ELEC / COMP 177 – Fall 2011

Computer Networking → Recap

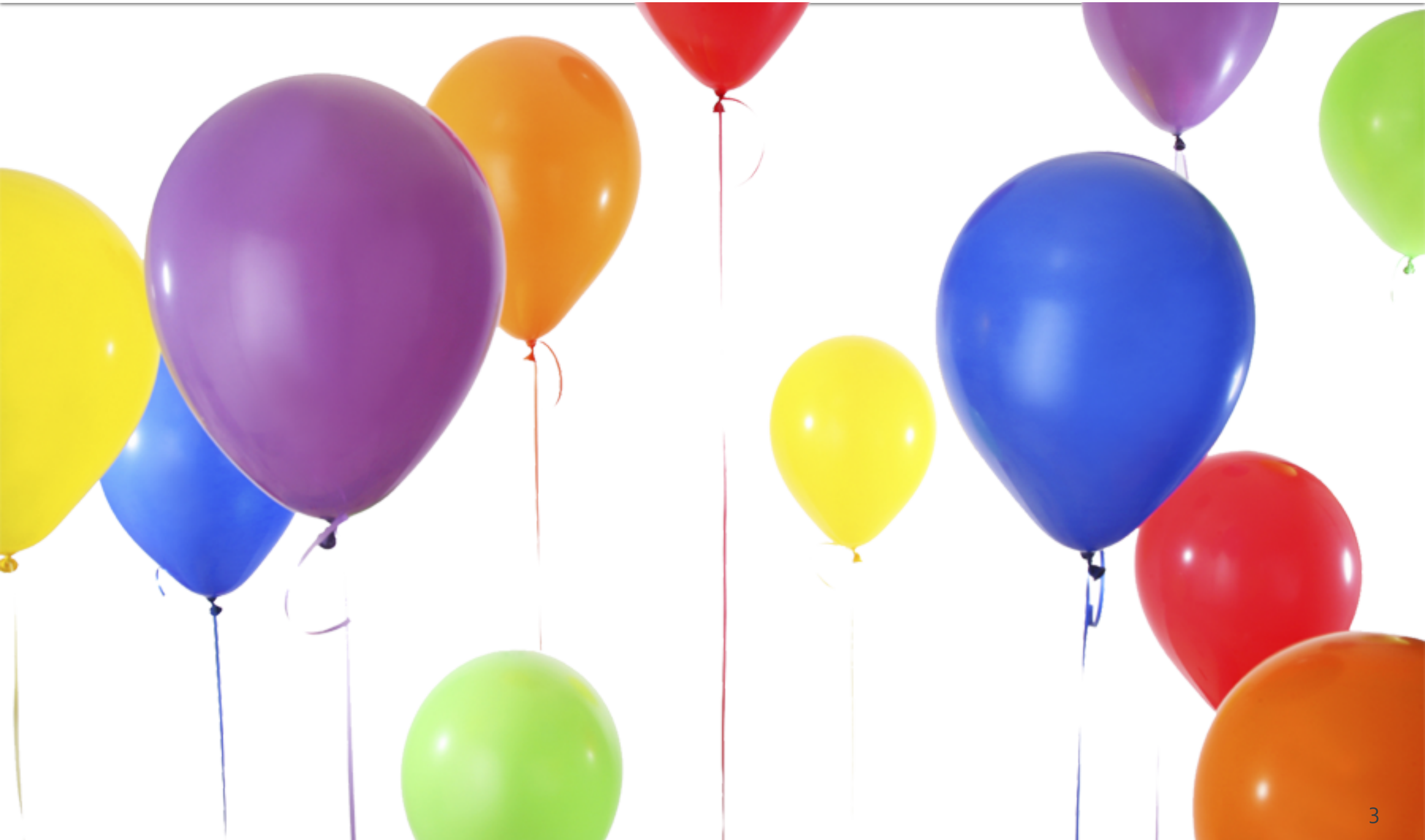
Some slides from Kurose and Ross, *Computer Networking*, 5th Edition

Schedule

- **Homework #6 - Presentation on security/ privacy**
  - **Topic selection** – Due Tuesday, Nov 22nd
  - **Slides** – Due Monday, Nov 28th
  - **Present!** – Tuesday, Nov 29th (and Thursday)

- **Project #3** – Due Tue, Dec 6th

Congratulations!



