Enabling Data Center SDN with Stateless Source Routing

Nathan Farrington

Semi-Automatic Ground Environment (SAGE)

- 1958 to 1984
- Provided unified view across North America
- Largest physical computer ever built
- Over 20 distributed computers
- Many firsts for computer science, including both *networking* and *software*

http://www.computerhistory.org/revolution/networking/19/399/2161

Networking and software grew up together.

SDN is not new; it's all we have.

Routing methodologies used today

- IP routing
  - By far the most popular methodology used in mega data centers
  - Extremely robust when combined with BGP for link up/down notification
  - Requires very careful address assignment; difficult for commodity hardware to support large routing tables needed for mega data centers
- Ethernet bridging
  - Since 2010, has not been used for large clusters due to scalability problems, lack of multipath support, and fragility of spanning tree protocol implementation and configuration
- MPLS label switch routing
  - Used extensively to interconnect different data centers

Commodity hardware is not keeping up

- Consider a 1,048,576-host data center
- A switch could require 1,048,576 forwarding table entries (E-W)
- More if you also want to reach the Internet (N-S traffic)
  - This is tens to hundreds of megabits of SRAM
  - High cost and power consumption; transistors used on forwarding tables are not used for other things (more ports, more features)
- Exact number of table entries depends on
  - Routing methodology (Ethernet, IP, or MPLS LSR)
  - Network topology
  - Address assignments to switches and hosts
 - Actual traffic patterns

Sourcey is stateless source routing

- Host chooses its own path through the (data center) network
  - Or alternatively an SDN controller tells a host which path to use
- Host puts the whole path in the packet header
- Each switch along the path simply obeys the next turn-by-turn instruction as written
 - No forwarding tables in the switches

Sourcey packet

- Addressed to next hop
- Ethertype is MPLS
- Dropped and recreated at each hop, just like IP

- Other MPLS headers treated as payload

| Ethernet Header | MPLS Header | MPLS Header | MPLS Header | IP Header | TCP Header | Payload |
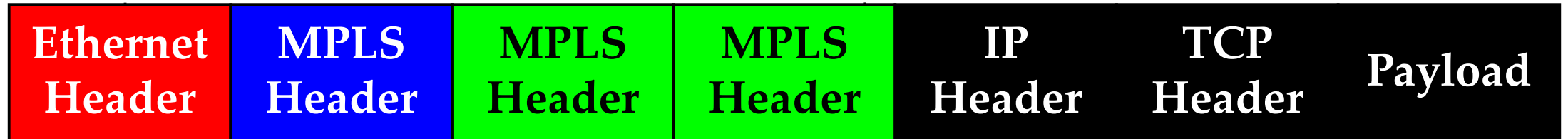
- Label value is the switch output port, e.g. 17 means output port 17
- MPLS header is dropped (popped)
- Number of MPLS headers equals number of hops through network

Data center network with Sourcey
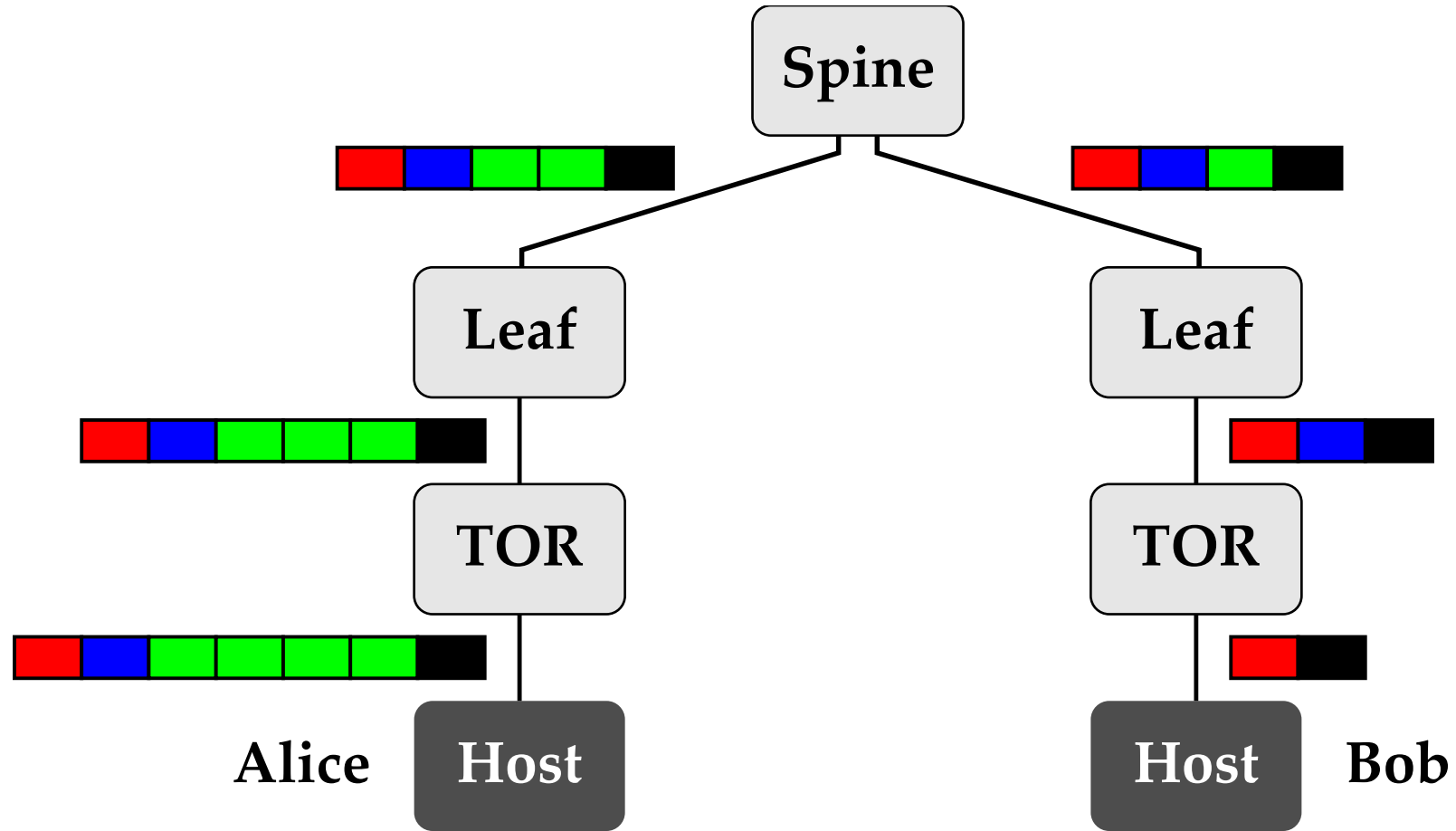
Do you realize how awesome it is
to get rid of the forwarding tables?

Challenges with Sourcey

- Multicast and broadcast?
  - Not common in mega data center networks

- Learning paths?
  - That's what SDN is for …

Pause for audience to point out more challenges

Related work

Microsoft, Cisco, Juniper:

Luyuan Fang, Fabio Chiussi, Deepak Bansal, Vijay Gill, Tony Lin, Jeff Cox, Gary Ratterree, *"Hierarchical SDN for the hyper-scale, hyper-elastic data center and cloud"*, in SOSR 2015.

Thank you!

It's a packet switched world.

But a packet ain't nothin' without a circuit.

My apologies to James Brown