One path is not enough anymore
Where are the feedback loops in the Internet?

Olivier Bonaventure

ICTEAM, IP Networking Lab
Université catholique de Louvain (UCL)
Place Sainte-Barbe, 2, B-1348, Louvain-la-Neuve (Belgium)
Internet in the early days

Client-server model
many clients, few servers
Internet in the early days

Client-server model
many clients, few servers

Each server hosts specific content
Internet in the early days

Client-server model
many clients, few servers

Each router computes one best path for each destination

Each server hosts specific content

© O. Bonaventure, 2010
Internet in the early days

Client-server model
many clients, few servers

Each router computes one best path for each destination

Each server hosts specific content
Issues

Key issues with this model

Network congestion when many flows compete to use the same bottleneck link

Server congestion
   How to serve a growing base of clients?
      A single server per service/content is not sufficient anymore

Network performance
   Latency matters for many applications
      How can we reduce the latency perceived by the user?
   Can we avoid network congestion by moving flows?
TCP congestion control

A

R1

R2

C

10 Mbps

2 Mbps

10 Mbps
TCP congestion control

[Diagram showing network traffic and buffer sizes]

© O. Bonaventure, 2010
TCP congestion control

10 Mbps
R1
2 Mbps
R2
10 Mbps

Buffers of R2

10 Mbps

2 Mbps
TCP congestion control

A

10 Mbps

R1

2 Mbps

R2

10 Mbps

C

10 Mbps

Buffers of R2

10 Mbps

Blue arrows:

2 Mbps

Red arrow:
TCP congestion control
TCP congestion control
TCP congestion control
TCP congestion control

A

R1

2 Mbps

R2

10 Mbps

C

10 Mbps

2 Mbps

Buffers of R2

© O. Bonaventure, 2010
TCP congestion control

Buffers of R2

A

10 Mbps

R1

2 Mbps

R2

10 Mbps

C

10 Mbps

2 Mbps
TCP congestion control

TCP self-clocking
Basic assumptions
Packet losses indicate network congestion
TCP sources adjust their transmission rate by adjusting their congestion window
TCP congestion control (2)

Basic assumptions
Packet losses indicate network congestion
TCP sources adjust their transmission rate by adjusting their congestion window

Host A
Rate

Host B Rate

Host A
Rate

Host B Rate

Efficiency line

Congested region

Fairness line

10 Mbps R1 100 Mbps R2 2 Mbps

© O. Bonaventure, 2010
TCP congestion control (3)

Additive increase and multiplicative decrease
Each TCP source maintains a congestion window

Congestion window **halved** after each packet loss
Congestion window **incremented by one packet** every round-trip time when there are no losses
**slow-start** at beginning of connection or after a severe loss

V. Jacobson, Congestion Avoidance and Control, SIGCOMM `88, © O. Bonaventure, 2010
TCP congestion control (4)

One of the most popular networking topic

Many, many variants to address various issues
  TCP over technology x
  TCP over wireless
    in wireless networks, packet losses do not necessarily indicate congestion
  High speed TCP
    included in recent Linux and Windows implementations
  TCP modelling
  TCP fairness
  TCP compatible congestion control for real-time apps

...
What is changing on the Internet?
What is changing on the Internet?

What is changing on the servers?

© O. Bonaventure, 2010
Internet traffic is changing

Arbor networks study, NANOG47
Google is now 6% of total Internet traffic
http://www.nanog.org/meetings/nanog47/
Origin of Internet content

Huge datacenters

Content distribution networks
Importance of CDNs

Arbor networks study, NANOG47
10% of Internet traffic
P2P decrease and web consolidation

<table>
<thead>
<tr>
<th>Rank</th>
<th>Application</th>
<th>2007</th>
<th>2009</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Web</td>
<td>41.68</td>
<td>52.00</td>
<td>+10.31</td>
</tr>
<tr>
<td>2</td>
<td>Video</td>
<td>1.58</td>
<td>2.64</td>
<td>+1.05</td>
</tr>
<tr>
<td>3</td>
<td>VPN</td>
<td>1.04</td>
<td>1.41</td>
<td>+0.38</td>
</tr>
<tr>
<td>4</td>
<td>Email</td>
<td>1.41</td>
<td>1.38</td>
<td>-0.03</td>
</tr>
<tr>
<td>5</td>
<td>News</td>
<td>1.75</td>
<td>0.97</td>
<td>-0.78</td>
</tr>
<tr>
<td>6</td>
<td>P2P</td>
<td>2.96</td>
<td>0.85*</td>
<td>-2.11</td>
</tr>
<tr>
<td>7</td>
<td>Games</td>
<td>0.38</td>
<td>0.49</td>
<td>+0.12</td>
</tr>
<tr>
<td>8</td>
<td>SSH</td>
<td>0.19</td>
<td>0.28</td>
<td>-0.08</td>
</tr>
<tr>
<td>9</td>
<td>DNS</td>
<td>0.20</td>
<td>0.17</td>
<td>-0.04</td>
</tr>
<tr>
<td>10</td>
<td>FTP</td>
<td>0.21</td>
<td>0.14</td>
<td>-0.07</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>2.56</td>
<td>2.67</td>
<td>+0.11</td>
</tr>
<tr>
<td></td>
<td>Unclassified</td>
<td>46.03</td>
<td>37.00</td>
<td>-9.03</td>
</tr>
</tbody>
</table>