Milestone

- Reached an important milestone in the last class
- **Successfully sent a single IP packet across the global Internet**
  - Now know all of the key protocols and standards necessary to accomplish that task
  - Let's review the semester to date…

Recap – Ethernet Frame

7	1	6	6	2	0-1500	0-46	4	
Preamble	SFD	**DA**	**SA**	**Type**	**Data**	Pad	CRC	*Gap…*

- Destination MAC address
- Source MAC address
- Type (of encapsulated data)
- The data!
- **Who assigns the source address?**
  - **Does it contain information on network location?**
- **If I just have an Ethernet frame, where can I send data to?**

Recap – Ethernet Switch

- **How is a Ethernet switch better than a hub?**
- **How does a switch learn the location of computers on the network? (what *field*)**
- **What is stored in the forwarding table?**
  - MAC address, output port

Recap – Ethernet

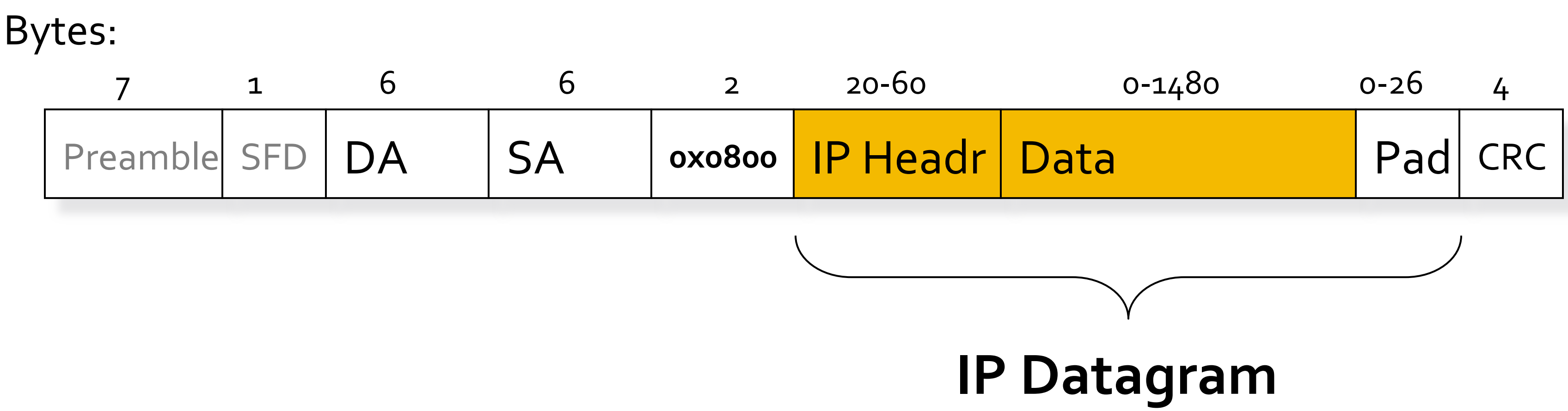
- **Why can't we use Ethernet for global communication?**
  - Broadcasts to find location of computers – too much bandwidth to do worldwide
  - Loops – Ethernet uses spanning tree to prevent loops
    - Can't have a single "root" of the Internet!
  - **Address contains no information about location on network**
    - Would need to have a forwarding table with one entry for every PC on the Internet we want to communicate with
    - i.e. a single worldwide "phonebook" with no shortcuts!

Recap – IP Datagram

<table>
<tr><th colspan="2">1 byte</th><th colspan="2">1 byte</th><th colspan="2">1 byte</th><th colspan="2">1 byte</th></tr>
<tr><td>Version</td><td>HdrLen</td><td colspan="2">Type of Service</td><td colspan="4">Total Length</td></tr>
<tr><td colspan="4">Identification</td><td>Flags</td><td colspan="3">Fragment Offset</td></tr>
<tr><td colspan="2">Time-To-Live</td><td colspan="2">Protocol</td><td colspan="4">Header Checksum</td></tr>
<tr><td colspan="8">Source IP Address</td></tr>
<tr><td colspan="8">Destination IP Address</td></tr>
<tr><td colspan="8">Options and padding (optional)</td></tr>
<tr><td colspan="8">Payload</td></tr>
</table>

Recap – IP Datagram

- Are IP packets separate from Ethernet frames?



- Time-to-live field: what's it used for?

Recap – IP encapsulated in Ethernet

<table>
<tr><td colspan="5">Destination MAC Address</td></tr>
<tr><td colspan="2">Destination MAC Address</td><td colspan="3">Source MAC Address</td></tr>
<tr><td colspan="5">Source MAC Address</td></tr>
<tr><td colspan="2">Type (`0x0800`)</td><td>Version</td><td>HdrLen</td><td>Type of Service</td></tr>
<tr><td colspan="2">Total Length</td><td colspan="3">Identification</td></tr>
<tr><td>Flags</td><td>Fragment Offset</td><td colspan="2">Time-To-Live</td><td>Protocol</td></tr>
<tr><td colspan="2">Header Checksum</td><td colspan="3">Source IP Address</td></tr>
<tr><td colspan="2">Source IP Address</td><td colspan="3">Destination IP Address</td></tr>
<tr><td colspan="2">Destination IP Address</td><td colspan="3">Options and Padding</td></tr>
<tr><td colspan="2">Options and Padding</td><td colspan="3">Payload</td></tr>
<tr><td colspan="5">Payload</td></tr>
<tr><td colspan="5">Ethernet CRC</td></tr>
</table>

Recap – IP Datagram

- **Where does the source IP address come from?**
  - DHCP (possibly running on the router)
- **Where does the destination IP address come from?**
  - DNS can be used to translate a host name from the *user* (e.g. www.pacific.edu) into an IP address (e.g. `138.9.110.12`)

