Experimenting TCP Performance with FABRIC

Ewa Deelman¹, Anirban Mandal², Prasanna Balaprakash⁵, Mariam Kiran⁵, Krishnan Raghavan³, Hongwei Jin³, Cong Wang², Komal Thareja², Imtiaz Mahmud⁴, George Papadimitriou¹

¹University of Southern California, ²Renaissance Computing Institute, ³Argonne National Laboratory, ⁴Lawrence Berkeley National Laboratory, ⁵Oak Ridge National Laboratory
Meet the Team

Ewa Deelman
USC (Lead PI)

Anirban Mandal
RENCI (Co-PI)

Prasanna Balaprakash
ORNL (Co-PI)

Mariam Kiran
ORNL (Co-PI)

George Papadimitriou
USC

Cong Wang
RENCI

Krishnan Raghavan
ANL

Imtiaz Mahmud
LBNL

Komal Thareja
RENCI

Hongwei Jin
ANL

http://poseidon-workflows.org
What are we doing?
Improving Network Performance for our data transfers

TCP Algorithms CCA?
AQM (Queueing) Algorithm?

Reliable data delivery?

Novel Innovations for Community:
● Interplay between Transfer and queuing in routers
● Build dataset for transfer protocols behavior
● Lead to a NEW transfer methods for “better and fair Internet”
Experimental Setup on FABRIC

RTT $\approx 62$ms
Experimental Setup on FABRIC

**Scenarios**

<table>
<thead>
<tr>
<th>CCA 1 - CCA 2</th>
<th>AQM</th>
<th>Queue Length</th>
<th>Bottleneck BW</th>
</tr>
</thead>
<tbody>
<tr>
<td>BBRv1 - CUBIC</td>
<td>FIFO</td>
<td>0.5 x BDP</td>
<td>100 Mbps</td>
</tr>
<tr>
<td>BBRv2 - CUBIC</td>
<td>FIFO</td>
<td>1 x BDP</td>
<td>500 Mbps</td>
</tr>
<tr>
<td>BBRv3 - CUBIC</td>
<td>FIFO</td>
<td>2 x BDP</td>
<td>1 Gbps</td>
</tr>
<tr>
<td>HTCP - CUBIC</td>
<td>FQ Codel</td>
<td>4 x BDP</td>
<td>10 Gbps</td>
</tr>
<tr>
<td>Reno - CUBIC</td>
<td>RED</td>
<td>8 x BDP</td>
<td>25 Gbps</td>
</tr>
<tr>
<td>CUBIC - CUBIC</td>
<td>RED</td>
<td>16 x BDP</td>
<td></td>
</tr>
<tr>
<td>BBRv1 - BBRv1</td>
<td>FIFO</td>
<td>0.5 x BDP</td>
<td>100 Mbps</td>
</tr>
<tr>
<td>BBRv2 - BBRv2</td>
<td>FIFO</td>
<td>1 x BDP</td>
<td>500 Mbps</td>
</tr>
<tr>
<td>BBRv3 - BBRv3</td>
<td>FIFO</td>
<td>2 x BDP</td>
<td>1 Gbps</td>
</tr>
<tr>
<td>HTCP - HTCP</td>
<td>FQ Codel</td>
<td>4 x BDP</td>
<td>10 Gbps</td>
</tr>
<tr>
<td>Reno - Reno</td>
<td>RED</td>
<td>8 x BDP</td>
<td>25 Gbps</td>
</tr>
</tbody>
</table>

**Iperf3 Configuration**

<table>
<thead>
<tr>
<th>Bottleneck BW</th>
<th>Total #Flows</th>
<th>iperf3 Configuration</th>
</tr>
</thead>
<tbody>
<tr>
<td>100 Mbps</td>
<td>2</td>
<td>1 iperf3 process/node 1 stream</td>
</tr>
<tr>
<td>500 Mbps</td>
<td>10</td>
<td>5 iperf3 processes/node 1 stream each</td>
</tr>
<tr>
<td>1 Gbps</td>
<td>20</td>
<td>10 iperf3 processes/node 1 stream each</td>
</tr>
<tr>
<td>10 Gbps</td>
<td>200</td>
<td>10 iperf3 processes/node 10 parallel streams each</td>
</tr>
<tr>
<td>25 Gbps</td>
<td>500</td>
<td>25 iperf3 processes/node 10 parallel streams each</td>
</tr>
</tbody>
</table>
Experimental Setup on FABRIC

**Pseudocode**

```plaintext
foreach aqm_type:
  foreach cca_config:
    foreach speed_config:
      Calculate bdp
      Calculate buffer_size list based on BDP
    foreach buffer_size:
      Apply aqm and buffer_size
      for 1..5:
        Start iperf3 servers
        Start iperf3 clients for 200 seconds
        Wait for 210 seconds
        Kill iperf3 servers
```

**BDP Formula**

\[
BDP = \frac{BW_{bottleneck} \times RTT}{8 \times \text{bytcs}}
\]

**iPerf3 Server CMD:**

```
iPerf3 Server CMD: iperf3 -s -p {port} -f m
```

**iPerf3 Client CMD:**

```
iPerf3 Client CMD: iperf3 -c {server_ip} -p {port_num} -C {cca} -t 200 -f m -P {flows} -M 8900
```

**tc qdisc add**

```
tc qdisc add dev {iface} root handle 1: tbf rate {speed} burst 1570000 limit {applied_bdp}
tc qdisc add dev {iface} parent 1:1 handle 10: fq_codel limit {packets}
```
Generated Figures

BBRv1 vs CUBIC
8 BDP buffer size
AQM = FIFO
Bandwidth = 500 Mbps

http://poseidon-workflows.org
Grafana dashboard
Summary

- Details of Implementation and Code Access
  - Full Fabric implementation - automated code.
  - Scripts for processing pcap files and extracting necessary data.
  - A script to process iperf data and compile information for each flow across experiments.
  - Tools to create both PNG and interactive HTML visualizations for the collected data.
- Access to the gathered data, figures, and raw files.
- A grafana dashboard for easy data access and visualization.
Acknowledgements

DOE ASCR Award (DE-SC0022328): Integrated Computational and Data Infrastructure (ICDI) Program