Ultra-Low Latency Shaper

Christian Boiger
christian.boiger@fh-deggendorf.de
IEEE 802 Plenary
July 2011
San Francisco, CA

Why Is a Ultra-Low Latency Class Necessary?

- Current priority of traffic classes
  1. AVB Classes (Class A and B)
  2. Strict Priority

- Without a new shaper/class concept the current Class A with preemption would guarantee the lowest possible latency
  - ➔ per hop worst case latencies > 125µs
  - This does not meet the proposed automotive and industrial goals in the Gen2 assumptions document

- Legacy traffic with the priority of the Ultra-Low Latency (UL) traffic (from outside the Ultra-Low Latency domain) has to be remapped

  => A Ultra-Low Latency Class is necessary. This can be realized with a three class concept (3 AVB Classes) or a Class A' and B' (2 AVB Classes).

New Class vs. Two Class Concept

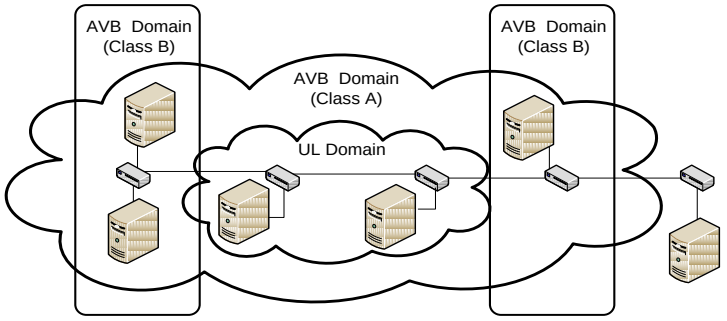
Possible new order of traffic classes (highest priority first)

- With a new Ultra-Low Latency Class
  1. Ultra-Low Latency Class
  2. AVB Class A
  3. AVB Class B
  4. Strict Priority

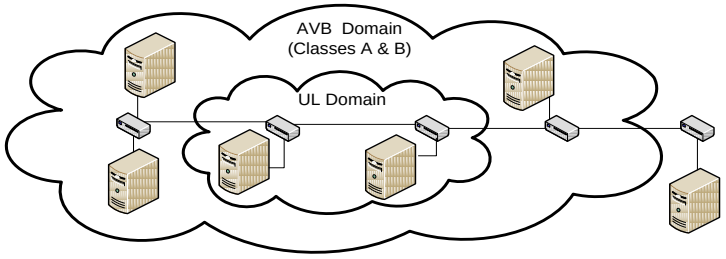
- Without a new Ultra-Low Latency Class
  1. Class A’ (could be Ultra-Low Latency or Gen 1 Class A)
  2. Class B’ (could be Gen 1 Classes A or B)
  3. Strict Priority

Is a New Class Necessary?

- Disadvantages of a two class concept (only Classes A' and B')
  - Not backward compatible with AVB Gen1 networks
    The traffic of a Gen1 AVB network which is using Class A and B cannot be transported over a Gen 2 AVB network which is using the ultra-low latency class
  - No support for wireless AVB traffic in an UL network(assuming AVB Gen1 Class A is used in the network)



- Disadvantages of a three class concept
  - Additional PCP is necessary



Is a "Shaper" Necessary?

- The frames of the Ultra-low Latency (UL) Class have the highest priority
- Therefore the bandwidth has to be limited and controlled in order to allow other traffic
- Strict priority with preemption
  - No bandwidth control
  - Which Class should be used?
  - Unpredictable latencies for AVB streams if strict priority has a higher priority than AVB Classes

