Q&A Session

Alexander Dietmüller
Tobias Bühler

https://comm-net.ethz.ch/

ETH Zürich
August 6 2019
How do you guide IP packets from a source to the destination?
Essentially, there are three ways to compute valid routing state:

<table>
<thead>
<tr>
<th>Intuition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1 Use tree-like topologies</td>
<td>Spanning-tree</td>
</tr>
<tr>
<td>#2 Rely on a global network view</td>
<td>Link-State</td>
</tr>
<tr>
<td>#3 Rely on distributed computation</td>
<td>Distance-Vector</td>
</tr>
<tr>
<td></td>
<td>SDN</td>
</tr>
<tr>
<td></td>
<td>BGP</td>
</tr>
</tbody>
</table>
Essentially, there are three ways to compute valid routing state:

<table>
<thead>
<tr>
<th>Intuition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1 Use tree-like topologies</td>
<td>Spanning-tree</td>
</tr>
<tr>
<td>#2 Rely on a global network view</td>
<td>Link-State</td>
</tr>
<tr>
<td>#3 Rely on distributed computation</td>
<td>Distance-Vector</td>
</tr>
</tbody>
</table>

- **Intuition**:
  - **#1** Use tree-like topologies
  - **#2** Rely on a global network view
  - **#3** Rely on distributed computation

- **Example**:
  - **Spanning-tree**
  - **Link-State**
  - **Distance-Vector**

**BGP**

**SDN**
Distance-vector protocols are based on Bellman-Ford algorithm
Every router maintains a table of all received distance vectors.
Every router maintains a table of all received distance vectors.

Let $d_x(y)$ be the cost of the least-cost path known by $x$ to reach $y$.

<table>
<thead>
<tr>
<th></th>
<th>dest.</th>
<th>via</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y$</td>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>$Z$</td>
<td>13</td>
<td>1</td>
</tr>
<tr>
<td>$U$</td>
<td>30</td>
<td>2</td>
</tr>
</tbody>
</table>
Let $d_x(y)$ be the cost of the least-cost path known by $x$ to reach $y$.

Every router maintains a table of all received distance vectors until convergence.

Each node bundles these distances into one message (called a vector) that it repeatedly sends to all its neighbors.

<table>
<thead>
<tr>
<th>X routing table</th>
<th>dest.</th>
<th>via</th>
<th>Z</th>
<th>X vector</th>
<th>dest. cost</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$Y$</td>
<td>3</td>
<td>11</td>
<td></td>
<td>$Y$ 3</td>
</tr>
<tr>
<td></td>
<td>$Z$</td>
<td>13</td>
<td>1</td>
<td></td>
<td>$Z$ 1</td>
</tr>
<tr>
<td></td>
<td>$U$</td>
<td>30</td>
<td>2</td>
<td></td>
<td>$U$ 2</td>
</tr>
</tbody>
</table>
Every router maintains a table of all received distance vectors.

Let $d_x(y)$ be the cost of the least-cost path known by $x$ to reach $y$.

Each node bundles these distances into one message (called a vector) that it repeatedly sends to all its neighbors until convergence.

Each node updates its distances based on neighbors’ vectors:

$$d_x(y) = \min \{ c(x,v) + d_v(y) \} \quad \text{over all neighbors } v.$$
Whenever a router uses another one, it will announce it an infinite cost.

The technique is known as poisoned reverse.
Internet routing comes into two flavors: *intra-* and *inter-domain* routing.

- **inter-domain routing**: Find paths between networks
- **intra-domain routing**: Find paths within a network
Find paths *between* networks
Internet

\downarrow

Border Gateway Protocol (BGP)
BGP is the routing protocol “glueing” the entire Internet together.
BGP announcements carry complete path information instead of distances
Each AS appends itself to the path when it propagates announcements.
There are 2 main business relationships today:

- customer/provider
- peer/peer

*many less important ones (siblings, backups,...)*
These policies are defined by constraining which BGP routes are *selected* and *exported*.