Recap – IP Routers

- Ethernet switches forward packets based on destination MAC address
- **What do routers forward packets based on?**
  - Destination IP address
- **What is in the router's forwarding table?**
  - Prefixes, e.g. `138.16.9/24`
  - Next hop IP + exit port
- **What happens if more than one prefix matches the destination IP address?**
  - Longest prefix match determines winner

Recap – Forwarding versus Routing

- Forwarding
  - Move packets from router's input to appropriate router output
  - *Longest prefix match* (LPM)
- Routing
  - Determine path (route) taken by packets from source to destination
  - Routing algorithms such as RIP and OSPF

Example

- Send a single IP packet from Pacific to the main Moscow State University web server
- My IP:
  - 138.9.253.252
- MSU's IP:
  - 93.180.0.18

Traceroute

How does this actually **work**?

```
traceroute to www.msu.ru (93.180.0.18), 30 hops max, 60 byte packets
 1  10.10.5.252 (10.10.5.252) [AS65534]  0.679 ms  0.736 ms  0.760 ms
 2  10.0.0.93 (10.0.0.93) [AS1]  0.746 ms  0.823 ms  0.929 ms
 3  10.0.0.90 (10.0.0.90) [AS1]  0.675 ms  0.695 ms  0.684 ms
 4  138.9.253.252 (138.9.253.252) [*]  1.900 ms  1.894 ms  1.885 ms
 5  74.202.6.5 (74.202.6.5) [*]  5.580 ms  5.575 ms  5.565 ms
 6  nyc2-pr2-xe-1-2-0-0.us.twtelecom.net (66.192.253.150) [AS4323]  79.473 ms  76.655 ms 76.484 ms
 7  ae6-2.RT.TC1.STO.SE.retn.net (87.245.233.249) [AS9002]  177.970 ms  178.061 ms  178.052 ms
 8  GW-RUNNet.retn.net (87.245.249.50) [AS9002]  179.340 ms  179.508 ms  179.369 ms
 9  kt12-1-gw.spb.runnet.ru (194.85.40.141) [AS3267]  195.699 ms  195.706 ms  195.709 ms
10  b57-1-gw.spb.runnet.ru (194.85.40.153) [AS3267]  195.688 ms  195.680 ms  195.655 ms
11  m9-1-gw.msk.runnet.ru (194.85.40.133) [AS3267]  195.658 ms  195.643 ms  195.600 ms
12  m9-2-gw.msk.runnet.ru (194.85.40.214) [AS3267]  197.311 ms  194.277 ms  194.209 ms
13  msu.msk.runnet.ru (194.190.255.234) [AS3267]  194.287 ms  194.203 ms  194.611 ms
14  93.180.0.191 (93.180.0.191) [AS2848]  194.248 ms  194.751 ms  194.580 ms
15  93.180.0.187 (93.180.0.187) [AS2848]  194.950 ms  194.936 ms  194.888 ms
16  www.msu.ru (93.180.0.18) [AS2848]  194.633 ms !X  194.527 ms !X  194.776 ms !X
```

Companies Handling Our Packet

Name
1.) Time Warner Telecom
2.) RETN
3.) Runnet - State Institute of Information Technologies & Telecommunications
4.) Moscow State University

Assumptions

- Assume that I know
  - My own MAC address (hardwired on the NIC)
  - My own IP address (assigned via DHCP to be within my local subnet)
  - The subnet mask for my local network
  - The IP address of my gateway router leading "outside"
  - The IP address of MSU that I want to send a message to

Step 1

- **What happens first?**
  - Compare destination IP with my IP and subnet mask
    - My IP: `138.9.110.104`
    - My subnet mask: `255.255.255.0`
    - Thus, my subnet is `138.9.253/24`
  - Destination IP of `93.180.0.18` is (way!) outside my LAN

Step 2

- **The destination is outside of my LAN. What happens next?**
  - Need to send packet to gateway router
- **What does the Ethernet/IP packet look like?**
  - Destination MAC: ???
  - Source MAC: My MAC
  - Destination IP: MSU's IP
  - Source IP: My IP
  - TTL: 64 (a reasonable default)

Step 3

- **How do I get the MAC address of the router port attached to my LAN?**
  - I know my gateway router's IP address
  - Use ARP (Address Resolution Protocol)
- **Who receives my ARP request?**
  - Everyone – broadcast to all hosts on LAN
  - *"Who has 138.16.110.1? Tell 138.9.110.104"*
- **Who replies to my ARP request?**
  - Only the host (if any) with the requested IP address. This should be the router

Step 4

- Assume there is an Ethernet switch between you and the router
- **What happens if the switch has seen the MAC address of the router before?**
  - Packet is sent out only the port that faces the router
- **What happens if the switch has *not* seen the MAC address before?**
  - Packet is broadcast out all ports
- Switch **always** learns (or re-learns) from each packet

