ECE 435 – Network Engineering
Lecture 15

Vince Weaver

http://web.eece.maine.edu/~vweaver

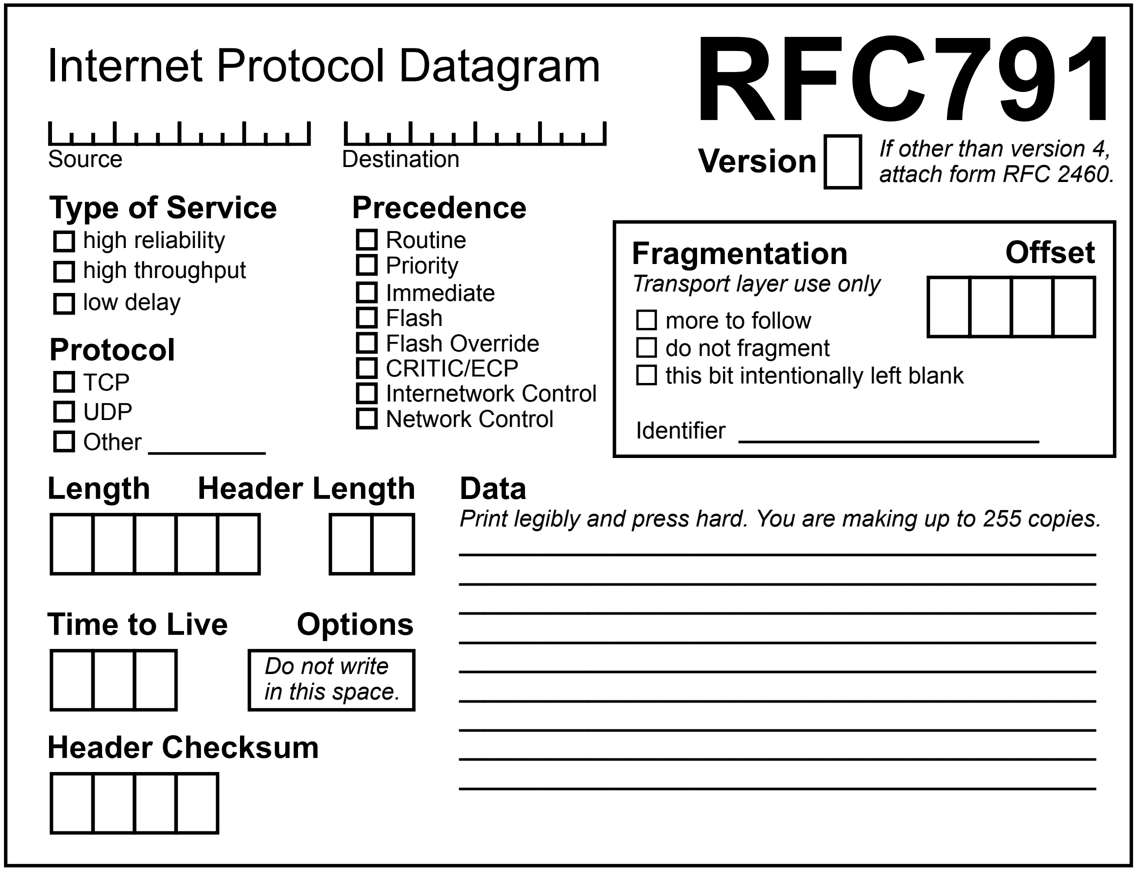
vincent.weaver@maine.edu

25 March 2021

Announcements

- Note, this lecture has no video recorded due to problems with UMaine zoom authentication at class start time
- HW#6 graded
- Don't forget HW#7
- Project Topics due

RFC791 Post-it-Note



Internet Protocol Datagram

RFC791

Source

Destination

Version □ *If other than version 4, attach form RFC 2460.*

Type of Service

□ high reliability
□ high throughput
□ low delay

Protocol

□ TCP
□ UDP
□ Other ________

Precedence

□ Routine
□ Priority
□ Immediate
□ Flash
□ Flash Override
□ CRITIC/ECP
□ Internetwork Control
□ Network Control

Fragmentation **Offset**

Transport layer use only

□ more to follow
□ do not fragment
□ this bit intentionally left blank

Identifier ____________________

Length **Header Length**

Time to Live **Options**

Do not write in this space.

