Proposal on Fault Tolerant Time Propagation

Richard Tse, Microchip Technology
Rob Donnelly, NASA Jet Propulsion Laboratory

IEEE P802.1DP / SAE AS6675
Nov 2022

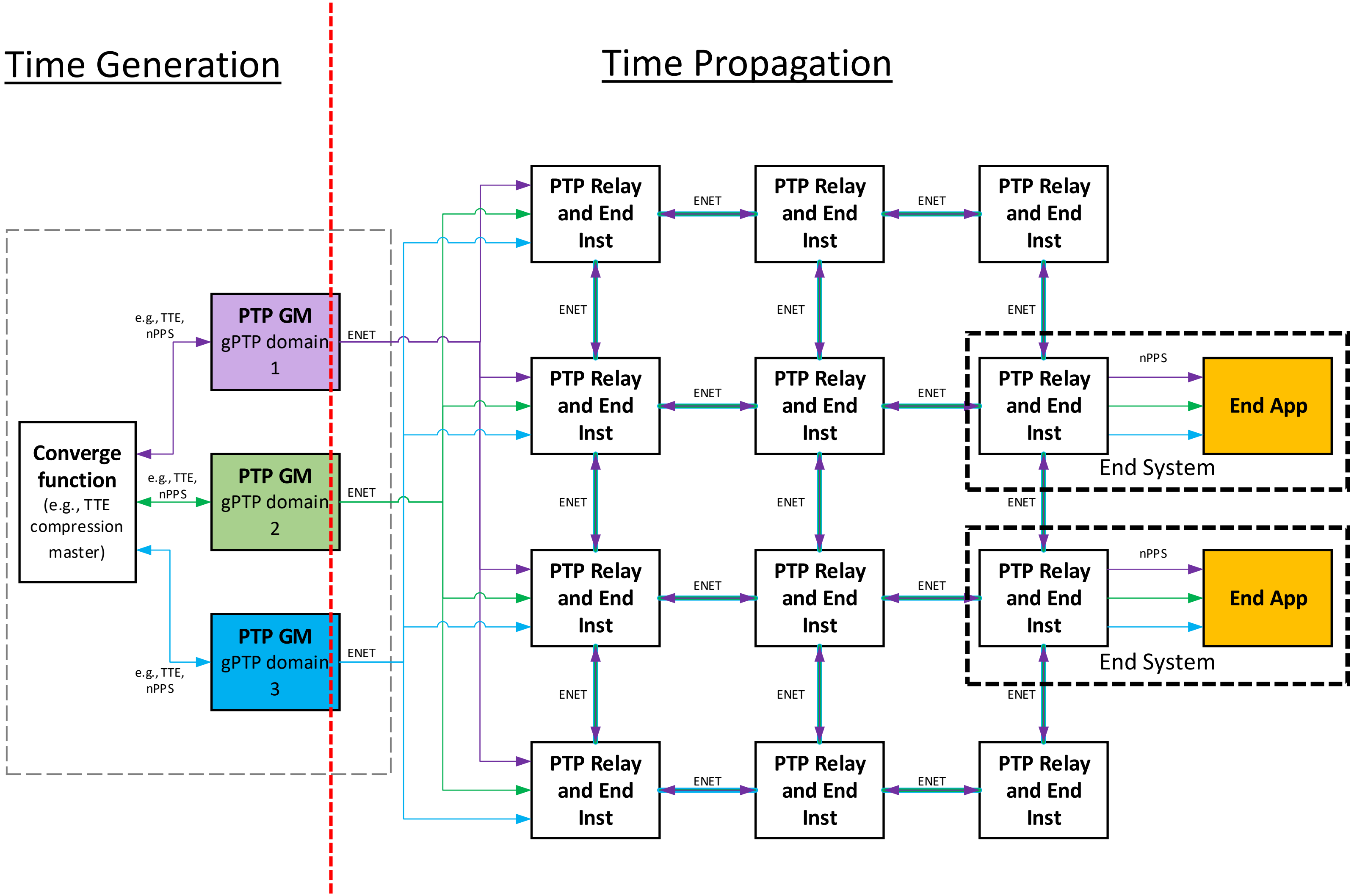
Outline

- Agreement Generation and Agreement Propagation
- Time Generation fault tolerance concepts
- Background use cases and needs
- Time Propagation
  - Assumptions, properties, & guarantees
  - Fault Tolerance Concepts
  - TMR and DMR fault Detection at End Station
  - ToD Selection at End Station
  - TSN Use of “Best” Time Domain
  - Guarantees
- Conclusions
- Examples for Fault Tolerant Time Generation and Propagation System
- Appendix