- **Selection**: Which path to use?
- **Export**: Which path to advertise?
which path to use?
control outbound traffic

which path to advertise?
Business relationships conditions

route selection

For a destination $p$, prefer routes coming from

- customers over
- peers over $route$ type
- providers
which path to use?

which path to advertise?
control inbound traffic
Routes coming from peers and providers are only propagated to customers

<table>
<thead>
<tr>
<th></th>
<th>customer</th>
<th>peer</th>
<th>provider</th>
</tr>
</thead>
<tbody>
<tr>
<td>customer</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>peer</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>provider</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
On the wire, BGP is a rather simple protocol composed of four basic messages

<table>
<thead>
<tr>
<th>type</th>
<th>used to</th>
</tr>
</thead>
<tbody>
<tr>
<td>OPEN</td>
<td>establish TCP-based BGP sessions</td>
</tr>
<tr>
<td>NOTIFICATION</td>
<td>report unusual conditions</td>
</tr>
<tr>
<td>UPDATE</td>
<td>inform neighbor of a new best route</td>
</tr>
<tr>
<td></td>
<td>a change in the best route</td>
</tr>
<tr>
<td>KEEPALIVE</td>
<td>inform neighbor that the connection is alive</td>
</tr>
</tbody>
</table>
UPDATE

inform neighbor of a new best route

a change in the best route
BGP UPDATEs carry an IP prefix together with a set of attributes

- **IP prefix**
- **Attributes**
  - Describe route properties used in route selection/exportation decisions
  - are either local (only seen on iBGP)
  - or global (seen on iBGP and eBGP)
<table>
<thead>
<tr>
<th>Attributes</th>
<th>Usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEXT-HOP</td>
<td>egress point identification</td>
</tr>
<tr>
<td>AS-PATH</td>
<td>loop avoidance</td>
</tr>
<tr>
<td></td>
<td>outbound traffic control</td>
</tr>
<tr>
<td></td>
<td>inbound traffic control</td>
</tr>
<tr>
<td>LOCAL-PREF</td>
<td>outbound traffic control</td>
</tr>
<tr>
<td>MED</td>
<td>inbound traffic control</td>
</tr>
</tbody>
</table>
Prefer routes…

with higher LOCAL-PREF

with shorter AS-PATH length

with lower MED

learned via eBGP instead of iBGP

with lower IGP metric to the next-hop

with smaller egress IP address (tie-break)
Each BGP router processes UPDATEs according to a precise pipeline
Life of a BGP router is made of three consecutive steps

while true:

■ receives routes from my neighbors

■ select one best route for each prefix

■ export the best route to my neighbors
In practice, multiple URLs can be mapped to the same IP

How does a web server receiving an HTTP request know, which website you want to access?
The host field tells the server which website it should serve.

HTTP request:

GET / HTTP/1.1
Host: www.google.com
Anycast, Unicast, Multicast

„one-to-one-of-many“

Important, discussed in lecture

Used for scalability, load-balancing (e.g. DNS root server)

Routing finds shortest-paths

Seamless replication

But, potential problems for stateful applications
Anycast, Unicast, Multicast
Anycast, Unicast, Multicast
Anycast, Unicast, Multicast

„one-to-one“

Destination address uniquely identifies a single receiver

No replication
Anycast, **Unicast**, Multicast
Anycast, Unicast, **Multicast**

„one-to-many-of-many“ („many-to-many-of-many“)

E.g. useful to stream the same video to multiple receivers
Anycast, Unicast, **Multicast**
### Exercise 4 - Detective Work