Step 5

- The packet reaches your gateway router (first router between here and MSU)
- **What does the router do?**
  - Verify checksums
  - Longest prefix match on destination IP address
- **What information is returned from router's forwarding table?**
  - Next hop IP address
    - (of subsequent router, or final host)
  - Output port

Step 6

- Assume the next hop is also connected to this router via Ethernet
- **What do we need to know to send a message to this router?**
  - Its MAC address
- **How do we find this?**
  - Router does ARP (just like hosts do ARP)

Step 7

- **How does the router modify the packet when retransmitting?**
  - Destination MAC = change to be MAC of next hop
  - Source MAC = change to be MAC of this router
  - Destination IP = unchanged
  - Source IP = unchanged
  - TTL = decrement by 1
  - Checksum = recalculated

Step 8

- This process of re-transmitting a packet repeats for many routers across the network
  - *16 in this example*
- Eventually, however, the "next hop" in the forwarding table is the actual destination computer
  - Packet has arrived!
- **Is that all the complexity in the Internet?**
  - No – forwarding tables in the router aren't created by magic!

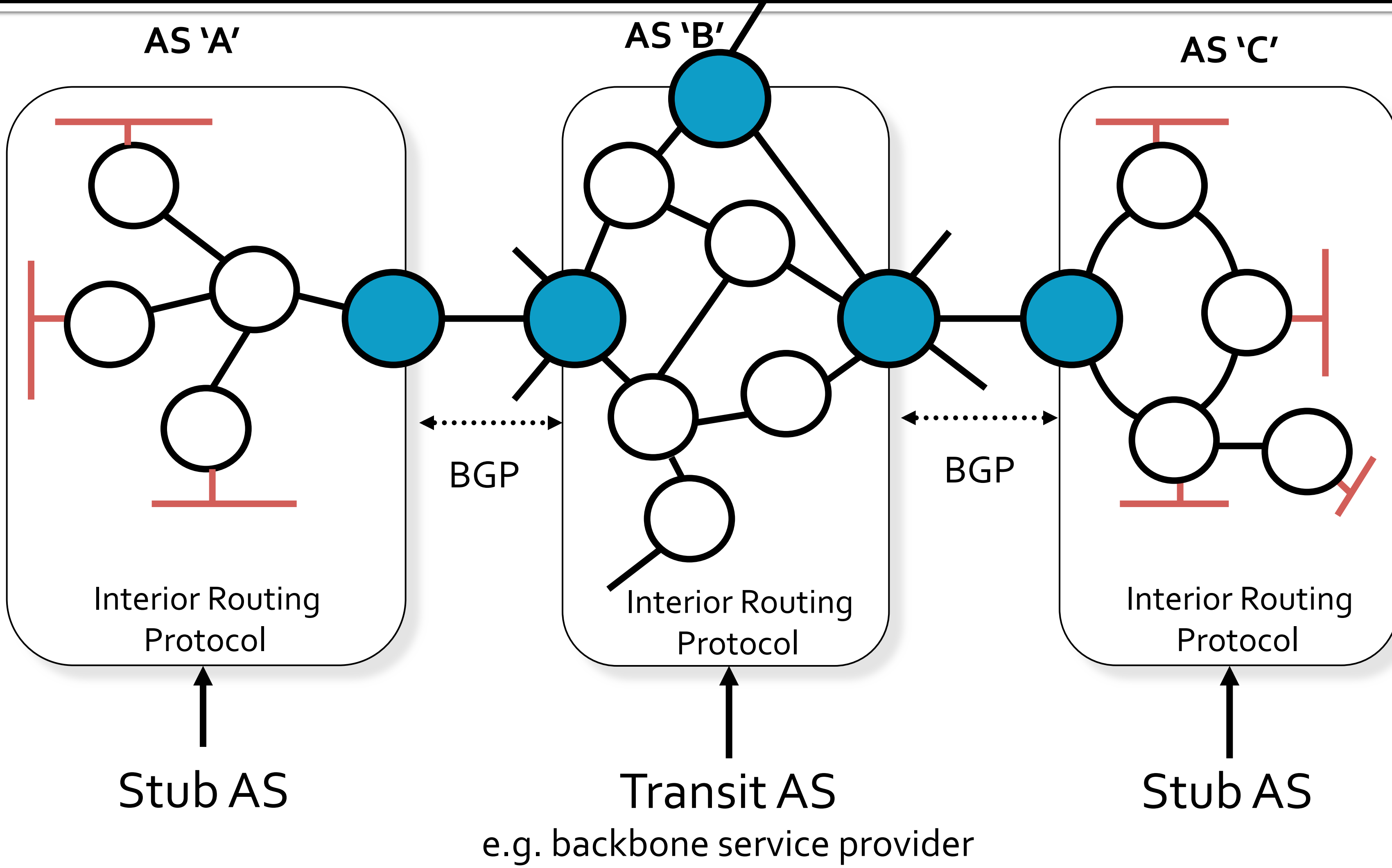
Routing

- In addition to forwarding packets, routers are busy (*asynchronously*) calculating **least-cost** routes to destinations
  - Goal: Have the forwarding table ready by the time your packet arrives with a specific destination
- **What happens if the forwarding table isn't ready, and there is no entry for your destination?**
  - Packet is dropped – you lose