Agreement Generation & Agreement Propagation

- Agreement Generation (for time)
  - *How multiple GrandMasters come to agreement on the time*
  - This involves multiple clocks continuously “collaborating” with one another to agree on a common time
    - This is not a leading and following mechanism
  - A protocol like Time Triggered Ethernet (TTE) performs this type of Agreement Generation
- Agreement Propagation (for time)
  - *How time is propagated from GrandMasters to end systems*
  - Clocks do not collaborate to agree on a common time
  - Per IEEE 1588 and IEEE 802.1AS, clocks collaborate to decide which single clock will lead all the others (e.g., BMCA)
    - If external port configuration mode is used, then this decision is made by whomever did the configuration
  - Once the TSN Grandmaster (GM) clock has been decided, the IEEE 1588 and IEEE 802.1AS protocols perform an Agreement Propagation operation
- The terms Time Generation and Time Propagation will be used in this presentation

Proposal Overview



Time Generation: Fault Tolerance Concepts

- Time Generation requires > 3 participants to overcome a failure in any one participant unless, from [3]:

  "Theorem 3: If there is a bound on the rate at which messages can be generated or if there is a protocol for signing unforgeable signatures that can be authenticated, then the Clock Synchronization Condition can be achieved as long as the faults do not disconnect the network."

- A Time Generation mechanism with a bound on the message rate could be used to synchronize the Grandmaster (GM) clocks of 3 PTP time domains for TSN
  - Since all 3 time domains are aligned, an end station's use of any non-faulty time domain cannot be distinguished from its use of another non-faulty time domain
- This presentation will not delve any more into the Time Generation topic
  - Outside the scope of IEEE 802.1AS
  - Another topic for study by IEEE 802.1DP and SAE AS6675?
    - State theoretical requirements but don't specify implementation?

Background: Use Cases and Needs

- From TSN Time Synchronization NASA's Use Cases and Needs [1]:
  - "[Need to tolerate **arbitrary failures** of **end systems**]
    Note: Includes timing failures. Drives the need for **fault-tolerant averaging** rather than a single trusted high-priority master"
  - "[Need to] tolerate multiple device failures – at least one simultaneous worst-case failure of an **end system and switch**"
  - "Ideally the network should tolerate the failure of any network component, … without any non-faulty devices transitioning out of a stable synchronized state"

Time Propagation: Assumptions, Properties, & Guarantees

- Assumptions:
  - GMs are in agreement, within some tolerance
  - When there are no faults, all PTP End Instances (all time domains) will be aligned within some tolerance
- Properties:
  - PTP End Instances operate independently
  - PTP End Instances don’t need to agree on a faulty time domain
  - PTP End Instances don’t need to agree on “the best” time domain
- Guarantees:
  - What can be guaranteed by a solution? We’ll come back to this.

Time Propagation: Fault Tolerance Concepts

- Considerations (for TSN):
  - Once the GM has been determined and external port configuration mode is used to set the timing propagation paths, no decisions need to be made by the clocks.
  - The GM simply sends its time and the PTP End Instances simply follow it.
- Triple Mode Redundancy (TMR) can be used to detect and identify one faulty time domain at an end station
  - Use 3 PTP time domains to achieve TMR
  - The 3 GM clocks are synchronized, per Time Generation Theorem 3 from [3]
- Double Mode Redundancy (DMR) can be used to detect (but not identify) a faulty time domain at an end station
  - Use 2 PTP time domains to achieve DMR
  - Are 3 participants still needed for Time Generation of the GM clocks?

