

Problem and Consideration of Deterministic QoS for Enhanced DetNet

Peng Liu (CMCC)

liupengyjy@chinamobile.com

Scenario Development of Deterministic Networks

The development of 5G and Industrial Internet has further enriched the scenarios and requirements of deterministic networks by promoting the new emerging applications.



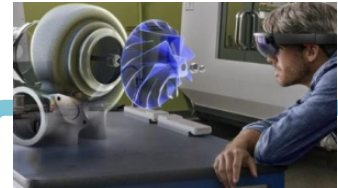
Power differential protection



Industrial control



Cloud gaming



AR/VR

ToB-Industrial low latency
(2000~2020)

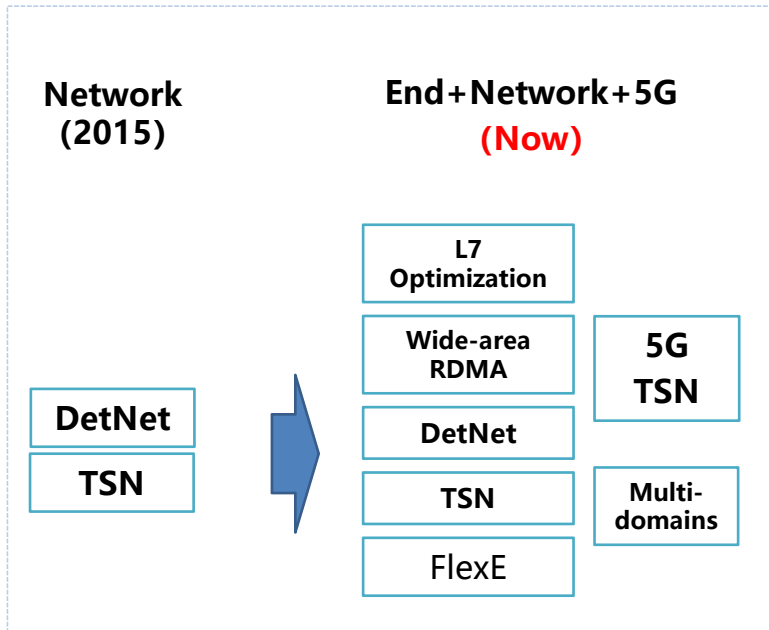


ToC-User Experience low latency
(2020~)

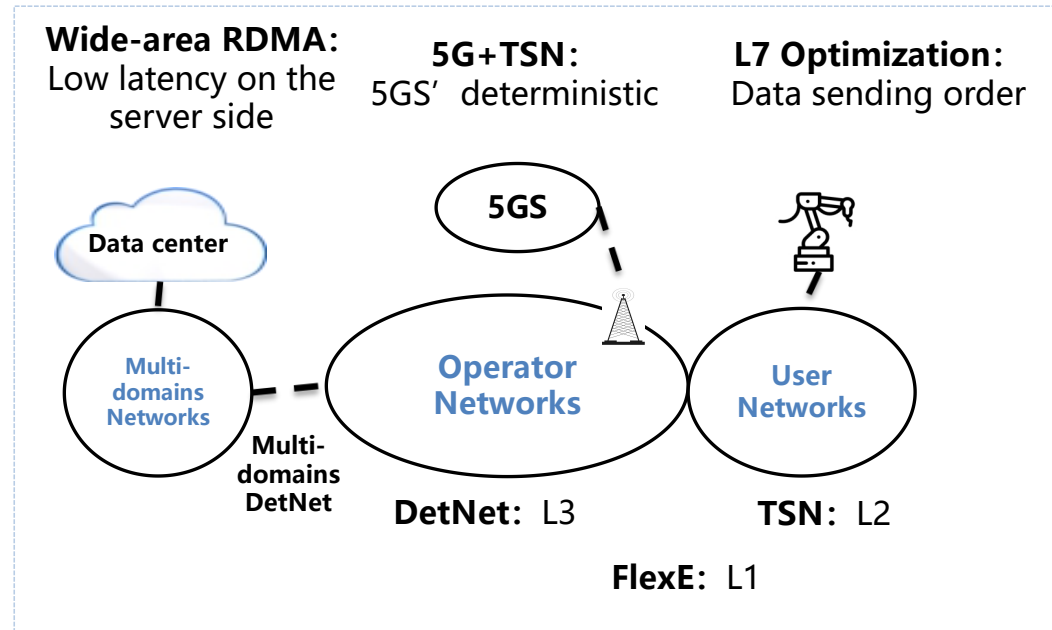
Technique Evolution of Deterministic Networks

With the demands of diversified applications, the technology of deterministic networks has gradually expanded from the network side to the end side, improving its technical system.

