Quantitative Performance Comparison of Various Individual and Combined Traffic Shapers in Time-Sensitive Networking

Luxi Zhao
Paul Pop
Sebastian Steinhorst
Necessity and Contributions

► Necessity
- Set of substandards (flow control):
  - 802.1Qbv – Time Aware Shaper (TAS);
  - 802.1Qav – Credit Based Shaper (CBS);
  - 802.1Qcr – Asynchronous Traffic Shaper (ATS);
  - 802.1Q – 2005 – Strict Priority (SP);
- Combinations …
- Independent studies;
- No quantitative comparison;
- Proper shapers selection – tricky

► Contributions
- Tutorial of NC-based analysis for TSN;
- Two new combined architectures (TAS+ATS+CBS, TAS+ATS+SP); extend NC approach;
- Plenty of quantitative comparison → surprising but interesting results;
- Provide a basis, select the suitable TSN traffic shapers.
Architecture – Individual Traffic Shapers

- IEEE 802.1Qbv – Time Aware Shaper (TAS);
- Global network clock synchronization (IEEE 802.1ASrev);
- Time-Triggered communication – GCL synthesis – Schedulability guarantee;
- GCL synthesis – NP-complete problem [1], [2].

Evaluation Parameters
- Schedulability – End-to-end latency bound
- Buffer size without frame loss – Backlog bound
- Stable Communication – Jitter bound

1. 802.1Qbv
2. Scheduling Synthesis
Architecture – Individual Traffic Shapers

1. 802.1Qbv
2. Scheduling Synthesis

IEEE 802.1Qbv – Individual Traffic Shapers

1. 802.1Qav
2. Network Calculus

IEEE 802.1Qav – Credit Based Shaper (CBS);
Allocate the bandwidth reservation for different classes (priority)
CBS algorithm – credit value (idleSlope / sendSlope)
– non-work conserving;
Schedulability guarantee – Network Calculus [3];

Evaluation Parameters

- Schedulability – End-to-end latency bound
- Buffer size without frame loss – Backlog bound
- Stable Communication – Jitter bound

Individual Traffic Shapers

1. 802.1Q - 2005
2. Network Calculus

IEEE 802.1Q - 2005 – Strict Priority (SP);
Low priority traffic can transmit only when the high priority queue is empty;
Schedulability guarantee – Network Calculus [4];
Architecture – Individual Traffic Shapers

- 1. 802.1Qbv
- 2. Scheduling Synthesis

Evaluation Parameters
- Schedulability – End-to-end latency bound
- Buffer size without frame loss – Backlog bound
- Stable Communication – Jitter bound

IEEE 802.1Qcr – Asynchronous Traffic Shaper (ATS);
- Asynchronous transmission, local clock;
- Two hierarchies of queues – shaped queue & shared queue;
- ATS – interleaved regulator – avoid burstiness cascades;
- Schedulability guarantee – Closed-form Formula [5], Network Calculus [6].
Evaluation – Individual Traffic Shapers

- Synthetic test cases – SRM, MR, MM, ST, MT
  - Each topology – 100 TCes;
  - Frame size – minimum (64 bytes) ~ maximum (1522 bytes);
  - Period (periodic) / Min time interval (sporadic) – T={1000, 2000, 5000, 10000} (μm);
  - 1 priority;
  - GCLs for TAS, Route – existing work [2];
  - Physical link rate 100 Mb/s.