TMR-Based Fault Detection at End Station

- Find median Time of Day (ToD) from the 3 time domains
- The maximum expected difference between the ToDs of 2 time domains, TDX and TDY, can be determined as follows:
  - $\text{maxdiff}_{TDXvTDY} = \text{accumTE}_{TDX} + \text{accumTE}_{TDY} + \text{otherTE}$
    - accumTE_{TDn} = maximum absolute expected time error accumulated in the PTP communication path from GM to PTP End Instance for time domain *n*
    - otherTE = non-PTP sources of time error (e.g., offsets between GM clocks of TDX and TDY)
- Time domain X (TDX) is deemed to be faulty if its ToD differs from the ToD of the median time domain (TDMED) by more than the following threshold:
  - $\text{threshold}_{TDXvTDY} = \text{maxdiff}_{TDXvTDY} + \text{margin} + \text{hysteresis}$
    - TDX = the time domain being tested, TDX ≠ TDMED
    - TDY = TDMED
- Time Propagation failure is declared if more than one time domain is deemed to be faulty
- Do not use hot-standby
  - Changing 2 working + 1 faulty time domain into 2 working time domains does not improve fault tolerance

DMR-Based Fault Detection at End Station

- Time Propagation failure is declared if the ToDs from the 2 time domains differ by more than a threshold
  - $threshold_{TD1vTD2} = maxdiff_{TD1vTD2} + margin + hysteresis$
- Do not use hot-standby
  - Hot-standby and DMR serve different purposes and do not work together
    - DMR’s goal is simply to detect a failure by comparing two time domains
    - Hot-standby tries to fix a faulty time domain, which eliminates DMR and prevents nodes downstream of the hot-standby from detecting a failure

ToD Selection at End Station

- Options to select from any set of non-faulty time domains
  - Select time domain per precedence level set by management layer
  - Select time domain with minimum expected accumTE
    - accumTE could be set by management layer
    - accumTE could be signaled in Announce messages using ENHANCED_ACCURACY_METRICs TLV (defined in p1588a draft amendment)
  - Select time domain whose GM is closest
    - smallest stepsRemoved value in Announce messages
- Combining results from multiple time domains
  - Might be more complex to analyze and to implement
  - Might not produce a better result than simple selection

TSN Use of “Best” Time Domain

- Can other TSN protocols make use of the selected “best” time domain?
  - IEEE 802.1Qbv Enhancements for Scheduled Traffic: Time-Aware Shaper (TAS)
    - Protect against disruption of time sensitive traffic flow due to a failed time domain
    - Since all 3 time domains are synchronized, switching from one time domain to another time domain should not cause any disruption and might not even be noticeable
  - Any others?

Guarantees

- With 3 time domains:
  - If up to one time domain is faulty at any/all end station(s), all end stations are guaranteed to remain aligned within the allowed tolerance (fail operational)
- With 2 time domains:
  - If up to one time domain is faulty at any end station, other end stations without faults are guaranteed to remain aligned within their allowed tolerance
  - Any single fault that causes a time domain to stray beyond its allowed tolerance at an end station is detected (fail stop)

