China Mobile’s HPWAN Services and Transport Optimizations

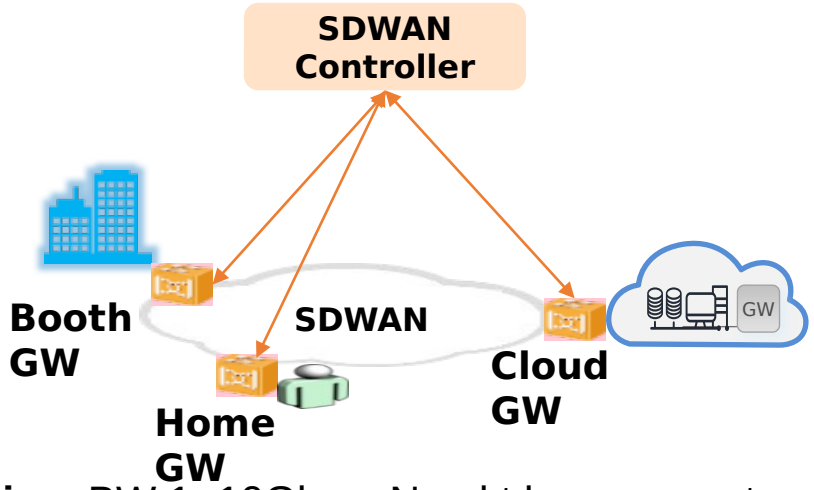
Kehan Yao, China Mobile

IETF 121 HPWAN BoF

Differential HPWAN Use Cases over China Mobile's Operator Network

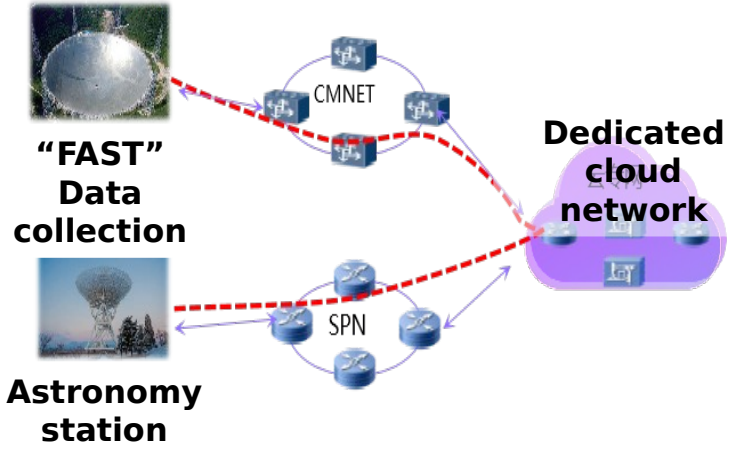
- **China Mobile provides two major types of HPWAN services.**
- **Each type of service is faced towards different customers and scenarios.**

Type1: BW(with SLA guarantee), Pkt loss rate(no SLA guarantee)



- **Metrics:** BW 1~10Gbps, No pkt loss guarantee
- **Primary customers:** To Home and small enterprise
- **Scenarios:** supermarker chains uploading operational data, movies and rendering data

