Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

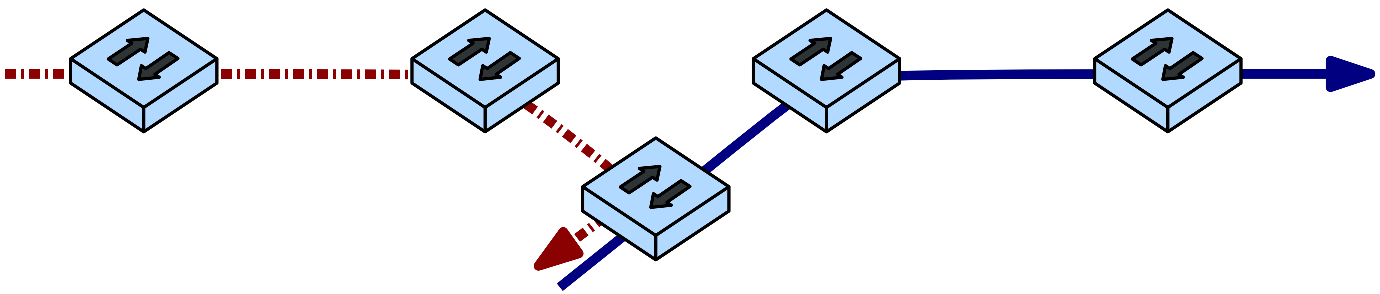
Institute of Computer Science
Chair of Communication Networks
Prof. Dr. Tobias Hoßfeld

Bridge-Local Guaranteed Latency with Strict Priority Scheduling

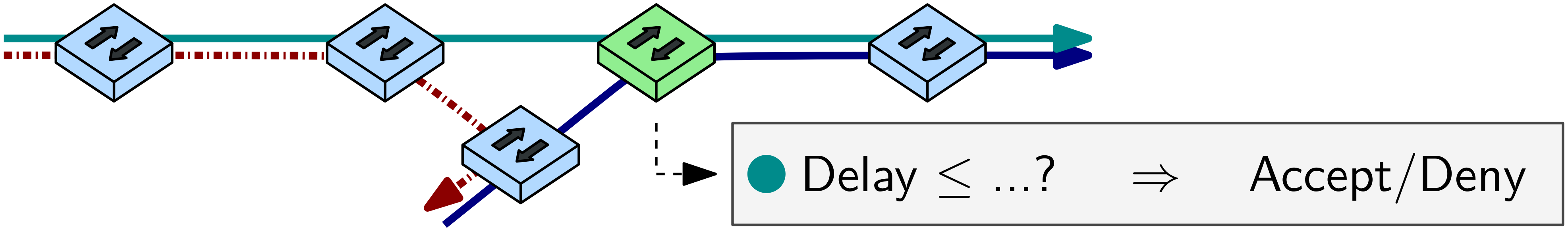
Alexej Grigorjew – March 02, 2020

comnet.informatik.uni-wuerzburg.de

Introduction – Distributed Admission Control



Introduction – Distributed Admission Control



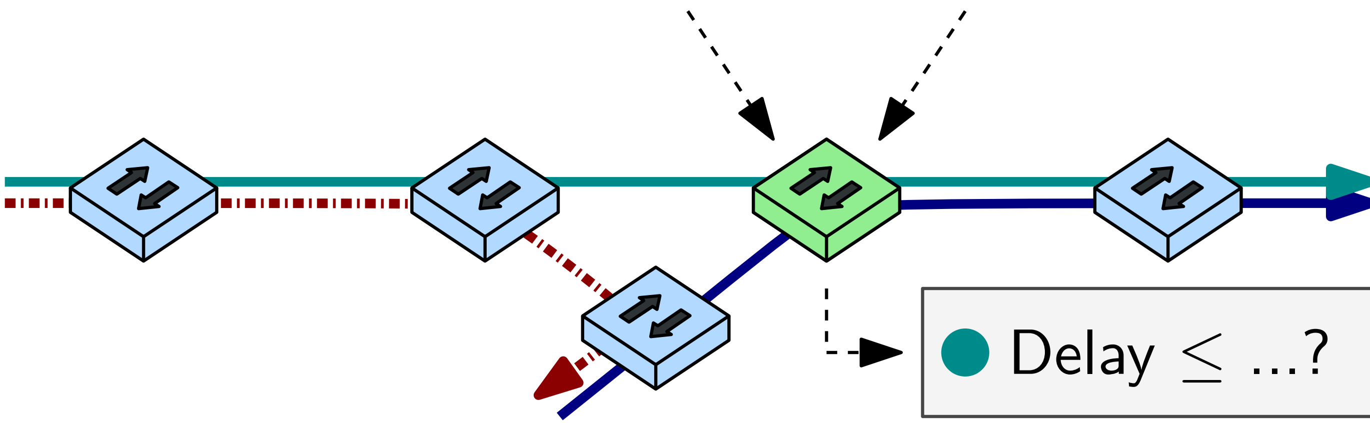
Introduction – Distributed Admission Control

Transmission Selection

SP	CBS	...
ATS	CQF	

Traffic Specification (SRP, RAP)

MaxFrameSize	Interval
MaxFramesPerInterval	...



● Delay $\leq$ ...? $\Rightarrow$ Accept/Deny

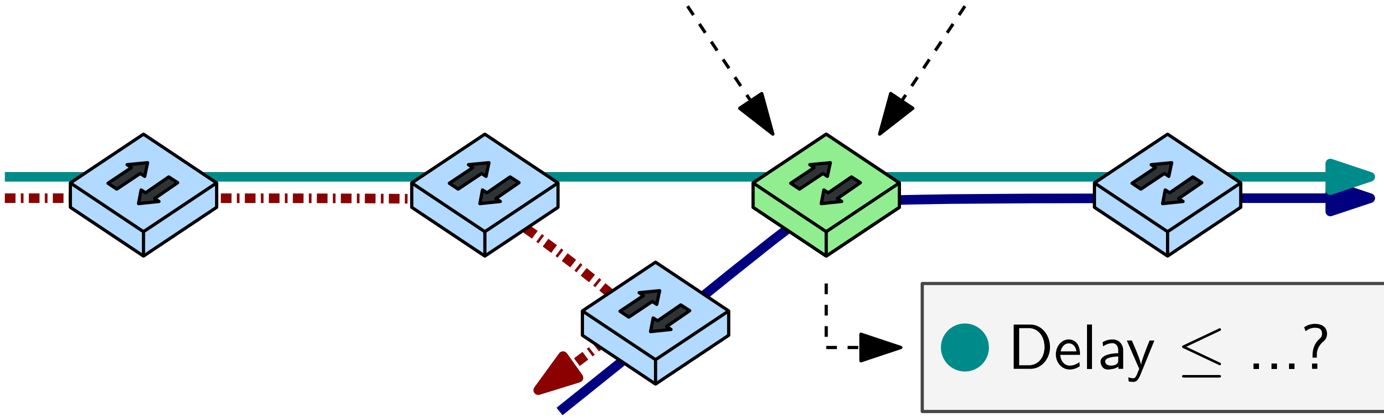
Introduction – Distributed Admission Control

Transmission Selection

SP	CBS	...
ATS	CQF	

Traffic Specification (SRP, RAP)

MaxFrameSize	Interval
MaxFramesPerInterval	...



● Delay $\leq$ ...? $\Rightarrow$ Accept/Deny

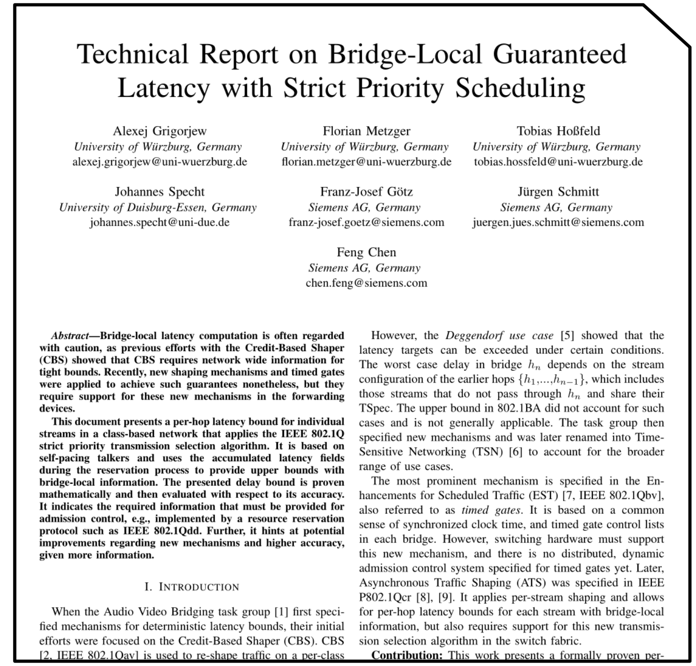
Desired Features:

- Computationally feasible
- Do not require global information (from ●)
- Support brownfield installations $\Rightarrow$ SP

Table of Contents

Preliminaries:

- Switch delay model
- Assumptions and constraints
  - Talker characteristics
  - Switch characteristics



Technical Report on Bridge-Local Guaranteed Latency with Strict Priority Scheduling

Alexej Grigorjew
University of Würzburg, Germany
alexej.grigorjew@uni-wuerzburg.de

Florian Metzger
University of Würzburg, Germany
florian.metzger@uni-wuerzburg.de

Tobias Hoßfeld
University of Würzburg, Germany
tobias.hossfeld@uni-wuerzburg.de

Johannes Specht
University of Duisburg-Essen, Germany
johannes.specht@uni-due.de

Franz-Josef Götz
Siemens AG, Germany
franz-josef.goetz@siemens.com

Jürgen Schmitt
Siemens AG, Germany
juergen.jues.schmitt@siemens.com

Feng Chen
Siemens AG, Germany
chen.feng@siemens.com

***Abstract*—Bridge-local latency computation is often regarded with caution, as previous efforts with the Credit-Based Shaper (CBS) showed that CBS requires network wide information for tight bounds. Recently, new shaping mechanisms and timed gates were applied to achieve such guarantees nonetheless, but they require support for these new mechanisms in the forwarding devices.**

This document presents a per-hop latency bound for individual streams in a class-based network that applies the IEEE 802.1Q strict priority transmission selection algorithm. It is based on self-pacing talkers and uses the accumulated latency fields during the reservation process to provide upper bounds with bridge-local information. The presented delay bound is proven mathematically and then evaluated with respect to its accuracy. It indicates the required information that must be provided for admission control, e.g., implemented by a resource reservation protocol such as IEEE 802.1Qdd. Further, it hints at potential improvements regarding new mechanisms and higher accuracy, given more information.

I. Introduction

When the Audio Video Bridging task group [1] first specified mechanisms for deterministic latency bounds, their initial efforts were focused on the Credit-Based Shaper (CBS). CBS [2, IEEE 802.1Qav] is used to re-shape traffic on a per-class

However, the *Deggendorf use case* [5] showed that the latency targets can be exceeded under certain conditions. The worst case delay in bridge h_n depends on the stream configuration of the earlier hops $\{h_1,\ldots,h_{n-1}\}$, which includes those streams that do not pass through h_n and share their TSpec. The upper bound in 802.1BA did not account for such cases and is not generally applicable. The task group then specified new mechanisms and was later renamed into Time-Sensitive Networking (TSN) [6] to account for the broader range of use cases.

The most prominent mechanism is specified in the Enhancements for Scheduled Traffic (EST) [7, IEEE 802.1Qbv], also referred to as *timed gates*. It is based on a common sense of synchronized clock time, and timed gate control lists in each bridge. However, switching hardware must support this new mechanism, and there is no distributed, dynamic admission control system specified for timed gates yet. Later, Asynchronous Traffic Shaping (ATS) was specified in IEEE P802.1Qcr [8], [9]. It applies per-stream shaping and allows for per-hop latency bounds for each stream with bridge-local information, but also requires support for this new transmission selection algorithm in the switch fabric.

Contribution: This work presents a formally proven per-

https://nbn-resolving.org/urn:nbn:de:bvb:20-opus-198310

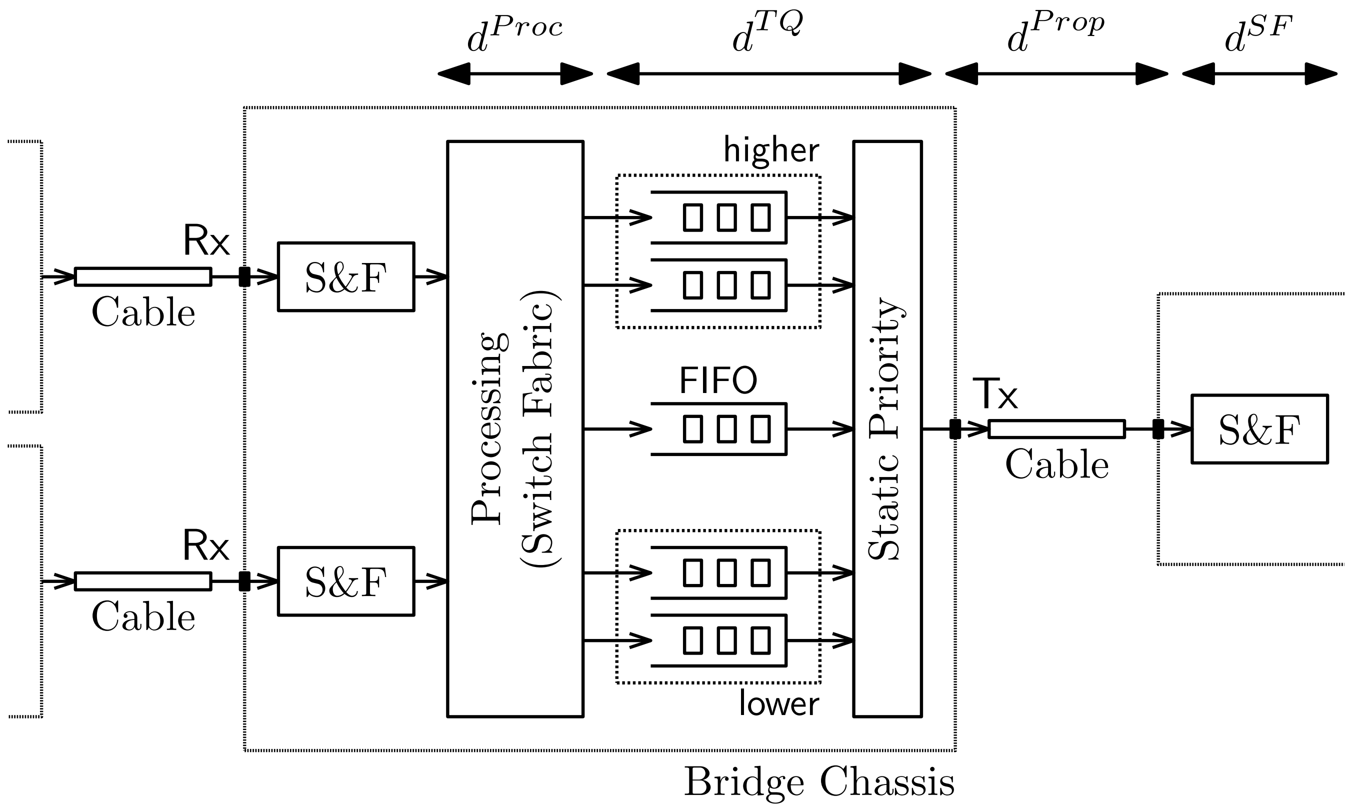
Contribution:

- Overview of required information from the Resource Allocation Protocol (RAP)
- Proven per-hop latency bound for Strict Priority (SP) transmission selection with only bridge-local information
- Initial evaluation of network capacity for an admission control system using this bound

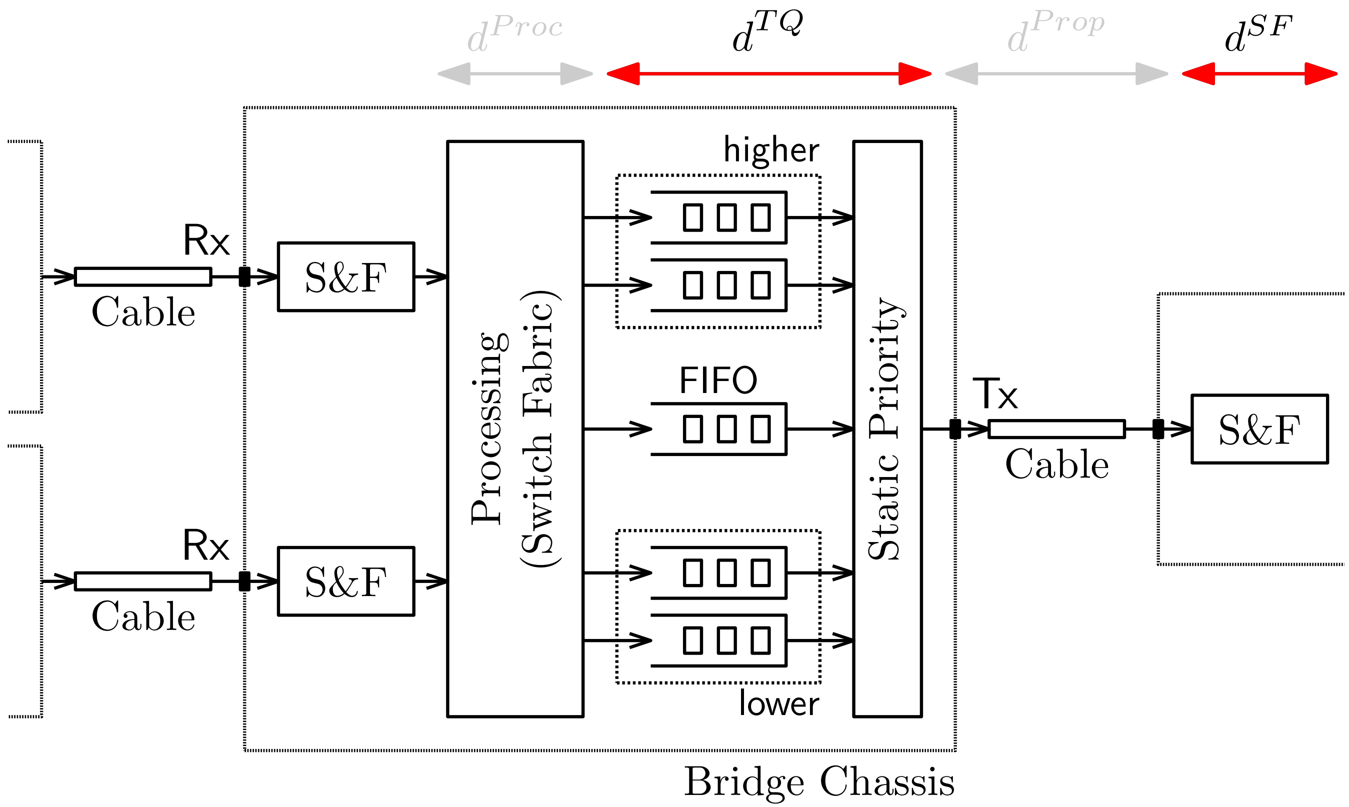
Preliminaries

Switch delay models, assumptions and constraints

Switch Delay Model



Switch Delay Model



- Processing delay d^{Proc} is device-specific and not considered
- Propagation delay d^{Prop} is bounded by max cable length
- Upper bound for $d^{TQ} + d^{SF}$ desired (queuing and transmission delay)