Autonomous Systems

- **What is an autonomous system?**
  - Grouping of routers owned/administered by a single entity
- **Can a company only have a single autonomous system?**
  - No, can have multiple AS's
- **Why might I have more than one?**
  - Better routing for geography (i.e. North America versus Europe versus Asia)
  - Other reasons…

Autonomous Systems



Traceroute

```
traceroute to www.msu.ru (93.180.0.18), 30 hops max, 60 byte packets
 1  10.10.5.252 (10.10.5.252) [AS65534]  0.679 ms  0.736 ms  0.760 ms
 2  10.0.0.93 (10.0.0.93) [AS1]  0.746 ms  0.823 ms  0.929 ms
 3  10.0.0.90 (10.0.0.90) [AS1]  0.675 ms  0.695 ms  0.684 ms
 4  138.9.253.252 (138.9.253.252) [*]  1.900 ms  1.894 ms  1.885 ms
 5  74.202.6.5 (74.202.6.5) [*]  5.580 ms  5.575 ms  5.565 ms
 6  nyc2-pr2-xe-1-2-0-0.us.twtelecom.net (66.192.253.150) [AS4323] 79.473 ms  76.655 ms  76.484 ms
 7  ae6-2.RT.TC1.STO.SE.retn.net (87.245.233.249) [AS9002]  177.970 ms  178.061 ms  178.052 ms
 8  GW-RUNNet.retn.net (87.245.249.50) [AS9002]  179.340 ms  179.508 ms  179.369 ms
 9  kt12-1-gw.spb.runnet.ru (194.85.40.141) [AS3267]  195.699 ms  195.706 ms  195.709 ms
10  b57-1-gw.spb.runnet.ru (194.85.40.153) [AS3267]  195.688 ms  195.680 ms  195.655 ms
11  m9-1-gw.msk.runnet.ru (194.85.40.133) [AS3267]  195.658 ms  195.643 ms  195.600 ms
12  m9-2-gw.msk.runnet.ru (194.85.40.214) [AS3267]  197.311 ms  194.277 ms  194.209 ms
13  msu.msk.runnet.ru (194.190.255.234) [AS3267]  194.287 ms  194.203 ms  194.611 ms
14  93.180.0.191 (93.180.0.191) [AS2848]  194.248 ms  194.751 ms  194.580 ms
15  93.180.0.187 (93.180.0.187) [AS2848]  194.950 ms  194.936 ms  194.888 ms
16  www.msu.ru (93.180.0.18) [AS2848]  194.633 ms !X  194.527 ms !X  194.776 ms !X
```

AS Numbers in Traceroute

AS	Name
65534	Reserved (private use) – *Pacific internal net*
0	Reserved (non-routed networks) – *Pacific internal net*
18663	University of the Pacific – *Traceroute didn't resolve this for some reason and reported [*]…*
4323	Time Warner Telecom
9002	RETN
3267	Runnet - State Institute of Information Technologies & Telecommunications (SIIT&T "Informika")
2848	Moscow State University

First AS

- First AS is Pacific's (AS18663)
- Do a lookup on the AS
  - http://www.ripe.net/data-tools/stats/ris/routing-information-service
  - https://www.dan.me.uk/bgplookup
  - http://www.peeringdb.com/
    - Among other places…
- Pacific's gateway(s) to the Internet advertise a BGP prefix (aka subnet)
  - **`138.9.0.0/16`**

First AS

- An advertisement is a *promise:*
  - If you give me packets destined for IP addresses in this range, I will move them closer to their destination.
  - In this case, we *are* the destination!
  - This advertisement *originates* from our AS

Second AS

- Pacific buys Internet service from Time Warner (AS4323) , which has border routers that speak BGP
  - Pacific's routers talk to their routers, and they learn of our advertisement for 138.9.0.0/16
  - Now, Time Warner knows how to reach Pacific's IPs
  - We also learn of their advertisements!
    - Both for prefixes *originating* at those ISPs, and prefixes *reachable* through those ISPs

Announcements

- **When Time Warner give our routers their BGP announcements, do we get lots of tiny entries like 138.9.0.0/16?**
  - Maybe
  - But, routes can be aggregated together and expressed with smaller prefixes, e.g. `138.0.0.0/8`
    - Reduces communication time plus router CPU and memory requirements

Second AS (continued)

- Pacific had only 1 announcement
- Time Warner *originates* ~1620 announcements (as of Nov 2011)
  - Some are large, e.g. `173.226.0.0/15`
  - Some are small, e.g. `159.157.233.0/24`
- Time Warner also provides transit to their *downstream* customers' prefixes, including Pacific's prefix
  - Total of ~6563 announcements (as of Nov 2011)
  - We get this full list, as does ever other (BGP-speaking) AS connected to Time Warner