* 18% via payload

Arbor networks study, NANOG47
P2P is being replaced by direct download

© O. Bonaventure, 2010
The new feedback loops

The same content can be served transparently by thousands of servers
The new feedback loops

The same content can be served transparently by thousands of servers

Where is www.g.com?
The new feedback loops

The same content can be served transparently by thousands of servers

Where is www.g.com?

Contact LB1
The new feedback loops

The same content can be served transparently by thousands of servers

Where is www.g.com?

Contact LB1
What is changing on the Internet?
What is changing on the Internet?

What is changing inside the network?
Equal Cost Multipath

If two paths have same exact cost, routers could use them both to reach destination requires small extension to Dijkstra’s shortest path.
Equal Cost Multipath

If two paths have same exact cost, routers could use them both to reach destination requires small extension to Dijkstra’s shortest path
Equal Cost Multipath

How to balance the packets over multiple paths?
- per packet round-robin
  - good load balancing
  - breaks TCP connections and causes retransmissions

- per-TCP connection splitting
  - all packets from one TCP connection follow same path
  - implemented by a hashing function on the router’s data path

Benefits
- improves load balancing
- routers know and can use alternate path in case of failures
Issues with ECMP

Coverage

<table>
<thead>
<tr>
<th>Network name</th>
<th>Size</th>
<th>Load Balancing Coverage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>E</td>
</tr>
<tr>
<td>Abilene</td>
<td>11</td>
<td>28</td>
</tr>
<tr>
<td>Geant</td>
<td>22</td>
<td>72</td>
</tr>
<tr>
<td>ISP1</td>
<td>26</td>
<td>56</td>
</tr>
<tr>
<td>ISP2</td>
<td>54</td>
<td>197</td>
</tr>
<tr>
<td>ISP3</td>
<td>110</td>
<td>340</td>
</tr>
<tr>
<td>ISP4</td>
<td>141</td>
<td>410</td>
</tr>
<tr>
<td>ISP5</td>
<td>210</td>
<td>784</td>
</tr>
<tr>
<td>ISP6</td>
<td>1167</td>
<td>4204</td>
</tr>
<tr>
<td>AS1221-Telstra</td>
<td>108</td>
<td>306</td>
</tr>
<tr>
<td>AS1239-Sprint</td>
<td>315</td>
<td>1944</td>
</tr>
<tr>
<td>AS1755-Ebone</td>
<td>87</td>
<td>322</td>
</tr>
<tr>
<td>AS3257-Tiscali</td>
<td>161</td>
<td>656</td>
</tr>
<tr>
<td>AS3967-Exodus</td>
<td>79</td>
<td>294</td>
</tr>
<tr>
<td>AS6461-AboveNet</td>
<td>141</td>
<td>748</td>
</tr>
</tbody>
</table>

Static - no reaction to traffic changes
Can we improve ECMP?

**Downstream Criteria**

A router can use a non shortest path by sending packets to a router that is closer to the destination than itself.
Can we improve ECMP?

Downstream Criteria

A router can use a non-shortest path by sending packets to a router that is closer to the destination than itself.
Can we improve ECMP?

Downstream Criteria

A router can use a non shortest path by sending packets to a router that is closer to the destination than itself.
Issues with downstream criteria

Computational cost
Naive implementation: run $k \times SPF$ on each router
Can be decreased down to about 2 SPF