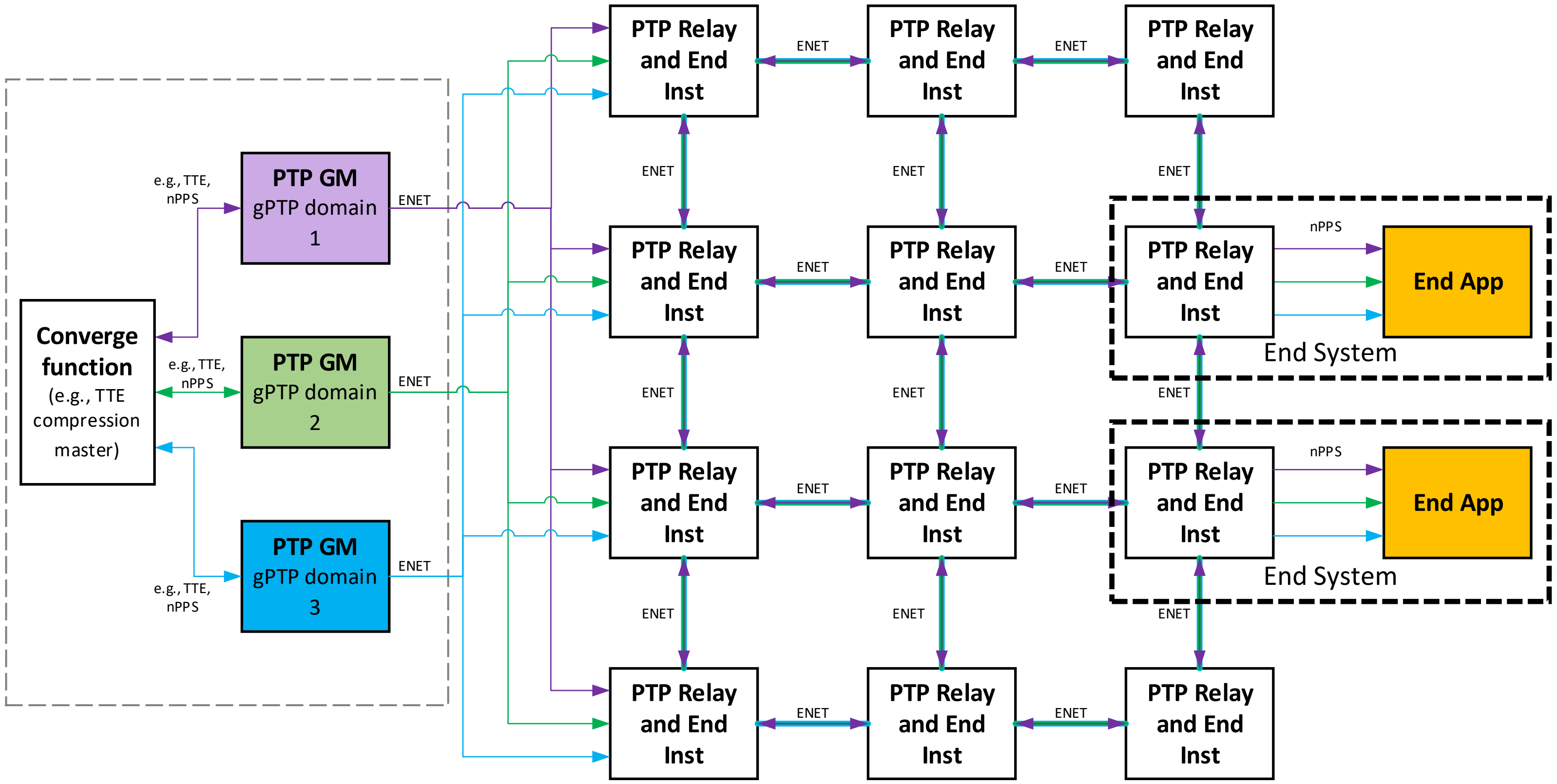
Conclusions

- Time Generation for GMs needs to be dealt with independently from Time Propagation
- TMR and DMR methods using multiple PTP time domains enhance fault tolerance for Time Propagation, and guarantee some behaviors (given certain assumptions and properties are true)
- Time domain selection is simple to implement and analyze
- Use of “best” time domain can improve fault tolerance of time sensitive traffic