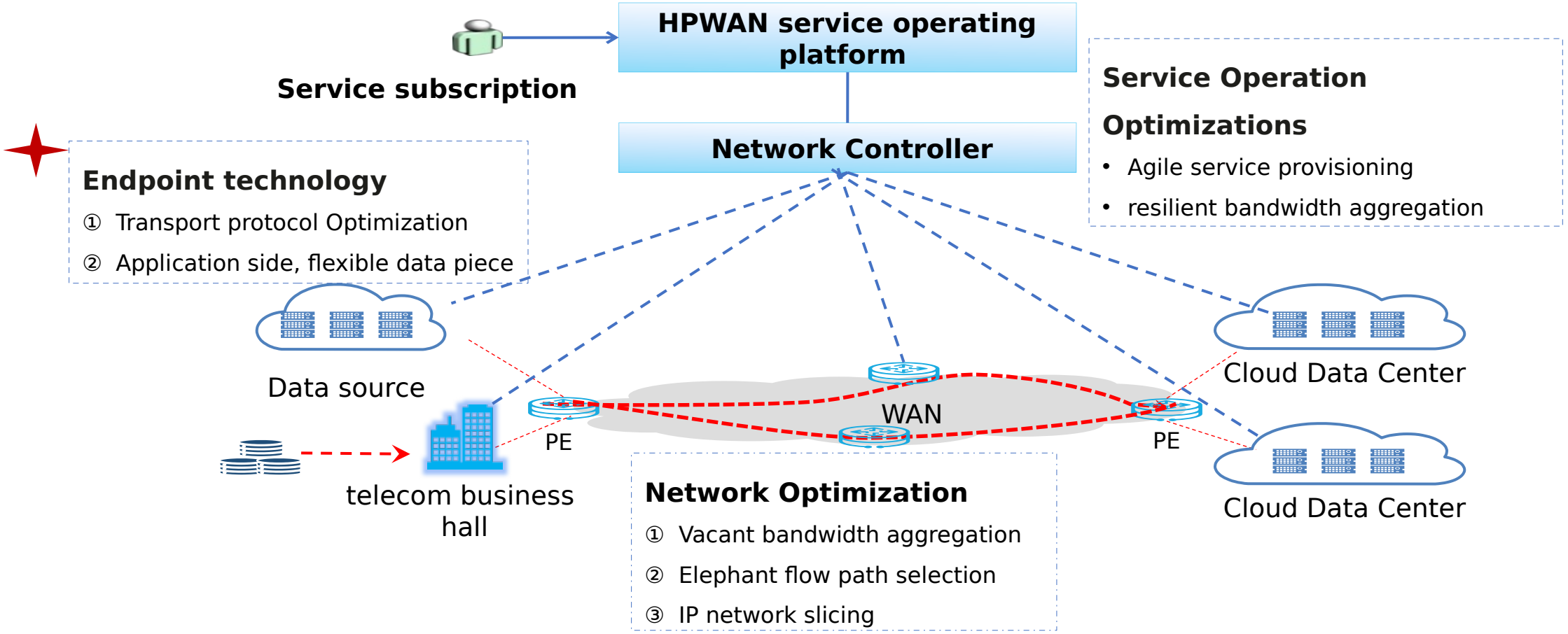
Type2: BW and Pkt loss rate (Both with SLA guarantee)



- **Metrics:** BW 10~100Gbps, Pkt loss ~0.01%
- **Primary customers:** to colleges, institutes, and large enterprise
- **Scenarios:** cross-clouds backup, astronomy data

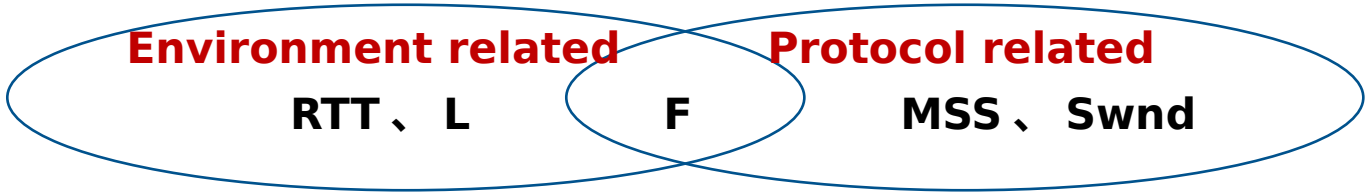
HPWAN Technical Requirements from Operators' View

- Improving **throughput** is the primary objective for HPWAN services
- From network operators' view, HPWAN services raise technical requirements on **network side, endpoint side, and system operations**.
- There are lots of exisiting network side optimizations, the new design space is primarily on endpoint side.



China Mobile's Experience with Different Transport Solutions

- **Factors that could impact the end-to-end transmission throughput:**
  - transmission distance(**RTT**), packet loss rate(**L**), Maximum Segment Size(**MSS**), sending window(**Swnd**), and other factors(**F**).
  - Other factors(**F**) include **CPU utilization, PCIe speed, hard disk IO speed,** etc.
  - Transport protocols should care about both hardware and software coordination.



- **China Mobile chooses two different technical roadmaps in transport solutions.**

Hardware-based solution

- **Product Integration:** network adapter
- **Performance:** depends on network adapter BW and offloading optimization
- **Scenarios:** high throughput requirements(~100Gbps)

Software-based solution

- **Product Integration:** Linux/Windows OS
- **Performance:** Depends on CPU types
- **Scenarios:** low and medium throughput requirements(<40Gbps

Performance Test under Different Transport Solutions

- **We analyzed the throughput performance of different transport protocols and solutions in draft** https://datatracker.ietf.org/doc/draft-yy-hpwan-transport-gap-analysis/
  - **Software-based solutions:** TCP + BBRv1, TCP + CUBIC, QUIC+ BBRv1(all with TOE enabled)
  - **Hardware-based solutions:** Modified RoCEv2

