MaxLatency Contribution

Contribution to IEEE 802.1Qdj

Astrit Ademaj, Juergen Wohlmuth

TTTech Industrial Automation AG Internal

December 14, 2020

IEEE 802.1Qcc Definitions

IEEE 802.1Qcc specifies a MaxLatency model and the CNC-CUC interface, and defines different reference points for:

- time-aware streams and
- non-time-aware streams

MaxLatency is a request parameter to the CNC for a particular stream.

There are practical issues with the current definitions as well as the YANG model (some of these are also presented and discusses in IEEE 802.1 Interim meeting in Sept Oslo 2018)

MaxLatency requirement at listeners

§3.118 Latency – is currently defined as:

> **The delay experienced by a frame in the course of its propagation between two points** in a network, measured from the time that a known reference point in the frame passes the **first point** to the time that the reference point in the frame passes the **second point.**

§46.2.3.6.2 MaxLatency – is currently defined as:

> Latency shall use the definition of 3.118, with additional context as follows: The 'known reference point in the frame' is the message timestamp point specified in IEEE Std 802.1AS for various media (i.e. start of the frame). The **'first point'** is in the Talker, at the reference plane marking the boundary between the network media and PHY (see IEEE Std 802.1AS). The **'second point'** is in the Listener, at the reference plane marking the boundary between the network media and PHY.

Issue 1: “MaxLatency” definition (1)

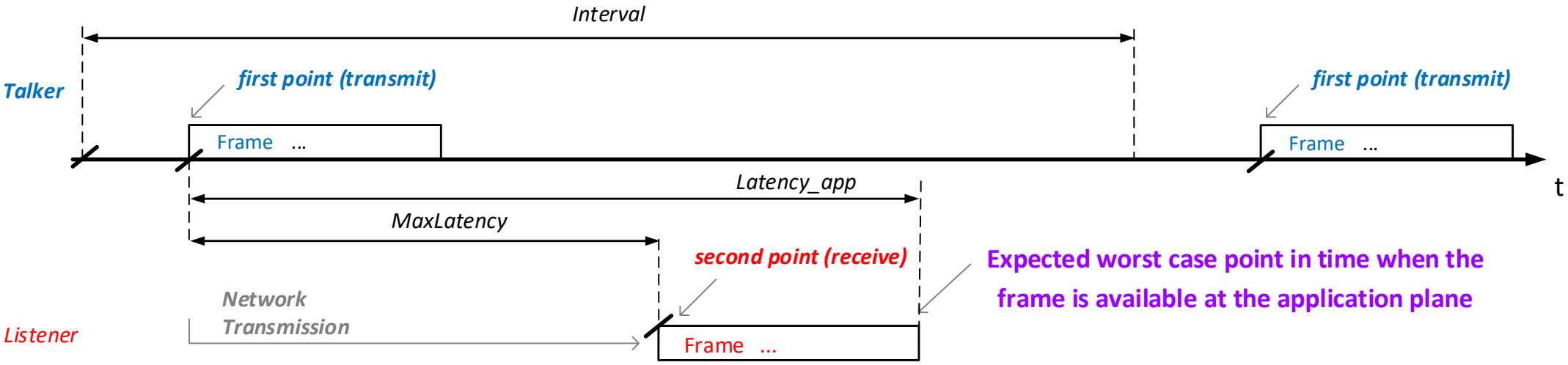
- When specifying a Latency requirements at the listener, from the user/application point of view – it specifies the time where the frame is available to the application layer/plane (**`Latency_app`**)
- With the current **`MaxLatency`** definition, the listener needs to consider
  - i) ***the maximum frame length***, and
  - ii) ***the links speed***

  in order to calculate the **`MaxLatency`** (to make the request to the control plane)