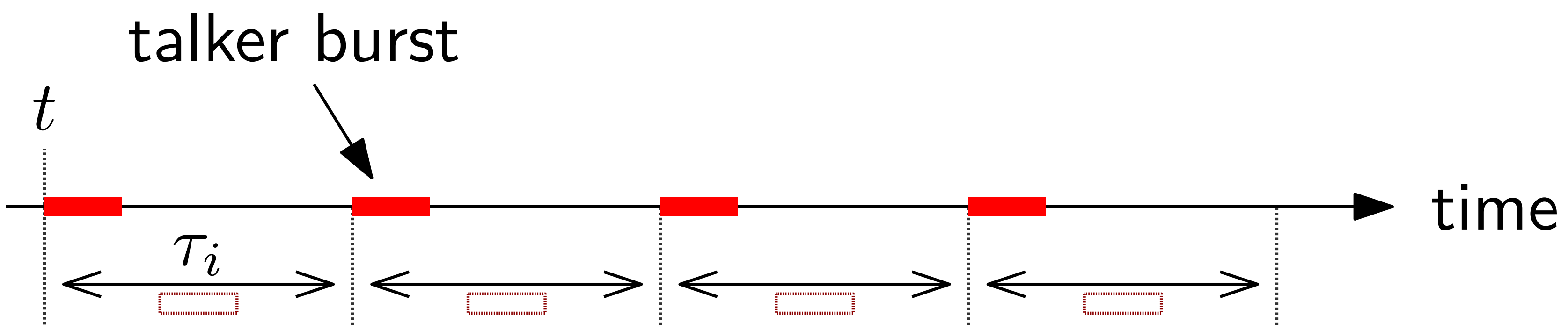
Assumptions and Constraints – Talkers

1. Frames of stream i do not exceed their **max frame size** $\hat{\ell}_i$ and **min frame size** $\check{\ell}_i$.

Assumptions and Constraints – Talkers

1. Frames of stream i do not exceed their **max frame size** $\hat{\ell}_i$ and **min frame size** $\check{\ell}_i$.

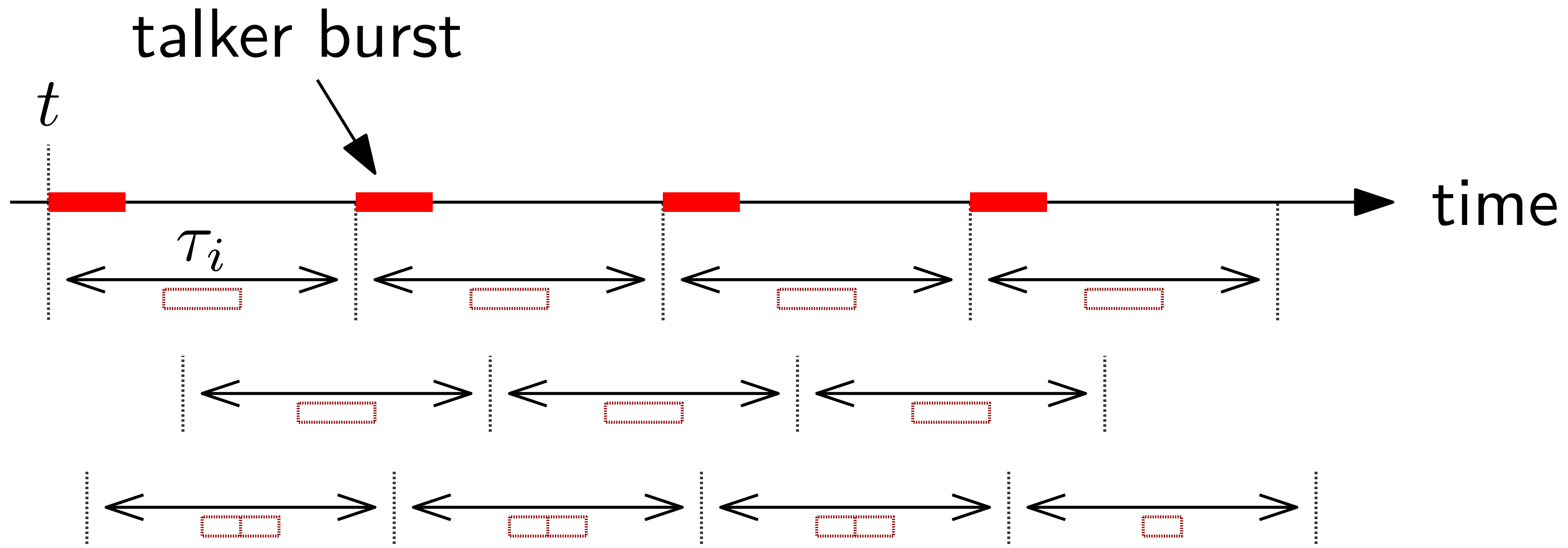
2. Talkers pace their traffic according to a **burst size** b_i and a **burst interval** τ_i. For any point t in time, the traffic sent by stream i in the time interval $[t, t + \tau_i]$ may not exceed b_i.



Assumptions and Constraints – Talkers

1. Frames of stream i do not exceed their **max frame size** $\hat{\ell}_i$ and **min frame size** $\check{\ell}_i$.

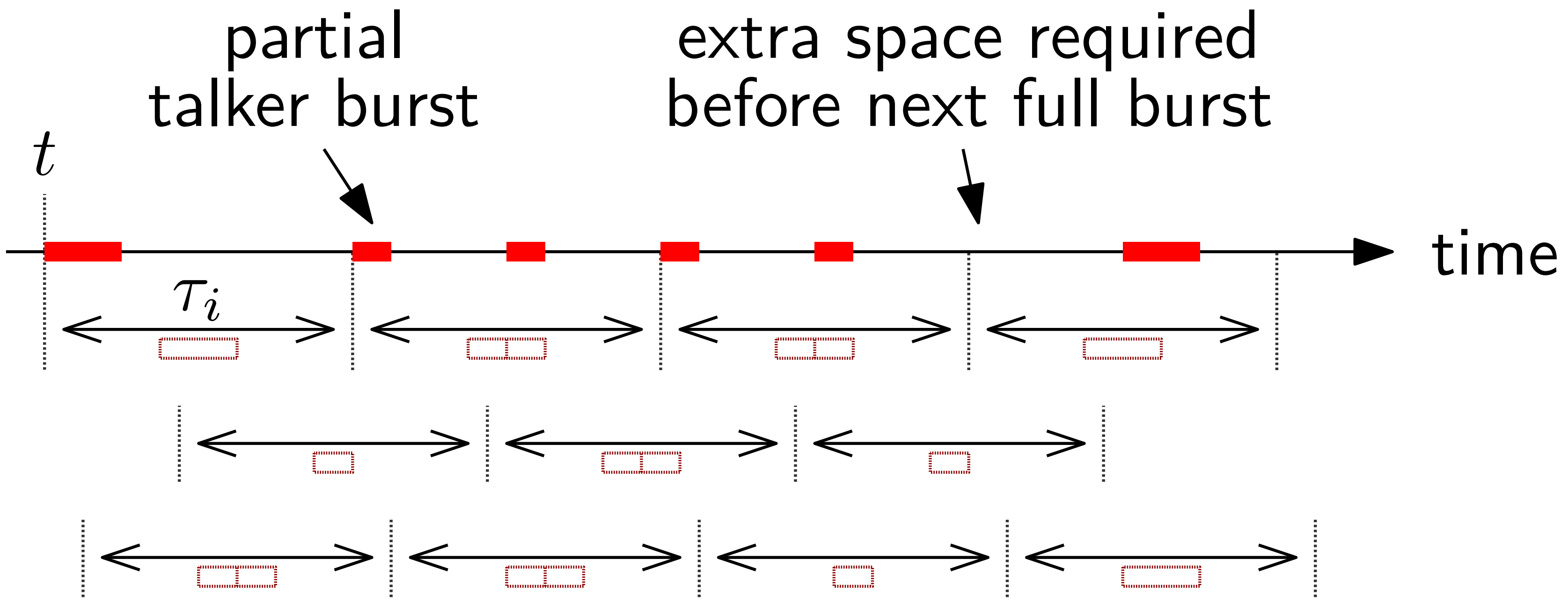
2. Talkers pace their traffic according to a **burst size** b_i and a **burst interval** τ_i. For any point t in time, the traffic sent by stream i in the time interval $[t, t + \tau_i]$ may not exceed b_i.



Assumptions and Constraints – Talkers

1. Frames of stream i do not exceed their **max frame size** $\hat{\ell}_i$ and **min frame size** $\check{\ell}_i$.

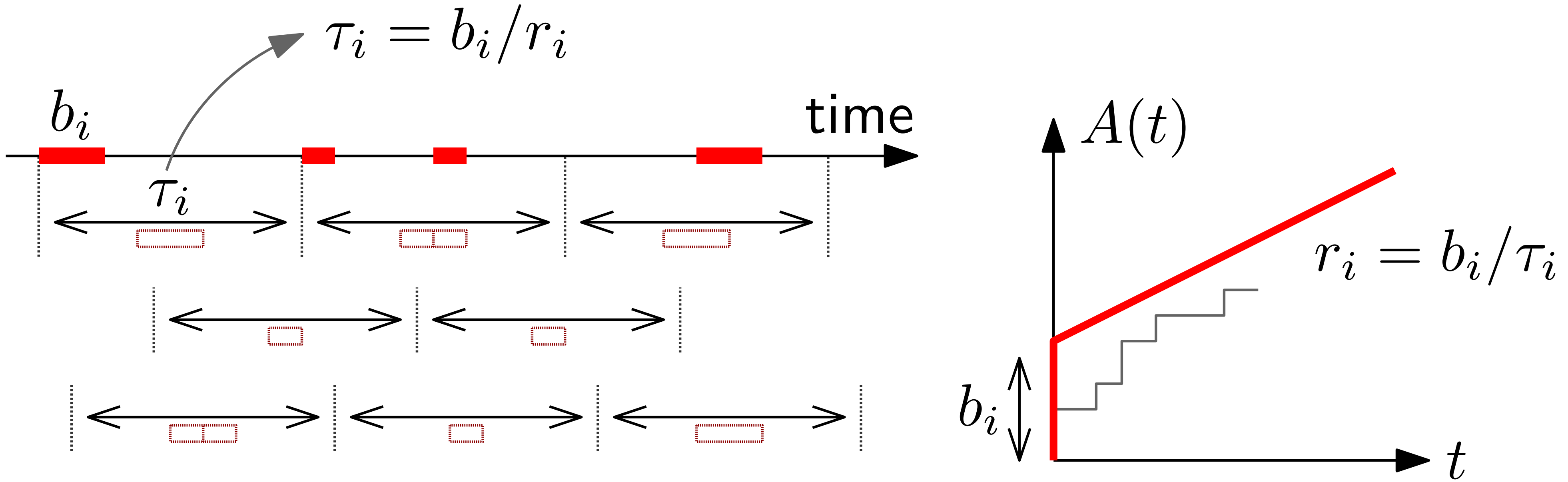
2. Talkers pace their traffic according to a **burst size** b_i and a **burst interval** τ_i. For any point t in time, the traffic sent by stream i in the time interval $[t, t+\tau_i]$ may not exceed b_i.



Assumptions and Constraints – Talkers

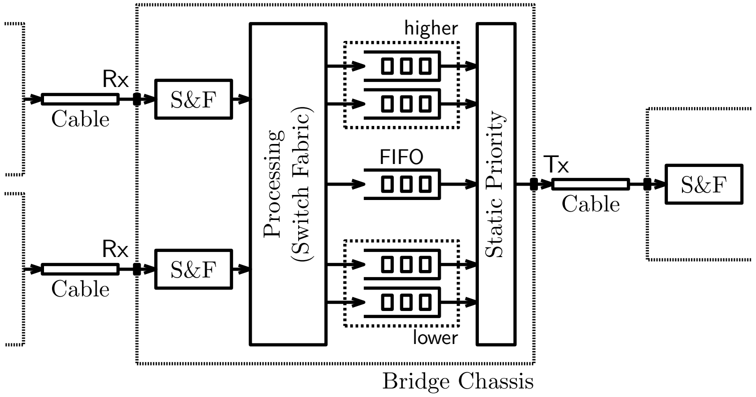
1. Frames of stream i do not exceed their **max frame size** $\hat{\ell}_i$ and **min frame size** $\check{\ell}_i$.

2. Talkers pace their traffic according to a **burst size** b_i and a **burst interval** τ_i. For any point t in time, the traffic sent by stream i in the time interval $[t, t+\tau_i]$ may not exceed b_i.



Assumptions and Constraints – Bridges

3. Bridges use IEEE 802.1Q **priority transmission selection**, i.e., frames with a higher traffic class are always selected for transmission before frames with lower traffic classes.
   (a) Within each traffic class, **FIFO** transmission selection is used.
   (b) **No shaping** mechanisms are used in any considered traffic class. The earliest frame of each class is always regarded eligible for transmission.

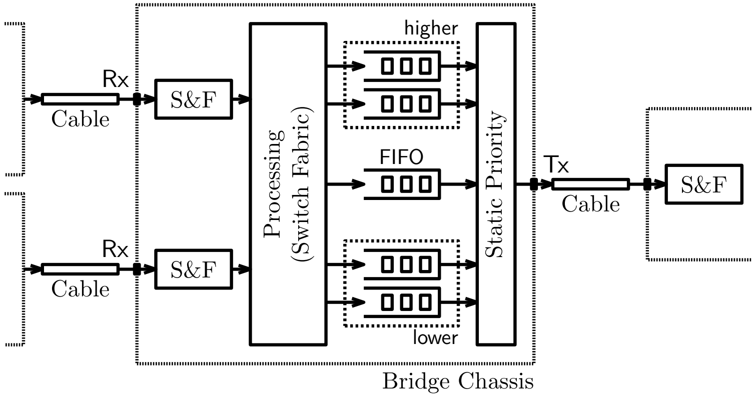


Assumptions and Constraints – Bridges

3. Bridges use IEEE 802.1Q **priority transmission selection**, i.e., frames with a higher traffic class are always selected for transmission before frames with lower traffic classes.

    (a) Within each traffic class, **FIFO** transmission selection is used.

    (b) **No shaping** mechanisms are used in any considered traffic class. The earliest frame of each class is always regarded eligible for transmission.



4. Each bridge h has a pre-configured maximum per-hop **delay guarantee** δ_p^h for each traffic class p.

    (a) **Admission control** prevents the deployment of new streams that would cause delay violations for any deployed stream.

Latency Bound

Required information, formula and reasoning

Required Information $\rightarrow$ TSpec

TSpec should include for stream i:

- Traffic class p_i
- Min frame size $\check{\ell}_i$ (e.g., 64 B)
- Max frame size $\hat{\ell}_i$ (e.g., 1542 B)
- Committed burst size b_i
- Burst interval τ_i

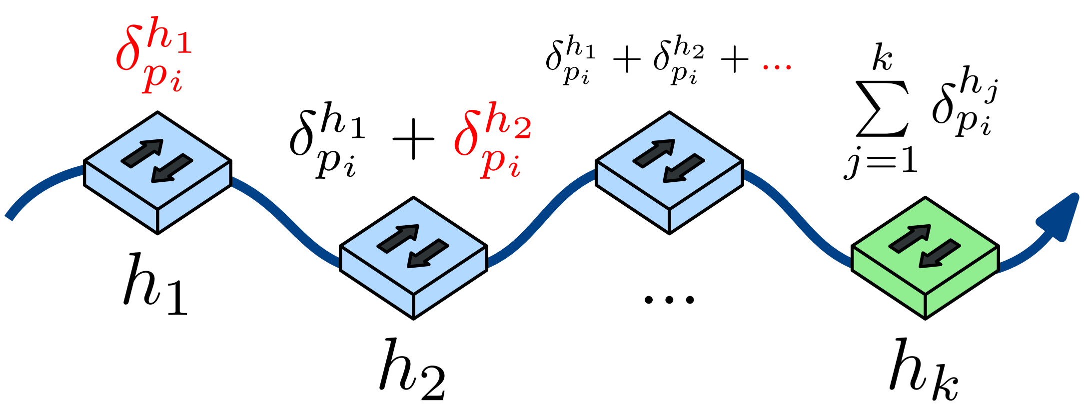
including preamble and IPG

Required Information → TSpec

TSpec should include for stream i:

- Traffic class p_i
- Min frame size $\check{\ell}_i$ (e.g., 64 B)
- Max frame size $\hat{\ell}_i$ (e.g., 1542 B)
- Committed burst size b_i
- Burst interval τ_i
- Accumulated max latency $accMaxD_i^{h_k}$
- Accumulated min latency $accMinD_i^{h_k}$

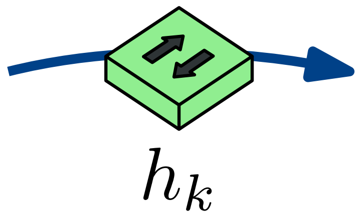
(Max frame size and committed burst size: including preamble and IPG)