	TCP+BBRv1 |0.1%Pkt loss	TCP+BBRv1 |1%Pkt loss	TCP+CUBIC |0.1%Pkt loss	TCP+CUBIC |1%Pkt loss
Single Stream	14Gbps	10Gbps	8.6Mbps	Null
3 Streams	41Gbps	24.5Gbps	28Mbps	Null
10 Streams	70Gbps	61Gbps	91Mbps	Null
25 Streams	84Gbps	84.7Gbps	Null	Null

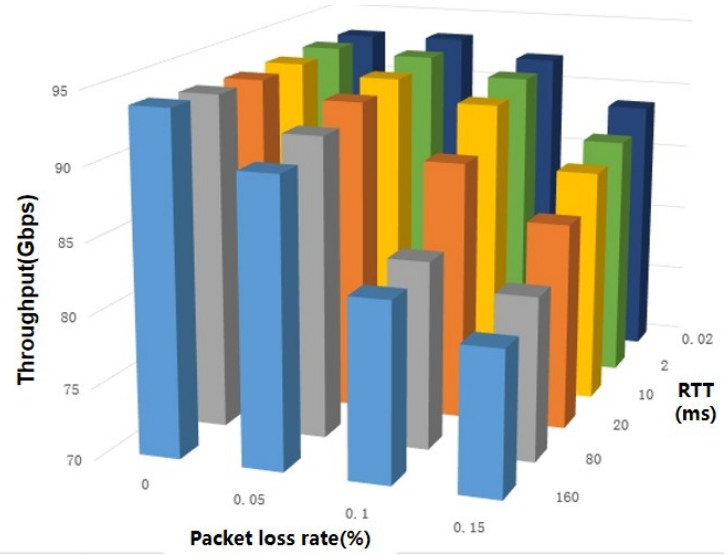
Figure 2: TCP throughput performance,RTT=70ms,MTU=1500

TCP throughput performance test (With TOE enabled)

	QUIC+BBRv1 |0.1%Pkt loss 40 cores	QUIC+BBRv1 |0.1%Pkt loss 80 cores
40 Streams	47.2Gbps	52.8Gbps
60 Streams	42.4Gbps	57.2Gbps
80 Streams	51.2Gbps	62.4Gbps
100 Streams	NULL	63.2Gbps

Figure 3: QUIC Throughput Performance, RTT=65ms, MTU=1500

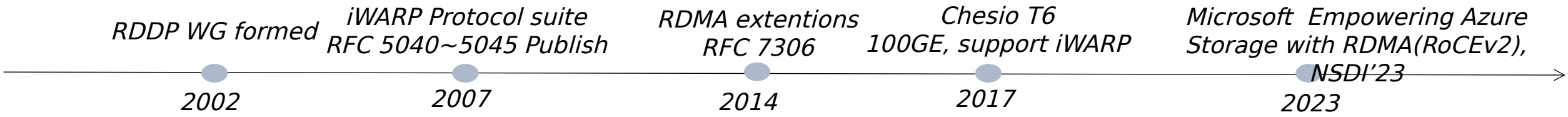
QUIC throughput performance test (With TOE enabled)