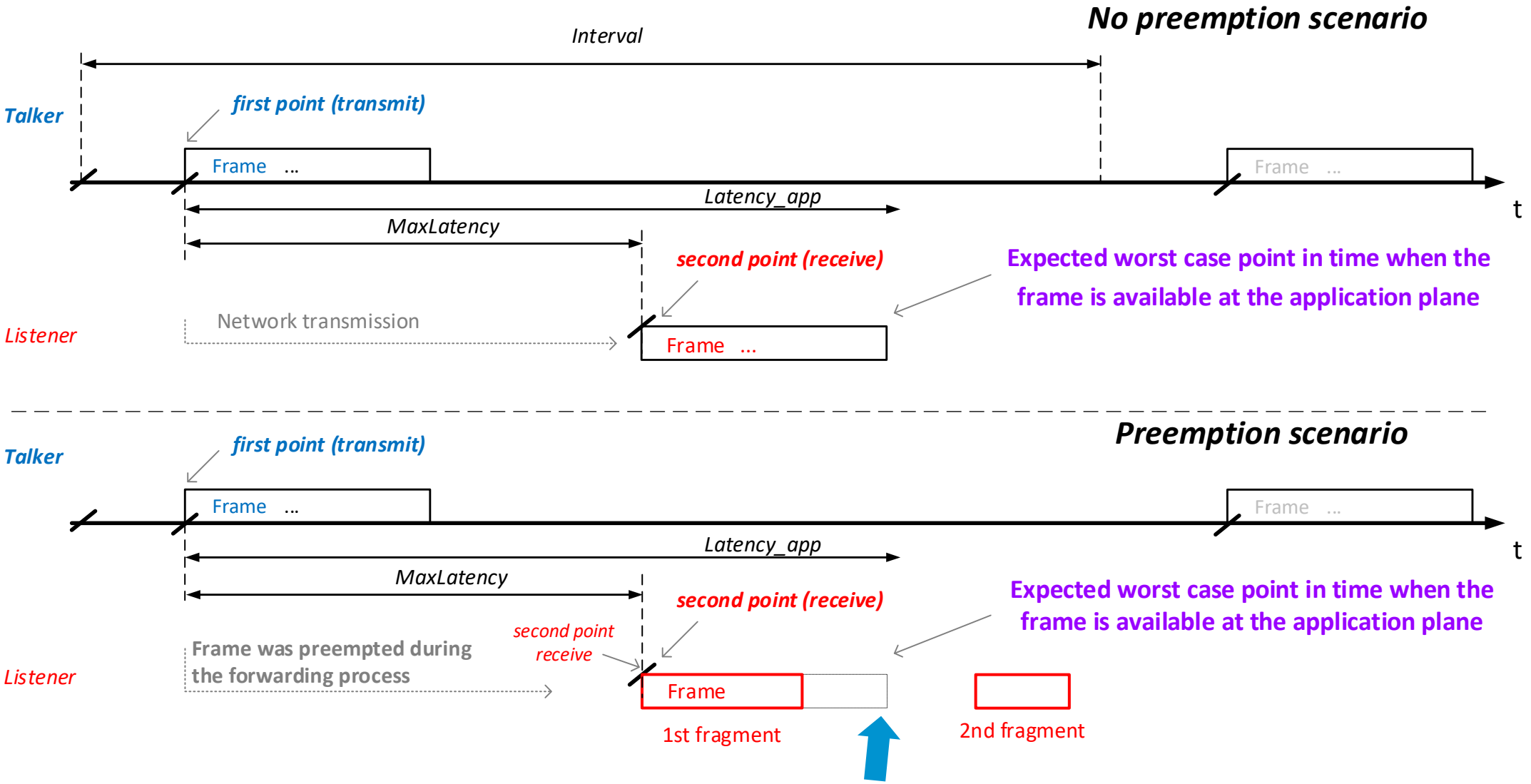
Example: 100Mbit/s link and frame size 1522

Latency_App = 1000µs

MaxLatency = 1000µs – 123µs



Issue 1: “MaxLatency” definition (2)



In case of preempted packets, with the actual “MaxLatency” definition, it is not possible for the listener to derive the point in time when the stream‘s frame is fully received, i.e. the point in time when the internal processing starts

Issue 1: “MaxLatency” definition (3)

With the current definition of MaxLatency and with preemption activated following “options” are possible:

a) CNC has to make sure that the preemption does not occur in the last hop (it means disabling preemption to all bridge ports connected to endpoints

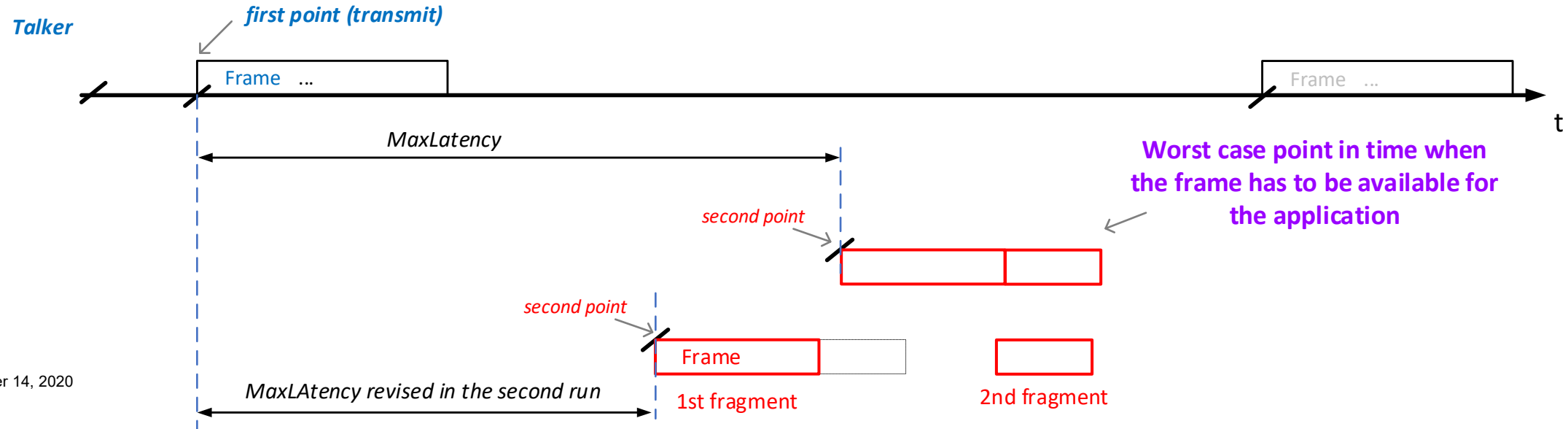
b) the application must account for the interleaving delay of the preempted fragment when calculating MaxLatency out of Latency_App

c) to account for the interleaved fragment delay, CNC must recalculate/revise internally the MaxLatency (redefine the second point) after doing the calculations in the first run (feedback loop – see example below)