Technological development



Technology application

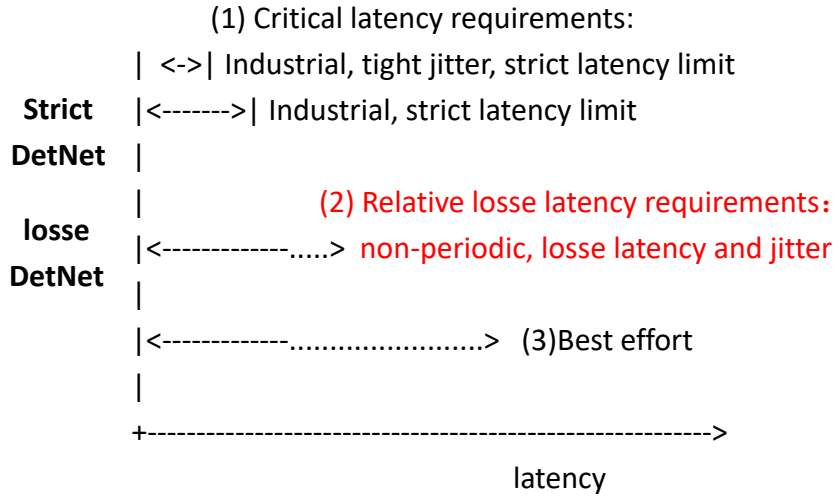


In the following 5 slides of ***problem statement***, the ‘DetNet’ or ‘Deterministic Networks’ doesn’t only refer to IETF DetNet WG, but is the general technique system, to avoid the misunderstanding.

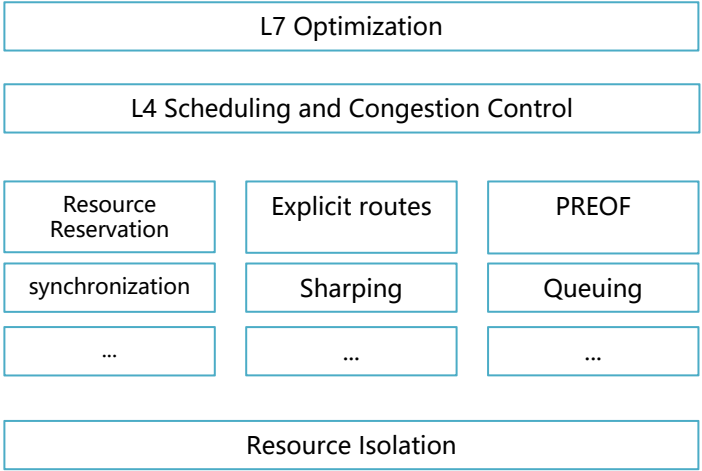
Problem 1: How to match Scenarios with Techniques?

The biggest problem faced by the deployment of deterministic networks is the balance between the application demand and detnet costs.