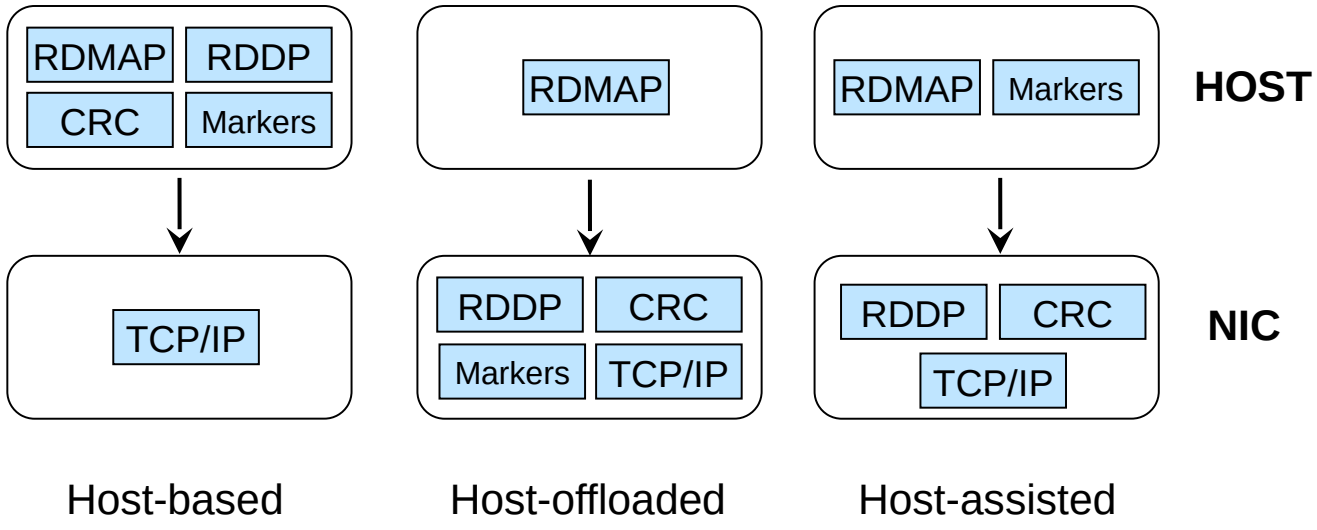
Modified RoCEv2 over WAN

- Software-only based solutions are **high-cost** because of **CPU overhead**.
- hardware-based solutions**(RDMA over WAN)** is promising**. (Please note that we don't want to standardize RoCEv2 in IETF, we just want to show its applicability for HPWAN applications.)**

Revisiting RDMA over WAN 20 Years Later



- **Why IETF Should care about RDMA today ?**
  - Because HPWAN is an Internet and enterprise network service.
  - Because IETF's standard **iWARP** doesn't perform well in throughput performance.



Three different offload methods of iWARP [SC'07]

	MSG size 128B		MSG size 1KB		MSG size 256KB	
	BW	CPU Util	BW	CPU Util	BW	CPU Util
host-based	100MB	12%	800MB	40%	1.8GB	75%
host-offloaded	1GB	10%	3.5GB	10%	3.5GB	10%
host-assisted	500MB	18%	3.5GB	50%	5.8GB	80%

iWARP throughput and CPU overhead performance Analysis[SC'07]

SC'07: Analyzing the Impact of Supporting Out-of-order Communication on In-order Performance with iWARP

RDMA over WAN New Requirements

- **Support RDMA**

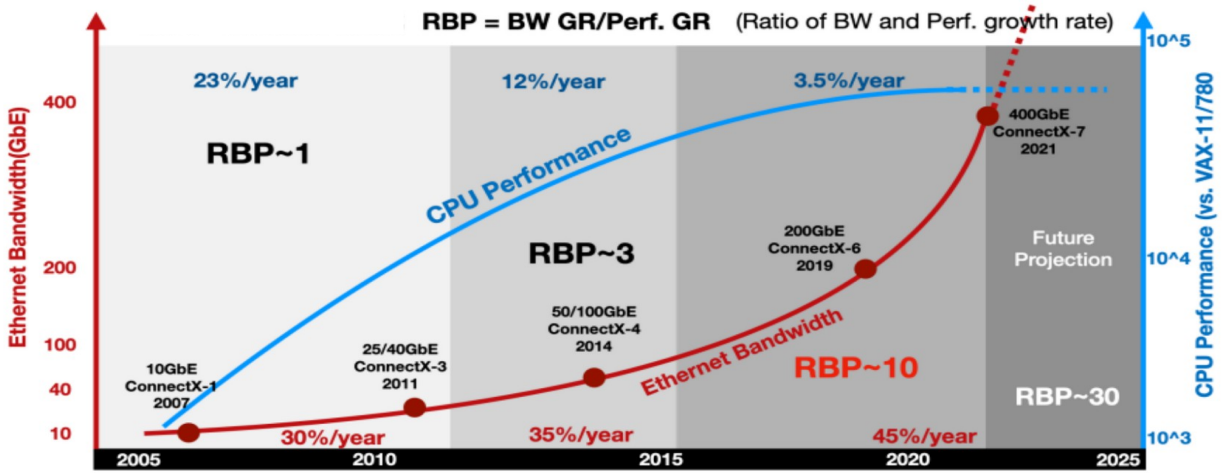
- **Lightweight and concise transport layer**
  - congestion control, rate control
  - reliability
  - security
  - multi-path transmission

- **Interfaces**
  - Compatibility between Verbs and Socket

Conclusions

- Beyond 400G **ultra high speed ethernet and Internet backbone** is speeding up and has become the industry trend.

- On the other hand, the improvement of **CPU performance** has been saturated.



Comparison between the growth rate of CPU performance and Ethernet Bandwidth

- **RDMA transport offloading in data center network sets a good example:**
  - like TTPoE, not only requires no modification on intermediate network(e.g, losslessness), but it also can maintain high throughtput and low latency.

- **Standardization on RDMA over WAN needs further attention, considering emerging large-scale HPWAN service.**