Third AS

- Time Warner (AS4323) can move this packet to New York City, where it enters the Equinix Internet Exchange
  - Private location to peer with dozens of other companies
  - Akamai, Amazon, Facebook, Google, Microsoft, many ISPs, etc…
- Time Warner connects with RETN (AS9002)
  - *Do they pay, or is this free?*
  - Same sharing of BGP announcements occurs here

Last AS

- The same thing is happening over in Eurasia
- Last AS of our path is Moscow State University (AS2848)
- MSU's gateway(s) to the Internet advertise a BGP prefix for `93.180.0.0/18` (along with 3 others that *originate* in this AS)
  - That encompasses the destination IP of `93.180.0.18`

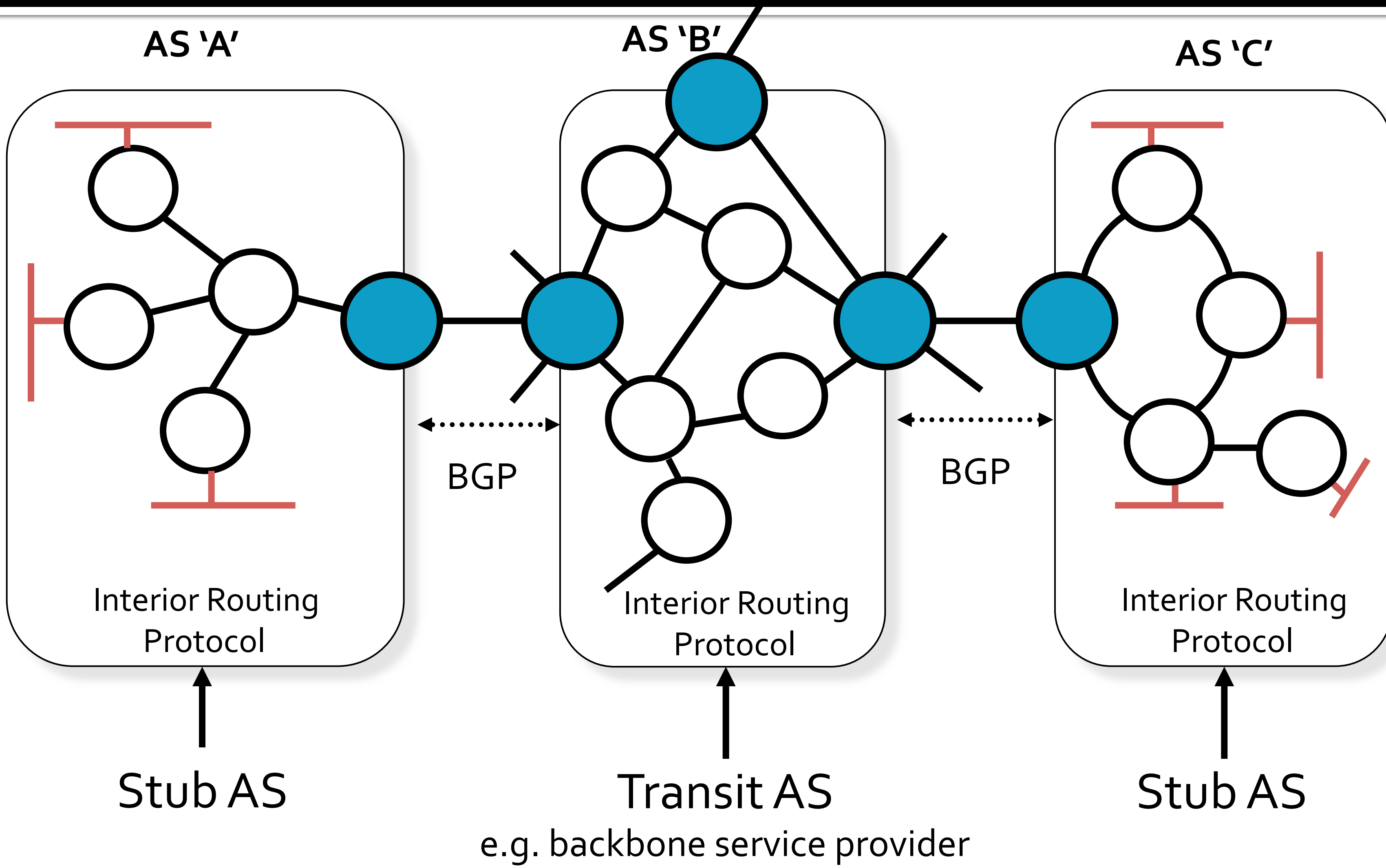
Next-to-last AS

- Moscow State University connects to Runnet (AS3267)
  - Runnet announces prefix `93.180.0.0/18` (along with 291 others reachable *downstream*, and 13 that *originate* in this AS)
  - Runnet now knows how to reach IPs that belong to MSU
- Runnet obtains transit through RETN, so our link is complete!

What's Missing?

