Towards a high-quality IP network

Huawei, IP Research Dept.
Jiayi Zhang
Oct.9, 2020
zhangjiayi11@huawei.com

Security Level:

HUAWEI

Requirement for high quality IP network

- In 5G era, various applications are enabled, with different requirements for network

V2X
Delay: 5 ms
Reliability: 10^{-4}

Smart Grid
Delay: 5 ms
Reliability: 10^{-5}

AR/VR
Delay: 50 ms
Reliability: 10^{-3}

8K Video
Delay: 150 ms
Reliability: 10^{-3}

Transportation
Delay: 10 ms
Reliability: 10^{-5}

Smart Factory
Delay: 10 ms
Reliability: 10^{-4}

- It requires multiple service level agreement (SLA) to network

5QI Value	Resource Type	Default Priority Level	Packet Delay Budget	Packet Error Rate	Default Maximum Data Burst Volume(NOTE 2)	Default Averaging Window	Example Services
1	GBR NOTE 1	20	100 ms	10^{-2}	N/A	2000 ms	Conversational Voice
2		40	150 ms	10^{-3}	N/A	2000 ms	Conversational Video (Live Streaming)
3		30	50 ms	10^{-3}	N/A	2000 ms	Real Time Gaming, V2X messages Electricity distribution – medium voltage, Process automation - monitoring
4		50	300 ms	10^{-6}	N/A	2000 ms	Non-Conversational Video (Buffered Streaming)
65		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]000 ms	Mission Critical user plane Push To Talk voice (e.g., MCPTT)
66		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]000 ms	Non-Mission-Critical user plane Push To Talk voice
67		15	100 ms	[illegible]	N/A	2000 ms	Mission Critical Video user plane
75		25	50 ms	10^{-2}	N/A	2000 ms	V2X messages
5	Non-GBR NOTE 1	10	100 ms	10^{-6}	N/A	N/A	IMS Signalling
6		60	300 ms	10^{-6}	N/A	N/A	Video (Buffered Streaming) TCP-based (e.g., www, e-mail, chat, ftp, p2p file sharing, progressive video, etc.)
7		70	100 ms	10^{-3}	N/A	N/A	Voice, Video (Live Streaming) Interactive Gaming
8		80	[illegible]	[illegible]	N/A	N/A	Video (Buffered Streaming) TCP-based (e.g., www, e-mail, chat, ftp, p2p file sharing, progressive video, etc.)
9		[illegible]					
69		[illegible]	[illegible]	[illegible]	N/A	N/A	Mission Critical delay sensitive signalling (e.g., MC-PTT signalling)
70		55	200 ms	10^{-6}	N/A	N/A	Mission Critical Data (e.g. example services are the same as QCI 6/8/9)
79		65	50 ms	10^{-2}	N/A	N/A	V2X messages
80		68	10 ms	10^{-6}	N/A	N/A	Low Latency eMBB applications Augmented Reality
81		11	5 ms	10^{-5}	160 B	2000 ms	Remote control (see TS 22.261 [2])
82	Delay Critical GBR	12	10 ms NOTE 5	10^{-5}	320 B	2000 ms	Intelligent transport systems
83		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]0 ms	Intelligent Transport Systems
84		[illegible]	10 ms	[illegible]	[illegible]	2000 ms	Discrete Automation
85		22	10 ms	10^{-4}	1358 B NOTE 3	2000 ms	Discrete Automation