Coverage

<table>
<thead>
<tr>
<th>Network name</th>
<th>Size</th>
<th>Load Balancing Coverage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>ECMP</td>
</tr>
<tr>
<td>Abilene</td>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>Geant</td>
<td>22</td>
<td>0</td>
</tr>
<tr>
<td>ISP1</td>
<td>26</td>
<td>0</td>
</tr>
<tr>
<td>ISP2</td>
<td>54</td>
<td>3</td>
</tr>
<tr>
<td>ISP3</td>
<td>110</td>
<td>10</td>
</tr>
<tr>
<td>ISP4</td>
<td>141</td>
<td>8</td>
</tr>
<tr>
<td>ISP5</td>
<td>210</td>
<td>20</td>
</tr>
<tr>
<td>ISP6</td>
<td>1167</td>
<td>22</td>
</tr>
<tr>
<td>AS1221-Telstra</td>
<td>108</td>
<td>6</td>
</tr>
<tr>
<td>AS1239-Sprint</td>
<td>315</td>
<td>22</td>
</tr>
<tr>
<td>AS1755-Ebone</td>
<td>87</td>
<td>12</td>
</tr>
<tr>
<td>AS3257-Tiscali</td>
<td>161</td>
<td>9</td>
</tr>
<tr>
<td>AS3967-Exodus</td>
<td>79</td>
<td>10</td>
</tr>
<tr>
<td>AS6461-AboveNet</td>
<td>141</td>
<td>11</td>
</tr>
</tbody>
</table>
The new feedback loops

Research issues
How to dynamically adjust splitting ratio on routers based on network-wide congestion?
Can we improve performance by dynamically adding new paths on routers?
The new feedback loops

Research issues
How to dynamically adjust splitting ratio on routers based on network-wide congestion? Can we improve performance by dynamically adding new paths on routers?
What is changing on the Internet?
What is changing on the Internet?

What is changing on end hosts?
The host-multihoming problem

More and more hosts have now two interfaces
WiFi and Ethernet
WiFi and 3G
...

Users would like to
  transparently switch over to one interface when the other fails without impacting the applications
  use both interfaces at the same time to send and receive packets faster
The complementary roles of IP addresses

The IP addresses currently used by endhosts play two complementary roles
The complementary roles of IP addresses

The IP addresses currently used by endhosts play two complementary roles

**Identifier role**: the IP address identifies (with port) the endpoint of transport flows
The complementary roles of IP addresses

The IP addresses currently used by endhosts play two complementary roles

**Identifier** role: the IP address identifies (with port) the endpoint of transport flows
The complementary roles of IP addresses

The IP addresses currently used by endhosts play two complementary roles:

- **Identifier** role: the IP address identifies (with port) the endpoint of transport flows.
- **Locator** role: the IP address indicates the paths used to reach the endhost; these paths are updated by routing protocols after each topology change.
The complementary roles of IP addresses

The IP addresses currently used by endhosts play two complementary roles:

**Identifier** role: the IP address identifies (with port) the endpoint of transport flows.

**Locator** role: the IP address indicates the paths used to reach the endhost; these paths are updated by routing protocols after each topology change.
The complementary roles of IP addresses

The IP addresses currently used by endhosts play two complementary roles:

**Identifier** role: the IP address identifies (with port) the endpoint of transport flows.

**Locator** role: the IP address indicates the paths used to reach the endhost. These paths are updated by routing protocols after each topology change.
The complementary roles of IP addresses

The IP addresses currently used by endhosts play two complementary roles:

**Identifier** role: the IP address identifies (with port) the endpoint of transport flows.

**Locator** role: the IP address indicates the paths used to reach the endhost. These paths are updated by routing protocols after each topology change.
The complementary roles of IP addresses

The IP addresses currently used by endhosts play two complementary roles:

- **Identifier** role: the IP address identifies (with port) the endpoint of transport flows.
- **Locator** role: the IP address indicates the paths used to reach the endhost. These paths are updated by routing protocols after each topology change.
Host multihoming today

State of the art in TCP/IP v4
Host multihoming today

State of the art in TCP/IPv4
Host multihoming today

State of the art in TCP/IPv4
Host multihoming today

State of the art in TCP/IPv4
Host multihoming today

State of the art in TCP/IPv4

Improvements
SCTP
Mobile IP
shim6
...

© O. Bonaventure, 2010
Multipath TCP

Application

Transport

R1

R2

RA

RB

RY

Application

Transport

© O. Bonaventure, 2010
Multipath TCP
Multipath TCP
Issues
When should a new TCP sub-flow be established?
How to schedule efficiently packets on sub-flows?
How to couple the congestion control on sub-flows without risking oscillations or low utilisation?
Interactions between multipath TCP and multipath forwarding inside the network
Multipath TCP should not be less secure than plain TCP
Incremental deployment
Conclusion

Internet is changing

The number of feedback loops is growing
Interactions among these feedback loops is not well understood and could cause problems

Traditional TCP congestion control
is being extended to allow hosts to efficiently use multiple paths

Server selection
Mainly done by proprietary algorithms
There is ongoing work for an open and standard server selection mechanism

Networks provide multiple paths
traffic distribution among available paths remains an online optimisation problem