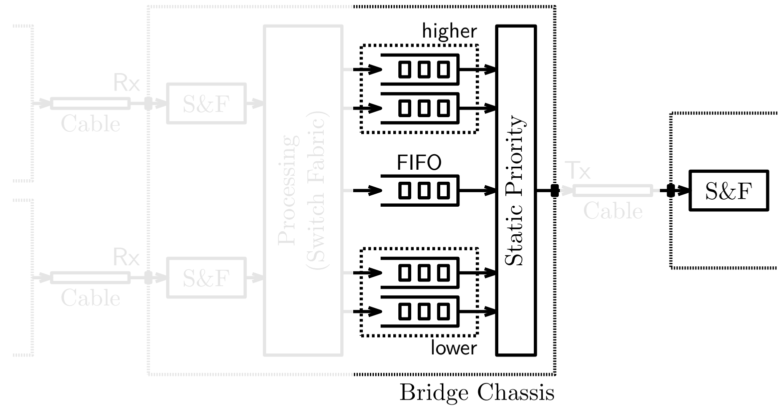


$$accMaxD_i^{h_k} = \sum_{j=1}^{k} \delta_{p_i}^{h_j}$$

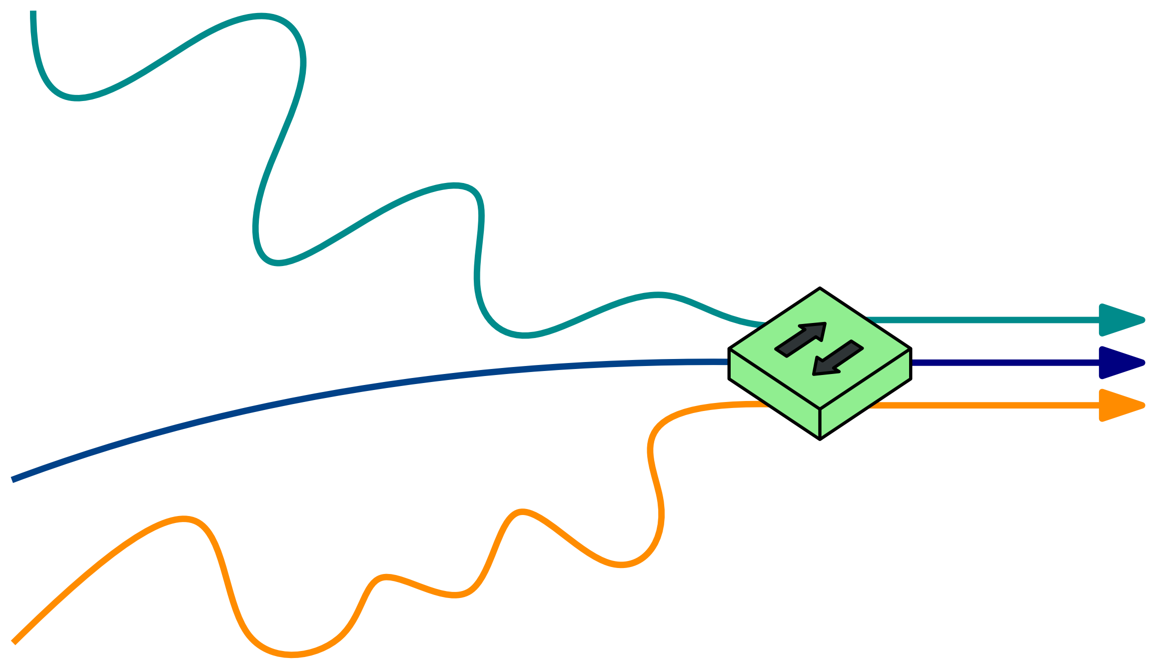
$$accMinD_i^{h_k} = \sum_{j=1}^{k} \frac{\check{\ell}_i}{link\ speed_{h_j}}$$

Latency Bound

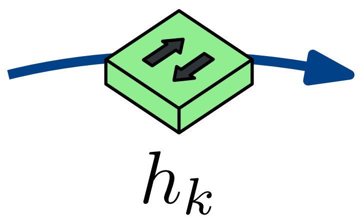


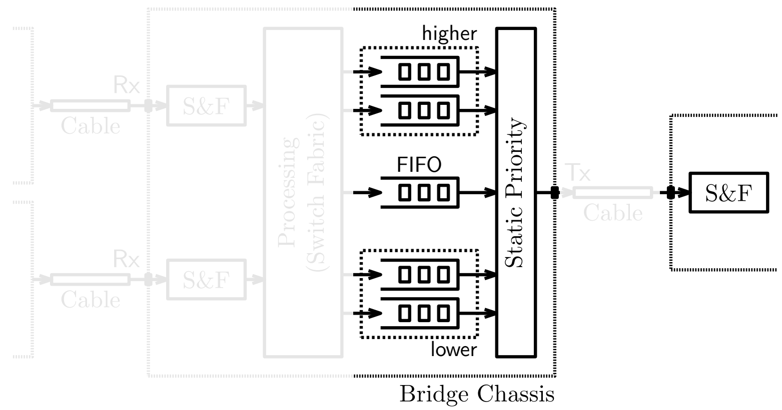


p_i	traffic class
$\hat{\ell}_i$	max frame size
b_i	burst size
$\delta_{p_i}^{h_k}$	delay guarantee
r	link speed
$\mathcal{S}$	set of all streams
$accMaxD_i^{h_k}$	
$accMinD_i^{h_k}$	

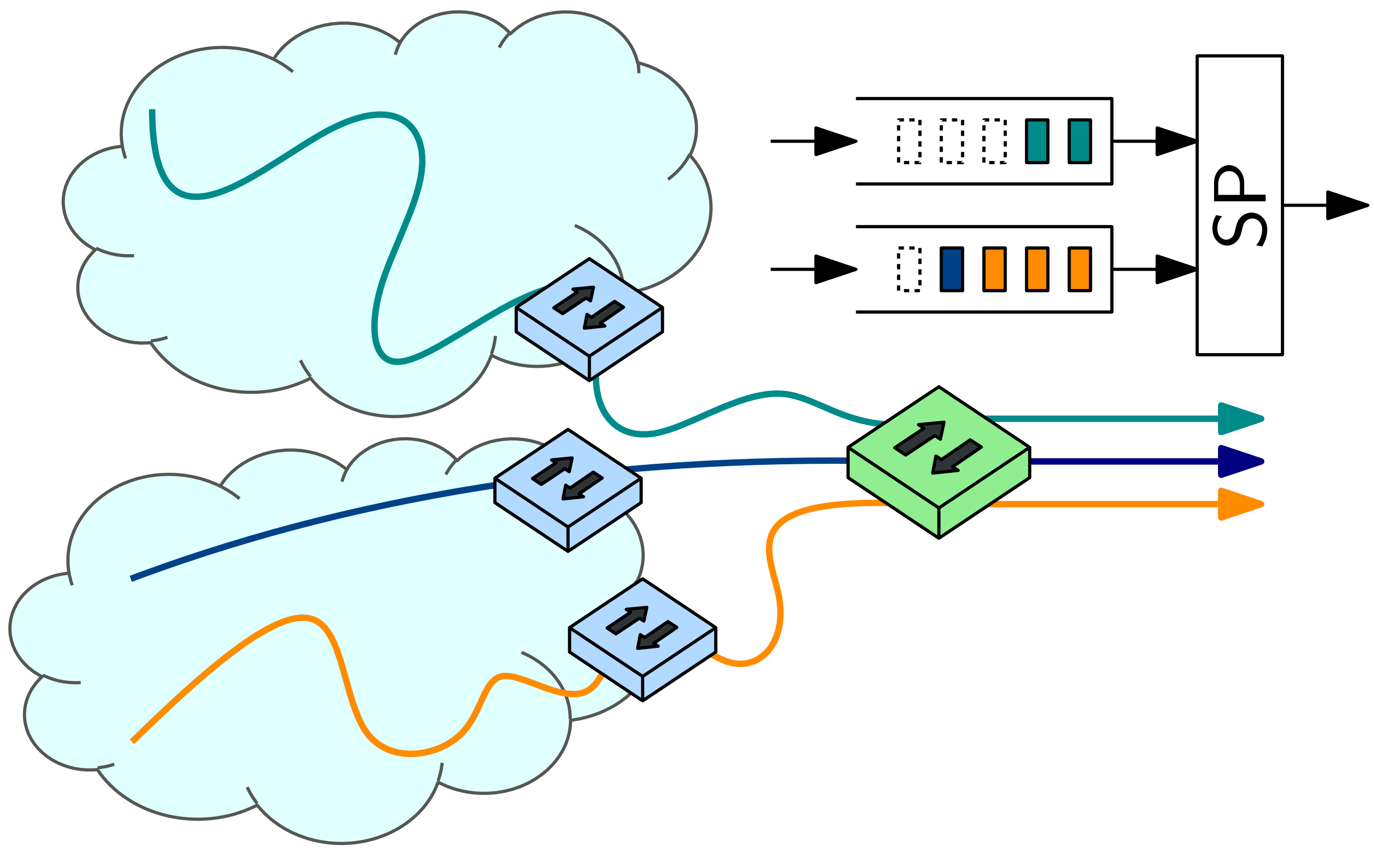


Latency Bound

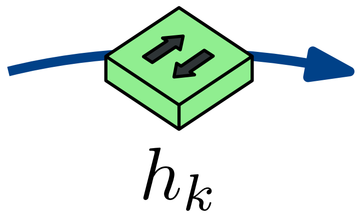


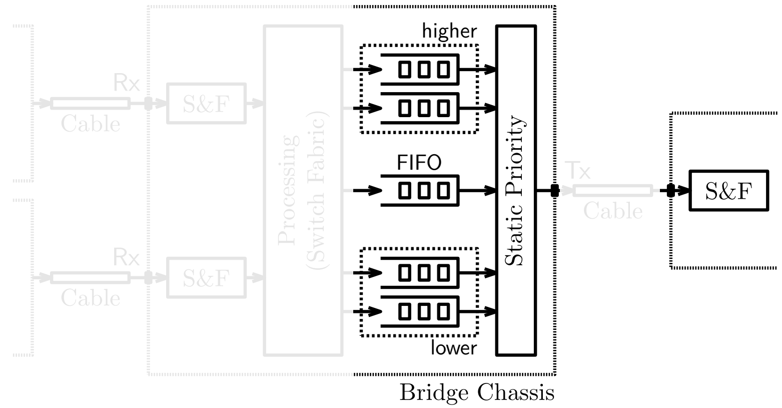


p_i	traffic class
$\hat{\ell}_i$	max frame size
b_i	burst size
$\delta_{p_i}^{h_k}$	delay guarantee
r	link speed
$\mathcal{S}$	set of all streams
$accMaxD_i^{h_k}$	
$accMinD_i^{h_k}$	



Latency Bound



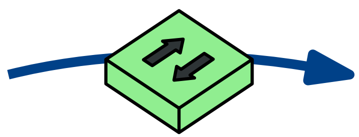


p_i	traffic class
$\hat{\ell}_i$	max frame size
b_i	burst size
$\delta_{p_i}^{h_k}$	delay guarantee
r	link speed
$\mathcal{S}$	set of all streams
$accMaxD_i^{h_k}$	
$accMinD_i^{h_k}$	

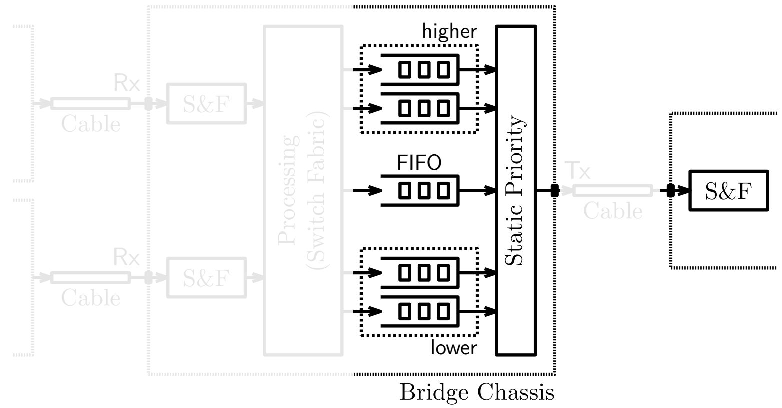
- Worst case latency of stream i at bridge h_k is bounded by:

$$d_i^{TQ,SF} \leq \sum_{\{x \in \mathcal{S} | p_x > p_i\}} y_{i,x} b_x / r \;+\; \sum_{\{x \in \mathcal{S} | p_x = p_i\}} z_x b_x / r \;+\; \max_{\{x \in \mathcal{S} | p_x < p_i\}} \hat{\ell}_x / r$$

Latency Bound



h_k



p_i	traffic class
$\hat{\ell}_i$	max frame size
b_i	burst size
$\delta_{p_i}^{h_k}$	delay guarantee
r	link speed
$\mathcal{S}$	set of all streams
$accMaxD_i^{h_k}$	
$accMinD_i^{h_k}$	

- Worst case latency of stream i at bridge h_k is bounded by:

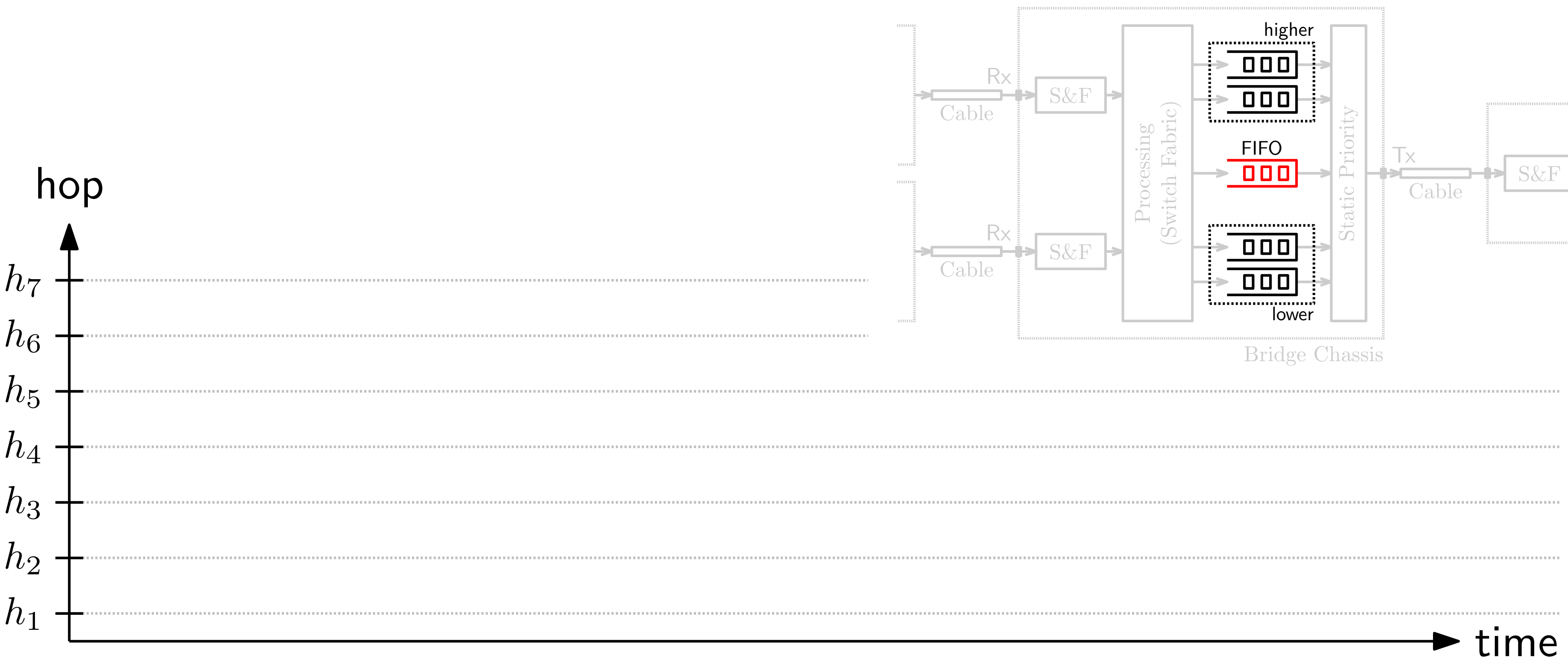
$$d_i^{TQ,SF} \leq \sum_{\{x \in \mathcal{S} | p_x > p_i\}} y_{i,x} b_x / r \; + \sum_{\{x \in \mathcal{S} | p_x = p_i\}} z_x b_x / r \; + \max_{\{x \in \mathcal{S} | p_x < p_i\}} \hat{\ell}_x / r$$

number of bursts from interfering streams

$$y_{i,x} \geq \left\lceil \frac{accMaxD_x^{h_k} - accMinD_x^{h_{k-1}} + \delta_{p_i}^{h_k}}{\tau_x} \right\rceil$$

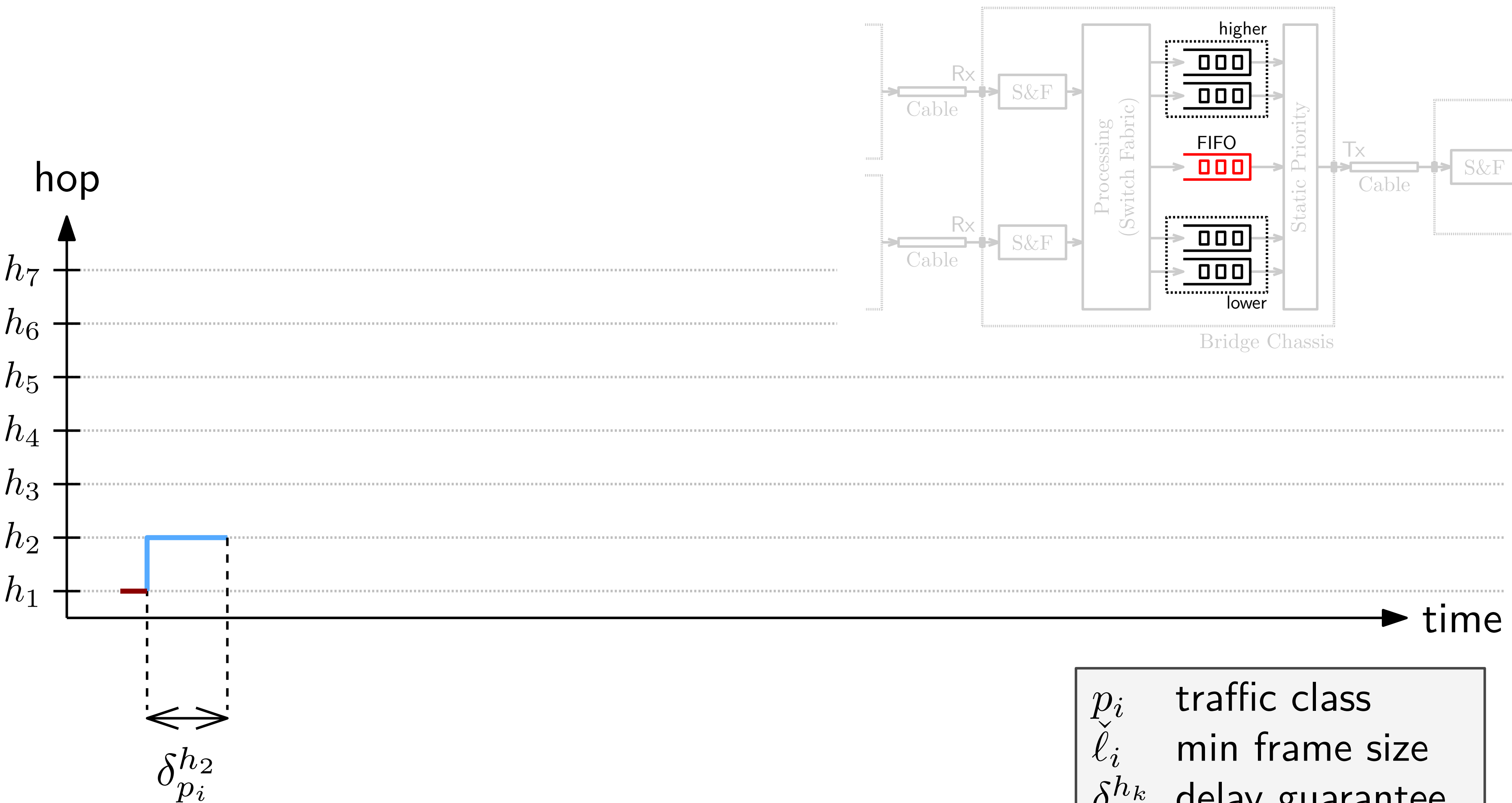
$$z_x \geq \left\lceil \frac{accMaxD_x^{h_k} - accMinD_x^{h_{k-1}}}{\tau_x} \right\rceil$$