<table>
<thead>
<tr>
<th>SRC MAC Address</th>
<th>DST MAC Address</th>
<th>SRC IP Address</th>
<th>DST IP Address</th>
</tr>
</thead>
<tbody>
<tr>
<td>6a:00:02:49:a1:a0</td>
<td>11:05:ab:59:bb:02</td>
<td>65.222.11.1</td>
<td>65.222.8.2</td>
</tr>
<tr>
<td>6a:00:02:49:a1:a0</td>
<td>da:15:00:00:01:11</td>
<td>65.222.11.1</td>
<td>65.222.16.1</td>
</tr>
<tr>
<td>da:15:00:00:01:11</td>
<td>11:05:ab:59:bb:02</td>
<td>129.132.103.40</td>
<td>65.222.8.2</td>
</tr>
<tr>
<td>11:05:ab:59:bb:02</td>
<td>40:34:00:7a:00:01</td>
<td>65.222.8.2</td>
<td>65.222.15.254</td>
</tr>
<tr>
<td>11:05:ab:59:bb:02</td>
<td>ac:00:0a:aa:10:05</td>
<td>65.222.8.2</td>
<td>65.222.9.99</td>
</tr>
<tr>
<td>ac:00:0a:aa:10:05</td>
<td>01:05:3c:34:00:02</td>
<td>65.222.9.99</td>
<td>65.222.13.255</td>
</tr>
<tr>
<td>6a:00:02:49:a1:a0</td>
<td>da:15:00:00:01:11</td>
<td>65.222.11.1</td>
<td>65.222.8.1</td>
</tr>
</tbody>
</table>

**a)** Can you identify all the hosts that are part of the local network?

**Solution:**

The local hosts are all the sources and destinations that do not have to go through the default gateway (e.g., their MAC address is not replaced by the MAC address of the router):

- 65.222.11.1
- 65.222.8.2
- 65.222.9.99
- 65.222.15.254
- 65.222.13.255

**b)** Can you reconstruct the IP subnet used to address the hosts within that local network?

**Solution:**

First, we should note that the router MAC address is only used for IP sources or destinations outside the local subnet (router is used as gateway) or for packets from/towards the router. With this in mind, we can identify the lowest subnet address from the packets (65.222.11.1 -> 65.222.8.1) and (65.222.11.1 -> 65.222.8.2) as 65.222.8.1. Furthermore, we can infer that 65.222.15.254 still belongs to the local subnet (65.222.8.2 -> 65.222.15.254) but 65.222.16.1 is a destination outside of the network (65.222.11.1 -> 65.222.16.1). We can therefore identify the used subnet as 65.222.8.0/21.
# Exercise 4 - Detective Work

<table>
<thead>
<tr>
<th>SRC MAC Address</th>
<th>DST MAC Address</th>
<th>SRC IP Address</th>
<th>DST IP Address</th>
</tr>
</thead>
<tbody>
<tr>
<td>6a:00:02:49:a1:a0</td>
<td>11:05:ab:59:bb:02</td>
<td>65.222.11.1</td>
<td>65.222.8.2</td>
</tr>
<tr>
<td>6a:00:02:49:a1:a0</td>
<td>da:15:00:00:01:11</td>
<td>65.222.11.1</td>
<td>65.222.16.1</td>
</tr>
<tr>
<td>da:15:00:00:01:11</td>
<td>11:05:ab:59:bb:02</td>
<td>129.132.103.40</td>
<td>65.222.8.2</td>
</tr>
<tr>
<td>11:05:ab:59:bb:02</td>
<td>40:34:00:7a:00:01</td>
<td>65.222.8.2</td>
<td>65.222.15.254</td>
</tr>
<tr>
<td>11:05:ab:59:bb:02</td>
<td>ac:00:0a:aa:10:05</td>
<td>65.222.8.2</td>
<td>65.222.9.99</td>
</tr>
<tr>
<td>ac:00:0a:aa:10:05</td>
<td>01:05:3c:34:00:02</td>
<td>65.222.9.99</td>
<td>65.222.13.255</td>
</tr>
<tr>
<td>6a:00:02:49:a1:a0</td>
<td>da:15:00:00:01:11</td>
<td>65.222.11.1</td>
<td>65.222.8.1</td>
</tr>
</tbody>
</table>

- **Can you identify all the hosts that are part of the local network?**
  - Solution: The local hosts are all the sources and destinations that do not have to go through the default gateway (e.g., their MAC address is not replaced by the MAC address of the router):
    - 65.222.11.1
    - 65.222.8.2
    - 65.222.9.99
    - 65.222.15.254
    - 65.222.13.255

- **Can you reconstruct the IP subnet used to address the hosts within that local network?**
  - Solution: First, we should note that the router MAC address is only used for IP sources or destinations outside the local subnet (router is used as gateway) or for packets from/towards the router. With this in mind, we can identify the lowest subnet address from the packets (65.222.11.1 -> 65.222.8.1) and (65.222.11.1 -> 65.222.8.2) as 65.222.8.1. Furthermore, we can infer that 65.222.15.254 still belongs to the local subnet (65.222.8.2 -> 65.222.15.254) but 65.222.16.1 is a destination outside of the network (65.222.11.1 -> 65.222.16.1). We can therefore identify the used subnet as 65.222.8.0/21.

