Resilient P2P Multicast from the Ground Up

Stefan Birrer & Fabián E. Bustamante
Department of Computer Science
Northwestern University
www.aqualab.cs.northwestern.edu
The Need for Group Communication

- The need for group communication
  - Online gaming (e.g. www.station.sony.com)
  - Video conferencing (e.g. Access Grid)
  - Bulk data dissemination (e.g. BitTorrent)
The Need for Group Communication

- The need for group communication
  - Online gaming (e.g. www.station.sony.com)
  - Video conferencing (e.g. Access Grid)
  - Bulk data dissemination (e.g. BitTorrent)
The Need for Group Communication

- The need for group communication
  - Online gaming (e.g. www.station.sony.com)
  - Video conferencing (e.g. Access Grid)
  - Bulk data dissemination (e.g. BitTorrent)
IP Multicast as one Solution

- Router replicate messages
- Efficient group communication
End System Multicast

- But, deployment issues with IP Multicast
  - Security, scalability, ...
- Application-layer or end-system multicast
The Problem with Transiency

- Median Session Uptime, a good indicator
  - 1 hour to 1 minute [Bustamante03,Gummadi03]
Nemo - Resilient Overlay Multicast

Achieve high delivery ratio w/o paying extra - in latency, duplicates, control traffic
Nemo - Resilient Overlay Multicast

Achieve high delivery ratio w/o paying extra - in latency, duplicates, control traffic
Nemo - Resilient Overlay Multicast

Achieve high delivery ratio w/o paying extra - in latency, duplicates, control traffic

Cluster based on proximity
Nemo - Resilient Overlay Multicast

Achieve high delivery ratio w/o paying extra -
in latency, duplicates, control traffic
Nemo - Resilient Overlay Multicast

*Achieve high delivery ratio w/o paying extra - in latency, duplicates, control traffic*

Leader participates in next higher layer
Nemo - Resilient Overlay Multicast

Achieve high delivery ratio w/o paying extra in latency, duplicates, control traffic
Nemo - Resilient Overlay Multicast

Achieve high delivery ratio w/o paying extra - in latency, duplicates, control traffic
Nemo's Data Forwarding

time
Nemo's Data Forwarding

2\textsuperscript{nd} Layer Leader
Nemo's Data Forwarding
Nemo's Data Forwarding

time
Peer Failure
Peer Failure

- Co-Leader shares forwarding responsibility with Leader
Peer Failure
Peer Failure

2\textsuperscript{nd} Layer Co-leader
Peer Failure

Forwarding alternates among Co-leaders
Peer Failure

time
Evaluation

● Measure effectiveness of protocol: Delivery ratio

● Cost of resilience: Latency and duplicate packets

● Methodology
  – Peers join the session in the warmup time
  – One publisher streams data

● Compare against
  – Nice [Banerjee02], Nice-PRM [Banerjee03], and Narada [Chu02]
## Benefits & Costs

**High Churn** *(MTTF 5')*

512 end hosts

<table>
<thead>
<tr>
<th>Protocol</th>
<th>Delivery [%]</th>
<th>Duplicates [packets/SeqNr]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nemo</td>
<td><strong>0.998</strong></td>
<td><strong>3.16</strong></td>
</tr>
<tr>
<td>Nice PRM(3,0.01)</td>
<td>0.993</td>
<td>12.47</td>
</tr>
<tr>
<td>Nice PRM(3,0.02)</td>
<td>0.994</td>
<td>18.20</td>
</tr>
<tr>
<td>Nice PRM(3,0.03)</td>
<td>0.994</td>
<td>24.22</td>
</tr>
<tr>
<td>Nice</td>
<td>0.992</td>
<td>7.10</td>
</tr>
<tr>
<td>Narada</td>
<td>0.852</td>
<td>0.00</td>
</tr>
</tbody>
</table>

*Best delivery ratio*
## Wide-Area Results

### High Churn (MTTF 5')

~72 end hosts

<table>
<thead>
<tr>
<th>Protocol</th>
<th>Delivery [%]</th>
<th>Duplicates [packets/SeqNr]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nemo</td>
<td>0.979</td>
<td>1.27</td>
</tr>
<tr>
<td>Nice PRM(3,0.02)</td>
<td>0.953</td>
<td>2.02</td>
</tr>
<tr>
<td>Nice</td>
<td>0.939</td>
<td>1.06</td>
</tr>
</tbody>
</table>

Best delivery ratio
Benefit & Cost

Nemo, Nice and Nice PRM overlap

No extra latency

Delivery Ratio [%]

Latency [s]
Conclusions

- Multicast for efficient group communication
  - Transiency can get in the way
- **Co-leaders offer a simple yet effective solution**
  - Improve resilience
  - Spread the load
- Nemo – Resilient overlay multicast
  - 14.6% higher delivery ratio than Narada
  - 50%-85% less Duplicates than Nice & Nice PRM
  - Comparable end-to-end latency
Nemo: Resilient Overlay Multicast
## Benefit & Cost

### Low Churn (MTTF 60')
512 end hosts

<table>
<thead>
<tr>
<th>Protocol</th>
<th>Delivery [%]</th>
<th>Duplicates [packets/SeqNr]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nemo</td>
<td>1.000</td>
<td>0.34</td>
</tr>
<tr>
<td>Nice PRM(3,0.01)</td>
<td>0.999</td>
<td>6.42</td>
</tr>
<tr>
<td>Nice PRM(3,0.02)</td>
<td>0.999</td>
<td>12.00</td>
</tr>
<tr>
<td>Nice PRM(3,0.03)</td>
<td>0.999</td>
<td>16.74</td>
</tr>
<tr>
<td>Nice</td>
<td>0.999</td>
<td>1.29</td>
</tr>
<tr>
<td>Narada</td>
<td>0.950</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Best delivery ratio

---

MTTF = Mean Time To Failure
Delivery Ratio under Churn

High Churn, 512 End Hosts

- Narada
- Nice
- Nice PRM(3,0.01)
- Nemo

Packet Sequence Number

Delivery Ratio [%]
Related Work

- Overlay multicast
  - Nice (Banerjee02)
  - ESM (Chu00, ...), Yoid (Francis00), ALMI (Pendarakis01), ...

- Resilient multicast
  - A lot of work on resilient IP Multicast
  - PRM - Probabilistic Resilient Multicast for Overlay (Banerjee03)

- Content Dissemination
  - Bullet (Kostic03)
  - SplitStream (Castro03)
  - BitTorrent (Cohen03)