Bandwidth Sensitive Services

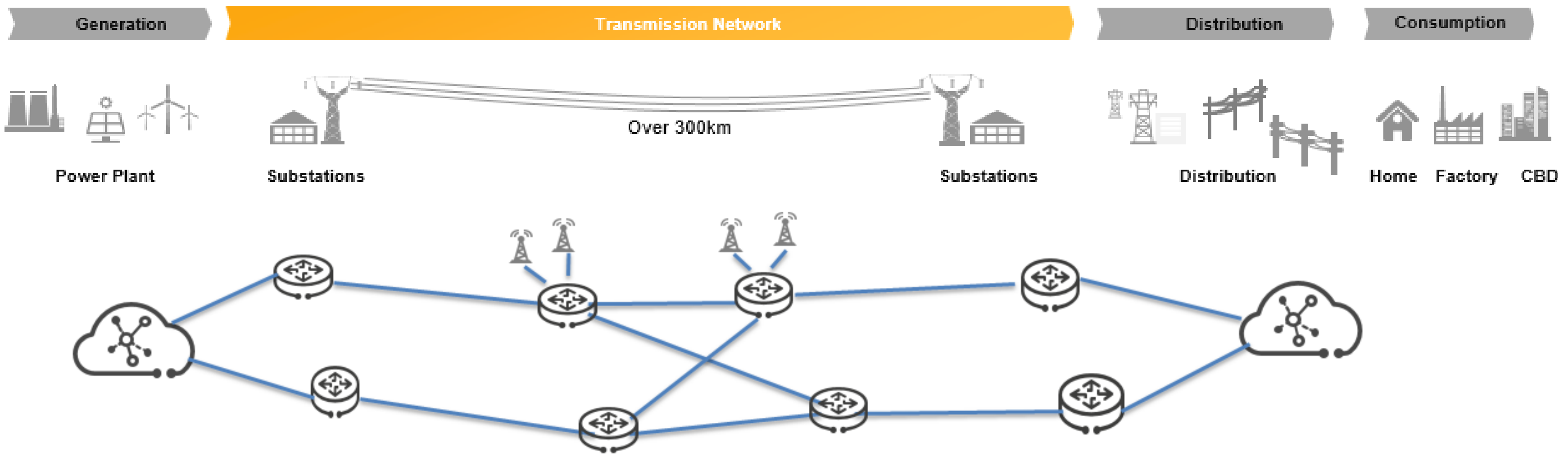
Connection Services

URLLC Latency Sensitive Services

- **Bandwidth sensitive services** care more about throughput as long as its short term bursting can be buffered and transit later.
- **Connection services** like the legacy services on IP networks.
- **Latency sensitive services** are defined for new applications required reliable bounded latency, appearing with new era carrier networks.

Challenge: how to guarantee bounded latency for different applications in various networks

Smart grid requirement over carrier networks



Strict requirements:

- ultra high reliability
- bounded low latency

Traditionally smart grid monitoring and controlling applications are connected by separate networks.

5G URLLC networking enables Smart Grid tele-protection over carrier network, which requires strict latency bound (5ms) according to 3GPP TS 23.501.

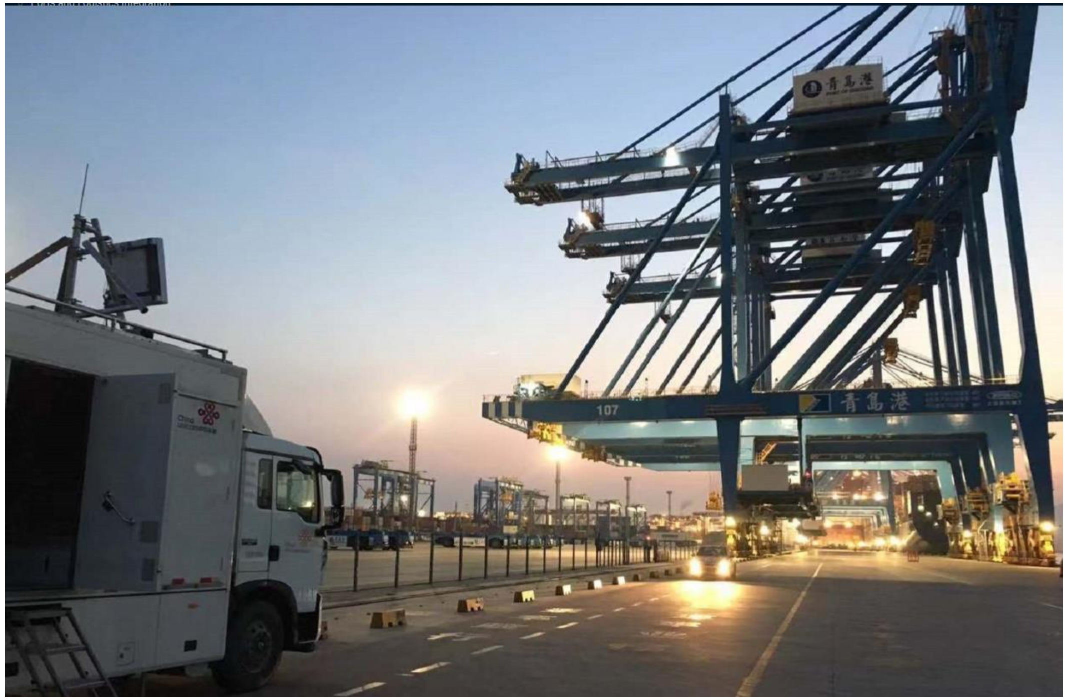
5QI Value	Resource Type	Default Priority Level	Packet Delay Budget (NOTE 3)	Packet Error Rate	Default Maximum Data Burst Volume (NOTE 2)	Default Averaging Window	Example Services
85		21	5 ms (NOTE 5)	10^{-5}	255 bytes	2000 ms	Electricity Distribution- high voltage (see TS 22.261 [2]). V2X messages (Remote Driving. See TS 22.186 [111], NOTE 16)

https://www.gsma.com/futurenetworks/wp-content/uploads/2020/02/5_Smart-Grid-Powered-by-5G-SA-based-Network-Slicing_GSMA.pdf