Header Checksum

Data

Print legibly and press hard. You are making up to 255 copies.

for more info, check IPv4 specifications at http://www.ietf.org/rfc/rfc0791.txt

HW#6 Review

- Header:

```
0x000e:  4500 = version(4), header length(5)=20 bytes
                ToS=0
0x0010:  0038 = packet length (56 bytes)
0x0012:  572a = identifier
0x0014:  4000 = fragment 0100 0000 0000 0000 =
                do not fragment, offset 0
0x0016:  40 =   TTL = 64
0x0017:  06 =   Upper layer protocol (6=TCP)
0x0018:  69cc = checksum
0x001a:  c0a80833 = source IP 192.168.8.51
0x001e:  826f2e7f = dest IP 130.111.46.127
```

- Valid IPs

  - 123.267.67.44 = N
  - 8.8.8.8 = Y
  - 3232237569 = 192.168.8.1
  - 0xc0a80801 = 192.168.8.1
- A class-A allocation is roughly $2^{24}/2^{32}$ which is 0.39%
- 192.168.13.0/24. subnet 255.255.255.0, lowest ip 192.168.13.1, highest 192.168.13.254 (traditionally can't use the host values with all 0 or 1s (so .0 and .255 on a /24). There's a push to allow .0 on Linux
- First hop not local (how to tell?) goes to router

Otherwise go direct (can you go direct? how).

- Ping google. 1e100.net?
- Traceroute. Some routers block?

  Neville hall

Some Midterm Notes

- http2 client with telnet

  Can't really unless you have fancy telnet client

  Can connect with "openssl" which will handle TSL connection

  Can't just do GET / though. You have to send a series of binary packets to INIT and request frames, headers and frame data comes back as a set of compressed interleaved binary data
- You might hear it's backwards compat but not

necessarily, mostly probably to convince people to switch at time?

So web-servers can drop back to 1.1 but they might do it by just passing back a message saying not supported

Remember Project Ideas Due

- Can work in groups
- Do something interesting network related.
- Can use any operating system and written in any language (asm, C, python, C++, Java, etc.)
- Coding, benchmarking
- Past projects: network games, firewall config, network attached storage, mesh networks
- Will be a final writeup, and then a 10 minute presentation and demo in front of the class during last week of classes.

Label Switching / MPLS

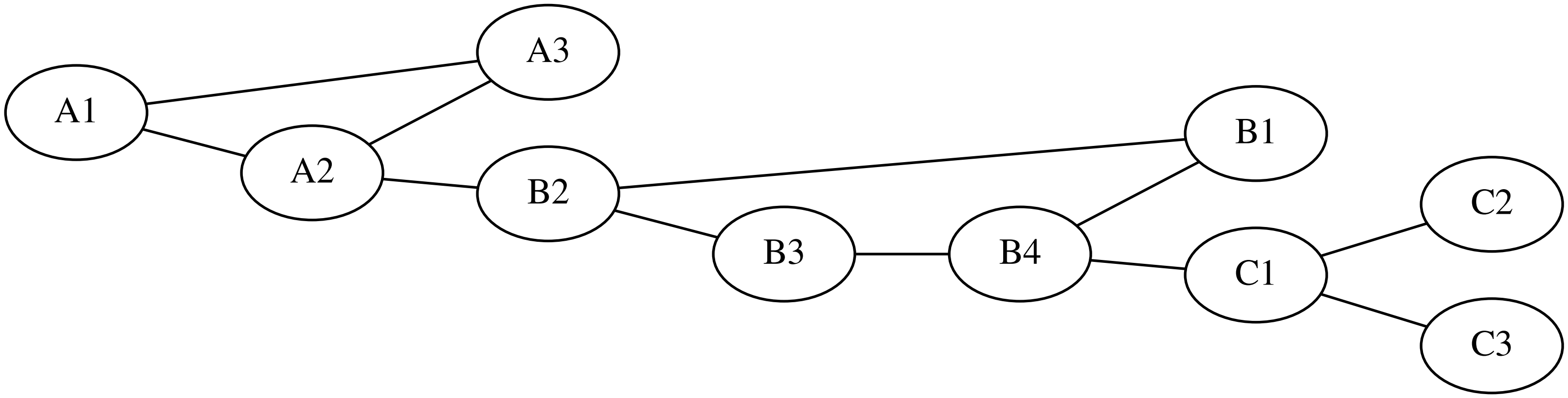
- MultiProtocol Label Switching
- Almost connection-oriented way to move packets across networks
- Needs to add MPLS header

Hierarchical Routing

- Would you want to have all routers in network on flat network? Crazy routing table
- Split into a hierarchy
- Autonomous System (AS) – a network under control of one group.