- The forwarding table!
  - We keep forgetting to generate the forwarding table!
- Need more information
  - BGP tells us links <u>between</u> autonomous systems
  - Other protocols (RIP, OSPF) tell us paths <u>within</u> autonomous systems

Autonomous Systems



Interior Routing Protocol

- Let's say (hypothetically) that Time Warner runs RIP, a distance-vector protocol
- **Does each router have a complete view of the network inside the AS?**
  - No
- **What algorithm is used to develop routes?**
  - Bellman-Ford (using distance vectors)

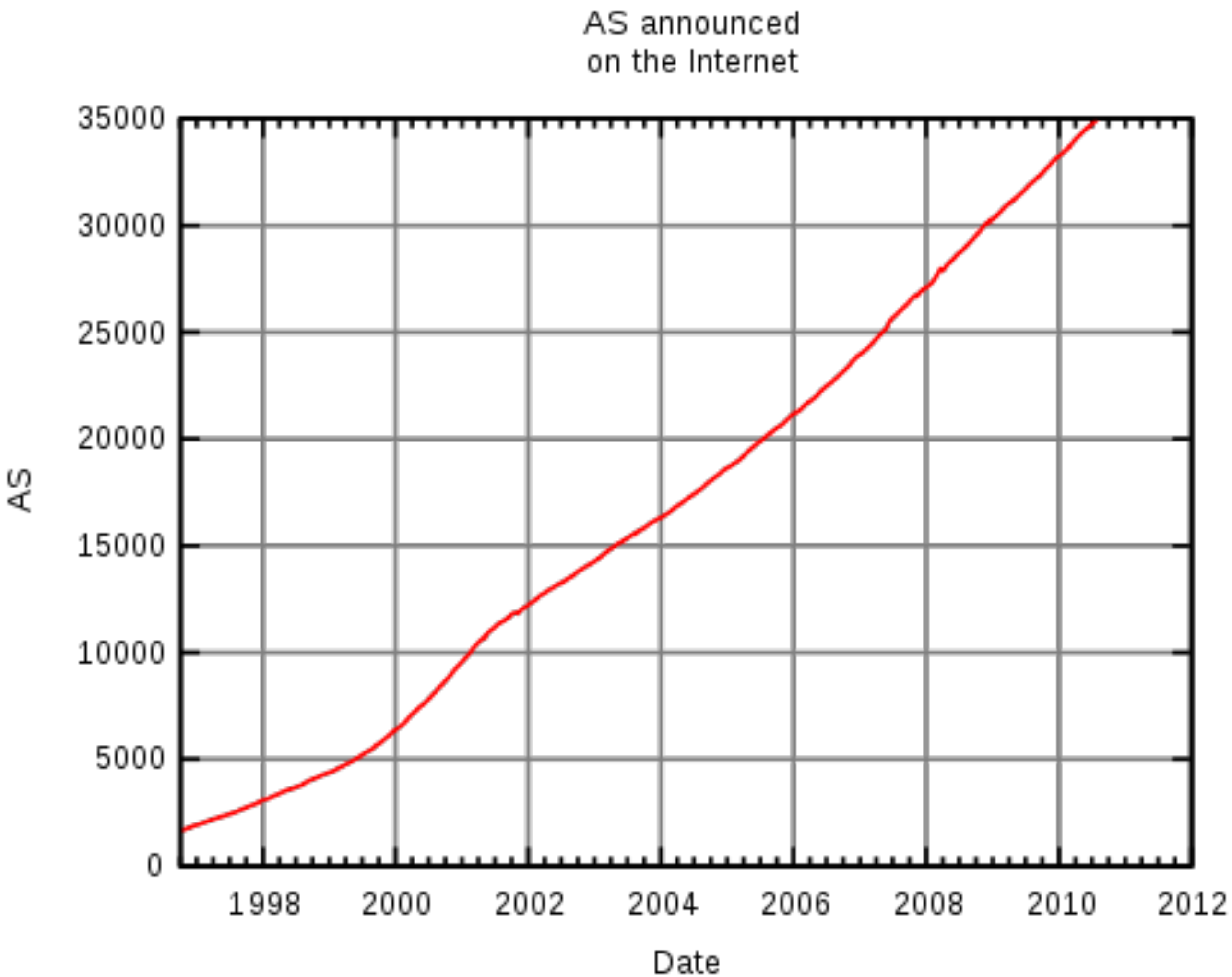
Interior Routing Protocol

- Let's say (hypothetically) that Time Warner runs OSPF, a link-state protocol
- **Does each router have a complete view of the network inside the AS?**
  - Yes
- **What algorithm is used to develop routes?**
  - Dijkstra's