not practical

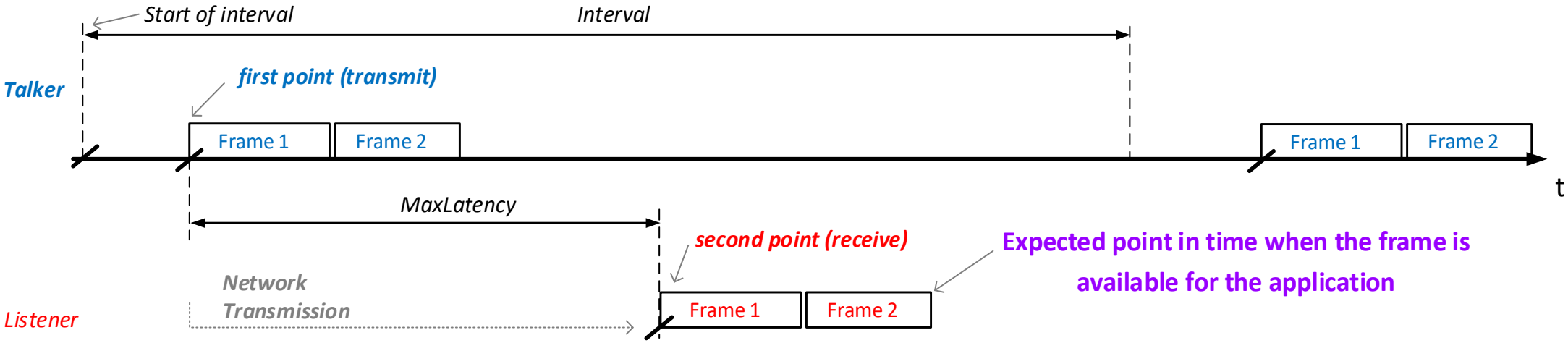
not possible as this knowledge is within the CNC and not in the local application)

significant increase of the CNC complexity

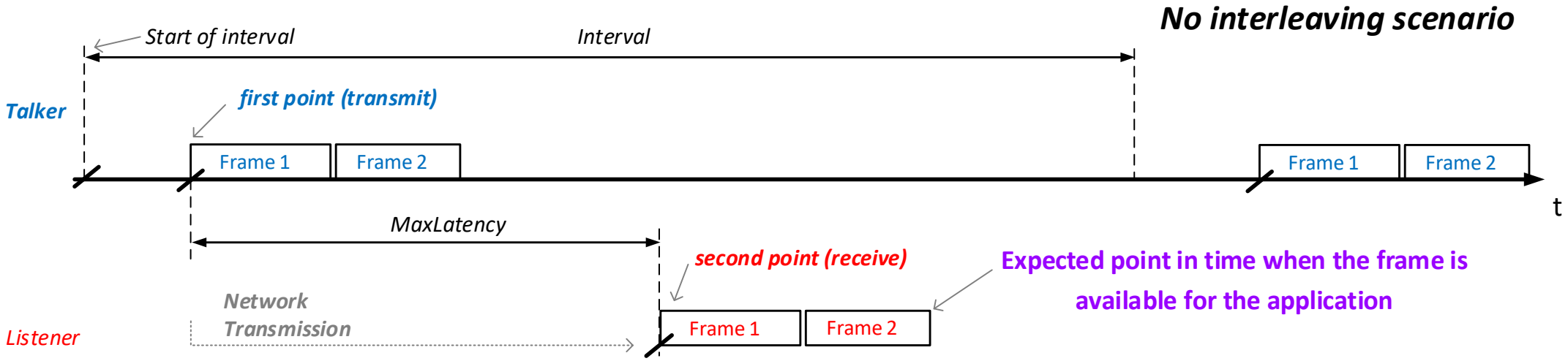


Issue 1: “MaxLatency” definition (4)

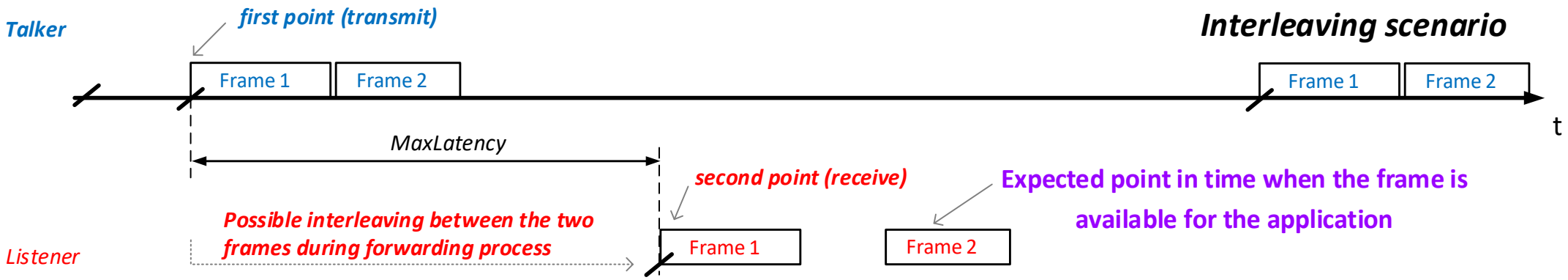
- Definition (46.2.3.6.2 MaxLatency): Maximum latency from Talker to Listener(s) for a single frame of the Stream.
- In case of streams with multiple frames (`max-frames-per-interval >1)`, the listener shall assume that the frames are send back-to-back



