DETERMINISTIC6G

DetNet

Latency Analysis of Mobile Transmission

draft-varga-detnet-mobile-latency-analysis

Balázs Varga, Joachim Sachs, Frank Dürr, Samie Mostafavi

DetNet WG

18th March, 2024, IETF 119

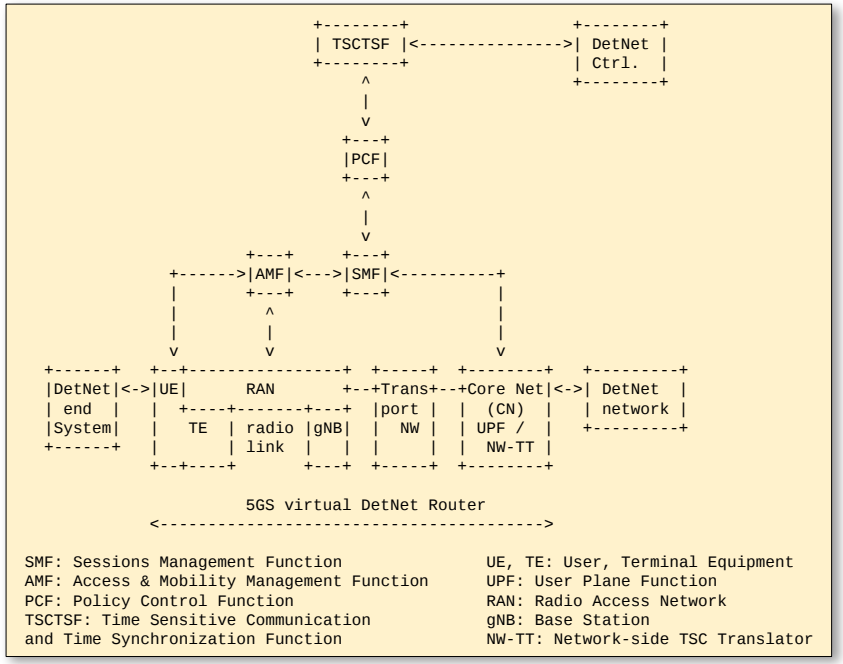
DetNet Latency Bound
Analysis of Mobile Systems Latency

- Intended status:
  - Informational
- Actual version:
  - draft-varga-detnet-mobile-latency-analysis-00

- Abstract:
  - Dependable time-critical communication over a mobile network has its own challenges. This document focuses on a comprehensive analysis of mobile systems latency in order to incorporate its specifics in developments of latency specific network functions. The analysis provides valuable insights for the development of wireless-friendly methods ensuring bounded latency as well as future approaches using data-driven latency characterization.

Introduction
Mobile Communication System

- 3GPP Releases:
  - Rel-16: Ultra-Reliable Low Latency Communications (URLLC)
    - Target: reliability, latency and QoS
      (e.g., automatic repetitions, antenna techniques, robust physical channels, Orthogonal Frequency Division Multiplex (OFDM) numerology,
      mini-slots, grant-free access, pre-emption, 5G QoS identifier (5QI) values for multiple time-critical services, QoS monitoring).
  - Rel-18:
    - DetNet support: based on the concept developed for Time-sensitive Networking (TSN) in former releases.
    - 5G system is represented in the end-to-end architecture as a set of virtual DetNet routers.
    - Note: In general bridging/routing service is out-of-scope for 3GPP specifications



Comparison
Wired vs. Wireless DetNet nodes

- 5G network
  - can form multiple virtual routers, each of which is realized via the UPF instance in the 5G core network.
  - significant differences in the characteristics of a wired DetNet router and such a virtual DetNet node:
    - Physical distance of ports: decimeter vs. 100's of meters or even kms (+ reduced hop counts)
    - Number of ports: fixed vs. dynamic
    - Latency characteristics: microsecond vs. milliseconds
    - Dynamicity of characteristics: designed in advance vs. determined during operation phase (radio environment changes)
- Intent:
  - able to efficiently ensure bounded latency in mixed (wired and wireless) DetNet scenarios -- > wireless-friendly functions to ensure latency bound

Mobile Transmission Analysis
Latency Breakdown

- In the 5G RAN the main latency contributors are:
  1. Time-domain reliability based on HARQ (Hybrid Automatic Repeat Request)
     - Latency, Robustness, Spectral efficiency
  2. Mobility with handover interruptions
     - Mobility is ensured by handover
  3. Time-division duplex structure
     - Various applicable configurations
  4. Congestion due to resource sharing and queuing
     - High load scenes
- Explained:
  - QoS architecture within the mobile network
  - Latency contributions in different layers of radio protocols

Example:
Observed characteristics in real network

- Packet delay distribution from a 5G system (OpenAirInterface 5G network)

	Packet Delay		
	10 ms	15 ms	20 ms
Cumulative probability	99%	99.99%	99.999%
HARQ Re-transmission	0.01%	15%	45%
RAN Transmission	27%	25%	10%
RAN Segmentation	43%	40%	30%
RAN Queuing Delay	29.99%	20%	15%

- In particular for high-reliability latency bounds (e.g. 99.999%) there is a substantial packet delay variation being introduced by a 5G system

Scheduling related future work
Facing the real problems

- Scheduling:
  - provides bounds on queuing delay (BUT the node internal forwarding delay is another integral part of end-to-end delay).
- The node internal forwarding delay:
  - mobile virtual DetNet nodes cause a packet delay that is stochastic and heavy-tailed.
- Challenges for end-to-end scheduling
  - Clock-driven scheduling scenarios:
    - Bad reliability-efficiency trade-off
    - Higher complexity of scheduling problem formulation and solution
  - Other (non-clock-driven) scheduling mechanisms:
    - Higher complexity of end-to-end delay analysis

Summary & Next Steps

- Summary
  - Wireless communication provides flexibility and simplicity, but with inherently stochastic components that lead to packet delay distributions metrics exceeding significantly those found in wired counterparts.
  - Some traffic shaping mechanisms, like time-scheduled transmission (i.e., IEEE 802.1Qbv), expect very deterministic latency behavior in every node on the transmissions path. The latency distribution of a 5G system makes it impracticable to implement some legacy time-schedule configurations.

- Next Steps
  - Comments / views are welcome
  - Discussion on wireless-friendly latency solutions

Thanks …