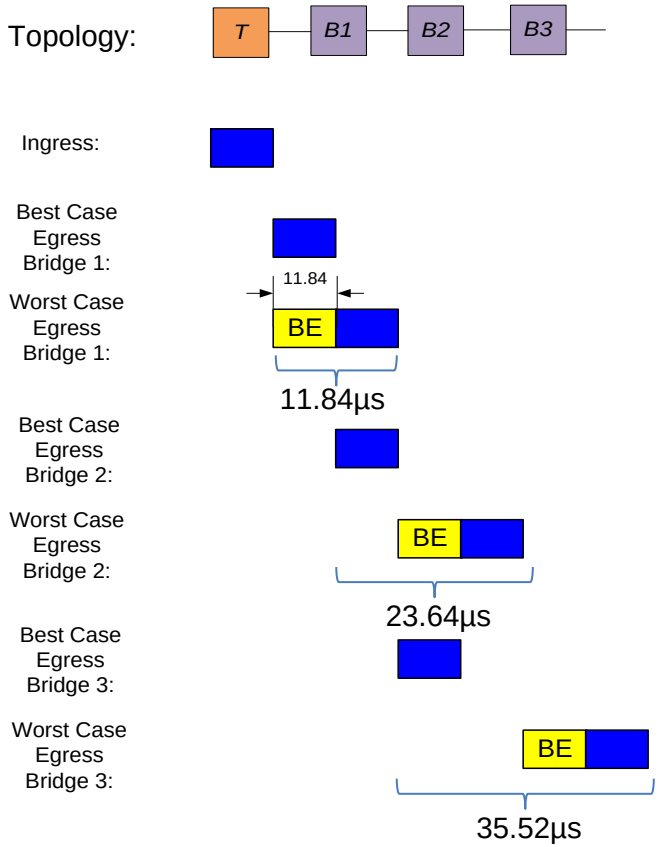
Possible Shaper

- Bursting Shaper (the shaper allows short bursts of ultra-low latency class traffic) (http://www.ieee802.org/1/files/public/docs2010/at-goetz-AVB-lowlatency-part1-0510.pdf)
- Time Aware Shaper (the shaper defines time slots in which the transmission of the ultra-low latency class traffic is guaranteed) (http://www.ieee802.org/1/files/public/docs2011/new-pannell-latency-options-0311-v1.pdf)
- Any other?

Bursting Shaper

- Bursting without further shaping and bandwidth observation (→babbling idiot can break the whole network)
- Bursting with bandwidth observation (drop frames when reserved bandwidth is exceeded?)
  - How big should the observation interval be?
  - Which frame gets dropped (oldest first?)
- Credit Based Shaper with “positive base”
- (Credit Based Shaper with an increased idle slope)
- Any other?

Problems With Non Time Aware Approaches



Assuming a minimum fragment frame size of 128 byte @ FE

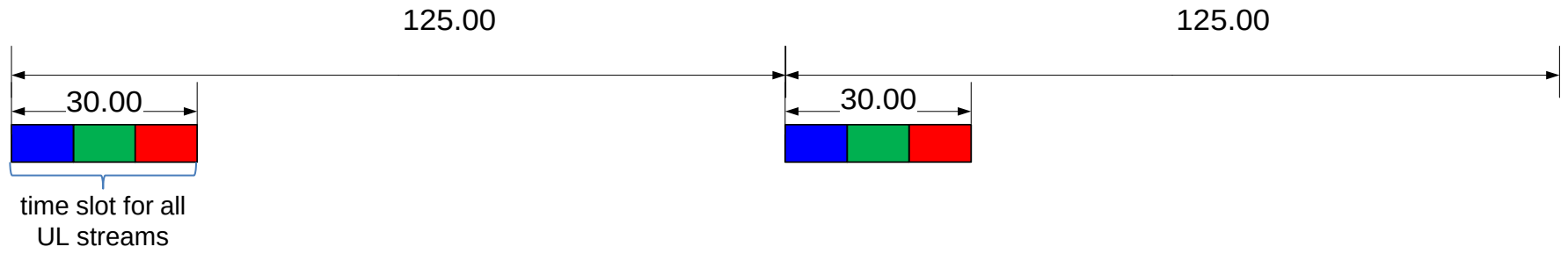
- ➔ Growing arriving window (jitter)
- ➔ No real determinism
- ➔ Collisions of UL frames possible (even with coordinated Talkers)
- ➔ No latency guarantee especially in bigger networks (assuming a transmission period of 125µs)