Reasoning – Residence Times of Frames in TQ



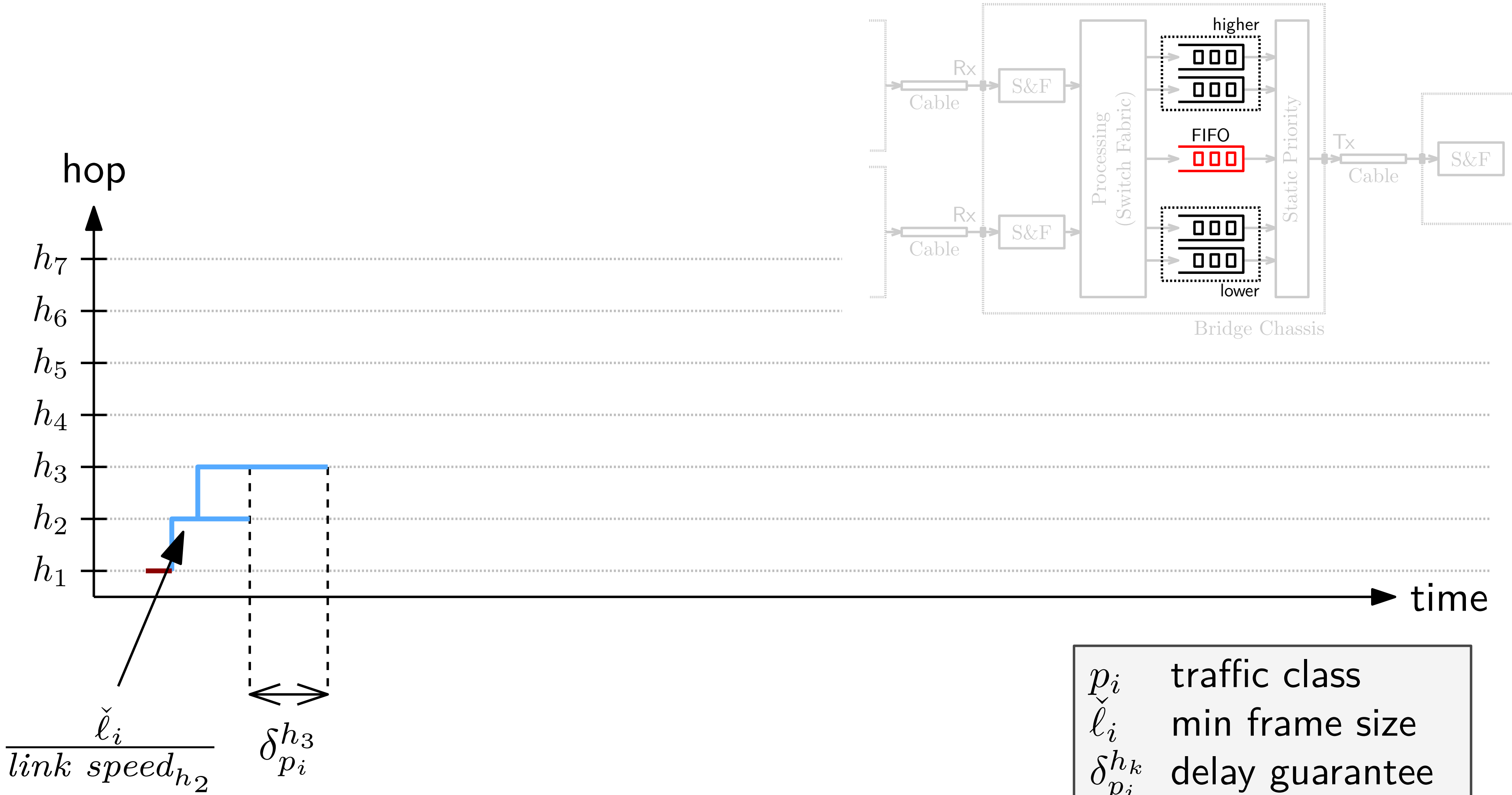
p_i	traffic class
$\check{\ell}_i$	min frame size
$\delta_{p_i}^{h_k}$	delay guarantee
τ_i	burst interval
$accMaxD_i^{h_k}$	
$accMinD_i^{h_k}$	

Reasoning – Residence Times of Frames in TQ



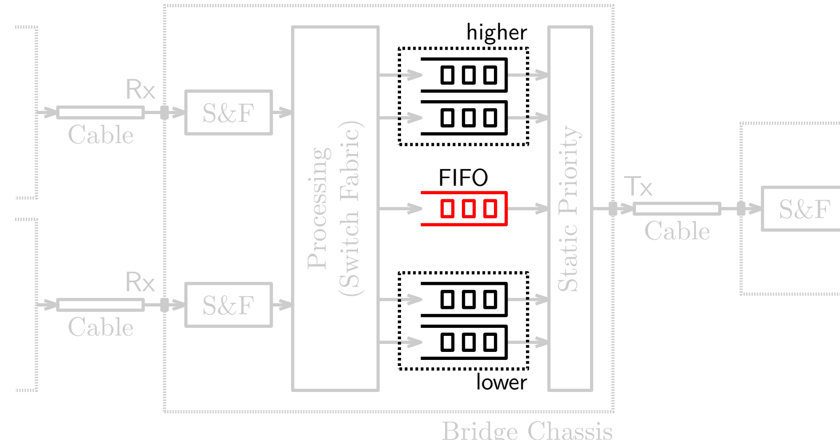
p_i	traffic class
$\check{\ell}_i$	min frame size
$\delta_{p_i}^{h_k}$	delay guarantee
τ_i	burst interval
$accMaxD_i^{h_k}$	
$accMinD_i^{h_k}$	

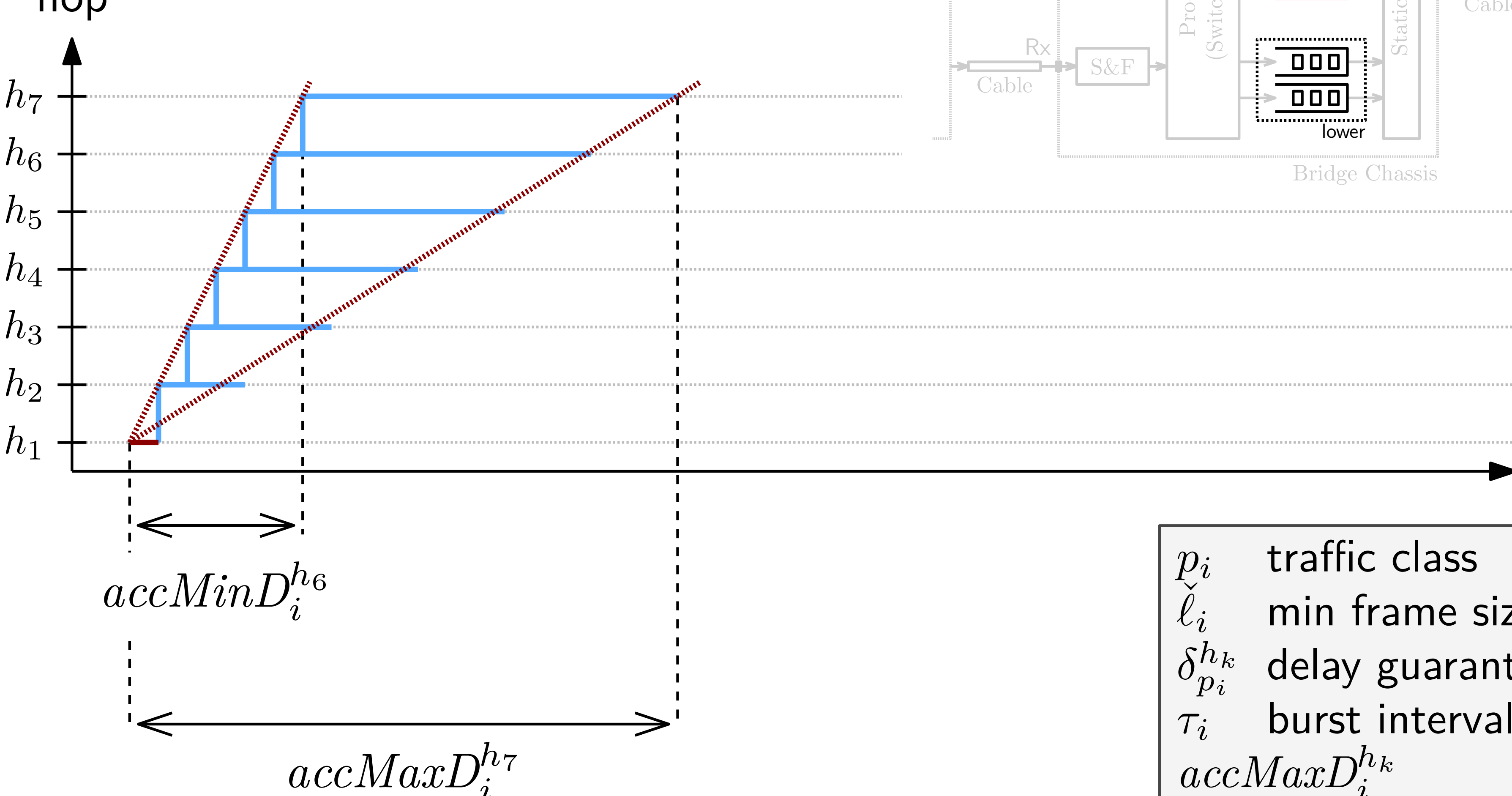
Reasoning – Residence Times of Frames in TQ



p_i	traffic class
$\check{\ell}_i$	min frame size
$\delta_{p_i}^{h_k}$	delay guarantee
τ_i	burst interval
$accMaxD_i^{h_k}$	
$accMinD_i^{h_k}$	

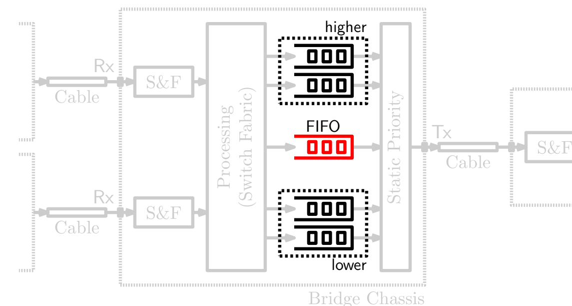
Reasoning – Residence Times of Frames in TQ

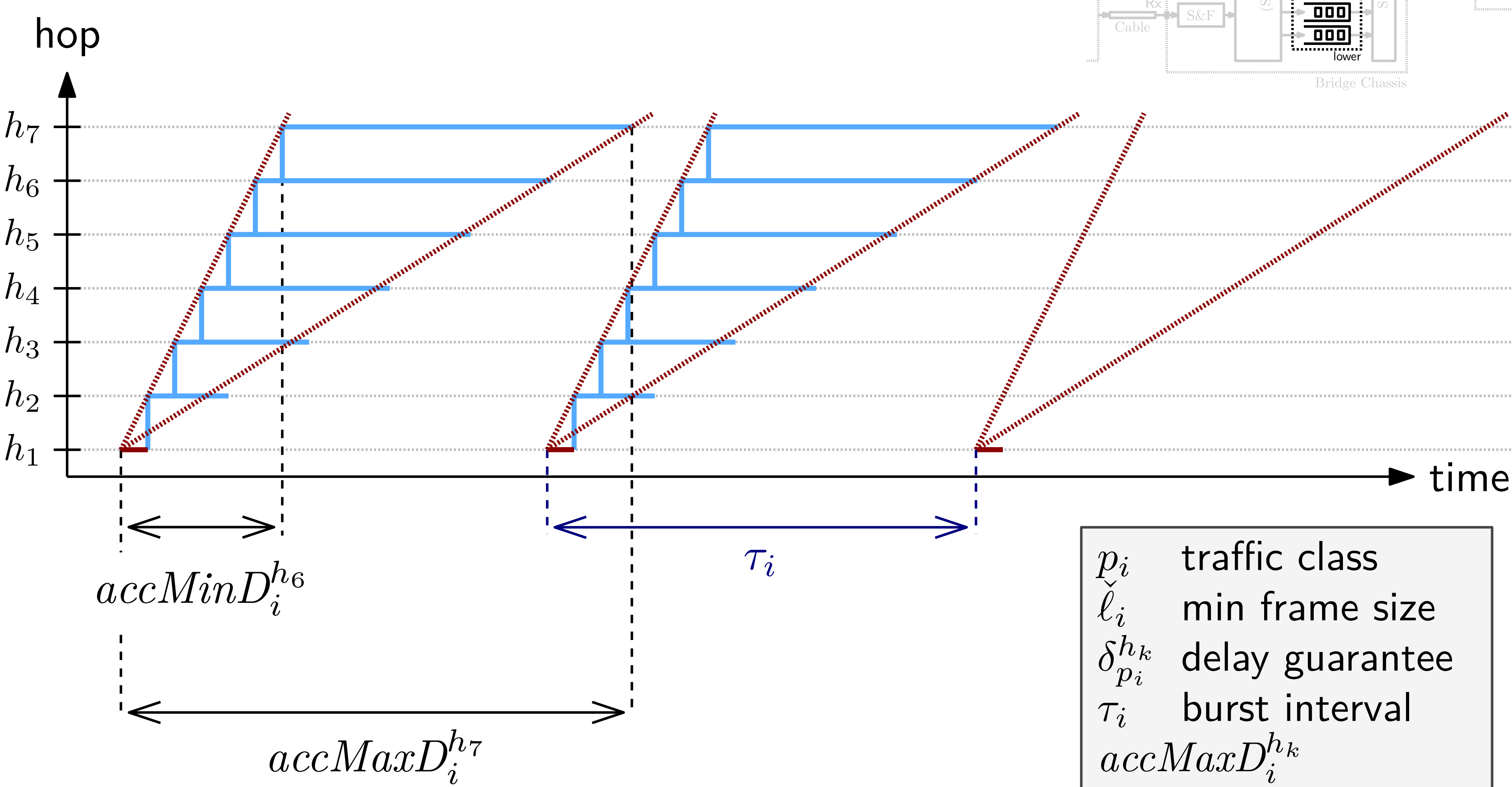




p_i	traffic class
$\check{\ell}_i$	min frame size
$\delta_{p_i}^{h_k}$	delay guarantee
τ_i	burst interval
$accMaxD_i^{h_k}$	
$accMinD_i^{h_k}$	

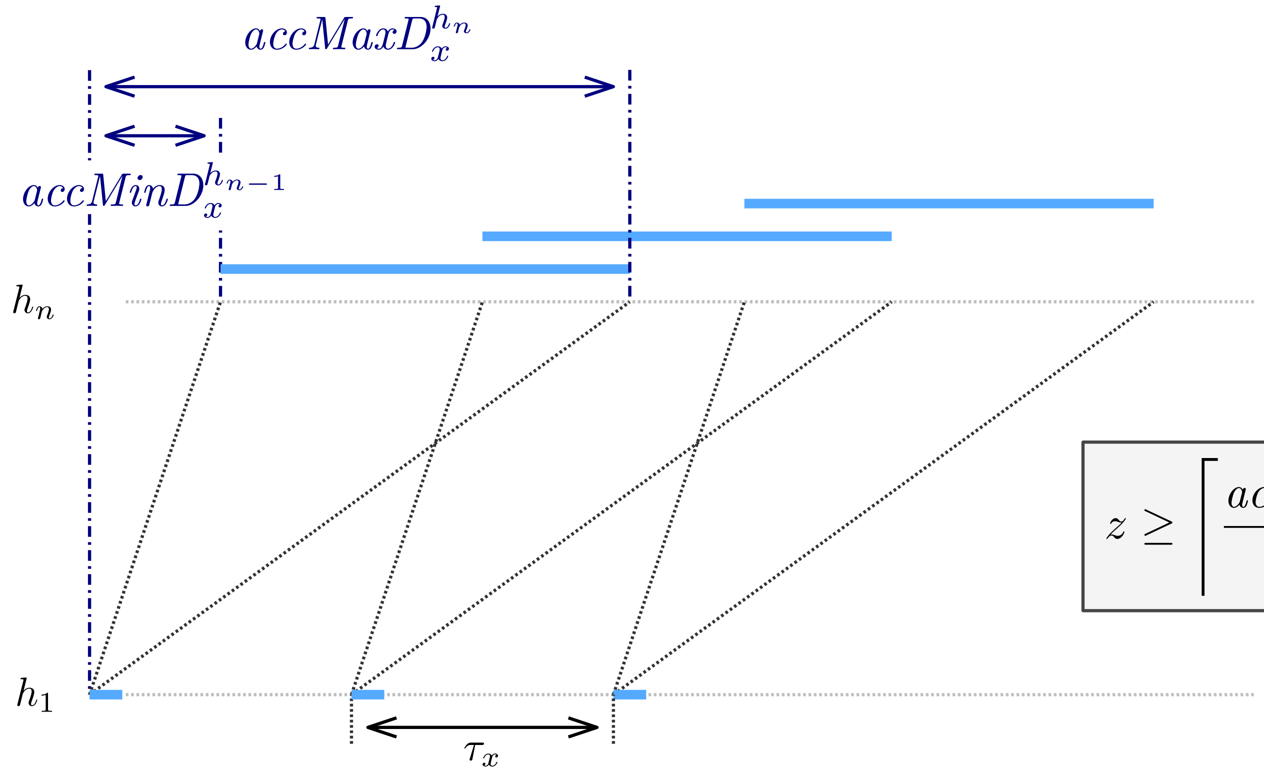
Reasoning – Residence Times of Frames in TQ





p_i	traffic class
$\check{\ell}_i$	min frame size
$\delta_{p_i}^{h_k}$	delay guarantee
τ_i	burst interval
$accMaxD_i^{h_k}$	
$accMinD_i^{h_k}$	