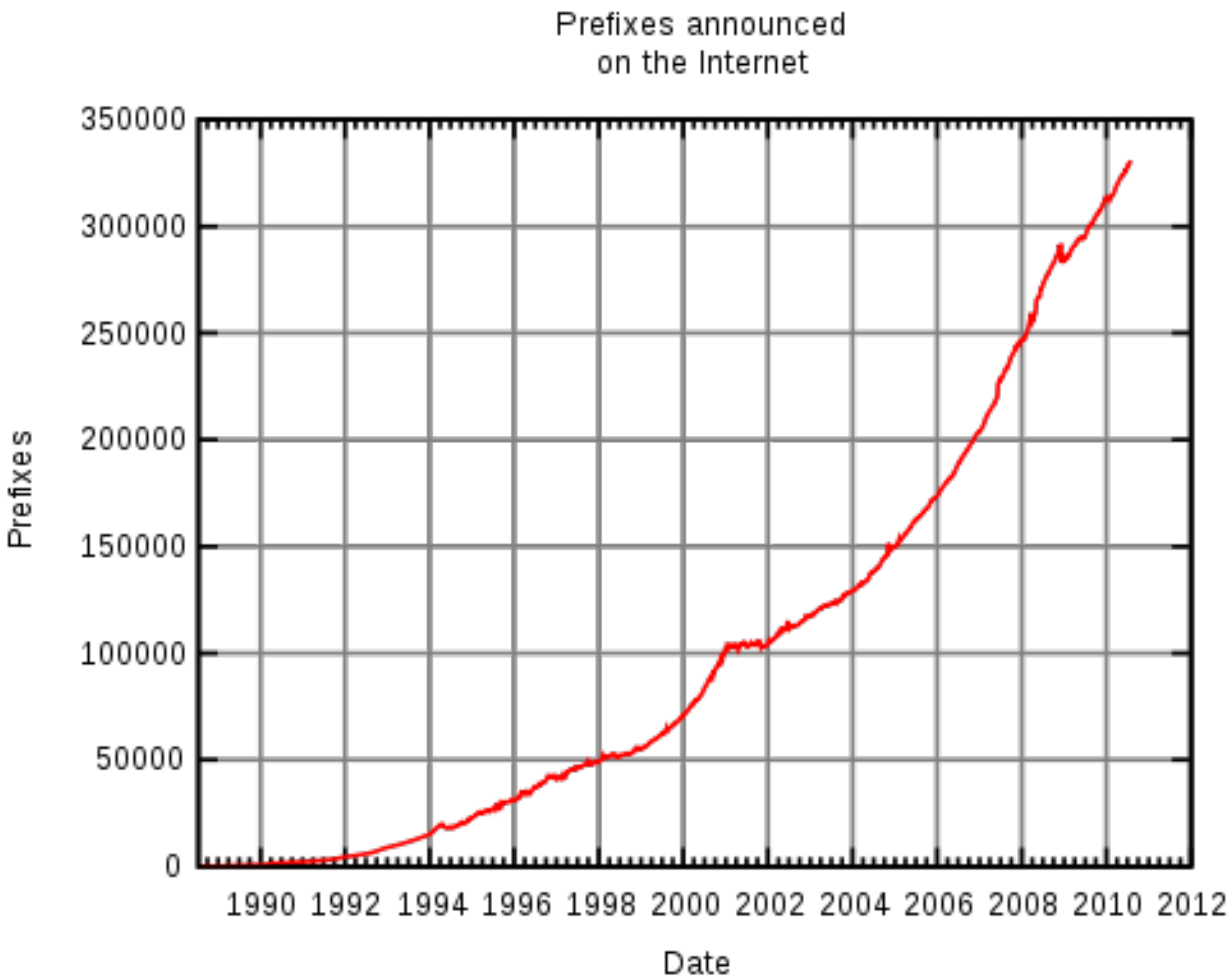
End Result is the Same

- Each router inside the AS updates its own forwarding table to direct BGP prefixes to the appropriate gateway router to the next AS
  - Rules might be very simple, i.e. just forward everything not destined to this AS to the same gateway router
  - Or rules might be complicated…
- **End result is a forwarding table for the router**
  - Prefix (for LPM)
  - Next-hop IP
  - Exit port

Growth of Internet – AS's



Growth of Internet - Prefixes



Growth

- **What does this growth mean for routers on the BGP-speaking Internet?**
  - They need to grow too! (more memory, faster CPUs, etc…)

What Does *My* Computer Do?

- Does *my* computer speak BGP?
  - No – your ISP's external gateway router does
- Does *my* computer speak RIP or OSPF?
  - No – your ISP's internal routers do
- Does *my* computer speak ARP?
  - Yes
- Does *my* computer speak IP?
  - Yes
- Does *my* computer speak Ethernet?
  - Yes

Milestone

- **Successfully sent a single IP packet across the global Internet**
  - Now know all of the key protocols and standards necessary to accomplish that task
- **Now can I waste time watching LOLcats?**

Milestone

- Not quite. One IP packet by itself is not enough to transmit an entire image
- **What else do we need?**
  - Method to **link multiple IP packets together** and deliver them to the **correct process** on the receiver
    - **Transport layer**: UDP, TCP (TCP also provides **reliability!**)
  - **Applications** need to be written to use this reliable network communication, and they need protocols of their own!
    - Web = HTTP, Email = POP / IMAP / SMTP, …