Examples for Fault Tolerant Time Generation and Propagation System

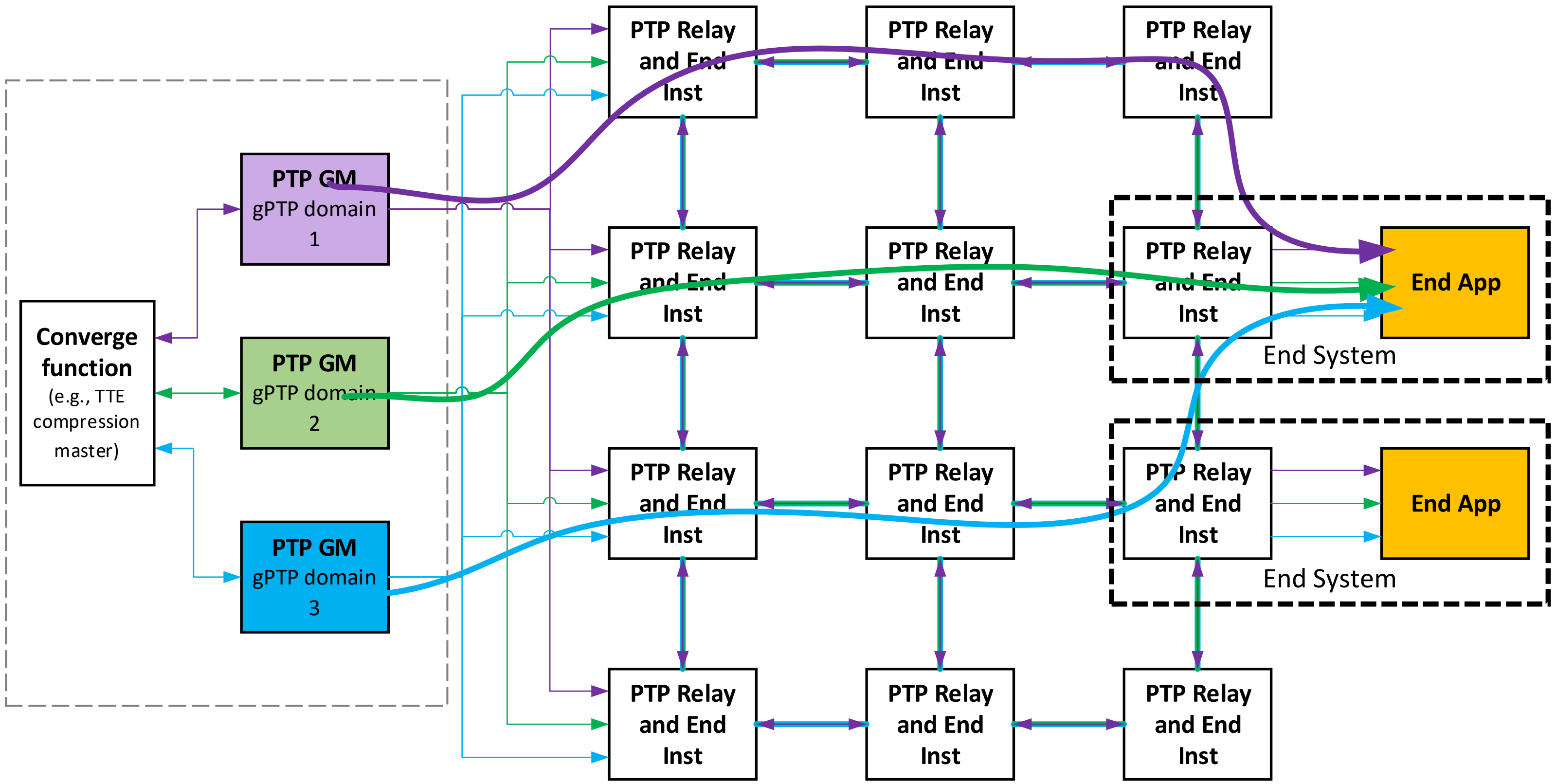
Example Fault-Tolerant Time Gen and Propagation System

- GMs of all three gPTP domains are synchronized with each other
- IEEE Std 802.1AS propagates time of three gPTP domains to all PTP End Instances via PTP Relay Instances
- End Applications use the "best" time recovered from the PTP End Instances
- Other TSN protocols (e.g., TAS) can also use the "best" time