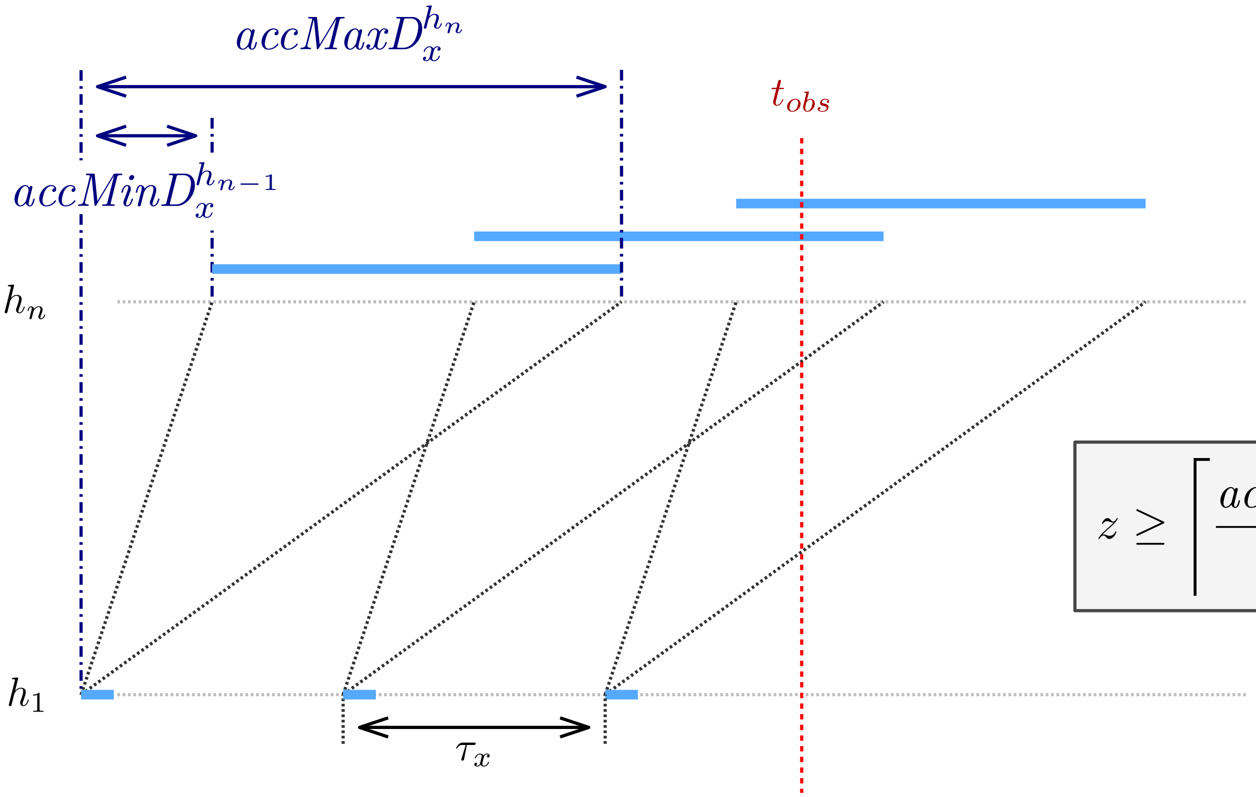
Reasoning – Same-Class Bursts $\mathbf{z_x}$



$$z \geq \left\lceil \frac{accMaxD_x^{h_n} - accMinD_x^{h_{n-1}}}{\tau_x} \right\rceil$$

- How many bursts from stream x can be in the queue of h_n at the same time?

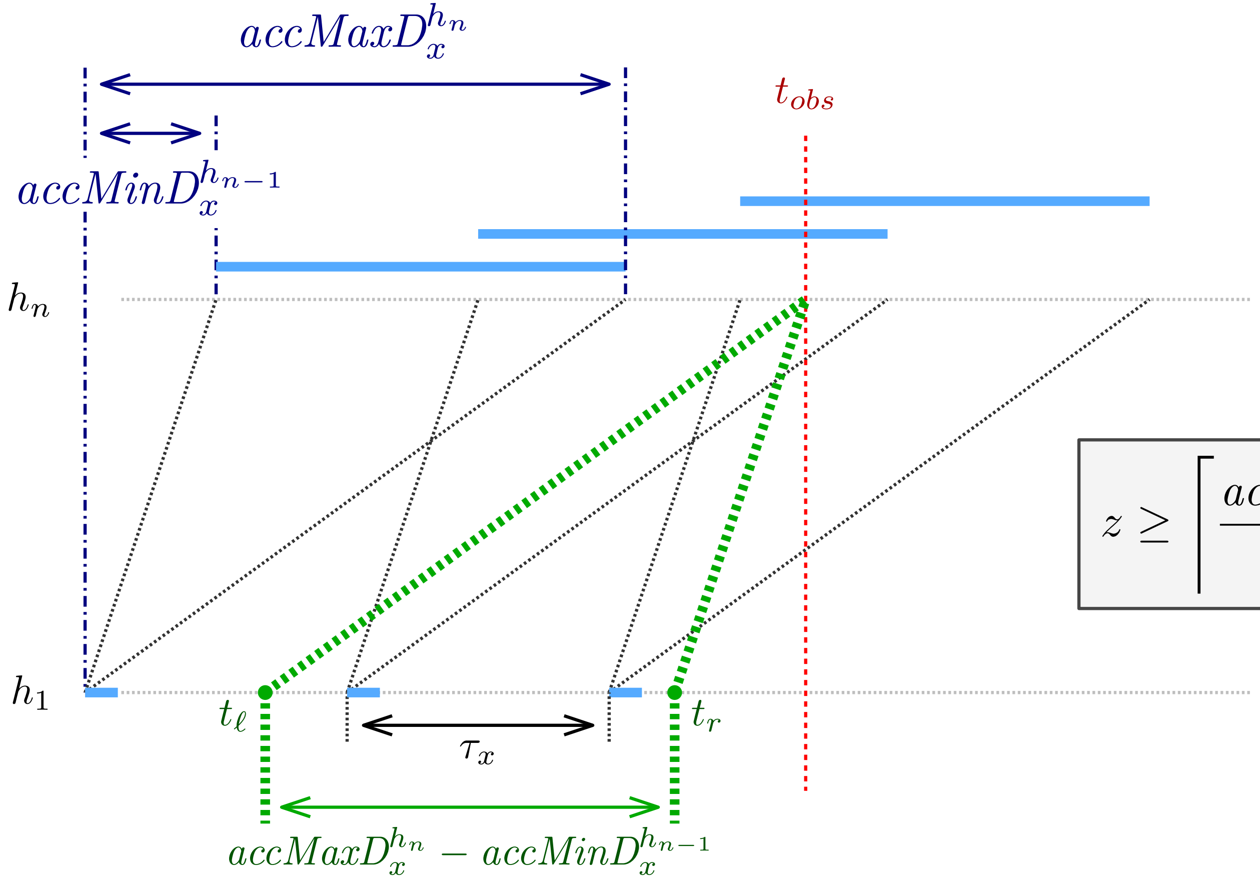
Reasoning – Same-Class Bursts $\mathbf{z_x}$



$$z \geq \left\lceil \frac{accMaxD_x^{h_n} - accMinD_x^{h_{n-1}}}{\tau_x} \right\rceil$$

- How many bursts from stream x can be in the queue of h_n at the same time?

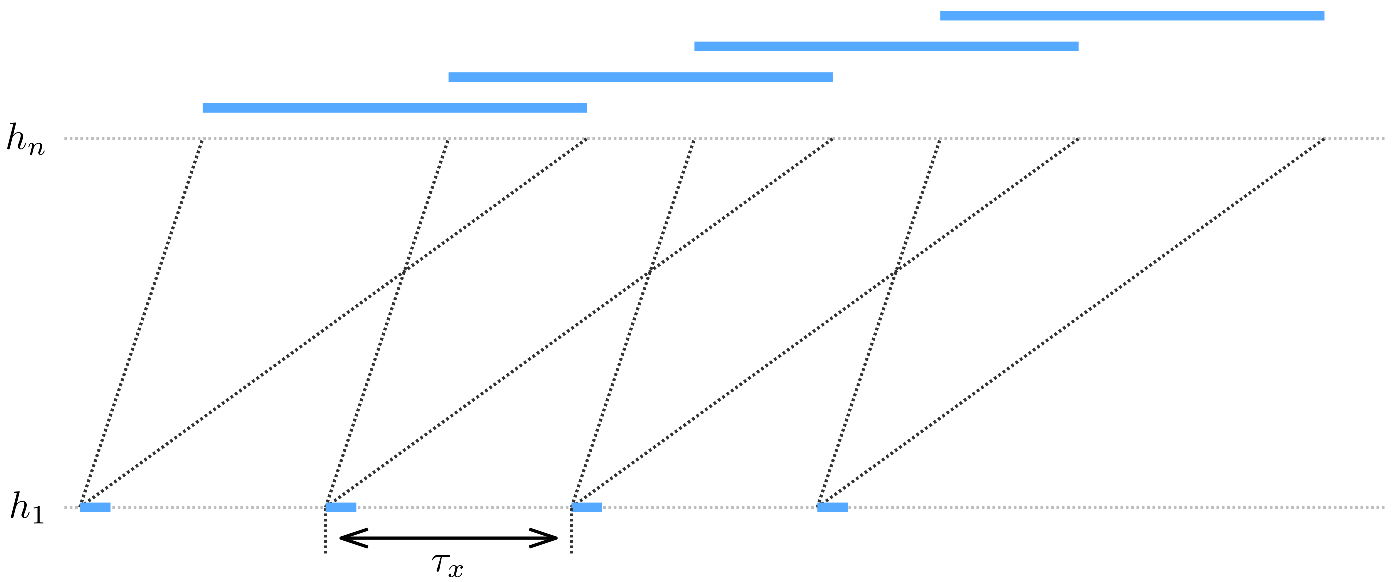
Reasoning – Same-Class Bursts $\mathbf{z_x}$



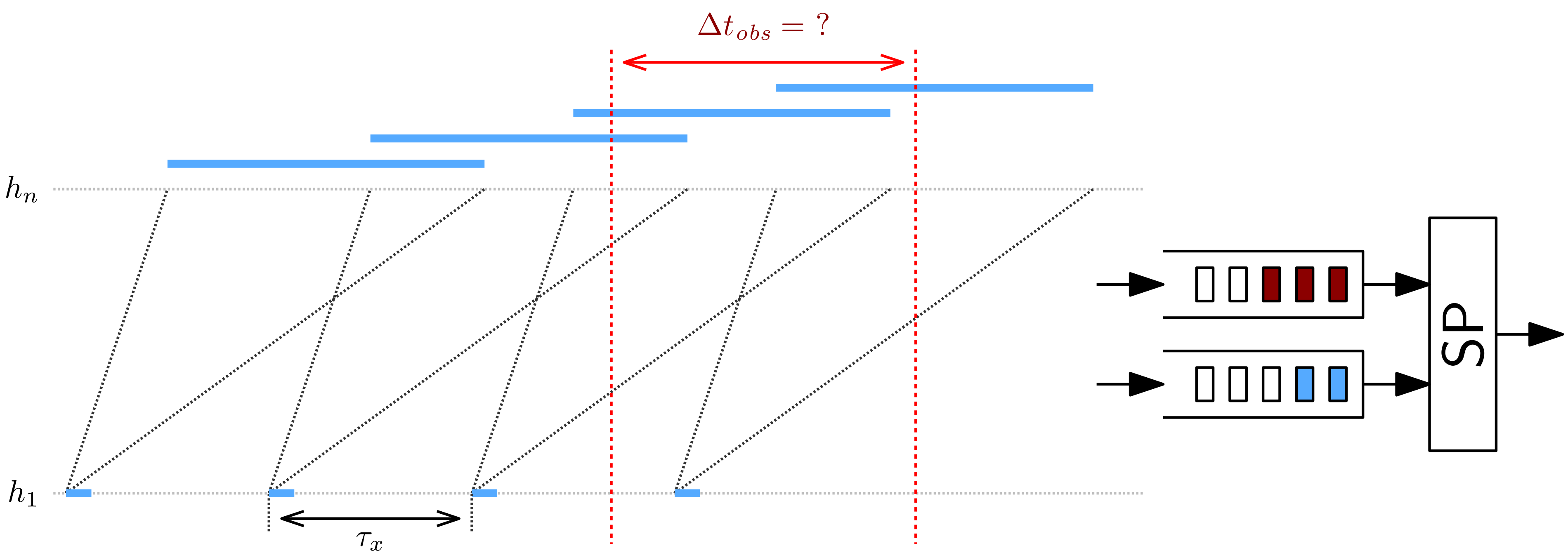
$$z \geq \left\lceil \frac{accMaxD_x^{h_n} - accMinD_x^{h_{n-1}}}{\tau_x} \right\rceil$$

- How many bursts from stream x can be in the queue of h_n at the same time?
- Project time t_{obs} to interval $[t_\ell, t_r]$ at the talker h_1

Reasoning – Higher-Class Bursts $\mathbf{y_{i,x}}$

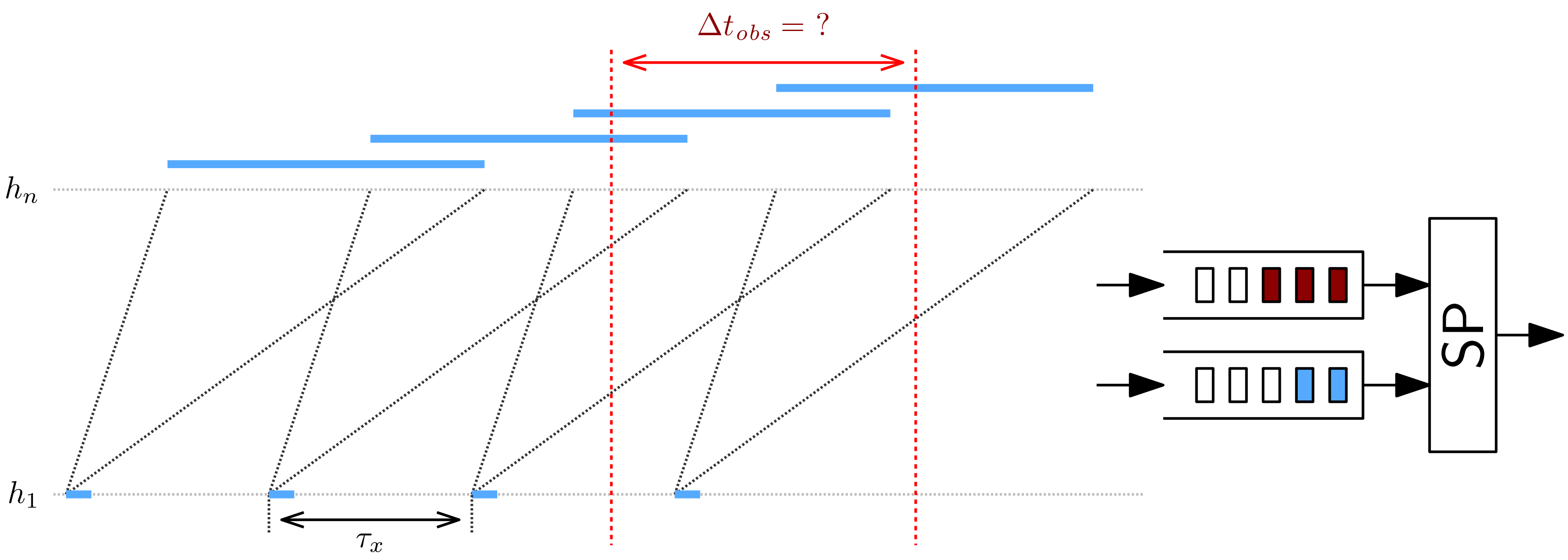


Reasoning – Higher-Class Bursts $\mathbf{y_{i,x}}$



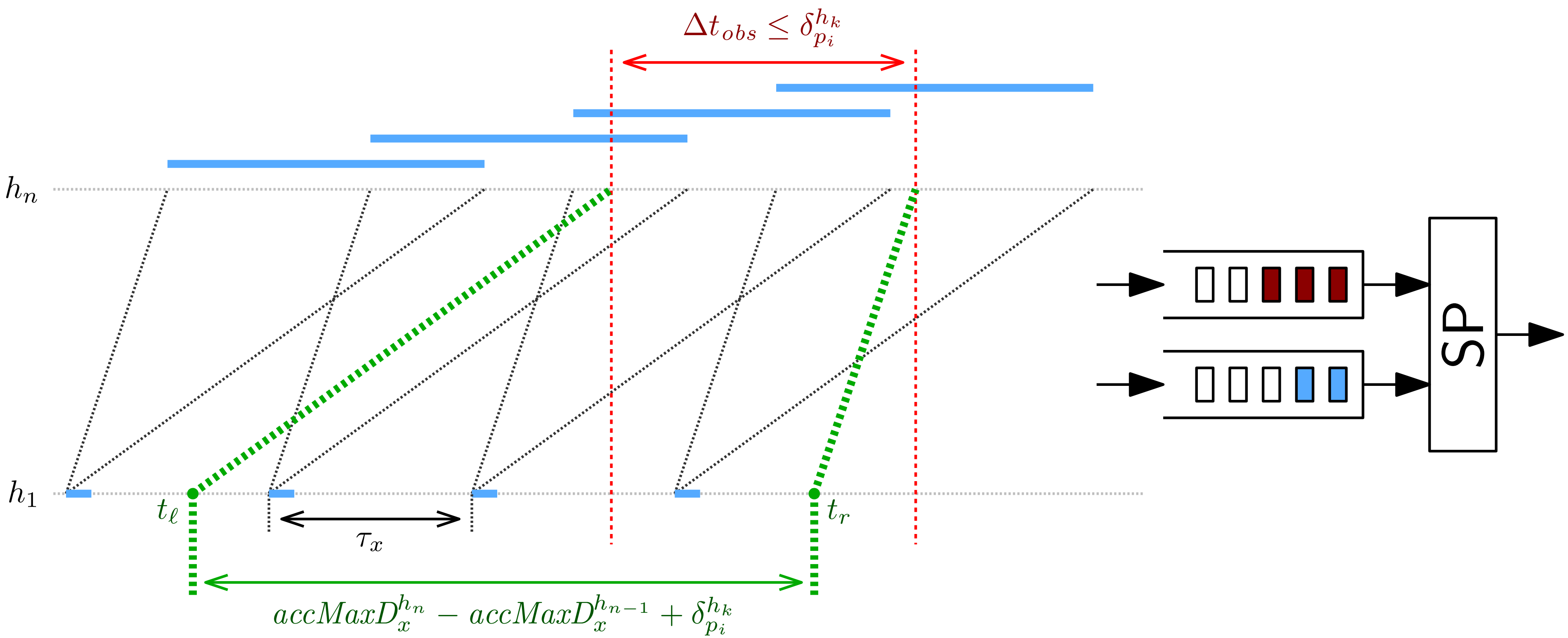
- Higher class frames ■ that arrive **later** can still interfere $\Rightarrow$ observe duration Δt_{obs} instead of a single moment