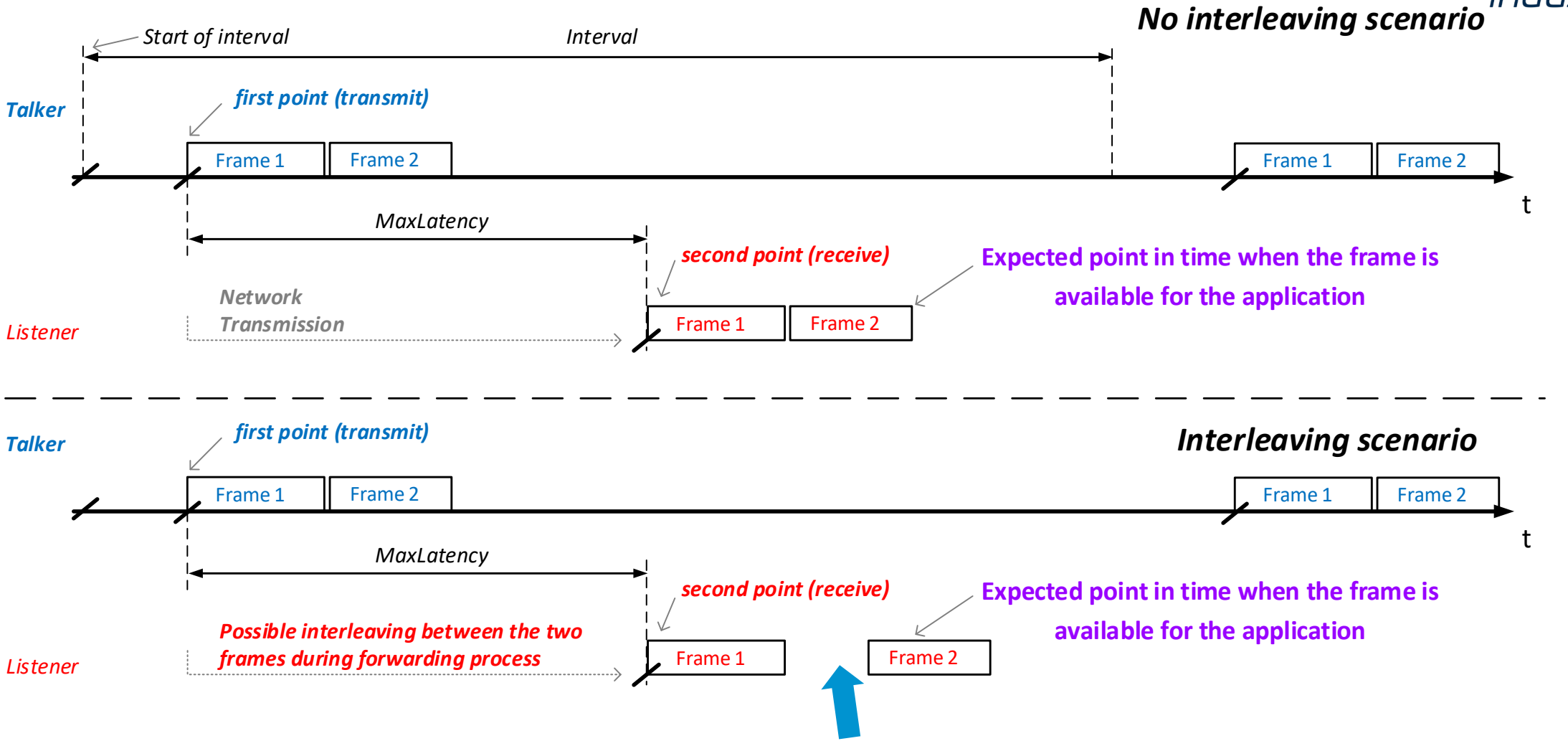
Issue 1: “MaxLatency” definition (5)



- In case of streams with multiple frames (`max-frames-per-interval >1`), the listener cannot be aware of possible interleaving frames (latency definition is for the single frame)
- Consequently, it is not possible for the listener to derive the point in time when the stream‘s packet is fully received, i.e. the point in time when the internal processing starts



Issue 1: “MaxLatency” definition (6)



- In case of streams with multiple frames the listener cannot be aware of possible interleaving
- Consequently it is not possible for the listener to derive the point in time when the stream‘s packet is fully received, i.e. the point in time when the internal processing starts

IEEE 802.1Qdj Contribution - 1

Proposal: change Sect. 46.2.3.6.2 to:

"MaxLatency shall use the definition of 3.118, with additional context as follows: The "known reference point in the frame" is the message timestamp point specified in IEEE Std 802.1AS for various media (i.e., start of the frame) for the "first point",

and the end of the last symbol of the FCS for the "second point".

The "first point" is in the Talker at the reference plane marking the boundary between the network media and PHY (see IEEE Std 802.1AS). The "second point" is in the Listener at the reference plane marking the boundary between the network media and PHY."