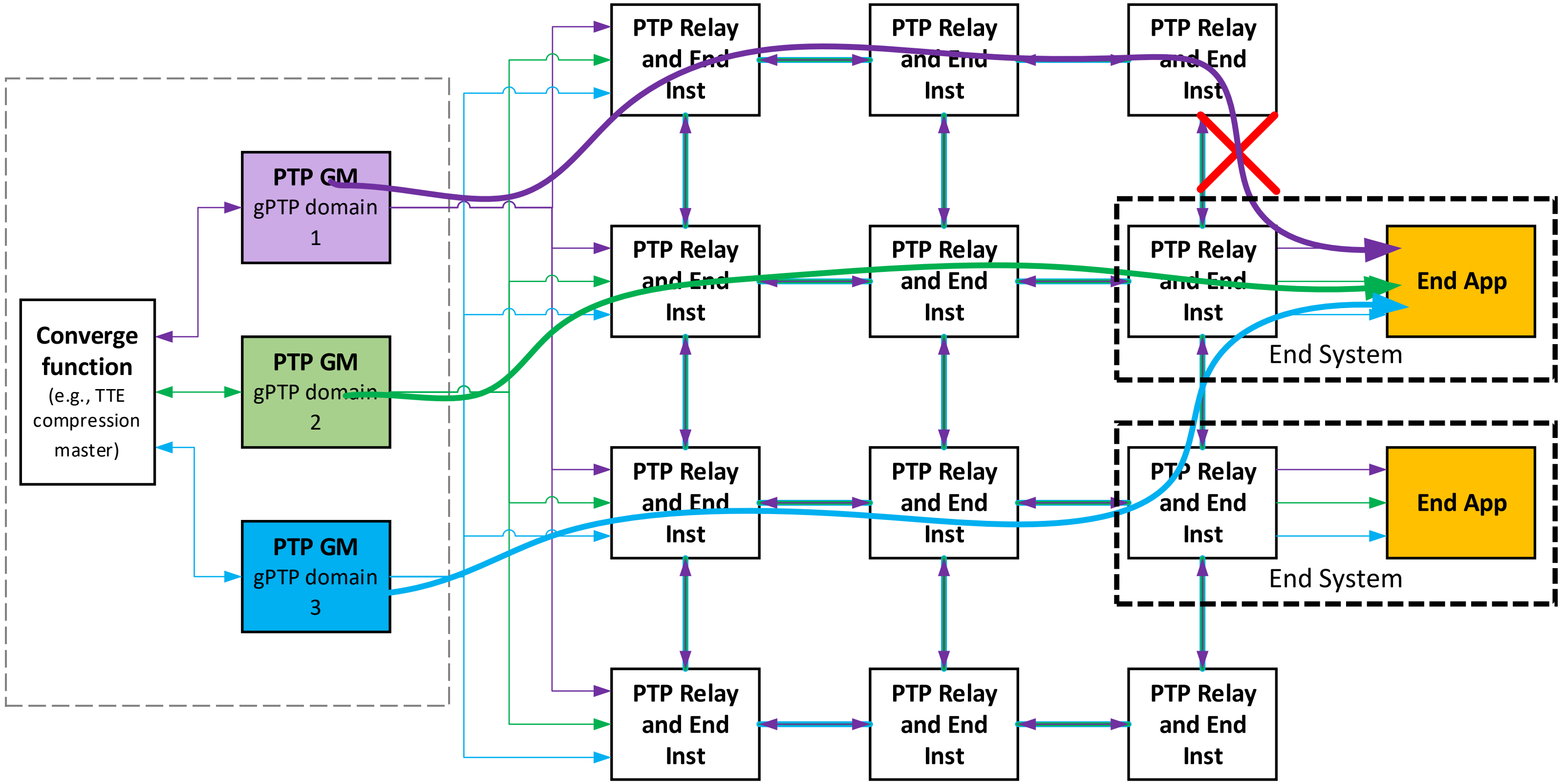
Example Time Propagation without Failure

- End App checks alignment of the 3 gPTP ToDs to determine if any are faulty
- Can use any non-faulty gPTP domain to get its local "best" gPTP ToD
- Can select "best" (non-faulty) gPTP domain based on various criteria



