PATHLET ROUTING

Brighten Godfrey
Scott Shenker
Ion Stoica
{pbg,shenker,istoica}@cs.berkeley.edu

UC Berkeley
Hotnets 2008
multipath internet routing

- **reliability**: source observes directly, reacts quickly
- **path quality**: source observes directly, knows what it wants
- **money**: network providers can sell new service

good for everyone!
Two routes diverged in a network, and I --
I took the one less transited,
And that has reduced latency by up to 41%.
Why don't we have it?

Even if everyone involved wants multipath, no way to do it in BGP!

All paths blocked except one.

Offer more? State explosion!
why don't we have it?

AS-level source routing...

...gives network owners no control.
why don't we have it?

Path vector (BGP)
great at blocking paths!
bad at allowing them.

AS-level source routing
great at allowing paths!
bad at blocking them.

Highly constrained routing policies.

goal: flexible policy control → many paths
pathlet routing

fragments of paths (pathlets) + source routing

goal: flexible policy control → many paths
• the protocol

• emulating other protocols

• local transit (LT) policies
1. **vnode**: virtual node within an AS

2. **pathlet**: sequence of vnodes

3. **announce** pathlets

4. **source lists** pathlets in packet

---

**example**

- **one per AS**
- **one per link**
- gossip all known pathlets to neighbors

= AS level source routing

```
128.2.0.0/16
```
- Pathlets tagged with Forwarding ID sequence
- Packet contains list of FIDs
- Forwarding table maps FID to, e.g., outgoing interface
• the protocol

• emulating other protocols

• local transit (LT) policies
emulating BGP

(not discussed: how to verify route follows advertised policy)
emulating other protocols

- **MIRO** [Xu, Rexford, SIGCOMM’06]

- **NIRA** [Yang, Clark, Berger, ToN’07]
• the protocol

• emulating other protocols

• local transit (LT) policies
Each ingress --> egress pair is either allowed or disallowed.

Subject to this, any path allowed!
LTP a common case?

- capture a network's direct costs

- valley freeness is a LT policy, and the common case in BGP export policies today
valley free routing as an LTP

“customers can route to anyone; anyone can route to customers”
## Scalability

<table>
<thead>
<tr>
<th>Forwarding Table Entries</th>
<th>BGP</th>
<th>Pathlet Routing, Class-Based LT Policies</th>
</tr>
</thead>
<tbody>
<tr>
<td>O(kn)</td>
<td>O(d+k)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Control Plane Entries</th>
<th>BGP</th>
<th>Pathlet Routing, Class-Based LT Policies</th>
</tr>
</thead>
<tbody>
<tr>
<td>O(dkn)</td>
<td>O((d+k)n)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Control Plane Messaging</th>
<th>BGP</th>
<th>Pathlet Routing, Class-Based LT Policies</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt; O(Lkn/d)</td>
<td>O(dn)</td>
<td></td>
</tr>
</tbody>
</table>

**Symbols:**
- \( n \) = number of ASes
- \( d \) = mean number of neighbors
- \( L \) = mean path len
- \( k \) = prefixes per AS
forwarding table entries

BGP one per destination (IP prefix) current Internet (CAIDA/APNIC):
266,073 entries

pathlet routing, LT policies one for each pathlet starting at the router 2,317 entries, max 6 entries, mean
• **pathlet routing:** flexible policies --> multipath with many choices, better scalability

• can't emulate everything, e.g. FBR [Zhu, Gritter, Cheriton '03]

• emulate others? path splicing [Motiwala, Elmore, Feamster, Vempala 2008], **Routing Deflections** [Yang, Wetherall 2006]

• challenge for all multipath protocols: different payment for different paths?

thanks: fonts by tom7