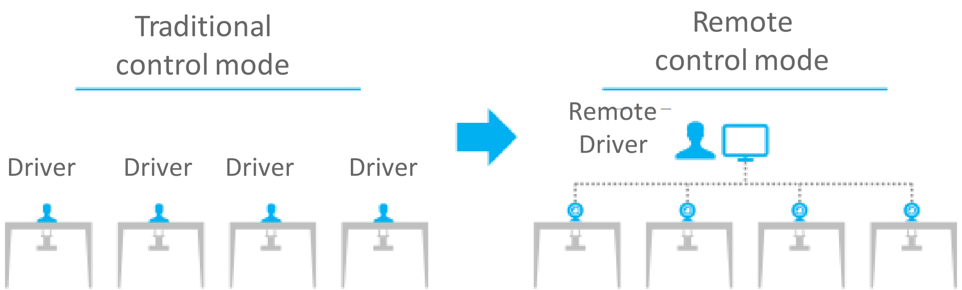
HUAWEI

From IEEE TSN Meeting, July 2020

Smart port over small regional network



"An automated ship-to-shore (STS) crane that was operated via a 5G link to the control center and used to lift containers."



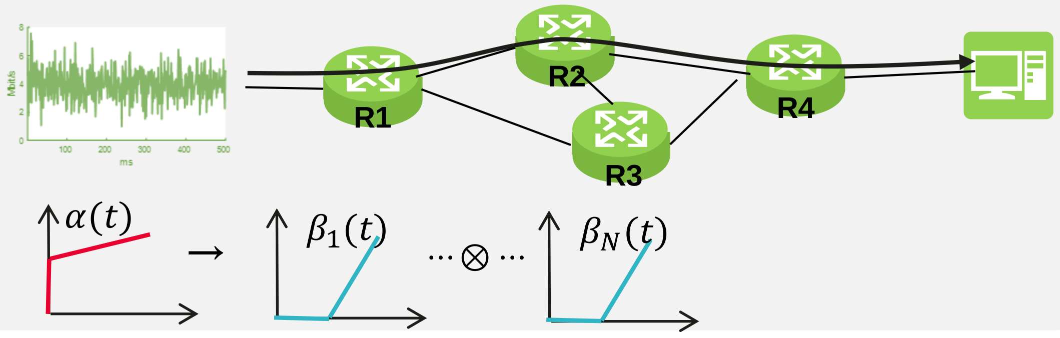
- To improve shipping efficiency and safety, video surveillance and AI detection are used to help controllers.
- Onsite remote controlling requires strict latency bound (30ms)
- Requirements
  - High Reliability
  - Bounded latency

https://www.maritime-executive.com/article/5g-smart-port-system-trialed-at-qingdao

Achieving SLA: by modeling the network and calculating performance bound

Three components to model the network

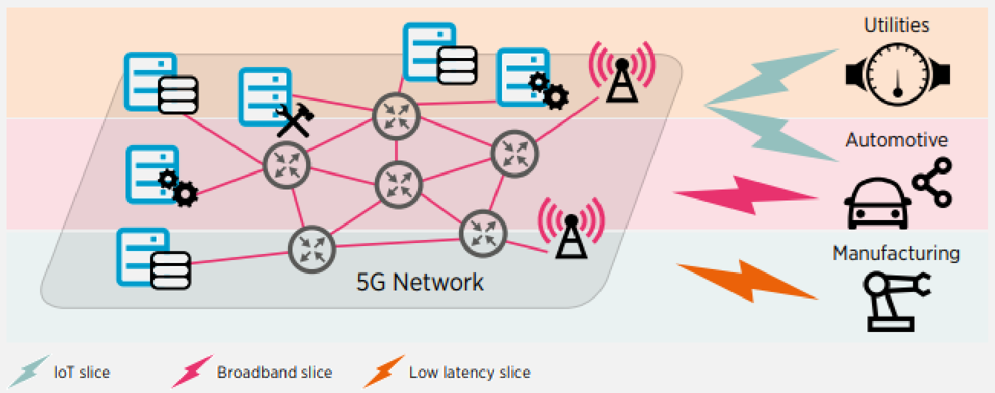
1. **Traffic fitting**
2. **Service model**
3. **Network analysis**



SLA guaranteed solution for different applications

- **Network Slicing** (resource allocation for multiple applications)
- **Network deployment on-demand**: delay constrained routing, for dynamic changing traffic**,** maybe with measurements
- Other solutions?....

