Considerations and questions regarding latency

A Leading Provider of Smart, Connected and Secure Embedded Solutions

Steve Gorshe

09/18/2024

Supporters

- Kirsten Mateus (BMW)
- Kamal Dalmia (Aviva Links)
- Scott Muma (Microchip)
- Claude Gauthier (NXP)

Outline

- High-level application considerations
- Consideration of latency components
- Network example
- Summary and recommendations

Application Considerations

- **Latency requirements are critical next step for the project**
  - Statements have been made during email list discussions along the lines of “Why would it not be in the interest of the committee to push the latency to the lowest achievable number?”
  - While well intentioned, such statements are not helpful to the project.
  - There are always tradeoffs, and the relationship between them is often non-linear. The last few % of optimization for one parameter can have a non-linearly higher impact on another one.
    - For example, if reaching the lowest achievable latency requires increasing circuit complexity and power at the camera module, the associated heat could degrade the image quality far more significantly than latency in an I2C parameter update.

Application Considerations (continued)

- **Latency requirements considerations (continued)**
  - The key engineering parameter of interest is the largest acceptable latency.
    - Further latency reductions beyond this provide little practical value, especially if they come at the expense of other parameters (e.g., power and complexity)
  - It appears that “lower latency” requirements are being used as a proxy for indirect side effects, e.g., for:
    - System bandwidth,
    - Ensuring that a maximum latency requirement is not exceeded,
    - Limiting skew between events
  - It is important to understand our degrees of freedom in the solution space in order to achieve a proper balance between performance and complexity.

Application Considerations (continued)

- **Bounded vs. deterministic latency**
  - The PHY by itself is not capable of ensuring deterministic latency if it is required by some applications.
    - A combination of PTP and other protocols such as IEEE 1722, etc. would be required
    - The job of the PHY is to provide bounded latency that could enable determinism

Application Considerations (continued)

- **Decision and reaction latency considerations**
  - The author assumes that a CAN-type network (or its replacement by, e.g., 10BASET1S) is still needed to enact the SoC/ECU decisions effecting things like breaking and engine control without the camera in the loop.
    - Alternatively, if there is a shared Ethernet network, decision commands have highest priority so that they are unaffected by camera-related packets
  - Consequently, the latency relevant to the P802.3dm project is essentially the decision-making time
    - I.e., the time from when an event is observed by the camera, communicated to the SoC, and the image is processed by the SoC so that it can determine a required action.

Application Considerations (continued)

- **Camera control/feedback loop latency**
  - The camera control/feedback loop is important, but should be considered separately, as explained by Kirsten in email thread comments
    - Camera adjustments are made to optimize the camera image in response to changing light conditions.
    - To impact the decision response time, the input lighting change to the sensor caused by an event would need to be so significant and rapid that the imager would deliver an unusable image to the SoC before it could be adequately reconfigured
    - The responses to Kirsten's comments gave examples of events occurring around the car with no tangible evidence that such scenarios could result in an unusable image

Application Considerations (continued)