### TABLE II

<table>
<thead>
<tr>
<th></th>
<th>SRM</th>
<th>MR</th>
<th>MM</th>
<th>ST</th>
<th>MT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Hops</td>
<td>2.7</td>
<td>4.2</td>
<td>3.8</td>
<td>3.5</td>
<td>5.5</td>
</tr>
<tr>
<td>Average Traffic Load</td>
<td>28.9%</td>
<td>20.5%</td>
<td>17.4%</td>
<td>29.0%</td>
<td>19.7%</td>
</tr>
<tr>
<td>Max Traffic Load</td>
<td>47%</td>
<td>40%</td>
<td>38%</td>
<td>47%</td>
<td>30%</td>
</tr>
<tr>
<td>Min Traffic Load</td>
<td>13%</td>
<td>8%</td>
<td>6%</td>
<td>13%</td>
<td>10%</td>
</tr>
</tbody>
</table>
Evaluation – Individual Traffic Shapers (1)

► Results
  ▶ Each TC,
    1) End-to-end latency upper bounds – flow;
    2) Backlog upper bounds – egress port;
    3) Jitter bounds – flow.
  ▶ Figure: each TC – metric – average value;
    100 TC – 100 dots - box plot.

► Comments
  ▶ Different topologies – similar trends while comparing different traffic shapers;
  ▶ TAS performs the best – latency, backlog, jitter;
  ▶ …
Evaluation – Individual Traffic Shapers (1)

► Results
  ► Table: $X_i$ metric value for flow $f_i$;
  \[
  \bar{X} = \text{aver}(X_i) = \text{aver}\left(\frac{X_{i1} - X_{i2}}{Y_{i1}}\right)
  \]

► Comments
  ► Latency & jitter – SP < CBS < ATS;
  Backlog – CBS < SP < ATS;
  ► Advantage of ATS ↓ \leftrightarrow concentration of flows ↑
  \leftrightarrow number of hops ↑
  ► ...

<table>
<thead>
<tr>
<th></th>
<th>(CBS – SP) /SP</th>
<th>(ATS – SP) /SP</th>
<th>(ATS – CBS) /CBS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average WCD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRM</td>
<td>30.7%</td>
<td>34.1%</td>
<td>2.8%</td>
</tr>
<tr>
<td>MR</td>
<td>28.2%</td>
<td>43.9%</td>
<td>12.5%</td>
</tr>
<tr>
<td>MM</td>
<td>26.2%</td>
<td>35.7%</td>
<td>7.7%</td>
</tr>
<tr>
<td>ST</td>
<td>31.2%</td>
<td>72.3%</td>
<td>31.4%</td>
</tr>
<tr>
<td>MT</td>
<td>27.3%</td>
<td>79.1%</td>
<td>40.8%</td>
</tr>
<tr>
<td>Average WCB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRM</td>
<td>-1.3%</td>
<td>38.0%</td>
<td>39.9%</td>
</tr>
<tr>
<td>MR</td>
<td>-2.5%</td>
<td>34.7%</td>
<td>38.1%</td>
</tr>
<tr>
<td>MM</td>
<td>-3.2%</td>
<td>31.0%</td>
<td>35.4%</td>
</tr>
<tr>
<td>ST</td>
<td>-0.8%</td>
<td>65.0%</td>
<td>65.5%</td>
</tr>
<tr>
<td>MT</td>
<td>-3.3%</td>
<td>64.4%</td>
<td>70.0%</td>
</tr>
<tr>
<td>Average WCJ</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRM</td>
<td>37.1%</td>
<td>41.4%</td>
<td>3.4%</td>
</tr>
<tr>
<td>MR</td>
<td>38.0%</td>
<td>60.7%</td>
<td>16.6%</td>
</tr>
<tr>
<td>MM</td>
<td>30.3%</td>
<td>43.0%</td>
<td>10.4%</td>
</tr>
<tr>
<td>ST</td>
<td>39.5%</td>
<td>91.9%</td>
<td>37.7%</td>
</tr>
<tr>
<td>MT</td>
<td>37.3%</td>
<td>108.3%</td>
<td>51.9%</td>
</tr>
</tbody>
</table>
Evaluation – Individual Traffic Shapers (1)

Comments

- ATS positive effect ↓ ↔ concentration of flows ↑
  ↔ number of hops ↑
- For example
  1. Flows concentration: MR > MM →
     ATS positive effect: MR < MM
  2. Number of hops: ST > SRM →
     ATS positive effect: ST < SRM