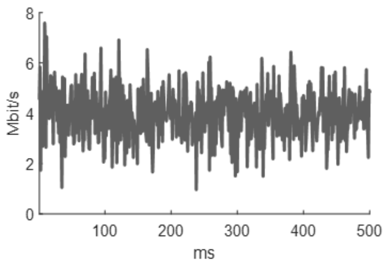
5G networks subdivided into virtual networks each optimised for one business case



https://www.gsma.com/futurenetworks/wp-content/uploads/2017/11/GSMA-An-Introduction-to-Network-Slicing.pdf

1. Traffic characterization

Traffic Arrival



Machine Type

Human to machine/ Human

Traffic characterization
(arrival curve/ distribution)

Deterministic characterization
- Period
- lower order moment

Stochastic Distribution
- Poisson
- MMOO
- (B)MAP , … …

Envelope / error bounding function
- Linear Envelope
- (Error) Bounding Function
- MGF (bound), ……

Other ways?

Network Performance Evaluation theories
- DNC
- SNC
- Queuing Theory

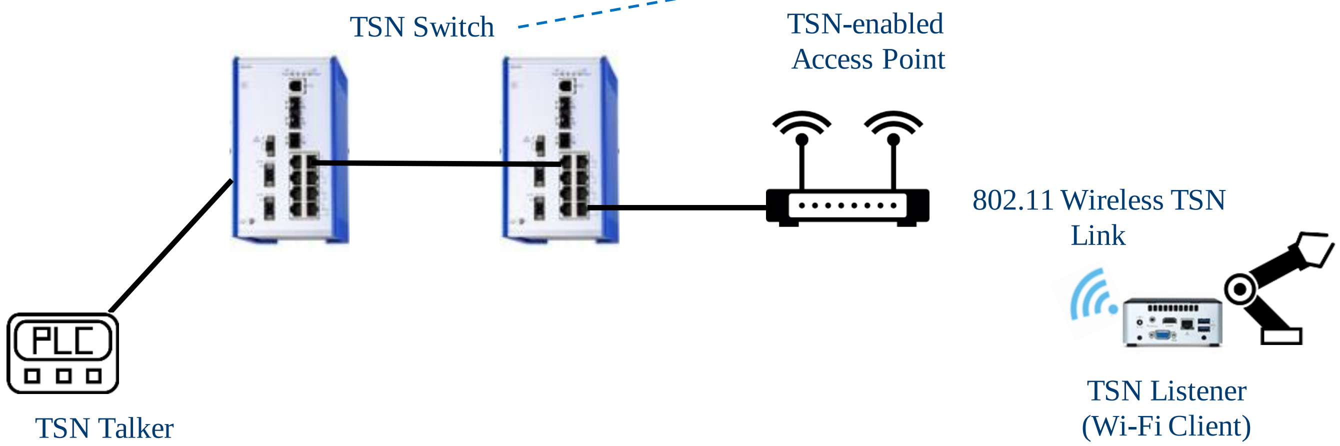
Difficulties: Bridge from real traffic to the fabulous NC toolkits

- For some traffic, maybe difficult to obtain a proper fitting (especially consider applications, transport layer protocol) → A database for traffic pattern?
- For some stochastic distribution, hard to derive bounding function / MGF (bound) – **From BMAP to MGF(bound)**?

HUAWEI

2. Service model

- Goal
  - Precise service model can notably improve the NC performance bound



① Advanced model for schedulers: FIFO / FlexE / TSN schedulers / WiFi / 5G

② Correction of theoretical service model for real device, based on measurement

③ Service model for unknown device/network: Measurement based modeling?

④ New scheduler for delay guarantee: with high throughput, low-complexity implementation. e.g. Deadline-based scheduling

HUAWEI

https://mentor.ieee.org/802.11/dcn/19/11-19-1287-01-00be-tsn-support-in-802-11-and-potential-extensions-for-tgbe.pptx

3. Network analysis

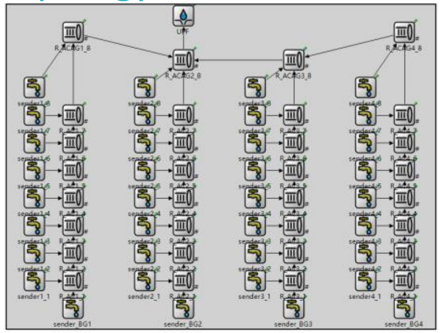
- Goal:
  - Tighter performance bound, to improve utilization
  - Reach the constructed worst-case delay
  - Availability for arbitrary topology, e.g. ring, join-fork, largescale