Scenarios



Techniques

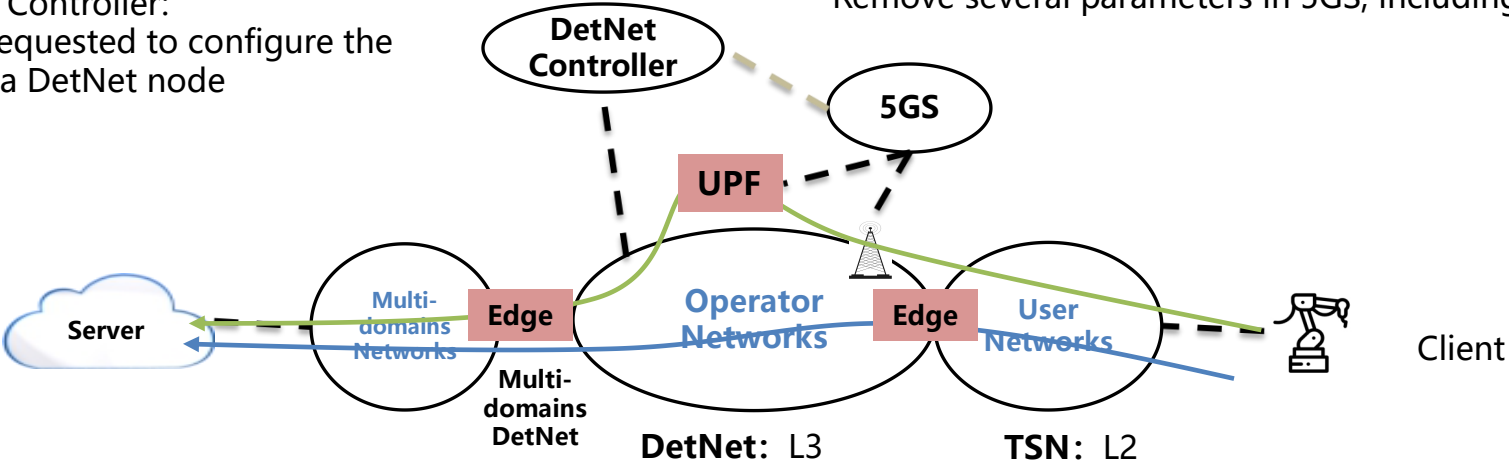


The detailed classification of the Applications demand and methods capability is needed.

Problem 2: How to Gurantee the E2E Latency of Multi-domain Networks?

DetNet Controller:
Being requested to configure the
5GS as a DetNet node

UPF:
Remove several parameters in 5GS, including Qfi, Qci



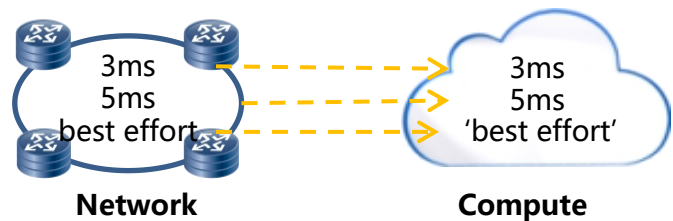
Edge between operators domains:
Different Queuing methods: TAS → ATS/CQF
Different granularity of clock
Different certifications

Edge between L2 and L3:
Different Queuing methods: TAS → ATS/CQF
Different granularity of clock

Problem 3: How to Gurantee the E2E Latency of Network and Computing?

The E2E latency of User experience includes the latency of packets/flows transmission(network) and the application processing(compute).

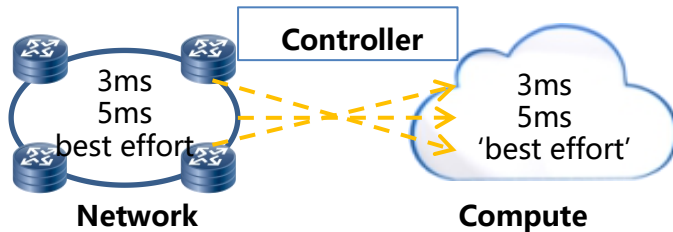
Option1: Determine network and/or computing separately



Network Resources: Dedicated resources and dedicated queues

Compute Resources: Realtime systems and thread lock

Option2: Determine network and computing together



Integrated resources: e.g. RSVP extension reserves computing resources??

Integrated scheduling: e.g. coordination of queue scheduling and thread scheduling??

Problem 4: How to Gurantee the Latency from the Perspective of Application?

To adapt to the traffic model of the application flows



Industrial Sensors

Periodic data collection



Video

Frame Transmission

Some data will be sent periodically, which is close to the data model of queuing methods.

Some applications' data will have mirco burst, while their demands of bandwidth is an 'average' number , which will cause the non-determinsitic latency.

To adapt to the running model of the applications



Cloud gaming

Multi-users in one game



AR/VR

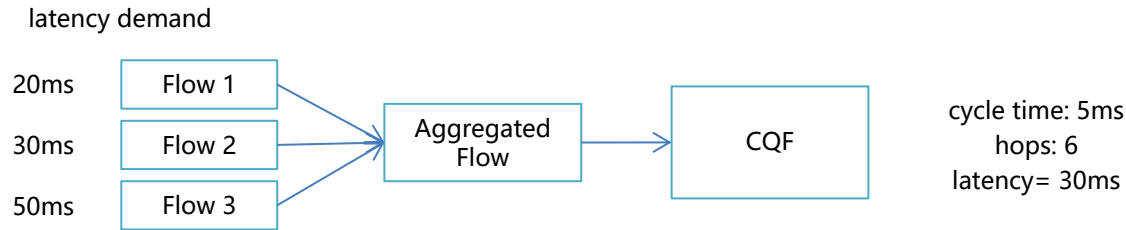
Single user in one game

Existing technology can to some extent ensure low latency and deterministic forwarding in the network.

It cannot guarantee service consistency for multiple users in the same service, which can lead to differences and "unfairness" in the experience.

Problem 5: How to Balance the Scalability and Performance?

