High-Performance Many-Core Networking: Design and Implementation

Jordi Ros-Giralt, PhD
giralt@reservoir.com

Innovating the Network for Data-Intensive Science (INDIS)
Nov 16th 2015

Reservoir Labs
632 Broadway
Suite 803
New York, NY 10012
TOC:
• Trade-Off Generality versus Performance
• Problem space: performance | algorithms
• DNAC: Dynamic Network Acceleration for many-Core
• Performance benchmark
• Providing HPC network visibility at SC15 (SCinet)
Trade-Off Generality Versus Performance

Layer n

(...)

Layer 3

Layer 2

Layer 1

Generality

Performance
Network Stack Packet Flow

Standard Network Stack:
Network Stack Packet Flow

Standard Network Stack:

HPC Optimized Network Stack:
Two orthogonal problems:

- High performance computing (performance)
- Dynamic packet forwarding (algorithms)
Two orthogonal problems:

- High performance computing (performance)
- Dynamic packet forwarding (algorithms)

DNAC: Dynamic Network Acceleration for multi-Core
<table>
<thead>
<tr>
<th>Feedback type</th>
<th>Description</th>
<th>Controller type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Packet congestion</td>
<td>Packets are dropped by the application as a consequence of the application being congested.</td>
<td>closed-loop control</td>
</tr>
<tr>
<td>Packet relevance</td>
<td>Packets are dropped by the application as a consequence of the packet not being relevant to the application.</td>
<td>open- or closed-loop control</td>
</tr>
</tbody>
</table>
(1) Packet DMA to memory
(2) Forwarder processes packet
(3) Forwarder passes packet to application
(4) Forwarder drops packet
(5) Closed loop feedback from ring
(6) Closed loop feedback from application
(7) Open loop configuration
Feedback type: packet congestion | controller type: closed-loop

**Algorithm:** *TED Queuing (Tail early dropping).* A queuing policy that, upon congestion, drops packets from the tails of each connection, preserving the heads, by dynamically computing a connection cut threshold.

Feedback type: packet relevancy | controller type: closed-loop

**Algorithm:** *Packet shunting.* Bro workers communicate packet shunting decisions to the forwarder.

Feedback type: packet relevancy | controller type: open-loop

**Algorithm:** *Packet prioritization.* Certain packets carry higher degrees of information. Example: dropping a FIN packet has both semantic implications at the protocol level and impacts performance as the upper layer needs to rely on expensive timeouts to free connection context.
DNAC: Performance Evaluation
R-Scope Appliance

- High-performance network analyzer running Bro*

<table>
<thead>
<tr>
<th>R-Scope Appliance Specifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td>Operational Mode</td>
</tr>
<tr>
<td>Maximum Throughput</td>
</tr>
<tr>
<td>Traffic Interfaces</td>
</tr>
<tr>
<td>Dimensions (H x W x D)</td>
</tr>
<tr>
<td>Weight</td>
</tr>
<tr>
<td>Power</td>
</tr>
<tr>
<td>Support</td>
</tr>
</tbody>
</table>

* Bro project: www.bro.org

Reservoir Labs
DNAC: Performance Evaluation

**Input rate 500Mbps**
- **http.log**: DNAC: 42000, no DNAC: 34000
- **files.log**: DNAC: 38000, no DNAC: 34000
- **http_track.log**: DNAC: 30000, no DNAC: 10000

**Input rate 5Gbps**
- **http.log**: DNAC: 40000, no DNAC: 15000
- **files.log**: DNAC: 38000, no DNAC: 13000
- **http_track.log**: DNAC: 30000, no DNAC: 10000
### Table 3. Number of events detected

<table>
<thead>
<tr>
<th></th>
<th>500Mbps input rate</th>
<th></th>
<th>5Gbps input rate</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>http</td>
<td>files</td>
<td>http_track</td>
<td>http</td>
</tr>
<tr>
<td>w/ DNAC</td>
<td>42449</td>
<td>39594</td>
<td>31300</td>
<td>38425</td>
</tr>
<tr>
<td>w/o DNAC</td>
<td>42434</td>
<td>39314</td>
<td>31200</td>
<td>15314</td>
</tr>
<tr>
<td>gain</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2.5</td>
</tr>
</tbody>
</table>
DNAC: Performance Evaluation

![Graph showing network performance metrics with and without DNAC](image)

- Packets received with DNAC
- Packets dropped with DNAC
- Packets forwarded without DNAC
- Packets forwarded with DNAC

*Reservoir Labs*
Trade-Off Generality Versus Performance

- Layer n
- Layer 3
- Layer 2
- Layer 1

Generality vs Performance:

- L_n
- L_3
- L_2
- L_1

Native API + dynamic packet control

Native API

mmap

General purpose OS
Providing Network Visibility at SC2015 (SCInet)

6 x R-Scope systems (6 x 10 Gbps) providing real time Bro* analytics

* Bro project: www.bro.org

Reservoir Labs
Providing Network Visibility at SC2015 (SCinet)

Sunday, November 15, 2015

World's Most Powerful Computer Network is Live in Austin

The SCinet network, SC's Supercomputing Internet, is now live! On November 14, the Austin Convention Center became home to the fastest and most innovative computer network in the world, delivering more than 1.6 terabits per second of network bandwidth to the SC conference (SC15).
Thank You

**Reservoir Labs**

632 Broadway  
Suite 803  
New York, NY 10012

812 SW Washington St.  
Suite 1200  
Portland, OR 97205