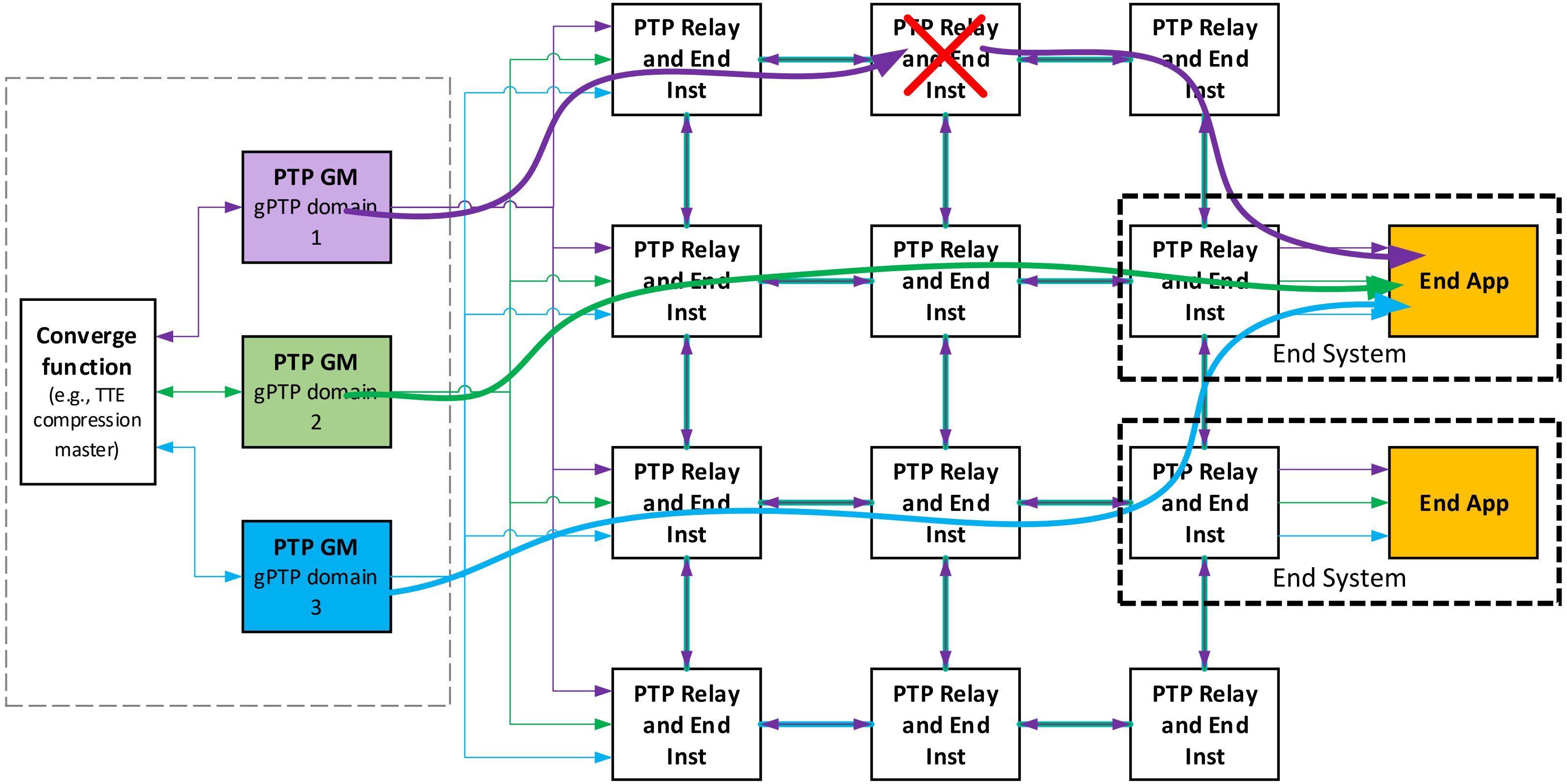
Example Fault-Tolerant Time Propagation with Failure (1/2)

- gPTP domain 1 fails (momentarily) at End Application and corresponding PTP Relay and End Instance
- Failure of gPTP domain 1 can be detected by:
  - Misalignment with gPTP domains 2 and 3
  - Lack of corresponding gPTP messages
- End Instance and End Application can continue use of either/both gPTP domains 2 and 3
- gPTP domain 1 is excluded from selection of "best" gPTP domain