Reasoning – Higher-Class Bursts $\mathbf{y_{i,x}}$



- Higher class frames ■ that arrive **later** can still interfere
  $\Rightarrow$ observe duration Δt_{obs} instead of a single moment
- **How long** can these frames interfere?
  $\rightarrow$ as long as ■ is in the queue: d_i^{TQ}
  $\Rightarrow$ recursive relationship: $d_i^{TQ} \leq \dots d_i^{TQ} \dots$

Reasoning – Higher-Class Bursts $y_{i,x}$

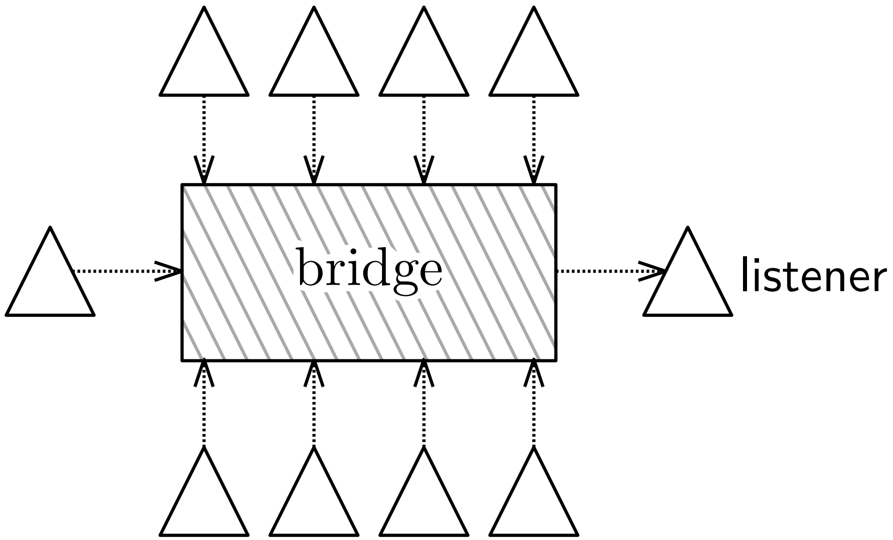


- Higher class frames ■ that arrive **later** can still interfere
  $\Rightarrow$ observe duration Δt_{obs} instead of a single moment
- **How long** can these frames interfere?
  $\rightarrow$ as long as ■ is in the queue: d_i^{TQ} upper bound $\delta_{p_i}^{h_k}$
  $\Rightarrow$ recursive relationship: $d_i^{TQ} \leq \ldots \cancel{d_i^{TQ}} \ldots$

Evaluation

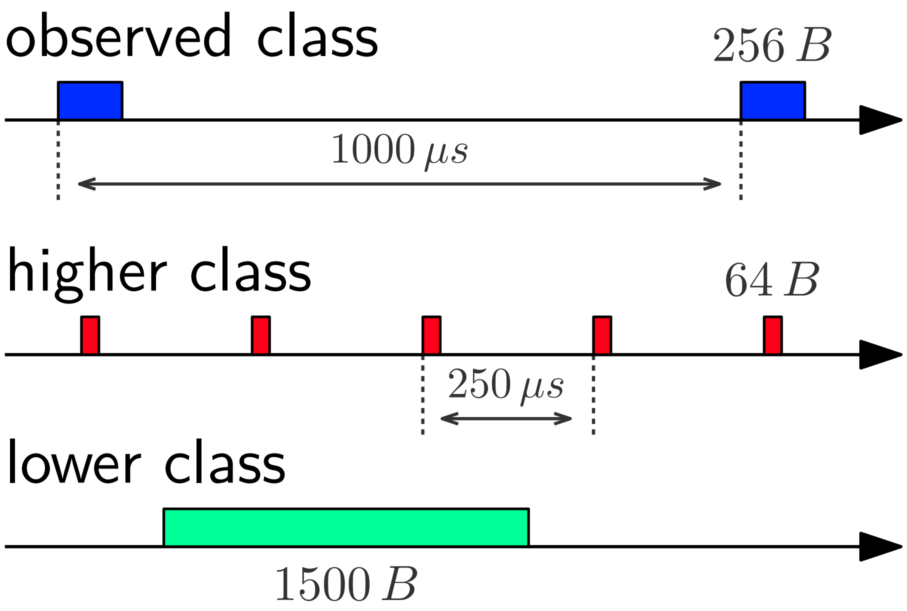
Inaccuracies and comparison to ATS

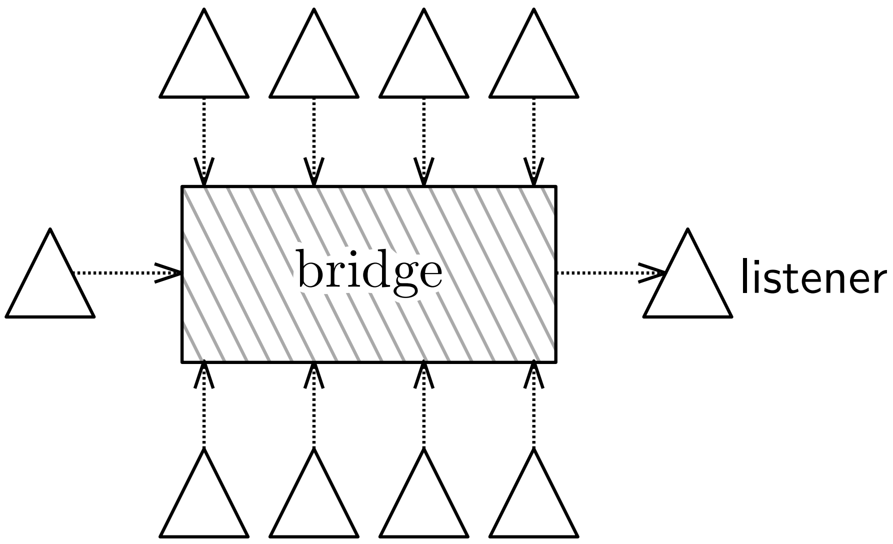
Worst Case Scenario Construction



- A single bridge is observed

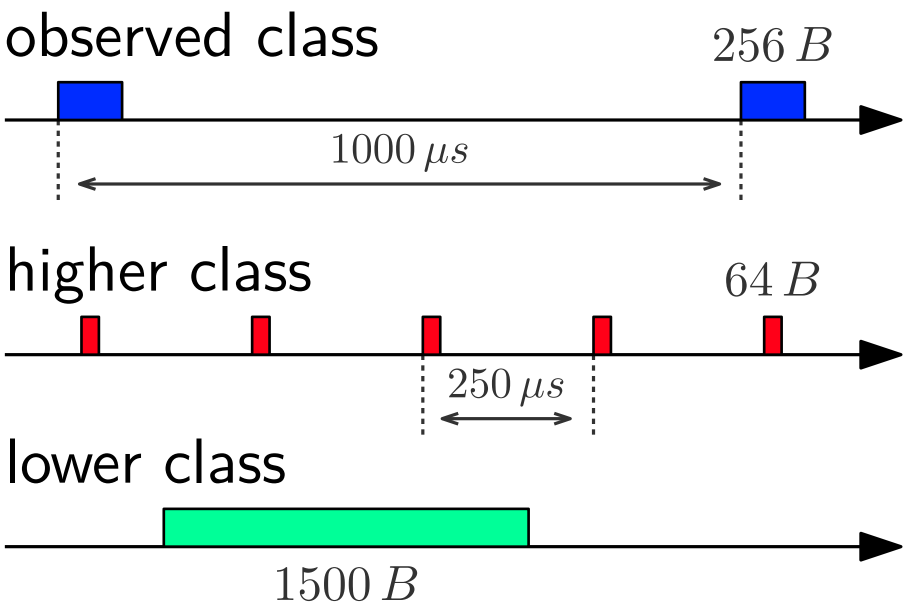
Worst Case Scenario Construction

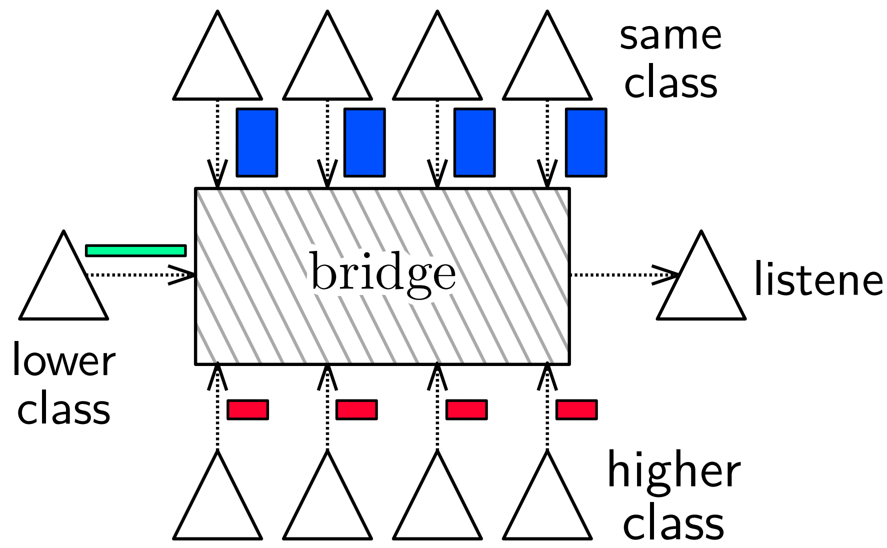




- A single bridge is observed
- Assuming periodic traffic (w.l.o.g.)

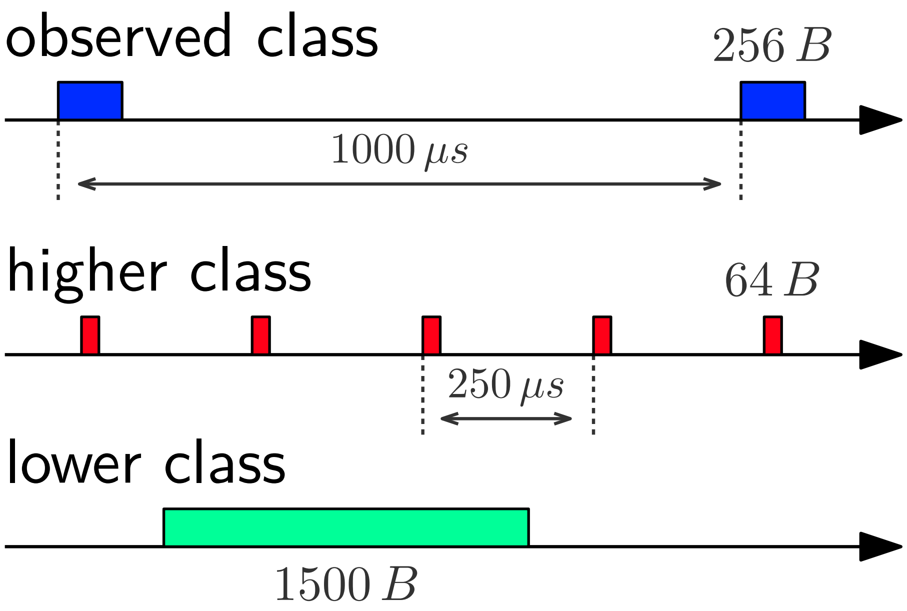
Worst Case Scenario Construction

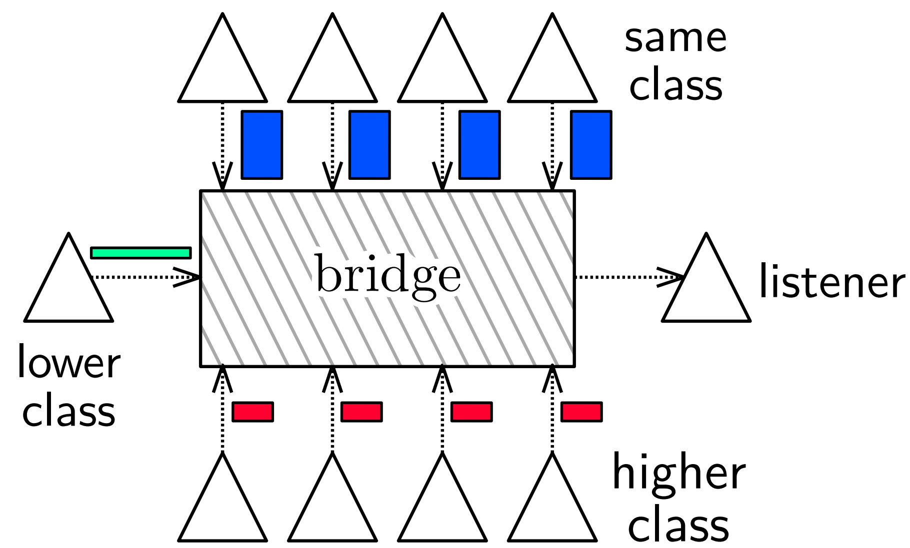




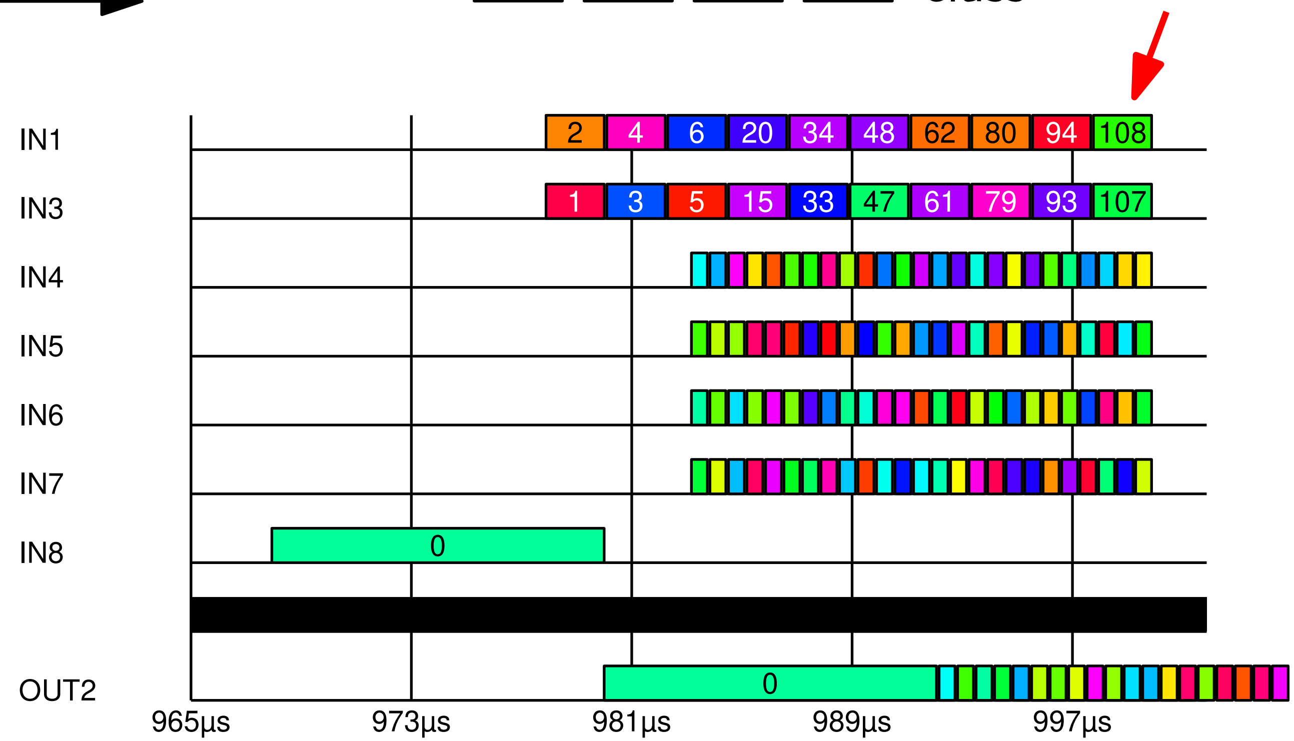
- A single bridge is observed
- Assuming periodic traffic (w.l.o.g.)
- Worst case construction and simulation