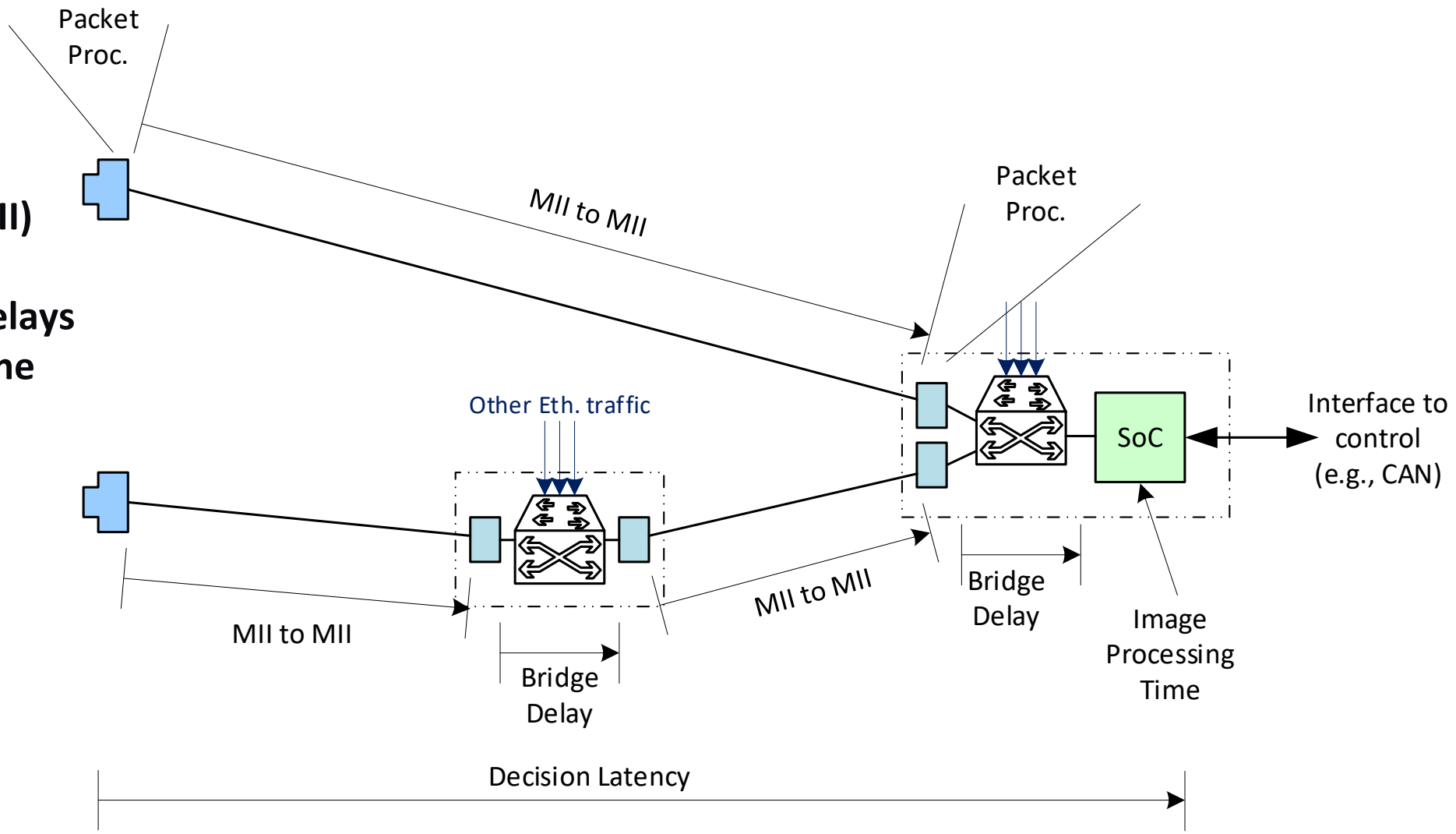
- **Additional latency considerations**
  - Do the latency requirements allow for the presence of an Ethernet switch in the network?
    - From the start of the Study Group, we have assumed that a zonal architecture is required, which implies an Ethernet switch at the zone controller. If the added latency of an Ethernet switch is unacceptable, then the Task Force needs to rethink some fundamental assumptions.

Network Example

Decision Delay:

- **Link Delay (MII to MII)**
- **Bridge Delay**
- **Packet processing delays**
- **Image processing time**

PHY Delays are the portion within the domain of IEEE P802.3dm



Summary

- **There are many open issues regarding latency with little consensus so far regarding the what and how of the requirements**
- **It is recommended that decision latency and camera control/feedback loop latency be considered separately**
- **Since we lack tangible evidence that camera/control latency is a critical factor, the author recommends that the Task Force focus on a bounded latency approach for the decision latency**
  - I.e., not a deterministic nor a lowest achievable latency approach
  - The upper bound must be determined by tangible application requirements.
  - If we have sufficient bandwidth for camera reconfiguration/I2C transactions we don't care about the latency of individual transactions, as long as the entire reconfiguration/transaction can be accomplished within the time limits of the camera control loop
- **The Task Force should avoid over constraint or suboptimal tradeoffs**
- **One important consideration is that the latency bound must allow for Ethernet switches and a range of latencies**

Thank You