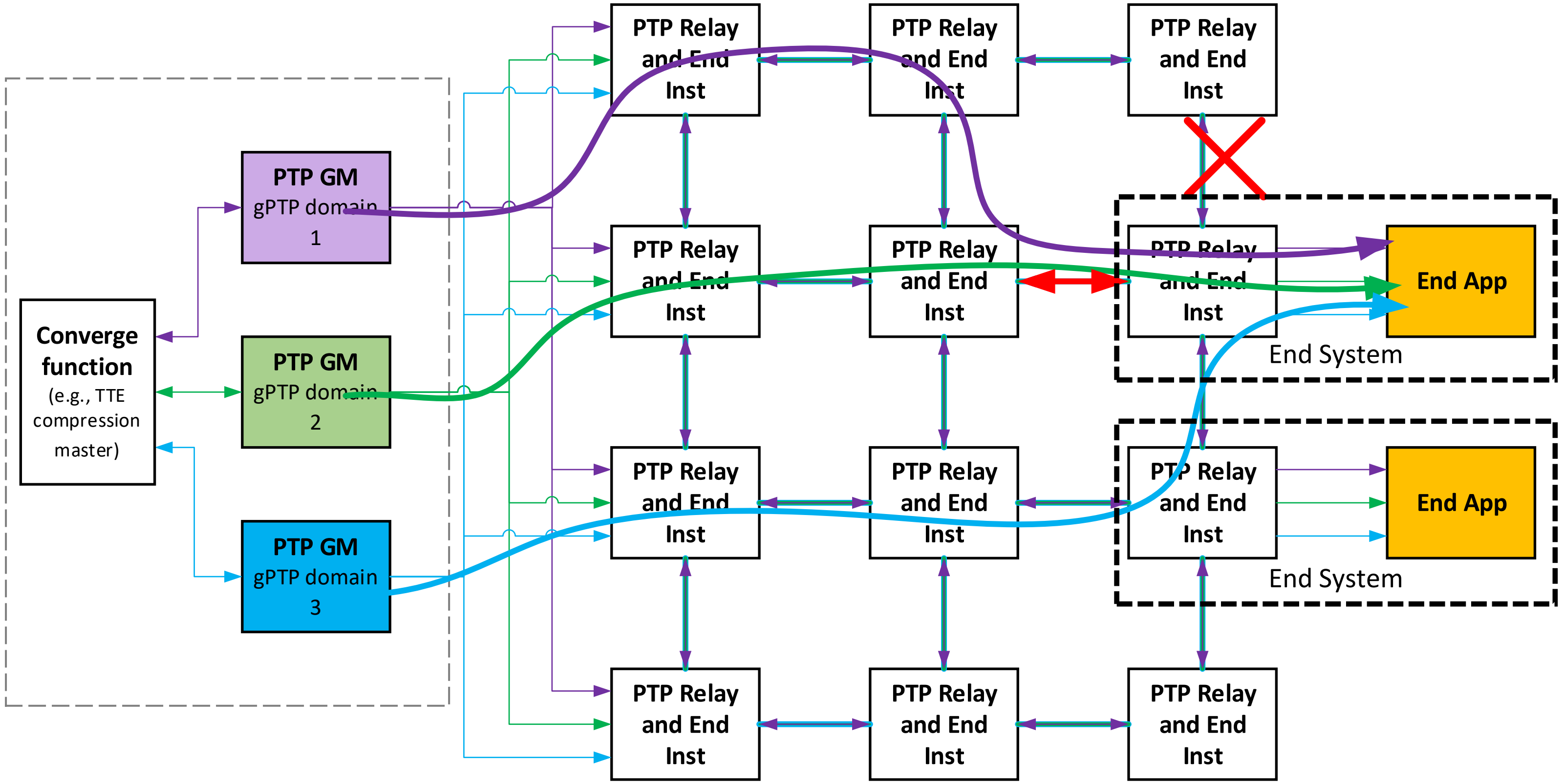
Example Fault-Tolerant Time Propagation with Failure (2/2)

- PTP Relay Instance for gPTP domain 1 fails and corrupts the time that it relays
- Failure of gPTP domain 1 can be detected by:
  - Misalignment with gPTP domains 2 and 3
- gPTP domain 1 is still followed by PTP End Instances but is excluded from selection of "best" gPTP domain by the End Application



Example Time Propagation Failure Restoration

- gPTP domain 1 is rerouted so it gets to End Application and corresponding PTP End Instance, bypassing the faulty link
- This new route shares a PTP Relay Instance and link with gPTP domain 2
- A second failure, on the highlighted link, can break both gPTP domains 1 and 2 and cause a sync failure at the End Application
  - This failure mode might be acceptable



Appendix

References

[1] IEEE P802.1DP/SAE AS6675 Ad HoC, TSN Time Synchronization – NASA's Use Cases and Needs, R Donnelly, May 2022

The synchronization of clocks in the presence of faults has been studied extensively in the following:

[2] Synchronizing Clocks in the presence of Faults, L Lamport, PM Melliar-Smith, 1982

[3] On the Possibility and Impossibility of Achieving Clock Synchronization, D Dolev, J Halpern, 1984

[4] A Unified Fault-Tolerance Protocol, P Miner, A Geser, L Pike, Jeffery Maddalon, 2004

[5] Fault-Tolerant Clock Synchronization in Distributed Systems, P Ramanathan, KG Shin, RW Butler, 1990

[6] Reaching Agreement in the Presence of Faults, M Pease, R Shostak, L Lamport, 1980

Example Time Propagation with Enhanced Failure Restoration

- A network with more extensive fault tolerance is shown
- gPTP domain 1 is rerouted so it gets to End Application and corresponding PTP End Instance, bypassing the faulty link without sharing any links with the gPTP domains 2 or 3
- TMR can detect realignment of gPTP domain 1 with gPTP domains 2 and 3
- End Instance and End Application can use either/any gPTP domains 1, 2, and 3 again
- gPTP domain 1 is now included back into the selection process of "best" gPTP domain