- **Router interface MAC address**
## Exercise 4 - Detective Work

### Solution

**a)** Can you identify all the hosts that are part of the local network?

Solution:
The local hosts are all the sources and destinations that do not have to go through the default gateway (e.g., their MAC address is not replaced by the MAC address of the router):

- 65.222.11.1
- 65.222.8.2
- 65.222.9.99
- 65.222.15.254
- 65.222.13.255

**b)** Can you reconstruct the IP subnet used to address the hosts within that local network?

Solution:
First, we should note, that the router MAC address is only used for IP sources or destinations outside the local subnet (router is used as gateway) or for packets from/towards the router. With this in mind, we can identify the lowest subnet address from the packets (65.222.11.1 -> 65.222.8.1) and (65.222.11.1 -> 65.222.8.2) as 65.222.8.1. Furthermore, we can infer that 65.222.15.254 still belongs to the local subnet (65.222.8.2 -> 65.222.15.254) but 65.222.16.1 is a destination outside of the network (65.222.11.1 -> 65.222.16.1). We can therefore identify the used subnet as 65.222.8.0/21.

---

### Table

<table>
<thead>
<tr>
<th>SRC MAC Address</th>
<th>DST MAC Address</th>
<th>SRC IP Address</th>
<th>DST IP Address</th>
</tr>
</thead>
<tbody>
<tr>
<td>6a:00:02:49:a1:a0</td>
<td>11:05:ab:59:bb:02</td>
<td>65.222.11.1</td>
<td>65.222.8.2</td>
</tr>
<tr>
<td>6a:00:02:49:a1:a0</td>
<td>da:15:00:00:01:11</td>
<td>65.222.11.1</td>
<td>65.222.16.1</td>
</tr>
<tr>
<td>da:15:00:00:01:11</td>
<td>11:05:ab:59:bb:02</td>
<td>129.132.103.40</td>
<td>65.222.8.2</td>
</tr>
<tr>
<td>11:05:ab:59:bb:02</td>
<td>40:34:00:7a:00:01</td>
<td>65.222.8.2</td>
<td>65.222.15.254</td>
</tr>
<tr>
<td>11:05:ab:59:bb:02</td>
<td>ac:00:0a:aa:10:05</td>
<td>65.222.8.2</td>
<td>65.222.9.99</td>
</tr>
<tr>
<td>ac:00:0a:aa:10:05</td>
<td>01:05:3c:34:00:02</td>
<td>65.222.9.99</td>
<td>65.222.13.255</td>
</tr>
<tr>
<td>6a:00:02:49:a1:a0</td>
<td>da:15:00:00:01:11</td>
<td>65.222.11.1</td>
<td>65.222.8.1</td>
</tr>
</tbody>
</table>

- [ ] Router interface MAC address
- [ ] Dst 65.222.8.2 does not go over router => internal
### Exercise 4 - Detective Work

<table>
<thead>
<tr>
<th>SRC MAC Address</th>
<th>DST MAC Address</th>
<th>SRC IP Address</th>
<th>DST IP Address</th>
</tr>
</thead>
<tbody>
<tr>
<td>6a:00:02:49:a1:a0</td>
<td>11:05:ab:59:bb:02</td>
<td>65.222.11.1</td>
<td>65.222.8.2</td>
</tr>
<tr>
<td>6a:00:02:49:a1:a0</td>
<td>da:15:00:00:01:11</td>
<td>65.222.11.1</td>
<td>65.222.16.1</td>
</tr>
<tr>
<td>da:15:00:00:01:11</td>
<td>11:05:ab:59:bb:02</td>
<td>129.132.103.40</td>
<td>65.222.8.2</td>
</tr>
<tr>
<td>11:05:ab:59:bb:02</td>
<td>40:34:00:7a:00:01</td>
<td>65.222.8.2</td>
<td>65.222.15.254</td>
</tr>
<tr>
<td>11:05:ab:59:bb:02</td>
<td>ac:00:0a:aa:10:05</td>
<td>65.222.8.2</td>
<td>65.222.9.99</td>
</tr>
<tr>
<td>ac:00:0a:aa:10:05</td>
<td>01:05:3c:34:00:02</td>
<td>65.222.9.99</td>
<td>65.222.13.255</td>
</tr>
<tr>
<td>6a:00:02:49:a1:a0</td>
<td>da:15:00:00:01:11</td>
<td>65.222.11.1</td>
<td>65.222.8.1</td>
</tr>
</tbody>
</table>