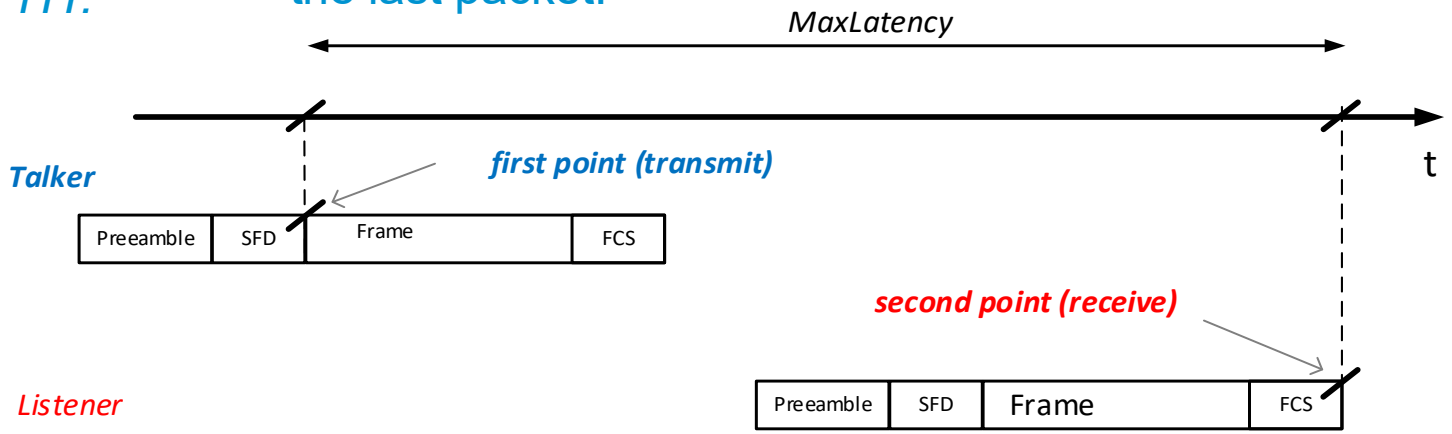
Effect:

The change represents the exact time during which the frame (or parts of it) under the responsibility of the CNC, i.e.,

- from the time the first symbol leaves the talker,
- to the time the last symbol entered the listener,

and at which time the listener can start its local processing.

This also applies to fragmented packets (which are possibly interleaved by other frames), since MaxLatency for talkers and listeners are calculated on the last packet.



Issue 2: “MaxLatency” definition for time aware talkers (1)

§3.118 Latency – is currently defined as:

> The delay experienced by a frame in the course of its propagation between two points in a network, measured from the time that a known reference point in the frame passes the first point to the time that the reference point in the frame passes the second point.

§46.2.3.6.2 MaxLatency – is defined as:

> Latency shall use the definition of 3.118, with additional context as follows: The ’known reference point in the frame’ is the message timestamp point specified in IEEE Std 802.1AS for various media (i.e. start of the frame)...

46.2.3.6.2 MaxLatency – redefinition in case of Tspec time-aware is present

> When TSpecTimeAware is present:
>
> The ’first point’ is assumed to occur at the start of the Interval, as if the Talker’s offsets (EarliestTransmitOffset and LatestTransmitOffset of 46.2.3.5) are both zero.

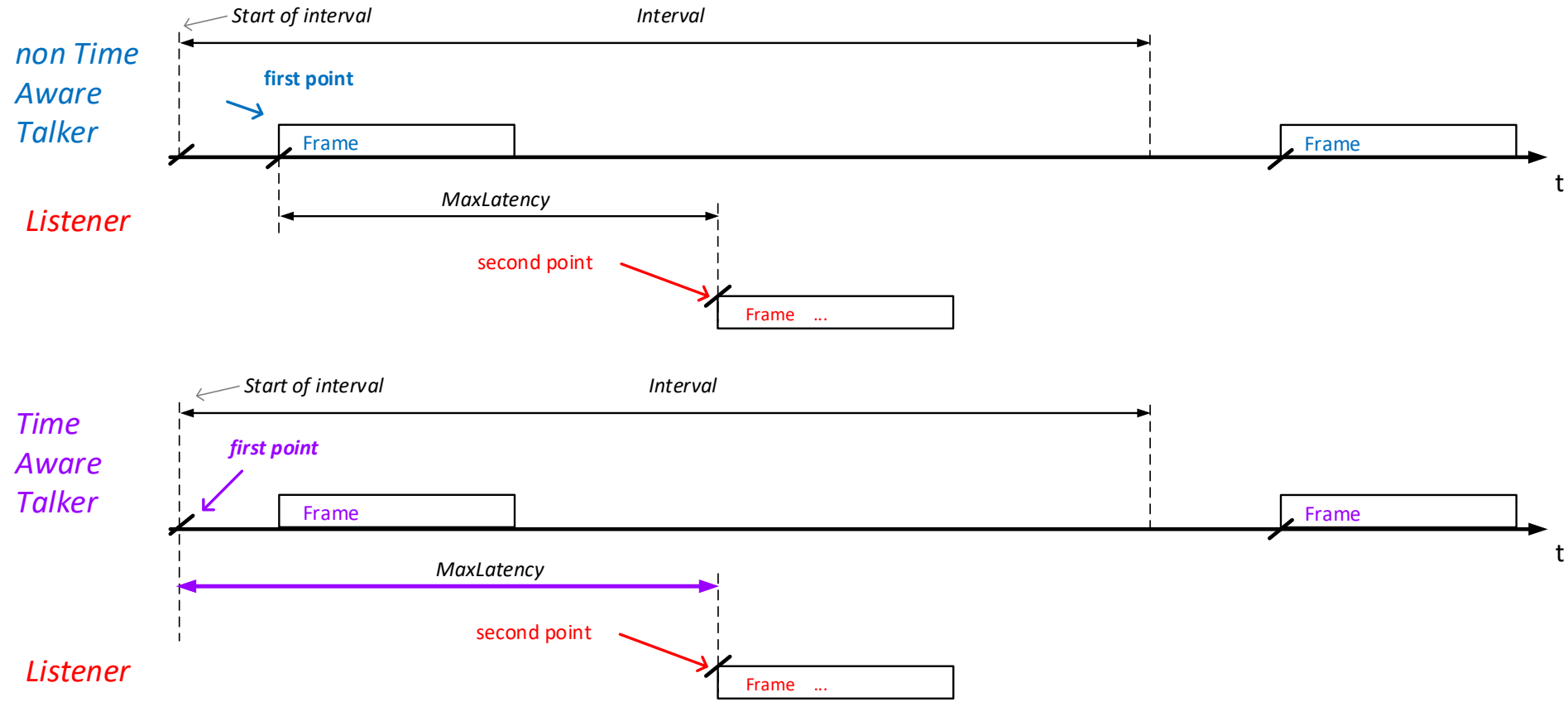
What does it mean?