  Inside an AS, interior routing, between is exterior routing.
- Systems under same command (same ISP) use intra-domain routing protocol, or interior gateway protocol (IGP)

- Border routers connect to border routers of others
- Inter-domain routing, EGP (exterior gateway protocol)
- types
  - Stub AS – like ISP with customers, one gateway to internet
  - Multihomed AS – multiple gateways (why?) redundancy. traffic generally doesn't flow through
  - Transit AS – traffic can flow through network



- Packet A1 - A3 internal A1 - B2 goes to border router and across, then local A1 - C2 goes to border router to B network, across local to B/C border, then finally to C
- If flat network, need to know 10 machines in routing table
- In hierarchical only need to communicate to 2-3 other routers, find way to border router

Intra-Domain Routing / Interior Gateway Protocols

RIP (Routing Information Protocol)

- Used by ARPAnet until 1979(?)
- by Xerox, included in BSD, routed RFC 2543
- distance vector routing, with hop count, max 15 hops m
  - RIP advertisements over UDP port 52
  - Send advertisement every 30s, or when changes
  - Only sends to neighbors
- Routing table: dest, next hop, distance
- Algorithm
  - Get table update

- ◦ Increment all hops by 1 (you're one hop away)
  - ◦ Go down list.

    If route not in table, add it

    If route there, and next hop same (but cost diff), replace it as this is new info

    If route there but cost less, replace it
- On power up, comes up with hard-coded routes and values of 1 and no next-hop. Can send packet to request immediate update from neighbors.
- Packet description
- Timers

- Periodic timer, technically 30s, reality randomized between 25 and 35 (why?)
- Expiration timer – 180s. If no update in this time, problem, hop count set to 16 (unreachable)
- Garbage collection – 120s – once unreachable, advertise it as such for a while before removing so others notice

- Issues
  - Slow Convergence – a change in routing tables takes 30s per hop to propagate through
Part of why limited to 15 hops

- ○ Instability – packets can be caught in loops. Ways to fix:

  Triggered update – send update info immediately, not wait 30s

  Split Horizon – if a router sends you update info, don't send this back to it Poison Reverse – like split horizon, but when send back, mark as 16 the routes received from that interface.
- There was a RIP2

OSPF (Open Shortest Path First)

- successor to RIP. RFC2328 (5340 for IPv6)
- Idea of Areas inside of an AS. Split up into areas.
- Each area connected by backbone router
- Router on two areas is area border router
- Link-state Routing
  - ○ State is flooded: when a change happens (and only then) it sends this state to all neighbors, which send to all neighbors, until the whole network receives it
  - ○ each router uses Dijkstra to find least cost for self,

builds table

- Types of link
  - Point-to-point – routers directly connected
  - Transient Link – network with several routers can be simplified?
  - Stub Link – a network connected to only one router
  - Virtual Link – a path between two routers that traverses other routers
- load balancing – supports equal-cost multipath routing (can equally use equal cost routes)
- supports CIDR routing

- support available for multicast
- 8-byte password for authentication
- supports hierarchical
- example? complex!

Inter-Domain Routing

- Can be complicated.

- Say company with network, and two connections to outside X,Y. Don't want to send packets out and back even if it looks like lower cost.

- Also don't want to transit packets between X and Y for outsiders. Policy.

BGP (border gateway protocol)

- Intro in 1989, four versions – BGP4 RFC 4271
- Uses TCP (reliable) port 179.
- Works for both IPv4 and IPv6 (the latter as an extension)
- Uses path vector rather than distance vector
  - ◦ full path, not just next-hop
  - ◦ exchanges info with neighbor, but includes complete path info to avoid looping.
  - ◦ Each AS has unique number, so if it sees itself in the path knows there is a loop.

  - Policy routing – can also reject new route based on policy
- Four types of messages – open, update, keepalive, notification
- Whole table not passed around (Due to size), only updates
- Due to size of internet, uses distance vector over link state.
- interior and exterior BGP
- iBGP makes sure that the setups for multiple gateway routers are kept synchronized

- eBGP used to talk between other exterior routers at peers.
- Keeps track of all feasible paths, but only advertises the “best” one

Routing table Size

- Example. Full BGP of internet backbone router might have more than 300,000 entries (2010) now over 700,000.
- `http://bgp.potaroo.net/`
- Some routers had limit of 512k so on August 12 2014 part of internet went down when crossed the border.
- Ipv6 currently only around 20k

Peering

- How companies agree to connect their networks together. There's not really a master connection, but instead companies agree to have routers talk to each other via BGP.

- Transit – pay money to pass through network.

- Peering – In many cases no money changes hands. Why? Well if you have a lot of users, but no content, people won't stay with you. Same if you have content but no access to users. Averages out and is mutually beneficial.

  - Increased redundancy
  - Increased capacity
  - Increased routing control
  - Improved performance
  - Fame (high-tier network)
  - Ease of requesting aid (?)