- **☐** Router interface MAC address
- **☐** Dst 65.222.8.2 does not go over router => internal
- **☐** Dst 65.222.8.1 reaches router and has to be in the same subnet as 192.168.8.2

**=> 65.222.8.1 is the IP of the router**
**VLAN**

**Access link**: part of only one VLAN
normally connects hosts with switches (to get „access“)

**Trunk link**: can carry traffic for multiple VLANs
normally connects switches to other switches or routers

The per-VLAN spanning tree still spans the **entire** network
even if some of the switches do not have hosts in all VLANs
=> better optimized paths for hosts in one VLAN
=> ready for new hosts in the future
VLAN (spanning tree from the slides)
Exercise 6 - Convergence

Consider this BGP network composed of 5 ASes. Each AS is assigned with a list of paths which indicates its preferences to reach AS 0.

Considering that only AS 0 originates prefixes, does that BGP network have a unique, stable solution?

a) If yes, indicate the path that each AS selects in the stable solution.

b) If not, describe an example of oscillation. For instance, by describing a sequence of messages that repeats itself.

Solution:

This BGP network does have a unique, stable solution in which:

- AS 1 selects [1,2,0] (preferred path);
- AS 2 selects [2,0];
- AS 3 selects [3,4,0] (preferred path);
- AS 4 selects [4,0] (preferred path).
Exercise 6 - Convergence

Does this network ever converge?

Consider this BGP network composed of 5 ASes. Each AS is assigned with a list of paths which indicates its preferences to reach AS 0. Considering that only AS 0 originates prefixes, does that BGP network have a unique, stable solution?

a) If yes, indicate the path that each AS selects in the stable solution.

b) If not, describe an example of oscillation. For instance, by describing a sequence of messages that repeats itself.

Solution:

This BGP network does have a unique, stable solution in which:

- AS 1 selects \([1,2,0]\) (preferred path);
- AS 2 selects \([2,0]\);
- AS 3 selects \([3,4,0]\) (preferred path);
- AS 4 selects \([4,0]\) (preferred path).
Exercise 6 - Convergence

Consider this BGP network composed of 5 ASes. Each AS is assigned with a list of paths which indicates its preferences to reach AS 0. Considering that only AS 0 originates prefixes, does that BGP network have a unique, stable solution?

a) If yes, indicate the path that each AS selects in the stable solution.

b) If not, describe an example of oscillation. For instance, by describing a sequence of messages that repeats itself.

Solution: This BGP network does have a unique, stable solution in which:

- AS 1 selects [1, 2, 0] (preferred path);
- AS 2 selects [2, 0];
- AS 3 selects [3, 4, 0] (preferred path);
- AS 4 selects [4, 0] (preferred path).
Consider this BGP network composed of 5 ASes. Each AS is assigned with a list of paths which indicates its preferences to reach AS 0. Considering that only AS 0 originates prefixes, does that BGP network have a unique, stable solution?

a) If yes, indicate the path that each AS selects in the stable solution.

b) If not, describe an example of oscillation. For instance, by describing a sequence of messages that repeats itself.

Solution:
This BGP network does have a unique, stable solution in which:

- AS 1 selects [1, 2, 0] (preferred path);
- AS 2 selects [2, 0];
- AS 3 selects [3, 4, 0] (preferred path);
- AS 4 selects [4, 0] (preferred path).
Does this network ever converge?