Issue 2: “MaxLatency” definition for time aware talkers (2)

§3.118 Latency: The delay experienced by a frame in the course of its propagation between two points in a network, measured from the time that a known reference point in the frame passes the first point to the time that the reference point in the frame passes the second point.

46.2.3.6.2 When TSpecTimeAware is present: The ’first point’ is assumed to occur at the start of the Interval, as if the Talker’s offsets (EarliestTransmitOffset and LatestTransmitOffset of 46.2.3.5) are both zero.

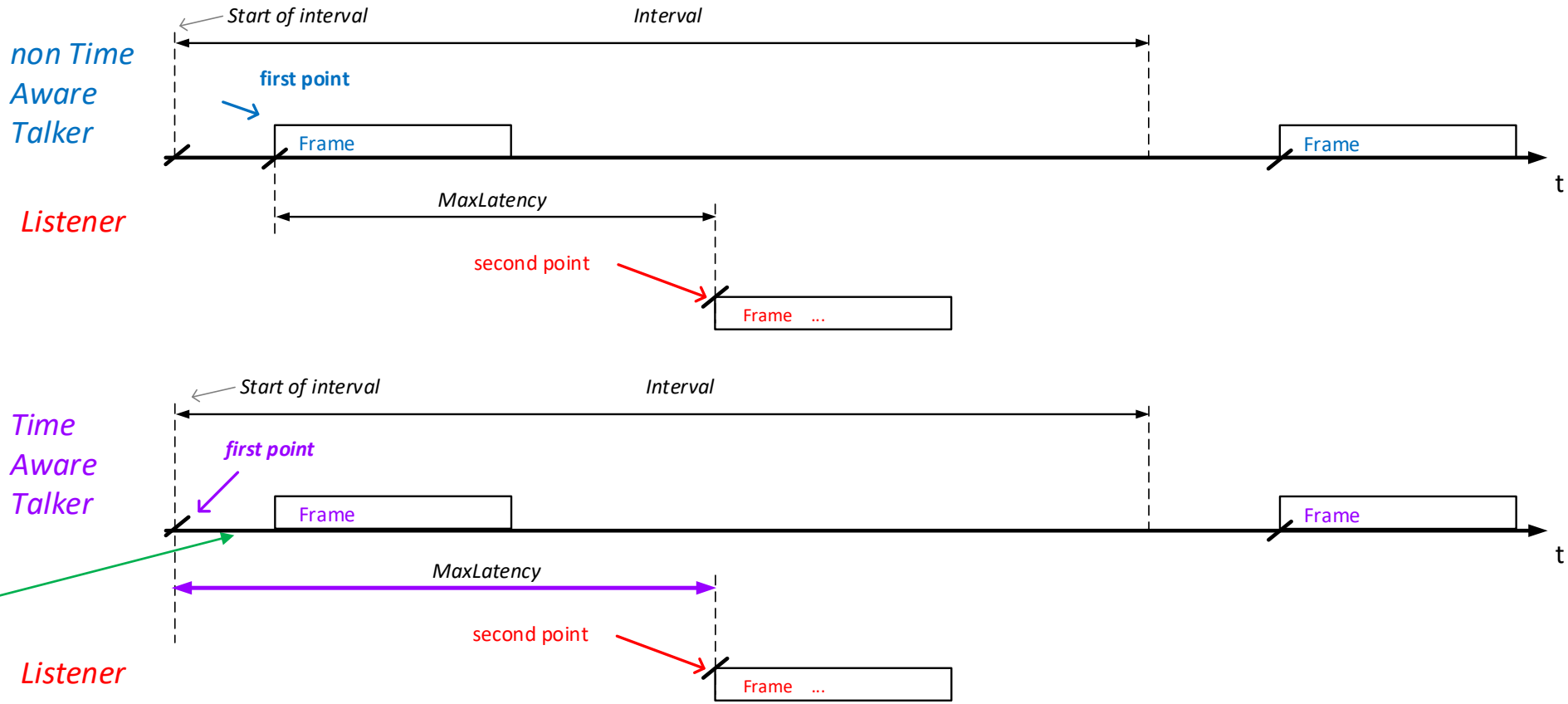
MaxLatency definition becomes a “deadline” semantics for a time aware talker



Issue 2: “MaxLatency” definition for time aware talkers (3)

§3.118 Latency: The delay experienced by a **frame in the course of its propagation** between two points in a network, measured from the time that a known reference point in the frame passes the first point to the time that the reference point in the frame passes the second point.

