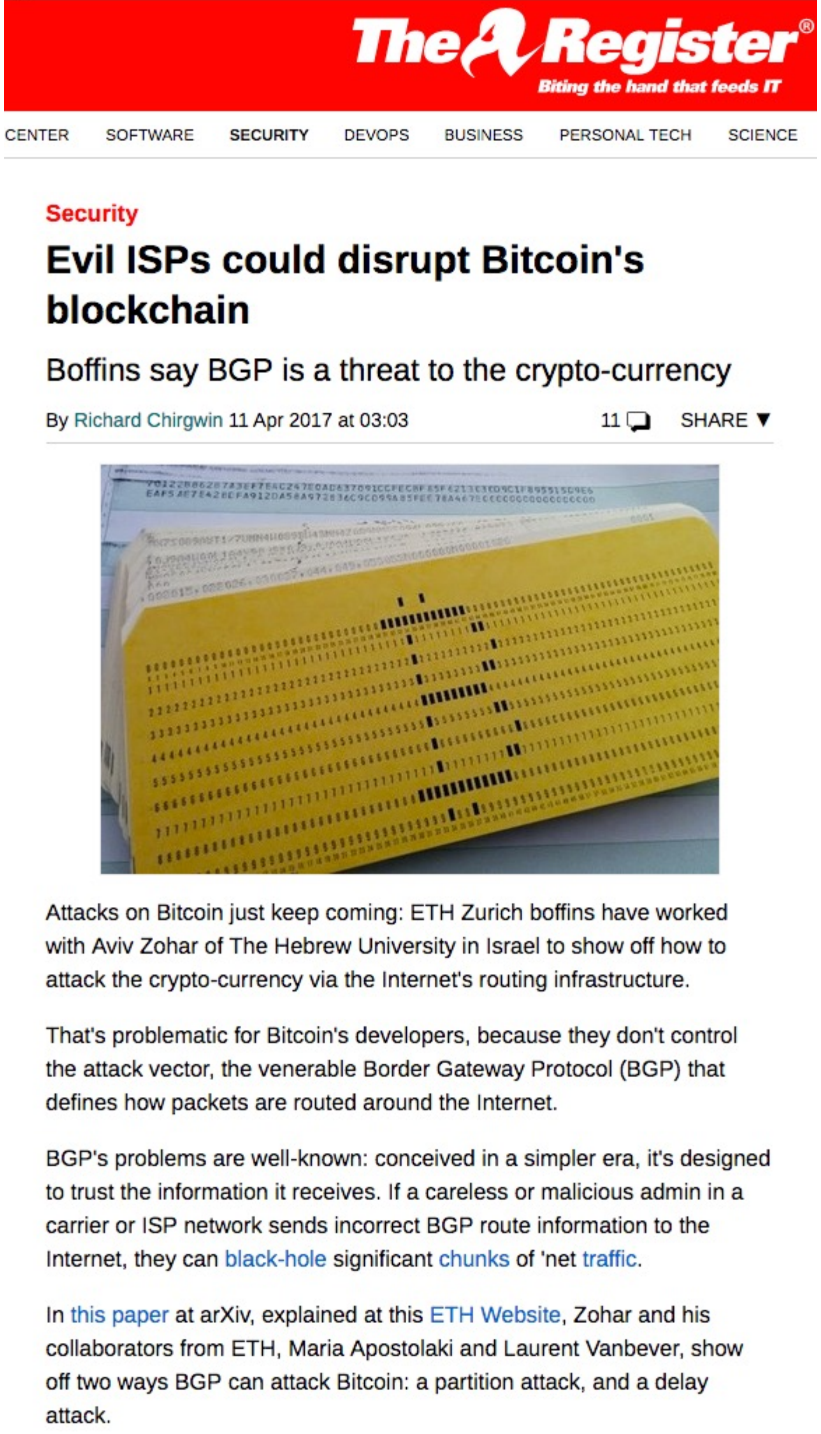
Last Friday, someone in Google fat-thumbed a border gateway protocol (BGP) advertisement and sent Japanese Internet traffic into a black hole.

The trouble began when The Chocolate Factory "leaked" a big route table to Verizon, the result of which was traffic from Japanese giants like NTT and KDDI was sent to Google on the expectation it would be treated as transit.

Since Google doesn't provide transit services, as BGP Mon explains, that traffic either filled a link beyond its capacity, or hit an access control list, and disappeared.

The outage in Japan only lasted a couple of hours, but was so severe that Japan Times reports the country's Internal Affairs and Communications ministries want carriers to report on what went wrong.

BGP Mon dissects what went wrong here, reporting that more than 135,000 prefixes on the Google-Verizon path were announced when they shouldn't have been.



Security

Evil ISPs could disrupt Bitcoin's blockchain

Boffins say BGP is a threat to the crypto-currency

By Richard Chirgwin 11 Apr 2017 at 03:03

Attacks on Bitcoin just keep coming: ETH Zurich boffins have worked with Aviv Zohar of The Hebrew University in Israel to show off how to attack the crypto-currency via the Internet's routing infrastructure.