Consider this BGP network composed of 5 ASes. Each AS is assigned with a list of paths which indicates its preferences to reach AS 0.

Considering that only AS 0 originates prefixes, does that BGP network have a unique, stable solution?

a) If yes, indicate the path that each AS selects in the stable solution.

b) If not, describe an example of oscillation. For instance, by describing a sequence of messages that repeats itself.

Solution:

This BGP network does have a unique, stable solution in which:

- AS 1 selects \([1, 2, 0]\) (preferred path);
- AS 2 selects \([2, 0]\);
- AS 3 selects \([3, 4, 0]\) (preferred path);
- AS 4 selects \([4, 0]\) (preferred path).

Exercise 6 - Convergence
Consider this BGP network composed of 5 ASes. Each AS is assigned with a list of paths which indicates its preferences to reach AS 0. Considering that only AS 0 originates prefixes, does that BGP network have a unique, stable solution?

a) If yes, indicate the path that each AS selects in the stable solution.

b) If not, describe an example of oscillation. For instance, by describing a sequence of messages that repeats itself.

Solution: This BGP network does have a unique, stable solution in which:

- AS 1 selects [1,2,0] (preferred path);
- AS 2 selects [2,0];
- AS 3 selects [3,4,0] (preferred path);
- AS 4 selects [4,0] (preferred path).
Does this network ever converge?

Consider this BGP network composed of 5 ASes. Each AS is assigned with a list of paths which indicates its preferences to reach AS 0.

Considering that only AS 0 originates prefixes, does that BGP network have a unique, stable solution?

a) If yes, indicate the path that each AS selects in the stable solution.

b) If not, describe an example of oscillation. For instance, by describing a sequence of messages that repeats itself.

Solution:
This BGP network does have a unique, stable solution in which:
• AS 1 selects [1, 2, 0] (preferred path);
• AS 2 selects [2, 0];
• AS 3 selects [3, 4, 0] (preferred path);
• AS 4 selects [4, 0] (preferred path).

Exercise 6 - Convergence
Does this network ever converge?

Consider this BGP network composed of 5 ASes. Each AS is assigned with a list of paths which indicates its preferences to reach AS 0.

Considering that only AS 0 originates prefixes, does that BGP network have a unique, stable solution?

a) If yes, indicate the path that each AS selects in the stable solution.

b) If not, describe an example of oscillation. For instance, by describing a sequence of messages that repeats itself.

Solution:
This BGP network does have a unique, stable solution in which:

- AS 1 selects [1, 2, 0] (preferred path);
- AS 2 selects [2, 0];
- AS 3 selects [3, 4, 0] (preferred path);
- AS 4 selects [4, 0] (preferred path).
Exercise 6 - Convergence

Consider this BGP network composed of 5 ASes. Each AS is assigned with a list of paths which indicates its preferences to reach AS 0. Considering that only AS 0 originates prefixes, does that BGP network have a unique, stable solution?

a) If yes, indicate the path that each AS selects in the stable solution.

b) If not, describe an example of oscillation. For instance, by describing a sequence of messages that repeats itself.

Solution:
This BGP network does have a unique, stable solution in which:
- AS 1 selects [1, 2, 0] (preferred path);
- AS 2 selects [2, 0];
- AS 3 selects [3, 4, 0] (preferred path);
- AS 4 selects [4, 0] (preferred path).
Consider this BGP network composed of 5 ASes. Each AS is assigned with a list of paths which indicates its preferences to reach AS 0.

Considering that only AS 0 originates prefixes, does that BGP network have a unique, stable solution?

a) If yes, indicate the path that each AS selects in the stable solution.

b) If not, describe an example of oscillation. For instance, by describing a sequence of messages that repeats itself.

Solution: This BGP network does have a unique, stable solution in which:

- AS 1 selects [1, 2, 0] (preferred path);
- AS 2 selects [2, 0];
- AS 3 selects [3, 4, 0] (preferred path);
- AS 4 selects [4, 0] (preferred path).
Important deadline

Use Slack or email to ask your last questions until 14.08.2019

Exam: 21.08.2019 – HIL D 15
Individual Questions