46.2.3.6.2 When TSpecTimeAware is present: The ’first point’ is assumed to occur at the start of the Interval, as if the Talker’s offsets (EarliestTransmitOffset and LatestTransmitOffset of 46.2.3.5) are both zero.



There are time intervals where the frame is not in propagation at all

Issue 2: “MaxLatency” definition for time aware talkers (4)

- *MaxLatency definition* is “transformed” to a “*Deadline” definition* for “time-aware” streams
  - *Latency requirements/constraints* and “*Deadline” requirements/constraints* are different values

Proposal 2.a: Introduce an explicit parameter for a “deadline” requirement for listeners

Proposal 2.b: In order to avoid issues with the preempted frames and multiple presented in this contribution use the redefined second point in (in this contributing) as time for the “deadline” as the proposed “second point” definition in this contributions to avoid the possible issues with the preemption and interleaving frames.

Issue 2: “Deadline” definition for time aware talkers (5)

§46.2.3.6.xy Deadline – is defined as:

The point in time within ‘Interval’ where the end of the last symbol of the FCS passes the reference plane marking the boundary between the network media and PHY.

Or

§46.2.3.6.xy Deadline – is defined as:

The delay between two points in time, measured from the time that a known reference point in the frame passes the first point to the time that the reference point in the frame passes the second point.

The ’first point’ is the start of the ‘Interval’.

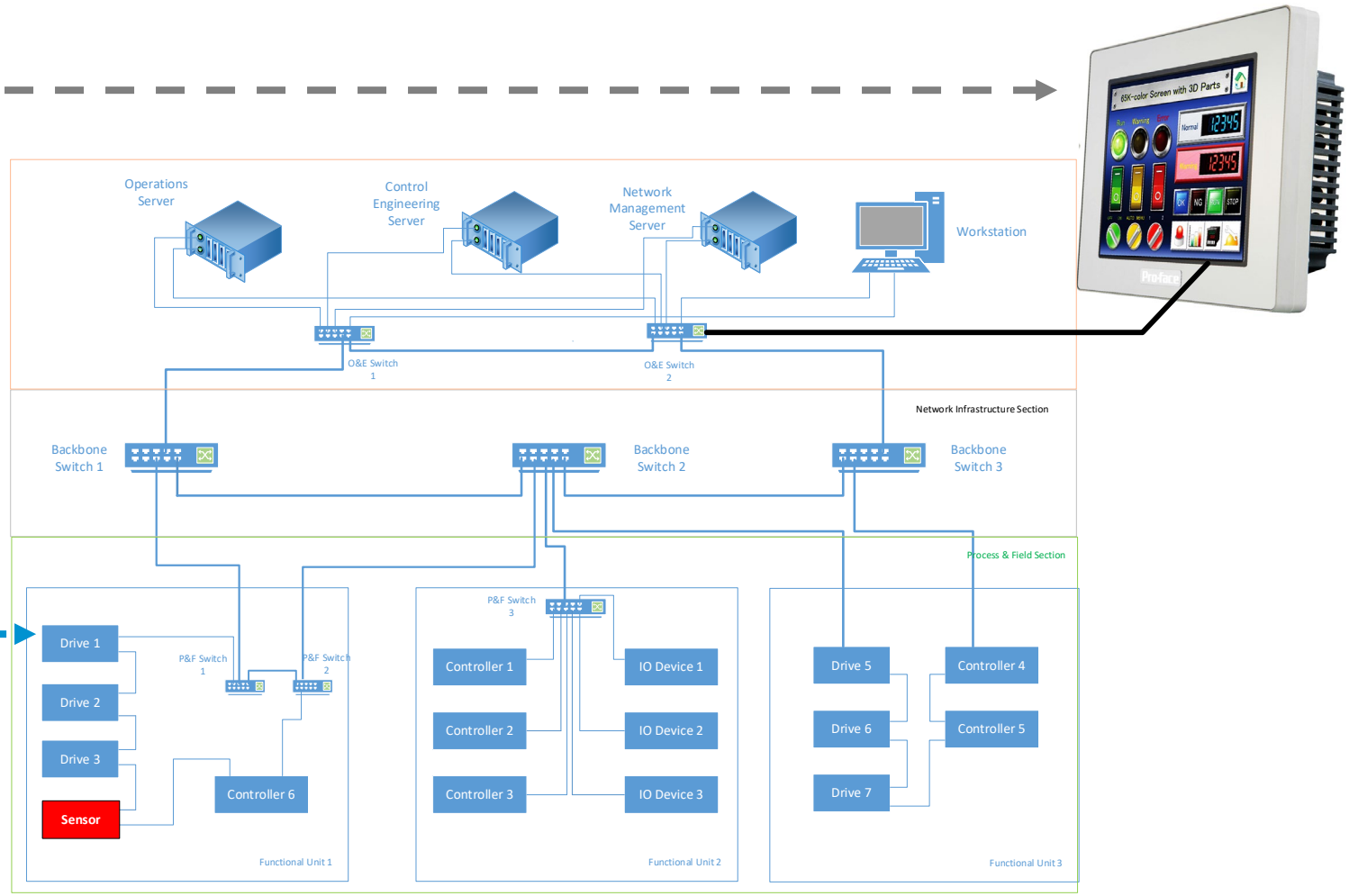
The ’second point’ is the point in time where the end of the last symbol of the FCS passes the reference plane marking the boundary between the network media and PHY.