Table 1
Evaluation – Individual Traffic Shapers (2)

► Synthetic test cases – MM
  ► Case 1 – 1 priority
  ► Average traffic load: 10% ~ 90%
  ► Each traffic load – 20 Tces

► Results
  ► \[ X = \text{aver}(X_i) = \text{aver}(\frac{(X_i^{SP} - X_i^{ATS})}{X_i^{ATS}}) \]

► Comments
  ► End-to-end latency bounds:
    – Average traffic load \( \uparrow \) \( \rightarrow \) comparison percentage \( X \uparrow \rightarrow \)
      ATS positive effect \( \uparrow \);
    – Average traffic load < 70% – SP performs better than ATS;
  ► Backlog bounds:
    – Average traffic load < 70% – unobvious change;
    – ATS performance is always inferior to SP.
  ► ...

Figure 2

ATS vs. without ATS (i.e., SP)
Evaluation – Individual Traffic Shapers (3)

► Synthetic test cases – MM
► Case 2 – 2 priorities + BE
► Average traffic load: 10% ~ 90%; High/low: 50% of overall traffic load
► Each traffic load – 20 Tces

► Results
► \[ X = \text{aver}(X_i) = \text{aver}\left(\left(\frac{X_i^{\text{SP}} - X_i^{\text{ATS}}}{X_i^{\text{ATS}}}\right)\right) \]

► Comments
► End-to-end latency bounds:
  High priority:
  – \approx Top 40% traffic load with single priority;
  – ATS no positive effect on high-priority traffic.
  Low-priority:
  – ATS positive effect on low-priority traffic \iff average overall traffic load > 30%.
  
  …

ATS vs. without ATS (i.e., SP)

Figure 3
Evaluation – Individual Traffic Shapers (3)

► Synthetic test cases – MM
  ► Case 3 – 2 priorities + BE
  ► Average traffic load: 10% ~ 90%; High/low: 50% of overall traffic load
  ► Each traffic load – 20 Tces

► Results
  ► $X = \text{aver}(X_i) = \text{aver}\left(\frac{X_i^{SP} - X_i^{ATS}}{X_i^{ATS}}\right)$

► Comments
  ► Backlog bounds:
    High priority:
    – ≈ Top 40% traffic load with single priority + BE;
    Low-priority:
    – ATS positive effect on low-priority traffic.

ATS vs. without ATS (i.e., SP)

Figure 3
Architecture – Combined Traffic Shapers

1. 802.1Qbv+802.1Q - 2005
2. Network Calculus

NEW
1. 802.1Qbv+802.1Qr
2. Network Calculus

► New: TAS+ATS(+SP); TAS+ATS+CBS
► TAS+SP vs. TAS+ATS(+SP)
► TAS+CBS vs. TAS+ATS+CBS

NEW: Tas+ATS(+SP); Tas+ATS+CBS

► TAS outperforms than all the others (latency, backlog, jitter);
Scalability problem.
► Promising combination model: TAS+X
Evaluation – Combined Traffic Shapers (1)

► Synthetic test cases – MM
  ► Case 1 – TT traffic load: 20%;
    – SP average traffic load: 10% ~ 70%
  ► Each traffic load – 20 TCes

► Results
  ► \( X = \text{aver}(X_i) = \text{aver}\left(\frac{X_i^{SP} - X_i^{ATS+SP}}{X_i^{ATS+SP}}\right) \)

► Comments
  ► With the influence of TT traffic (TAS)
  ► End-to-end latency bounds:
    – Average traffic load ↑ – ATS positive effect ↑;
    – ATS positive effect ← average overall traffic load > 40%;
  ► Backlog bounds:
    – ATS positive effect ← average overall traffic load > 30%;
Evaluation – Combined Traffic Shapers (2)

Realistic Test Cases – Orion CEV

- 31 ESeS, 15 SWs, 188 dataflow routes, 100 Mbps link rate;
- 99 TT flows (TAS), 87 rate constraint flows with the same priority → SP flows / AVB flows (CBS);
- TT traffic load in network → 1.5% on average & 5.5% in maximum.
- Overall traffic load in network → 3.5% on average & 10% in maximum.
- IdleSlope for AVB is set to 75% (default);

Results

- $100 \times \ln(X), X = (WCD, WCB)$;
- Sorted in increasing order by results (WCD, WCB).
Evaluation – Combined Traffic Shapers (2)

► Comments

► End-to-end latency bounds:
  – ATS no positive effect.
► Backlog bounds:
  – ATS positive effect.