Flow aggregation is an useful way to solve the problem of multiple flows in the scaling network. However, the aggregated flow will have the constant latency of different flows with different latency demands.



Setting multiple queues for different flows or setting dedicated resource in the aggregated flows seems return back and violate the 'aggregated' .

Summary of Problems and Considerations

- Problem 1: How to match Scenarios with Techniques?
- Problem 2: How to Gurantee the E2E Latency of Multi-domain Networks?
- Problem 3: How to Gurantee the E2E Latency of Network and Computing?
- Problem 4: How to Gurantee the Latency from the Perspective of Application?
- Problem 5: How to Balance the Scalability and Performance?



- May need the detailed classification.
- May need the enhancement of control plane.
- May need the enhancement of data plane. (Now the WG are working on this)
- May need be aware of application's demand.
- May need the fine granularity QoS.

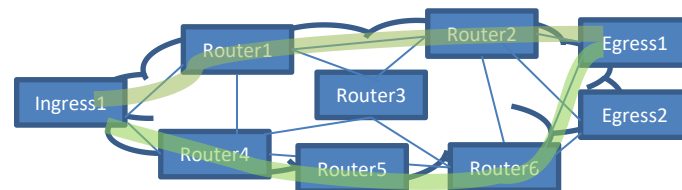
We may need more identifier and related enhanced functions to solve those problems, but we don't need to solve all the problems at the same time.

We are working on the enhancement of data plane and queuing mechanisms, we should further find out what we can do in DetNet and IETF.

Potential direction: Using the Outside QoS

- In a provider network, the flows are often tunneled. E.g. In the SRv6 case, we would have an inside IPv6 Header, and an outside IPv6 Header

Outside IPv6 Hdr	SA = Ingress IP, DA = SIDs, Outside QoS
SR Hdr	Segment List with Egress SID
Inside IPv6 Hdr	SA = Client IP, DA = Server IP, Inside QoS



- The outside QoS in the Outside IPv6 Header can be defined by the network operator, and only take effect within the operator's network
- Besides the differentiated DetNet QoS mentioned on the last page, other QoS marking mechanism of the outside QoS can also be considered
 - which would trigger different treatment on the nodes

Potential direction: Control Plane Extensions for DD-TE

- DD-TE means Differentiated DetNet-aware Traffic Engineering
 - which is mentioned in draft-xiong-detnet-teas-te-extensions
- Within the procedure of DD-TE, we need
 - the flow of the service can be identified and treated specifically on the edge nodes, and on the intermediate nodes
 - the identification could be a DD-QoS value (Differentiated DetNet QoS)

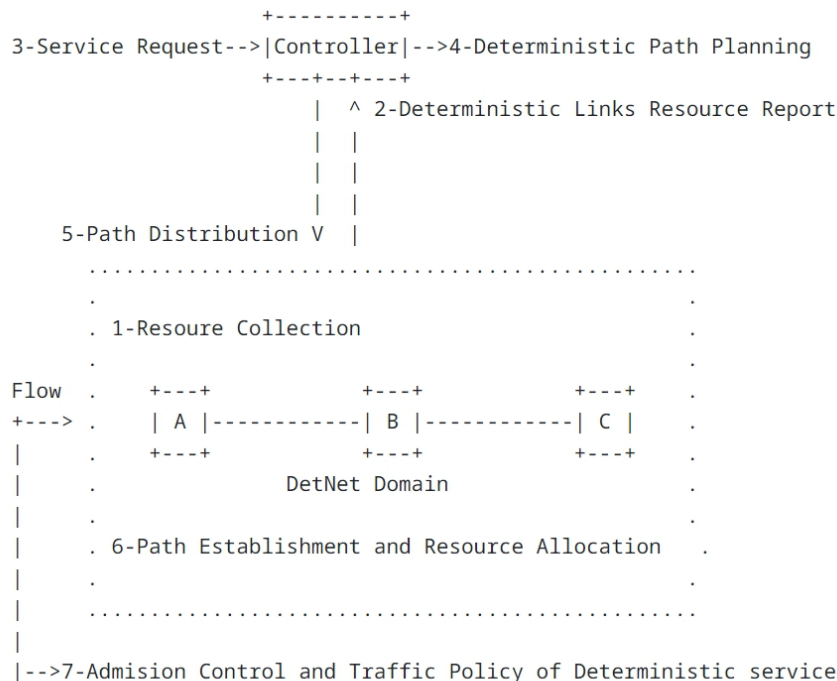


Figure 2: The Control Plane for DD-TE

Thanks! !