Optional

In case that Deadline value is larger than the Interval value, that means that the frame is being issued by the talker in the previous interval, and the frame is being expected at the: **Phase offset = Deadline % Interval**

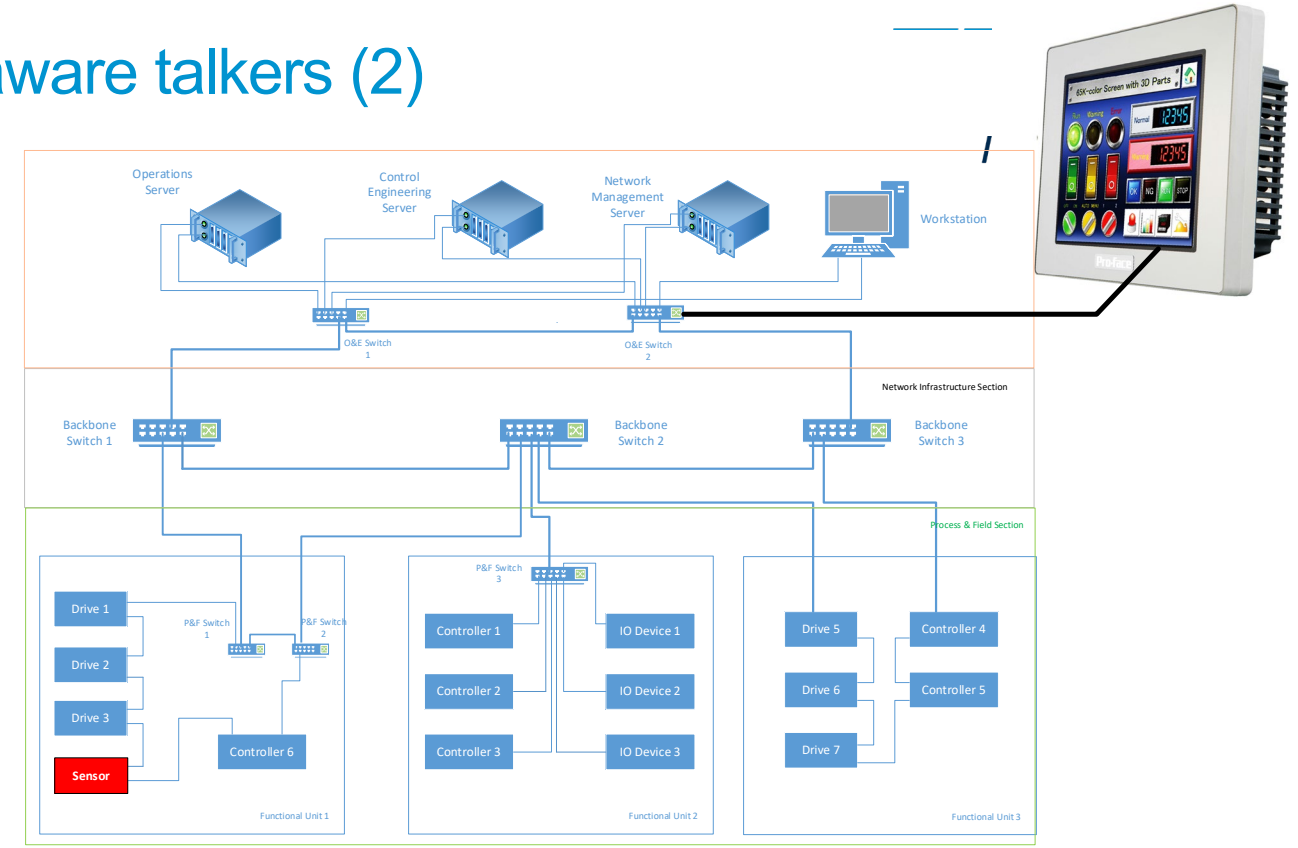
Issue 3: “MaxLatency” definition for time aware talkers (1)

- MaxLatency is the “relation” between a talker and one listener
- In a single Network there may be different MaxLatency values (one for each listener)
  - some listeners may have “tight” latency for the same stream
    - e.g. Drive 1 receiving data from the red Sensor
  - other listeners may have very relaxed ones or not at all
    - e.g., a monitoring HMI device



Issue 3: “MaxLatency” definition for time aware talkers (2)

- In the CNC-CUC UNI there is a MaxLatency for each listener
- MaxLatency can be specified at the talker (35.2.2.10.7). In that case the MaxLatency applies to all listeners
- The only use case in industrial where the MaxLatency makes sense is if it represents the validity of the data send with this stream
  - however, this is an application-level semantics and shall not be represented in the
- In some use cases setting the latency requirements at the talker may cause conflicts with the “fare away” listeners, specially in the case of converged networks
  - If the MaxLatency at the talker is set higher - that would invalidate the MaxLatency requirement for the listeners
  - If the MaxLatency at the talker is set lower - that would lead to some listeners not being able to receive the stream



Proposal 3: Remove the latency requirement at the talker

Summary of Contributions

- Redefine the second reference point for “MaxLatency”
  - Make the second point to be the last symbol of the frame check sequence (FCS)
- Add the “Deadline” parameter at the listeners (CNC-CUC UNI)
- Remove the MaxLatency parameter at the talker