Worst Case Scenario Construction

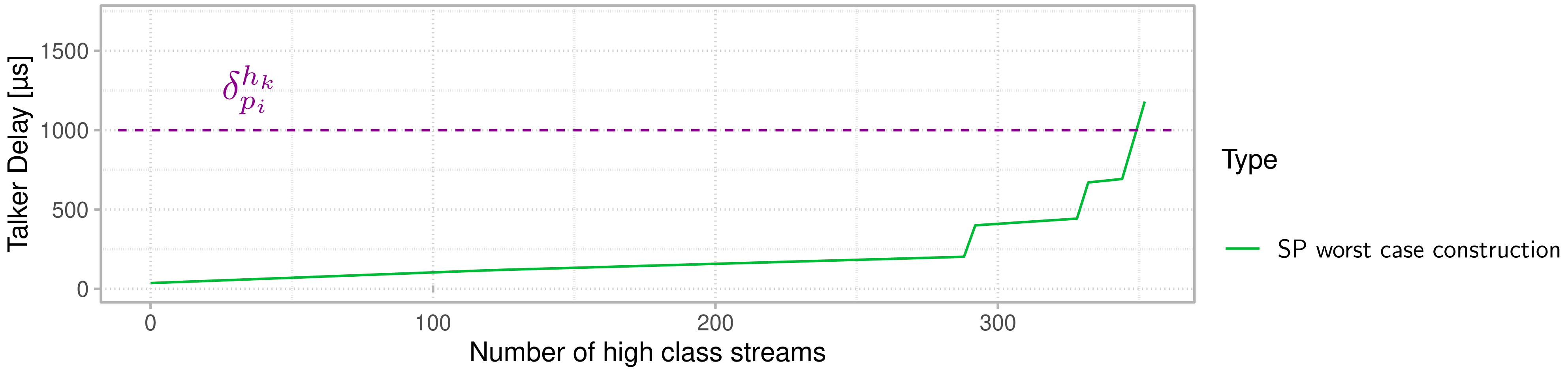




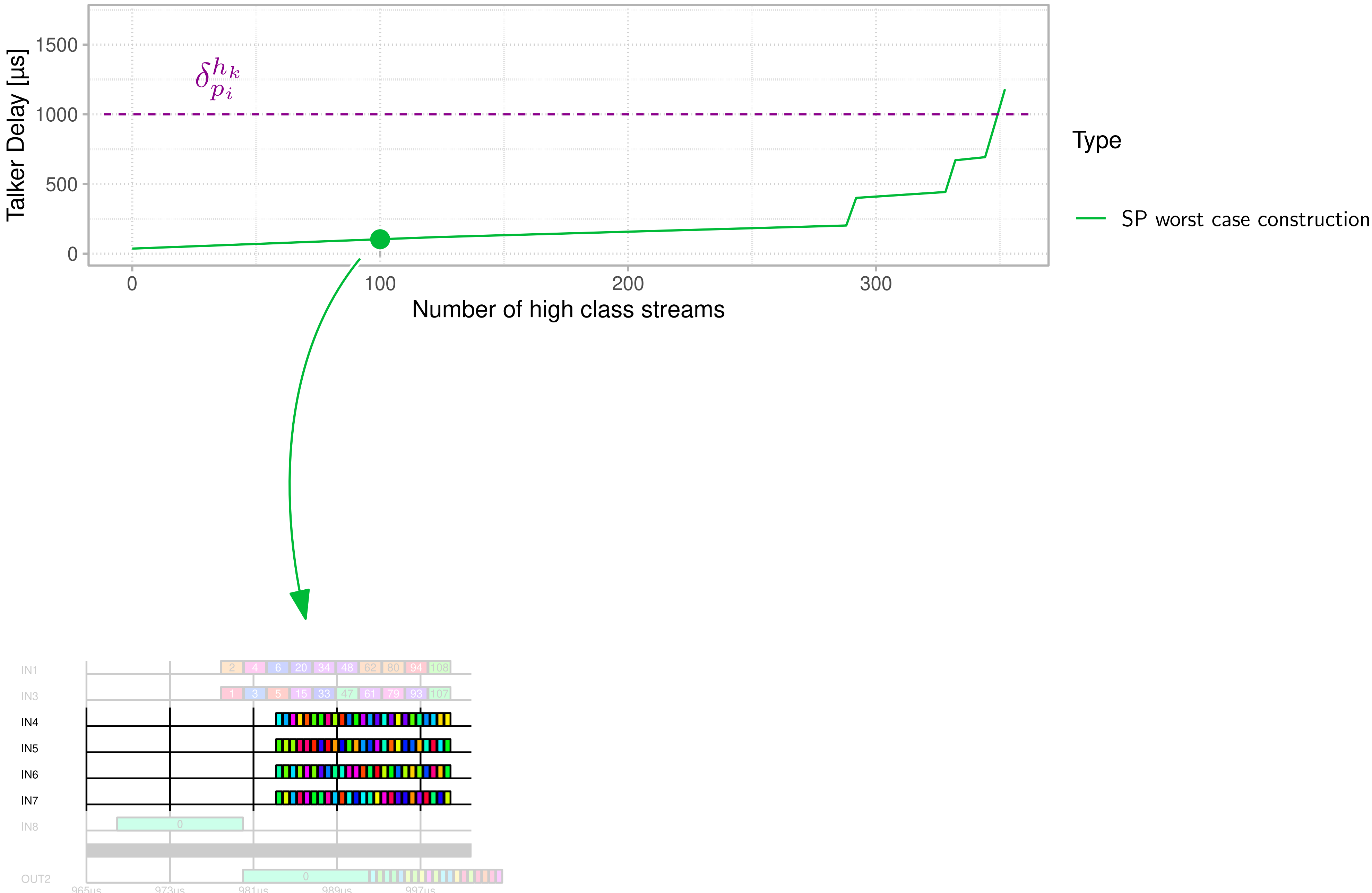
- A single bridge is observed
- Assuming periodic traffic (w.l.o.g.)
- Worst case construction and simulation



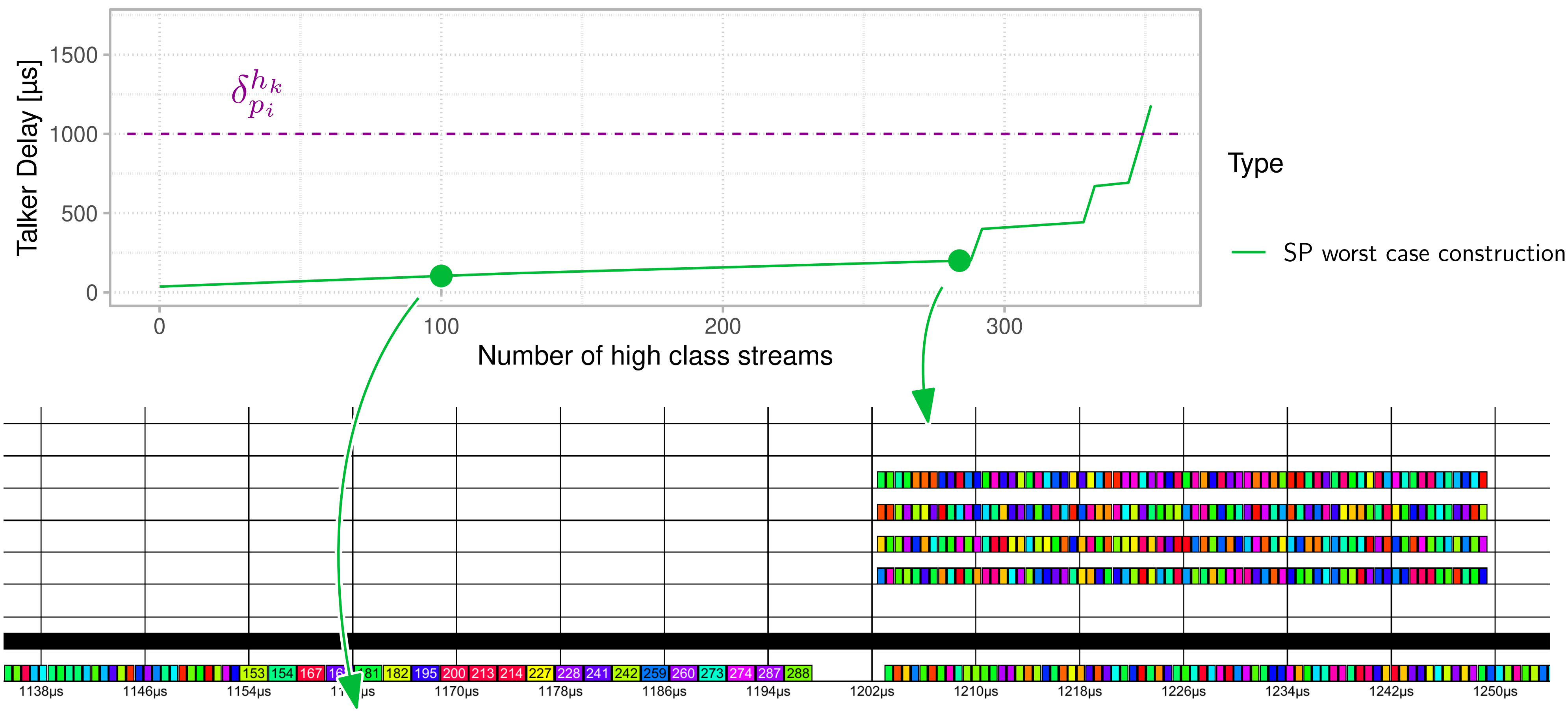
Worst Case Comparison – Higher Class Streams

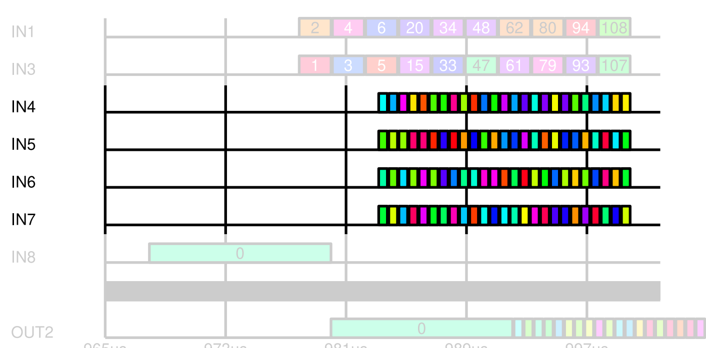


Worst Case Comparison – Higher Class Streams



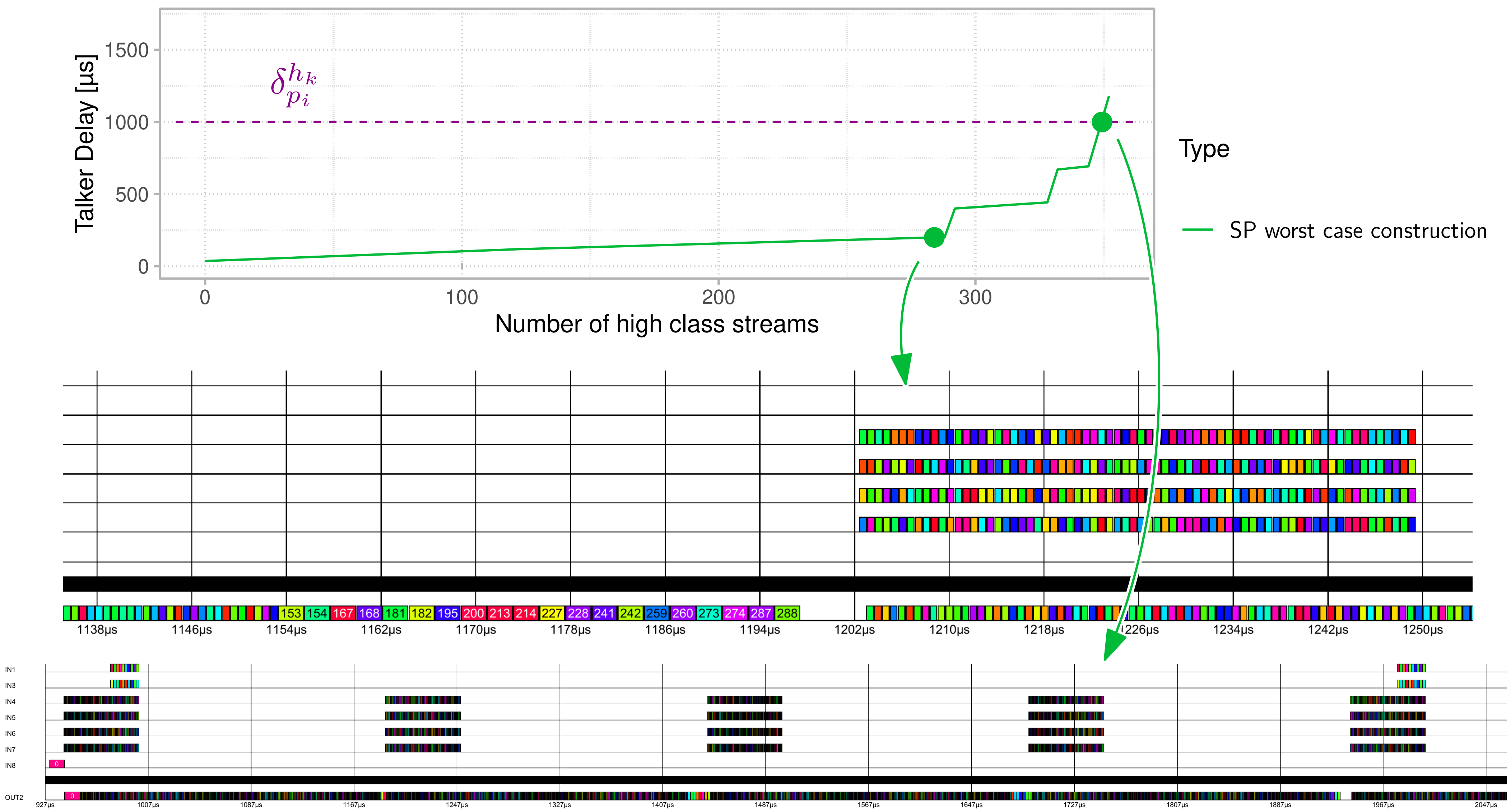
Worst Case Comparison – Higher Class Streams





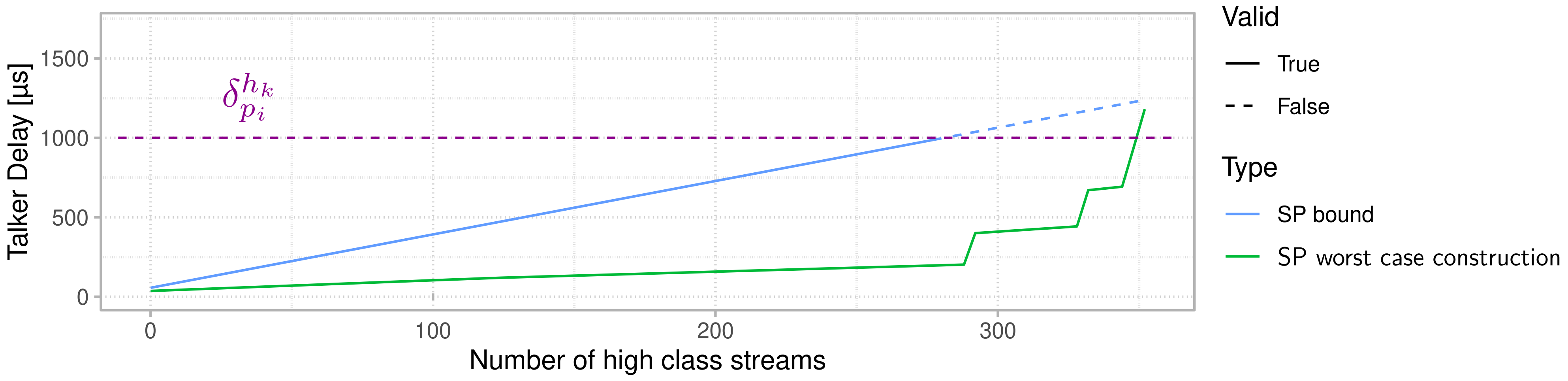
- Steep increase of worst case delay when close to τ_i

Worst Case Comparison – Higher Class Streams



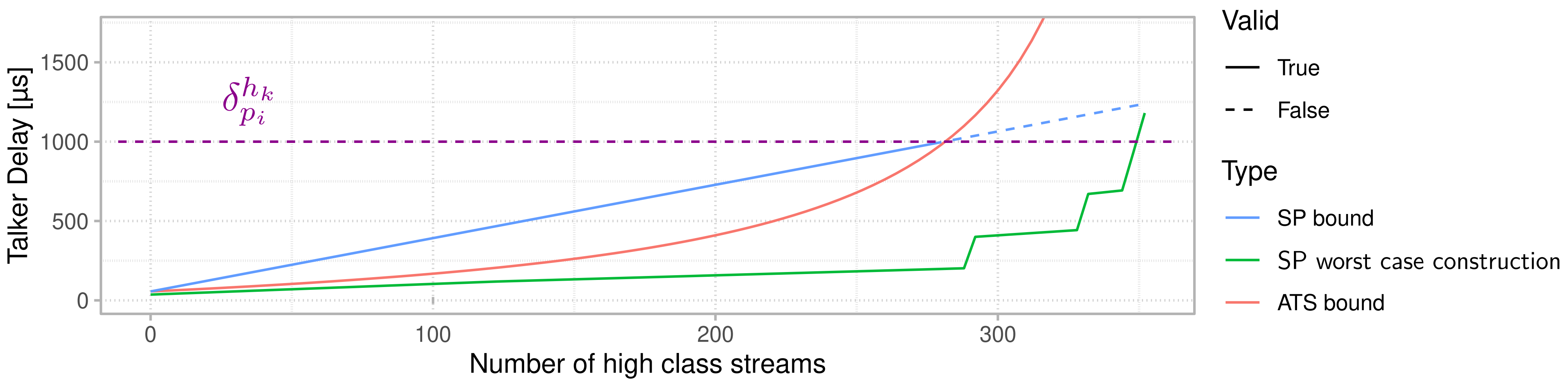
- Steep increase of worst case delay when close to τ_i
- Delay explodes near full bandwidth utilization

Worst Case Comparison – Higher Class Streams



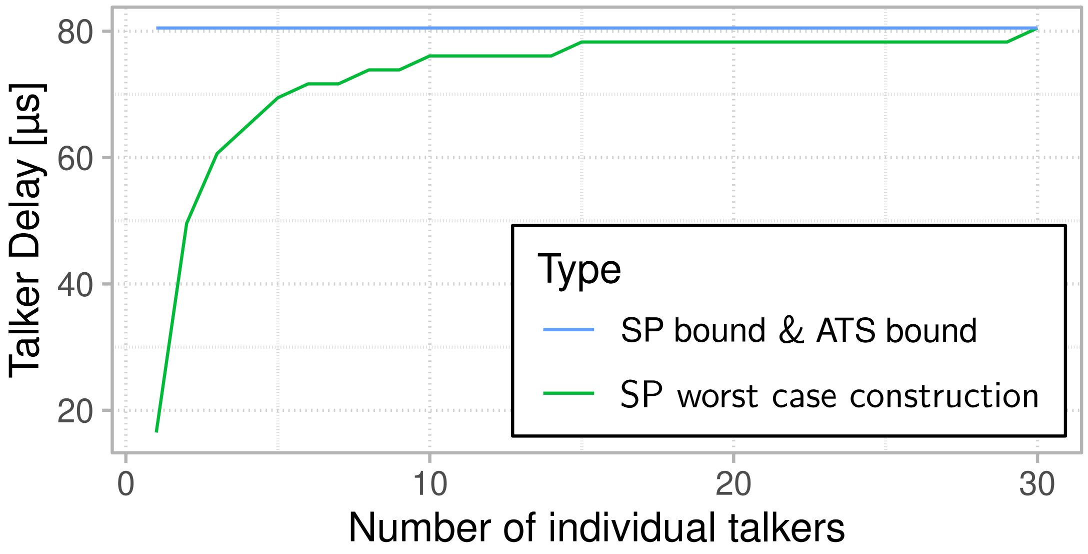
- Worst case construction
  - Steep increase of worst case delay when close to τ_i
  - Delay explodes near full bandwidth utilization
- SP bound moves linearly towards the full-utilization point
  - $\sim$280 streams can be deployed (instead of $\sim$350)

