Update on P4 Tofino Implementation Experiences with Advanced Stateless Multicast Source Routing

Steffen Lindner, Michael Menth, Toerless Eckert

http://kn.inf.uni-tuebingen.de
Agenda

► Recap
  ▪ Motivation
  ▪ Concept

► Updates: Segment-Encoded Explicit Trees (SEET)

► Evaluation

► Conclusion
Motivation

- **BIER bitstring typically 256 bits (32 B)**
  - Not large enough to cover large networks
- **Partition BIER domain into sets ≤ 256 nodes**
  - Use set identifier (SIs) to indicate bitstrings for that set
  - Helps to scale BIER to large domains
- **But …**
  - Large BIER domains require many sets to support all BFERs
  - One packet sent per set w/ a receiver
  - Leads to redundant packet copies if receivers are in different sets

- **Experiment**
  - BIER domain with 1024 core nodes and optimized SIs
  - Average node degree 4; each core node has 16 leaves
  - BIER bitstring length of $x$ bytes
  - Every source sends BIER packet to $n$ random receivers averaged over 20 runs
  - Rel. additional packets =

$$\text{Rel. additional packets} = \frac{\# \text{Packets in BIER domain on all links} - \# \text{Packets in IPMC domain on all links}}{\# \text{Packets in IPMC domain on all links}}$$

![Graph showing relative additional packets vs. number of receivers for different bitstring sizes: 32 B, 64 B, 128 B, and 256 B. The graph illustrates how the relative additional packets increase with the number of receivers and reach a peak before dropping off.]
Concept of Advanced Stateless Multicast Source Routing (1)

Idea
- Convert distribution tree into a list and encode it in the header
  - Instead of using a flat bitstring
- No need for sets or SIs
- Send multiple packets only if header size does not suffice

Forwarding principle for replication nodes
- Partition tree in header into subtrees
- Send packet copies encapsulated in headers with single subtree to next hops
- Packet header shrinks along the tree

Update on P4 Tofino Implementation Experiences with Advanced Stateless Multicast Source Routing
Concept for Advanced Stateless Multicast Source Routing (2)

Segment-Encoded Explicit Trees (SEET)

- Use domain-wide segment IDs to address next hops
  - New concept, included in
    - draft-eckert-pim-rts-forwarding

- SIDs can have domain-wide meaning
- Allows addressing of remote nodes several hops away

Diagram:

```
0
  1, 2
  1
  3
  4
  5
```

```
<table>
<thead>
<tr>
<th>SEET Header</th>
<th>Next protocol</th>
<th>Segment #1</th>
<th>Segment #2</th>
<th>...</th>
<th>Payload</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethernet</td>
<td>Identifier</td>
<td>D</td>
<td>B</td>
<td>P</td>
<td>L</td>
</tr>
<tr>
<td></td>
<td>n bit</td>
<td>1 bit</td>
<td>1 bit</td>
<td>y bit</td>
<td>8 bit</td>
</tr>
<tr>
<td></td>
<td>Byte aligned</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```
Segment-Encoded Explicit Trees (SEET) - Recap

Forwarding operation
- Check SID in first segment identifies processing node
- Check D-bit
  - Yes: deliver a copy to processing node
- Remove first segment
- Repeat until original packet empty
  - Copy next segment and the next L byte into a new packet
  - Forward new packet according to SID in first segment
  - Remove the segment and the next L byte from original packet
► Next protocol: Identifies protocol of payload
► Identifier: SID identifier with either local or global meaning
  ▪ First segment addresses receiving node
► D: Deliver-bit → Node is a receiver of the underlying packet
► B: Bitstring Indicator → Indicates that segment is followed by a BIER(-like) bitstring
► P: Padding → Identifier + B + D + P needs to be byte-aligned
► L: Number of bytes that are destined for node represented by the identifier
Updates: Segment-Encoded Explicit Trees (SEET) (2)

If B-bit set
- Length field split into
  - Bitstring length (4 bit): length of bitstring
  - Set Identifier (4 bit): used as in BIER

Advantages
- Combines TE capabilities of SEET with efficient replication of BIER
- Less OAM overhead compared to BIER(-TE)

If B-bit set
- Length field split into
  - Bitstring length (4 bit): length of bitstring
  - Set Identifier (4 bit): used as in BIER

Advantages
- Combines TE capabilities of SEET with efficient replication of BIER
- Less OAM overhead compared to BIER(-TE)

If B-bit set
- Length field split into
  - Bitstring length (4 bit): length of bitstring
  - Set Identifier (4 bit): used as in BIER

Advantages
- Combines TE capabilities of SEET with efficient replication of BIER
- Less OAM overhead compared to BIER(-TE)

If B-bit set
- Length field split into
  - Bitstring length (4 bit): length of bitstring
  - Set Identifier (4 bit): used as in BIER

Advantages
- Combines TE capabilities of SEET with efficient replication of BIER
- Less OAM overhead compared to BIER(-TE)
Experiment

- BIER domain with 1024 core nodes and optimized SIs
- Average node degree 4; each core node has 16 leaves
- BIER bitstring length of $x$ bytes
- Every source sends packet to $n$ random receivers averaged over 20 runs
- Rel. packets of SEET compared to BIER
Conclusion

► SEET allows encoding of generic multicast tree
  ▪ Can be combined with BIER(-like) bitstrings
  ▪ Scales better than BIER(-TE) in large domains
  ▪ Does not require subdomains/set identifiers

► Enables traffic engineering hop-by-hop or on shortest paths

► Implemented on P4/Intel Tofino™ with 100 Gb/s

► Paper with extensive evaluations & implementation description on the way