- Customer – you buy an internet connection

- Peering locations, often in large data centers.
  At one point there were 4 major ones (Metropolitan Area

Exchange) MAE-East (Virginia) [in basement of parking garage, at one point half of internet went through here], Chicago, NY, SF. All defunct now

- Exchange map: https://www.internetexchangemap.com/
- Depeering – if you think you aren't getting a good deal, break up. Some situations there is a fight, a hope that the customers lose enough performance will have to repeer.
- Related – net neutrality

- Tiers – Tier 1 network is one that can reach rest of internet without paying for transit; Tier 2 peers with some but purchases for other; Tier 3 only purchases

- IPv6 Peering issues – see `https://www.theregister.co.uk/2018/08/28/ipv6_peering_squabbles/`

Routing Security Issues

- Problems – routing black hole, use BGP to send addresses intentionally to 0.0.0.0 and get dropped. BGP will propagate

- router update mistakes can accidentally blackhole parts of the internet

- In 2008 Pakistan was trying to blackhole Youtube and accidentally announced to world via BGP and took it down world wide

- https://arstechnica.com/information-technology/
  2019/06/bgp-mishap-sends-european-mobile-traffi

- https://www.theregister.co.uk/2019/06/24/verizo
  bgp_misconfiguration_cloudflare/ Verizon accidental
  routed a lot of internet through Pittsburgh Steel Mill

Implementations

- Actual Router
- Can install on your Linux machine
- Zebra was traditionally, discontinued
- Quagga
- BIRD
- OpenBGPD and OpenOSPFD

- Potentially dangerous to mess around with unless you isolate your network well

Broadcast Routing

Unicast/Multicast/Broadcast

- Unicast – send from one machine to another
- What if want to send to multiple?
  - Multi-unicast – open direct connection to each destination. Inefficient
  - Broadcast – send to *every* destination? Waste bandwidth, but also need to know all possible destinations
  - Flooding? Also too much bandwidth
  - Multi-destination routing

- Multicast Goals
  - Only send to users who want it
  - Each member only receives one copy
  - No loops
  - Path traveled should be optimal
- Spanning tree – tree with source as root and members as leaves
- Reverse-path forwarding

Why would you multicast?

- Live streams? Backups?
- Why not just multi-unicast?
  - More work on sender, many more packets sent
  - Latency between first and last packet sent

Multicast IP

- For IP, just join a class D network
- To both sender and receiver it's like sending/receiving a unicast packet
- all the hard work done by routers
- How do you join a multicast group?
- Router two tasks: group membership management, packet delivery.

Group Management

- IGMP (Internet Group Management Protocol)
  IGMPv3 RFC 3376
  query, report, leave
  querier and noquerier
  router with lowest IP is querier
  no real controls on who can join or send

Multicast Trees

- Steiner tree – NP complete, no one uses
- Heuristics, but none generate entire tree as need centralized and global knowledge
- DVMRP (Distance-Vector Routing Protocol)
  original protocol, MBONE (tell story)
- Reverse path Forwarding – flood packet out all interfaces except one it came in on. Can have loops; drop dupes.

Then forward on the one that has traveled the shortest path.
Is running the routing table backwards

- Reverse path Broadcast – avoid getting multiple packets
- Protocol Independent Multicast (PIM)
  DVRMP not scalable for multicast groups with sparse members
- MOSPF
- CBT

Local Network Broadcasts

- 224.0.0.0/4 was reserved from Class D for multicast
- 224.0.0.0 to 224.0.0.255 for local network broadcasts
- Things like cluster stats (ganglia, can never get to work?)
- Routing info protocol (RPIv2) OSPF, mDNS, etc.

mDNS

- Multicast local network hostname resolution
- Bonjour (mac), Avahi (Linux)
- Multicast to 224.0.0.251 (ipv4) or ff02::fb (ipv6)
- Issue if two machines have same name
- Broadcast name as connect to network, all devices on local net subscribe to broadcast at that address

Other types of Routing

- Mobile – what do you do when machines can come and go?
  have a "home" location. Packets go there. When you get on network, update with actual location. Network gets packets at home location, encapsulates and sends to actual location

- Ad Hoc Routing
  Bunch of machines in an area, routers and devices can come or go more or less randomly.

route discovery

- Peer to Peer File Sharing
  - Centralized server? Napster? Easy to take down.
  - Want Distributed, no central control.
  - Flooding: connect to one other connected node. Floods requests (sort of like broadcast) until it finds who has file, then direct connect to transfer.

  - distributed hash table

- Secret routing

TOR / The onion Router

Packet encrypted multiple times, in layers. Randomly sent to next machine which decrypts that layer, passed on

At end comes out random “exit node” and drops onto regular internet