← Average overall traffic load (TT, SP/AVB) low;
→ Consistent with results in Fig. 4.

Figure 5
Evaluation – Combined Traffic Shapers (3)

► Realistic Test Cases – Orion CEV

► Increase traffic load:
   TT traffic load in network
   → 15% on average & 54% in maximum.
   Overall traffic load in network
   → 25% on average & 69% in maximum.

► 4 priorities, 25 flows of P1, 25 flows of P2, 24 flows of P3, 13 flows of P4;

► IdleSlope ← actual bandwidth utilization,
  \[ idS_{l_i} = \frac{operIdleSlope(P_i)}{GateOpenTime} \]

► Results

► \( 100 \times \ln(X), X = (WCD, WCB) \);

► Sorted in increasing order by results (WCD, WCB).
Evaluation – Combined Traffic Shapers (3)

► Comments

► End-to-end latency bounds:
  Backlog bounds:
  – ATS positive effect.

► ATS positive effect
  TAS+ATS+CBS > TAS+ATS+SP
  ← service capability AVB < SP.

► In combination of ATS
  → performance of SP & CBS get closer to each other.
Conclusion

- **SP vs. CBS**
  - SP is more beneficial to the transmission delay of high-priority traffic, while CBS can specify bandwidth reservation for each priority traffic;
  - Due to the credit controlling by CBS, the long-term rate of AVB traffic arrival is reduced, backlog bounds of AVB traffic are possible lower than SP traffic.

- **ATS vs. SP, CBS**
  - ATS has limited advantages for high-priority traffic;
  - Only when the average traffic load of high priority traffic exceeds a certain value (around 70% in MM for example), ATS can show its superiority;
  - The positive effect of ATS on low priority traffic is more obvious.

- **TAS vs. SP, CBS, ATS**
  - TAS implement flow-based TT scheduling, has the highest performance with ultra low latency, jitter and backlog;
  - TAS requires the synthesis of optimized GCLs, to which is difficult to scale to large networks with many flows.

- **TAS+ATS+X vs. TAS+X**
  - Combined use of ATS with TAS will make ATS play a more active role, of which the effect is similar to the reshaping impact of ATS used individually on low priority traffic.
  - TAS will maintain unchanged its advantages of ultra low latency and jitter.
Bibliography

Scheduling real-time communication in IEEE 802.1 Qbv time sensitive networks.
In 24th International Conf. on Real-Time Networks and Systems (RTNS), 2016.

Design optimisation of cyber-physical distributed systems using IEEE time-sensitive networks.

Complete modelling of AVB in network calculus framework.
In 22nd International Conf. on Real-Time Networks and Systems, 2014.

Per-flow guarantees under class-based priority queueing.
In IEEE Global Telecommunications Conference (GLOBECOM'03), 2003.

Urgency-based scheduler for time-sensitive switched ethernet networks.
In 28th Euromicro Conf. on Real-Time Systems (ECRTS), 2016.

A theory of traffic regulators for deterministic networks with application to interleaved regulators.

[7] L. Zhao, P. Pop, Q. Li, J. Chen, and H. Xiong,
Timing analysis of rate-constrained traffic in TTEthernet using network calculus.

Timing analysis of AVB traffic in TSN networks using network calculus.
In IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS), 2018.

Latency Analysis of Multiple Classes of AVB Traffic in TSN with Standard Credit Behavior using Network Calculus.
Thank you!