Time Aware Shaper

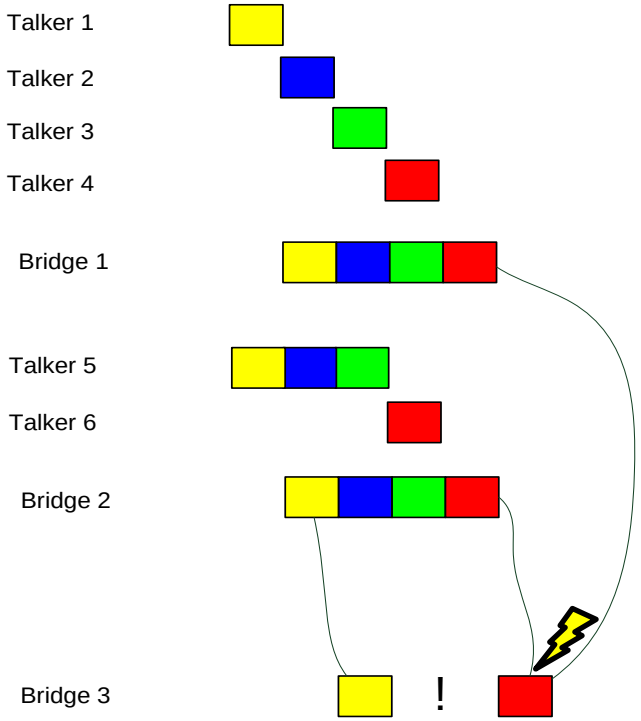
- One time slot for all UL streams
- One time slot for each UL stream
- Any other?

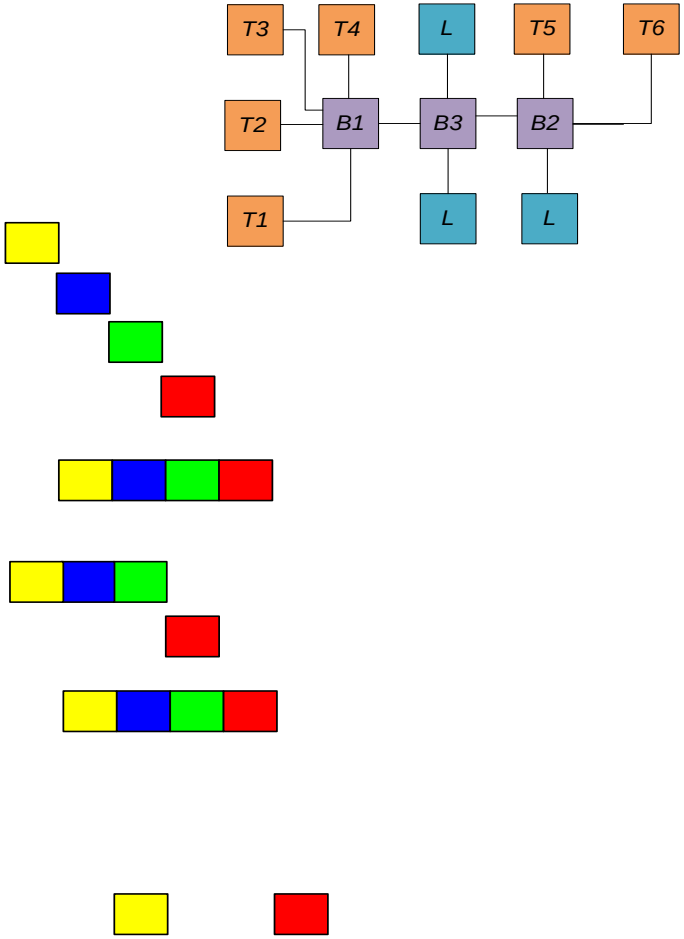
One Time Slot For All UL Streams

- Talkers have to be coordinated
- Not always possible (topology, stream)
- Fixed transmission period necessary (e.g. 125µs)
- Wasting bandwidth