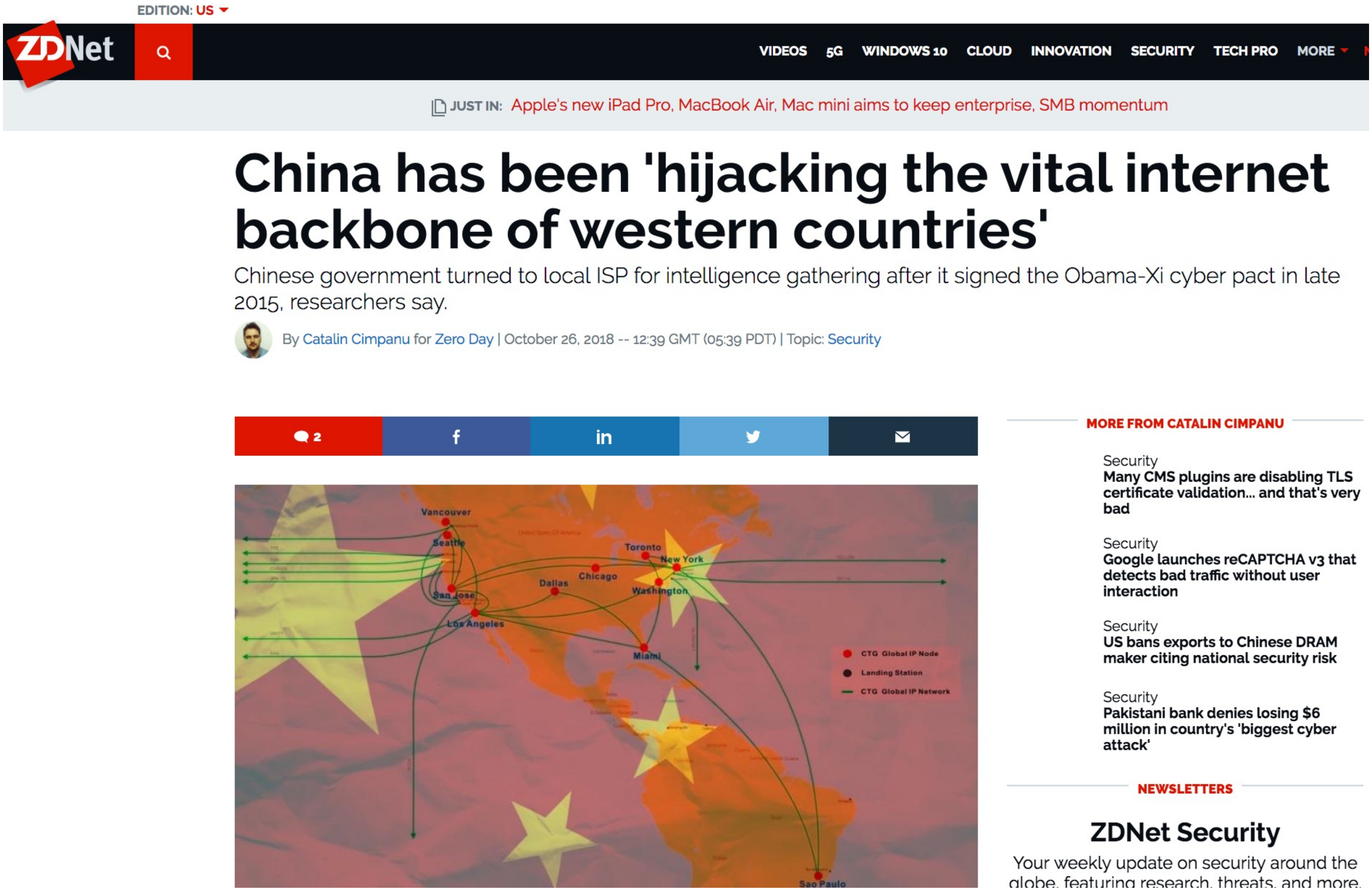
That's problematic for Bitcoin's developers, because they don't control the attack vector, the venerable Border Gateway Protocol (BGP) that defines how packets are routed around the Internet.

BGP's problems are well-known: conceived in a simpler era, it's designed to trust the information it receives. If a careless or malicious admin in a carrier or ISP network sends incorrect BGP route information to the Internet, they can black-hole significant chunks of 'net traffic.

In this paper at arXiv, explained at this ETH Website, Zohar and his collaborators from ETH, Maria Apostolaki and Laurent Vanbever, show off two ways BGP can attack Bitcoin: a partition attack, and a delay attack.

BGP hijacking

https://www.zdnet.com/article/china-has-been-hijacking-the-vital-internet-backbone-of-western-countries/



EDITION: US

ZDNet

VIDEOS 5G WINDOWS 10 CLOUD INNOVATION SECURITY TECH PRO MORE

JUST IN: Apple's new iPad Pro, MacBook Air, Mac mini aims to keep enterprise, SMB momentum

China has been 'hijacking the vital internet backbone of western countries'

Chinese government turned to local ISP for intelligence gathering after it signed the Obama-Xi cyber pact in late 2015, researchers say.

By Catalin Cimpanu for Zero Day | October 26, 2018 -- 12:39 GMT (05:39 PDT) | Topic: Security

MORE FROM CATALIN CIMPANU

Security
Many CMS plugins are disabling TLS certificate validation... and that's very bad

Security
Google launches reCAPTCHA v3 that detects bad traffic without user interaction

Security
US bans exports to Chinese DRAM maker citing national security risk

Security
Pakistani bank denies losing $6 million in country's 'biggest cyber attack'

NEWSLETTERS

ZDNet Security

Your weekly update on security around the globe, featuring research, threats, and more.