Worst Case Comparison – Higher Class Streams



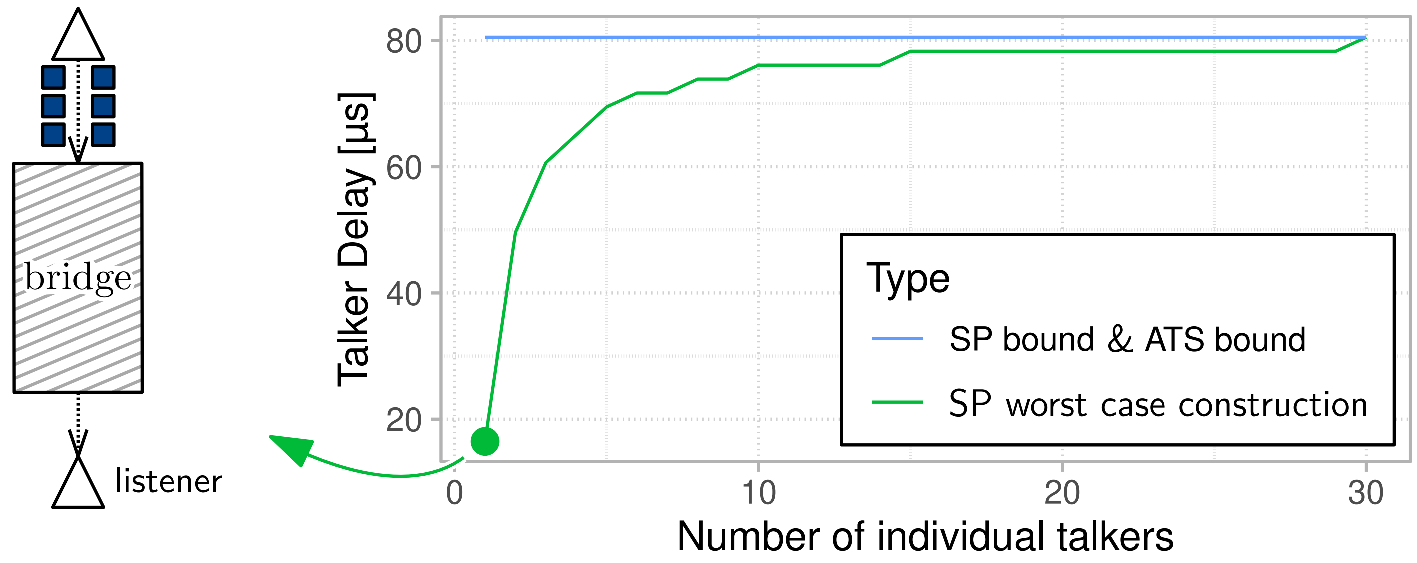
- Worst case construction
  - Steep increase of worst case delay when close to τ_i
  - Delay explodes near full bandwidth utilization
- SP bound moves linearly towards the full-utilization point
  - $\sim$280 streams can be deployed (instead of $\sim$350)
- Comparison to Asynchronous Traffic Shaping (Qcr)
  - ATS bound starts lower, but reaches $\delta_{p_i}^{h_k}$ at the same point

Worst Case Comparison – Frame Locality



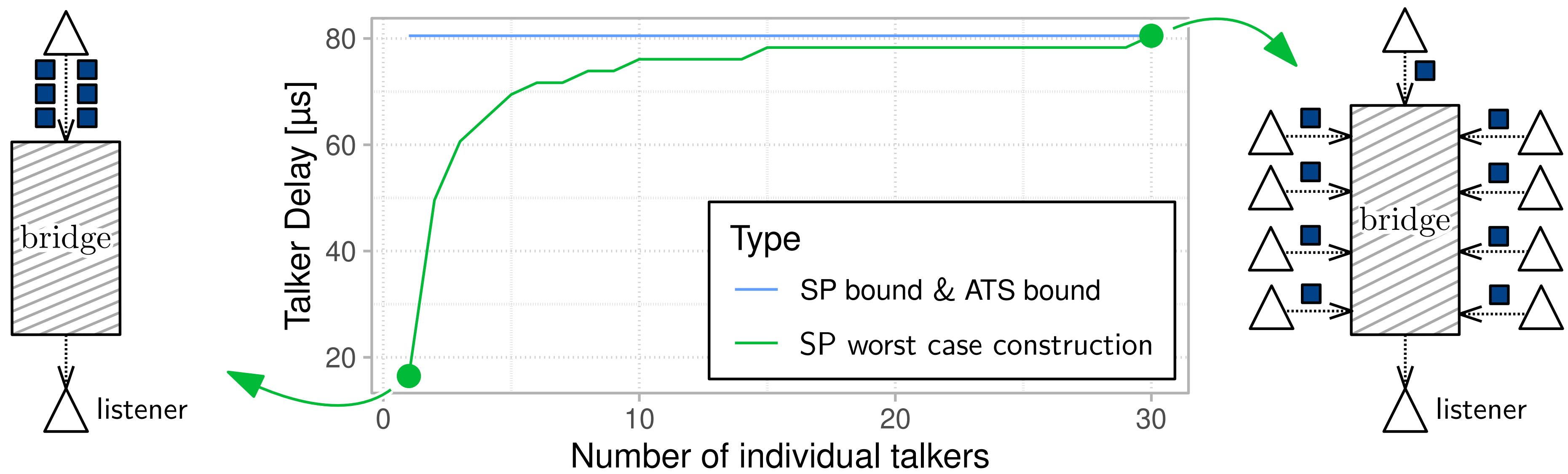
- Influence of network topology
  → Frames arriving from a single port vs. many in-ports

Worst Case Comparison – Frame Locality



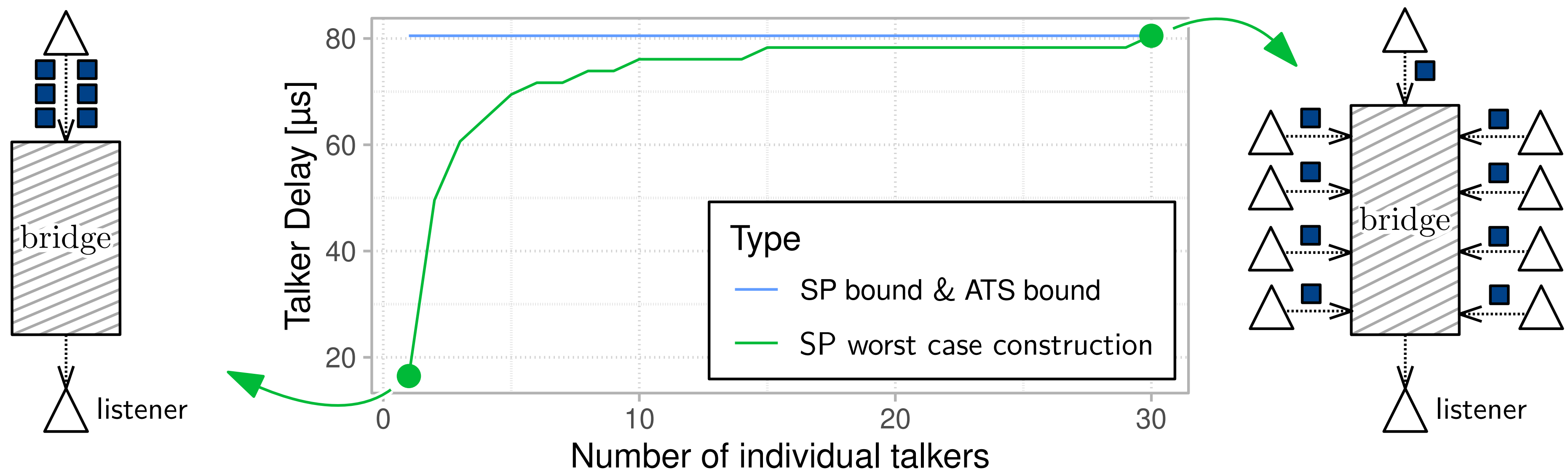
- Influence of network topology
  → Frames arriving from a single port vs. many in-ports
- Frames show less interference if they arrive from the same in-port

Worst Case Comparison – Frame Locality



- Influence of network topology
  $\rightarrow$ Frames arriving from a single port vs. many in-ports
- Frames show less interference if they arrive from the same in-port
- SP latency bound reached when all frames arrive from different ports

Worst Case Comparison – Frame Locality



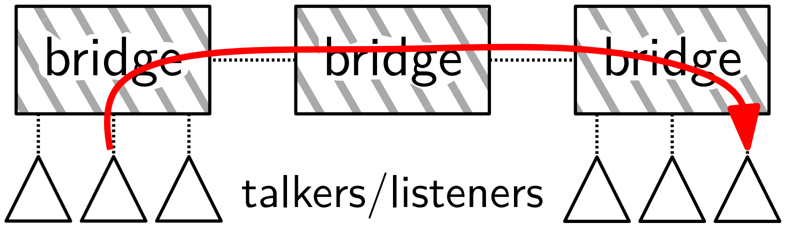
- Influence of network topology

  → Frames arriving from a single port vs. many in-ports
- Frames show less interference if they arrive from the same in-port
- SP latency bound reached when all frames arrive from different ports
- This is not solved by a simple subtraction!

  → The time interval, during which higher class frames can interfere, would become larger than the per-hop delay. $\boxed{\Delta t_{obs} \not\leq \delta_{p_i}^{h_k}}$

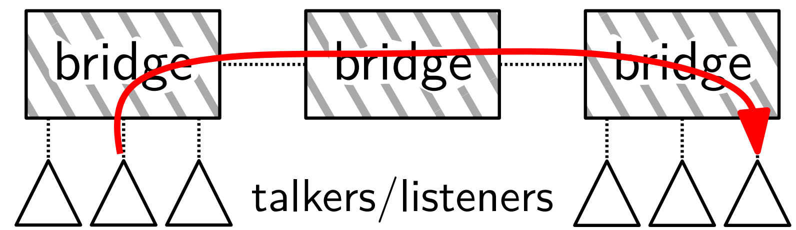
Comparison of Network Capacities – Setup

- Deployment of random* streams in a small network
- Admission control: check whether $d_i^{TQ,SF} \leq \delta_{p_i}^{h_k}$ for every hop



Comparison of Network Capacities – Setup

- Deployment of random* streams in a small network
- Admission control: check whether $d_i^{TQ,SF} \le \delta_{p_i}^{h_k}$ for every hop

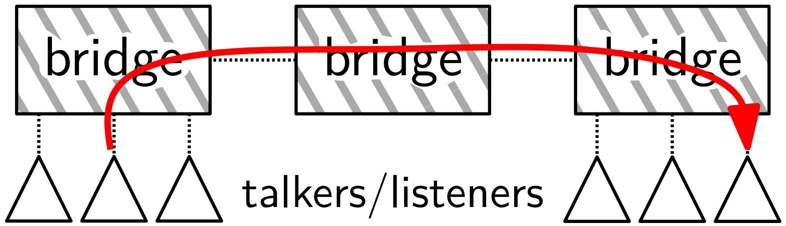


- *Random streams...
  - Random talker
  - Random listener
  - Random configuration from table →
  - 20 instances of each parameter set
  - Mean capacity with 99.5% confidence intervals reported

Traffic class p_x	**Burst** $b_x = \hat{\ell}_x$	**Burst interval** τ_x
3 (high)	128 B	250 µs
3 (high)	256 B	500 µs
3 (high)	512 B	1000 µs
2 (low)	1024 B	2000 µs
2 (low)	1522 B	4000 µs

Comparison of Network Capacities – Setup

- Deployment of random* streams in a small network
- Admission control: check whether $d_i^{TQ,SF} \leq \delta_{p_i}^{h_k}$ for every hop

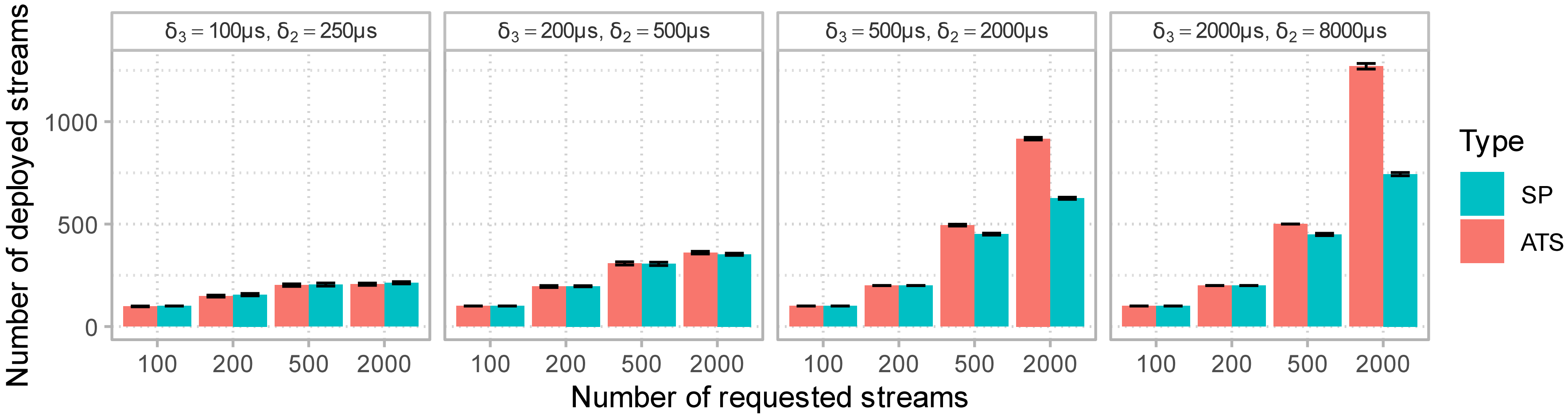


- *Random streams...
  - Random talker
  - Random listener
  - Random configuration from table →
  - 20 instances of each parameter set
  - Mean capacity with 99.5% confidence intervals reported

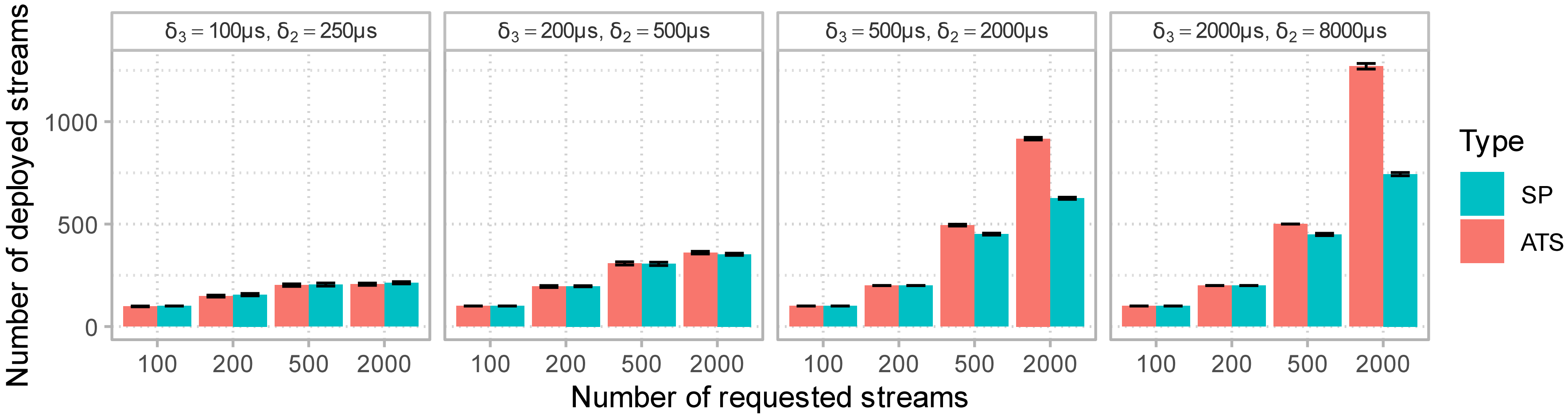
Traffic class p_x	**Burst** $b_x = \hat{\ell}_x$	**Burst interval** τ_x
3 (high)	128 B	250 µs
3 (high)	256 B	500 µs
3 (high)	512 B	1000 µs
2 (low)	1024 B	2000 µs
2 (low)	1522 B	4000 µs

- Parameters
  - Number of deployed streams: 100 – 2000
  - Per-hop delay guarantees for both traffic classes (δ_3, δ_2): 100 µs – 8000 µs

Comparison of Network Capacities – SP vs ATS

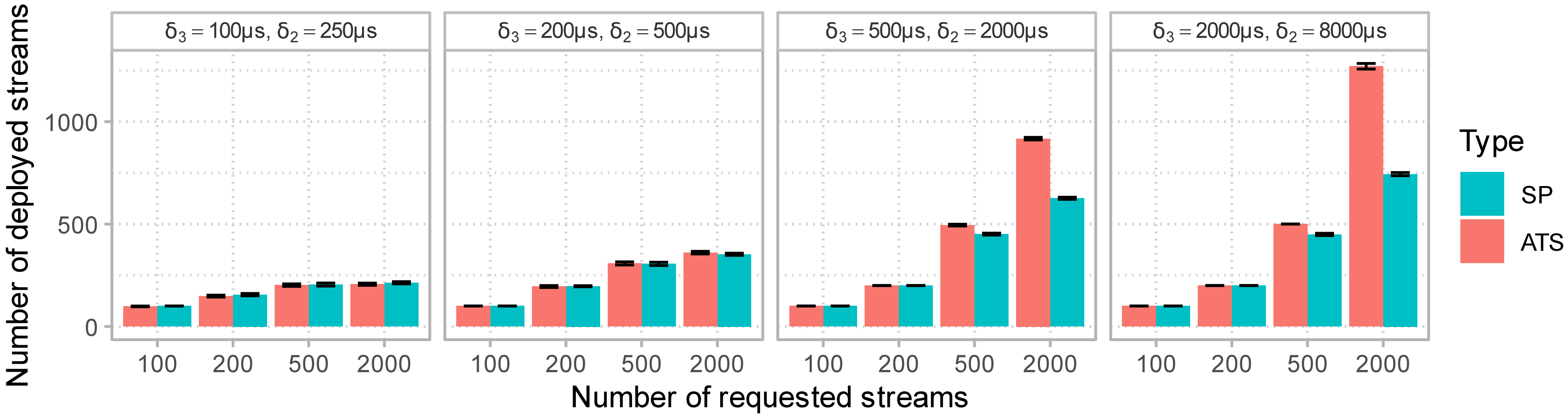


Comparison of Network Capacities – SP vs ATS



- No significant difference with small per-hop delay guarantees δ_{p_i}
  - Per-hop reshaping shows little effect if only one burst of each stream is in the network at the same time (cf. residence times)
  - SP is a viable alternative if burst intervals are large in comparison to δ_{p_i}

Comparison of Network Capacities – SP vs ATS



- No significant difference with small per-hop delay guarantees δ_{p_i}
  - Per-hop reshaping shows little effect if only one burst of each stream is in the network at the same time (cf. residence times)
  - SP is a viable alternative if burst intervals are large in comparison to δ_{p_i}
- ATS shows better network utilization than SP for large guarantees δ_{p_i}
  - Multiple bursts of the same stream in the network if $\tau_i \geq$ end-to-end delay
  - Per-hop beneficial for less impairment
  - SP may still be a viable: remaining bandwidth can be used by best effort traffic

Conclusion

- Bridge-local bounded latency with SP is feasible
  - **Proven** delay guarantee with low complexity for **distributed** systems
  - Bound only applicable in admission control scenarios
  - Streams whose latency exceeds their guarantee **must** be denied
- SP shows good network utilization in many situations
  - Capacity comparable to ATS for "large" transmission intervals
  - Still viable with "small" intervals $\rightarrow$ remaining bandwidth can be used by BE
- Requirements are similar to other mechanisms
  - Most information is already contained in current TSpec fields of Qcc
  - Accuracy can be improved by `accMinLatency` and `minFrameSize`
- Can be adapted depending on the scenario
  - Improving inaccuracies due to frame locality
  - Adaptation for other mechanisms (e.g. distributed admission control with TAS)