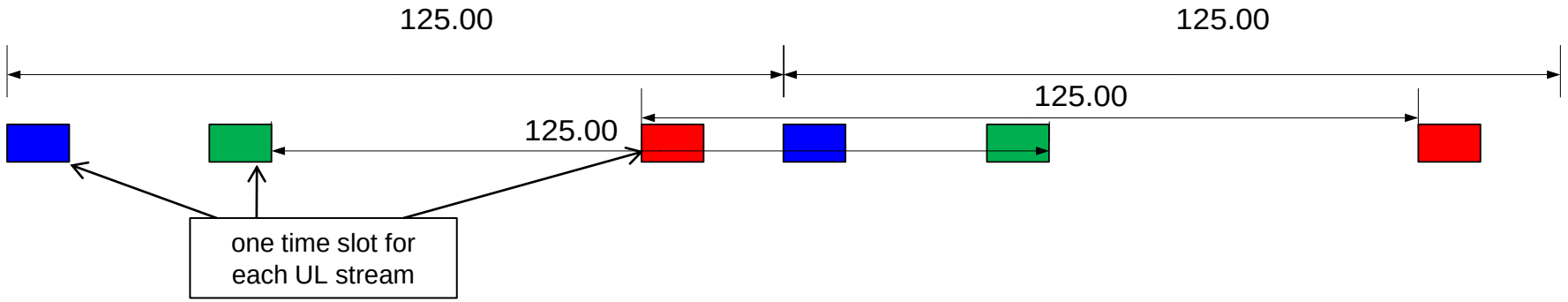
One Time Slot





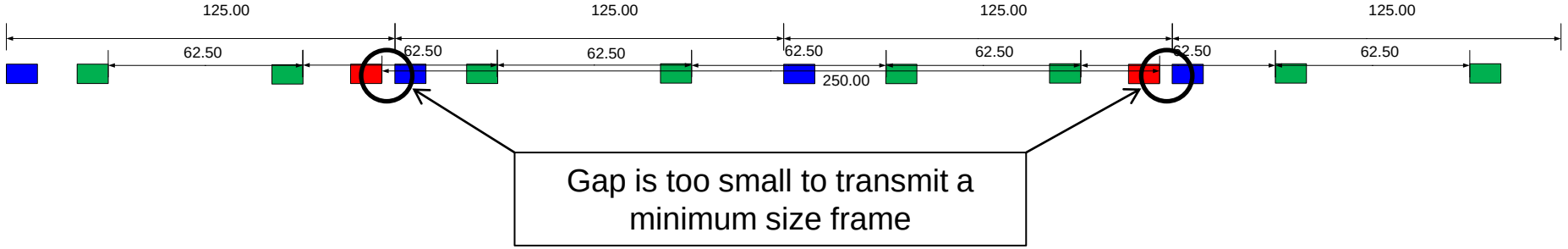
One Time Slot For Each UL Stream

- Reservation through SRP Gen2 possible (for very simple networks and very few streams)
- Engineering in bigger networks necessary (i.e. configuration of bridges and end stations via management)
- Variable transmission periods possible

