Analyzing the Influence of Resource Prioritization on HTTP/3 HOL Blocking and Performance

Constantin Sander, Ike Kunze, Klaus Wehrle

{sander, kunze, wehrle}@comsys.rwth-aachen.de
- **HTTP/2**: multiplexing via TCP connection
  - TCP unaware of streams, transport HOL blocking

- **HTTP/3**: multiple QUIC streams
  - Independent, no inter-stream HOL blocking
Motivation

- **HTTP/2**: multiplexing via TCP connection
  - TCP unaware of streams, transport HOL blocking

- **HTTP/3**: multiple QUIC streams
  - Independent, no inter-stream HOL blocking
• **Resource Prioritization: Browser signals server preferred scheduling**
  - E.g., send HTML first, then images
  - Different prioritization strategies per browser

<table>
<thead>
<tr>
<th>Prioritization Strategies</th>
<th>Chrome (sequential)</th>
<th>Firefox (WRR+sequential)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Round Robin</td>
<td>worst</td>
<td>best</td>
</tr>
<tr>
<td>Weighted Round Robin</td>
<td>worst</td>
<td>better</td>
</tr>
</tbody>
</table>
Background & Related Work: Resource Prioritization

- Resource Prioritization: Browser signals server preferred scheduling
  - E.g., send HTML first, then images
  - Different prioritization strategies per browser

<table>
<thead>
<tr>
<th>Prioritization Strategies</th>
<th>Wijnants et al. HTTP/2 Performance</th>
<th>Marx et al. HTTP/3 Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Round Robin</td>
<td>worst</td>
<td>worst</td>
</tr>
<tr>
<td>Weighted Round Robin</td>
<td>worst</td>
<td>worst</td>
</tr>
<tr>
<td>Chrome (sequential)</td>
<td>best</td>
<td>better</td>
</tr>
<tr>
<td>Firefox (WRR+sequential)</td>
<td>best</td>
<td>better</td>
</tr>
</tbody>
</table>
### Background & Related Work: Resource Prioritization

**Resource Prioritization: Browser signals server preferred scheduling**

- E.g., send HTML first, then images
- Different prioritization strategies per browser

<table>
<thead>
<tr>
<th>Prioritization Strategies</th>
<th>Wijnants et al. HTTP/2 Performance</th>
<th>Marx et al. HTTP/3 Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Round Robin</td>
<td>worst</td>
<td>worst</td>
</tr>
<tr>
<td>Weighted Round Robin</td>
<td>worst</td>
<td>worst</td>
</tr>
<tr>
<td>Chrome (sequential)</td>
<td>best</td>
<td>better</td>
</tr>
<tr>
<td>Firefox (WRR+sequential)</td>
<td>best</td>
<td>better</td>
</tr>
</tbody>
</table>

Website specific

E.g., send HTML first, then images

Different prioritization strategies per browser
Background & Related Work: Resource Prioritization

- **Resource Prioritization: Browser signals server preferred scheduling**
  - E.g., send HTML first, then images
  - Different prioritization strategies per browser

<table>
<thead>
<tr>
<th>Prioritization Strategies</th>
<th>Wijnants et al. HTTP/2 Performance</th>
<th>Marx et al. HTTP/3 Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Round Robin</td>
<td>worst</td>
<td>worst</td>
</tr>
<tr>
<td>Weighted Round Robin</td>
<td>worst</td>
<td>worst</td>
</tr>
<tr>
<td>Chrome (sequential)</td>
<td>best</td>
<td>better</td>
</tr>
<tr>
<td>Firefox (WRR+sequential)</td>
<td>best</td>
<td>better</td>
</tr>
</tbody>
</table>

Worst for HOL?
### Background & Related Work: Resource Prioritization

- **Resource Prioritization: Browser signals server preferred scheduling**
  - E.g., send HTML first, then images
  - Different prioritization strategies per browser

<table>
<thead>
<tr>
<th>Prioritization Strategies</th>
<th>Wijnants et al. HTTP/2 Performance</th>
<th>Marx et al. HTTP/3 Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Round Robin</td>
<td>worst</td>
<td>worst</td>
</tr>
<tr>
<td>Weighted Round Robin</td>
<td>worst</td>
<td>worst</td>
</tr>
<tr>
<td>Chrome (sequential)</td>
<td>best</td>
<td>better</td>
</tr>
<tr>
<td>Firefox (WRR+sequential)</td>
<td>best</td>
<td>better</td>
</tr>
</tbody>
</table>

HTTP/2 ≠ HTTP/3

Worst for HOL?

Premature QUIC stack constant rate / no loss
• Evaluate impact of prioritization on HTTP/3 performance under loss
  ▶ Change Loss, Loss Burstsize, RTT, Bandwidth
  ▶ Test (W)RR, Chrome, Firefox, Firefox (EPS adapted)
    ■ Identify HOL blocking + performance
Contribution

- **Evaluate impact of prioritization on HTTP/3 performance under loss**
  - Change Loss, Loss Burstsize, RTT, Bandwidth
  - Test (W)RR, Chrome, Firefox, Firefox (EPS adapted)
    - Identify HOL blocking + performance

<table>
<thead>
<tr>
<th>Loss</th>
<th>RTT</th>
<th>BW</th>
<th>Loss Burstsize</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
<td>10ms</td>
<td>1Mbps</td>
<td>1</td>
</tr>
<tr>
<td>1%</td>
<td>50ms</td>
<td>2Mbps</td>
<td>5</td>
</tr>
<tr>
<td>2%</td>
<td>100ms</td>
<td>5Mbps</td>
<td>10</td>
</tr>
<tr>
<td>5%</td>
<td>10Mbps</td>
<td>10Mbps</td>
<td>15</td>
</tr>
</tbody>
</table>

- Download 35 websites (from 1 & 2)
- Replay websites (30 times per setting)
- Measure SpeedIndex & HOL bytes

![Diagram of website prioritization and performance metrics]
Results – Bandwidth Influence

• Δ: Relative median difference to Chrome / sequential baseline

• HOL: (lower=better)
  ▶ Reduced with parallelism
  ▶ Vanishing differences for higher bandwidths

• SpeedIndex: (lower=faster/better)
  ▶ Fewer benefits of parallelism for higher bandwidth / cwnd
Results – Loss Influence

- **HOL:**
  - Less HOL blocking for higher loss (as expected)
  - Loss stopping many streams for sequential scheduling

- **SpeedIndex:**
  - Growing benefits for higher loss
  - Not as strongly as for HOL
Results – Correlation Loss

Correlation $\Delta$SpeedIndex and $\Delta$HOL

- $2\text{Mbps BW}$
- $100\text{ms RTT}$
- $1\text{BDP Queue}$

Sites:
- wikipedia.org
- nytimes.com
Results – Correlation Loss

- **Correlation SpeedIndex and HOL:**
  - Negative for smaller websites
    - HOL reduced, but only slightly
    - Negative effect of parallel prioritization
  - Positive for larger websites
    - HOL reduced more strongly
    - Negative effect outweighed
Conclusion

- **Reduced HOL Blocking via QUIC**
  - Multiple streams need to be active in parallel
  - HTTP Prioritization influences active streams: use Round Robin
  - Related Work: Round Robin detrimental for performance

- **New performance interplay between prioritization and network**
  - Round Robin can improve HOL and thus performance
  - Mainly for large websites / small BW / high RTT / random loss
  - No strong difference when using EPS

- **Overall: HTTP/3 prioritization still website + now also network dependent**