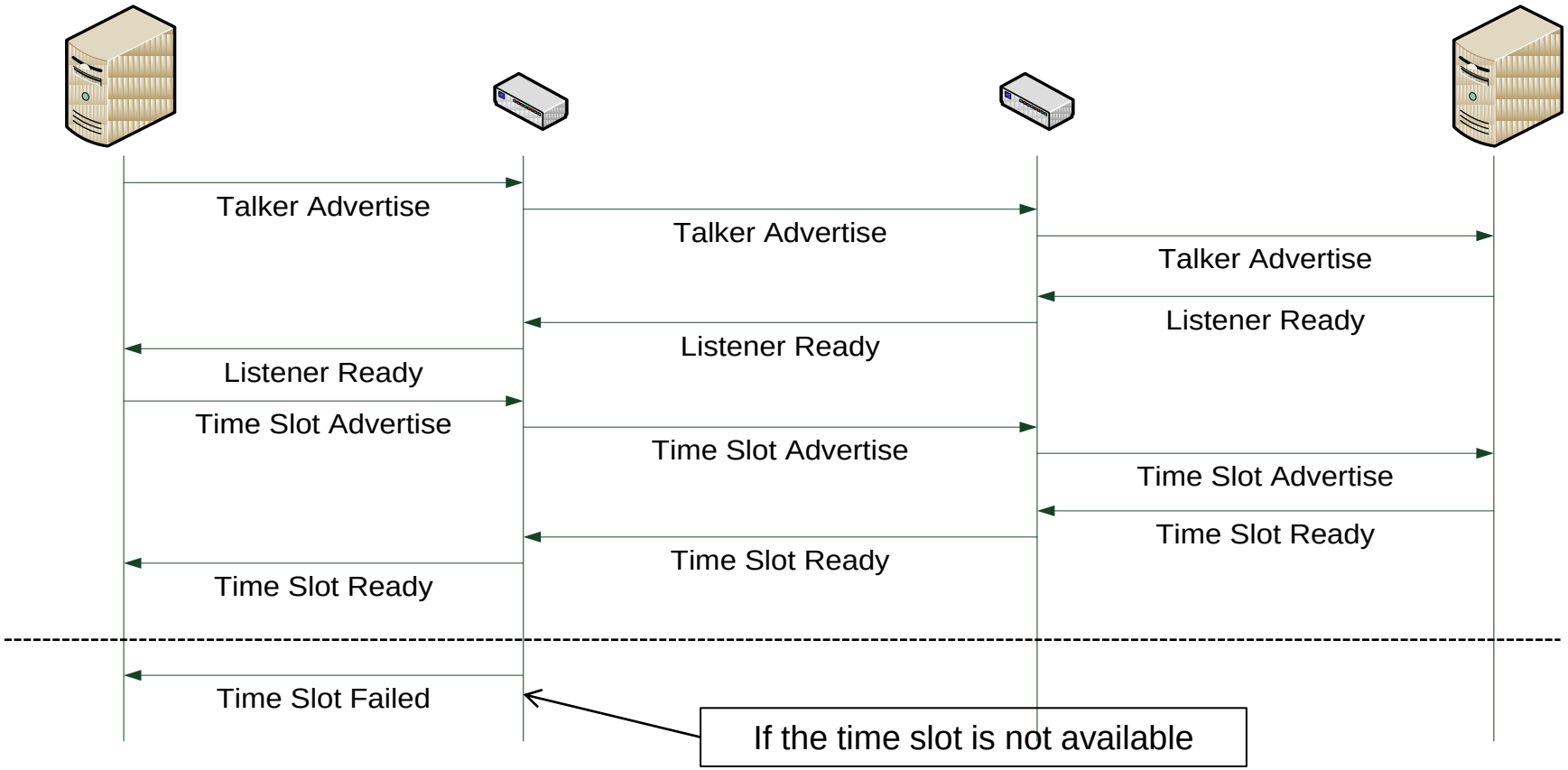


Variable Transmission Periods

- Especially in the time-aware approaches it is important to include a mechanism for variable transmission periods
- Otherwise transmission periods < the main one are not possible and transmission periods > the main one lead to a significant loss of bandwidth
- In order to avoid collisions of time slots transmission periods have to be multiples of each other (e.g. 31.25µs, 62.5µs, 125µs, 250µs, …)



Time Slot Reservation with SRP Gen2?



Possible Issues

- Problems in big complex networks
  - In some cases established UL streams might make it impossible for a new stream to be established even if it would be theoretically possible with a different scheduling of the streams
  - Streams with many listener in a complex network might be difficult to establish (but no problem in a line topology)
- Time between two reserved streams might be too short to transmit a min size frame or fragment
- => Big and complex networks or networks with many UL streams have to be engineered
- => Management interface necessary to predetermine the values, similar as proposed for stream reservation in Gen2

Thank You