- Research Directions
  - DNC:
    - Algebra-based DNC algorithm
    - Linear Programing (LP) method gives exact bound, need reduce computation complexity
    - Combined with Deep Learning
  - SNC:
    - MGF
    - (g)SBB
    - Tail bound
    - Doob's Martingale

    →

    - Abstract of Arrival / service model
    - E2e operation
    - Automated calculator

Topology

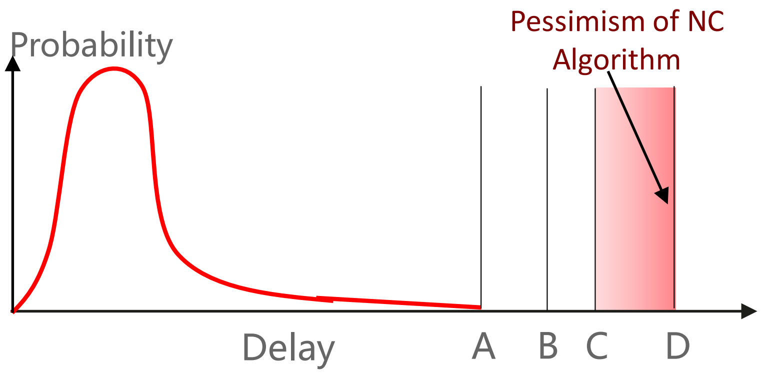


Theoretical delay bound

SFA	PMOO	TMA
820	804	314

Simulated maximum delay

Simulation Time	Accumulated Data (bit)	Average Delay (us)	Probability Delay			Maximum Delay
			99%	99.9%	99.99%	
1s	1401	83.7	106.0	106.4	106.4	106.4
10s	14008	84.1	106.0	107.2	107.5	107.5
1min	84008	84.2	122.4	127.7	128.0	128.1
10min	840044	84.5	197.0	235.5	239.7	240.9

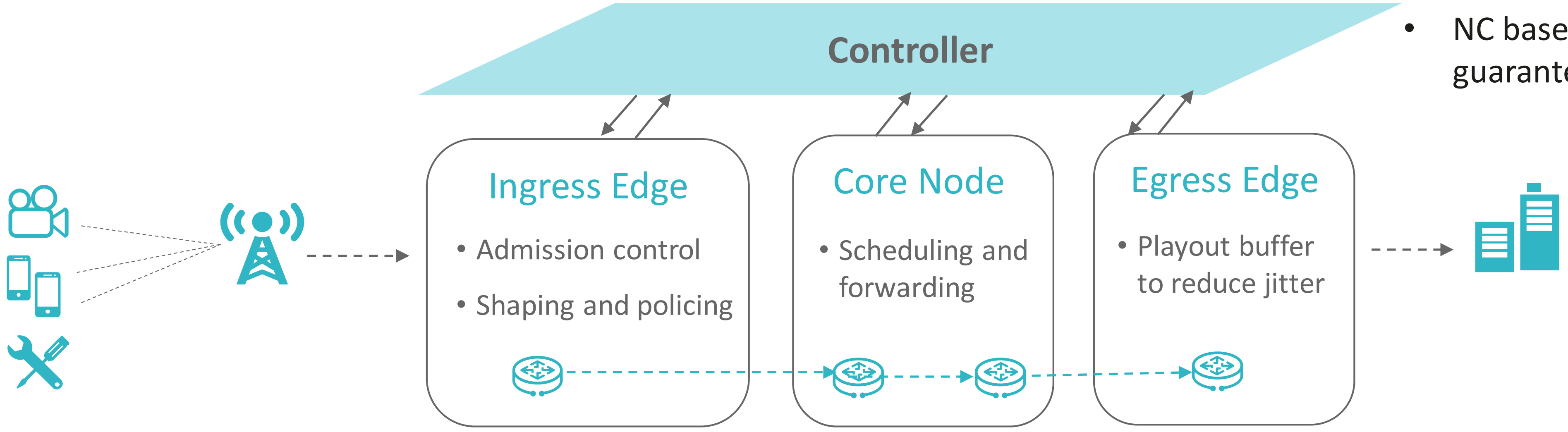


A： Simulation obtained maximum delay
B： Constructed corner-case delay
C： lowest of theoretical delay upper- bound
D： Delay upper-bound by some NC algorithm

Using NC to support a high quality IP network

Intra-carrier network

- Goal: Provide SLA guarantee for many service (slicing / resource allocation)



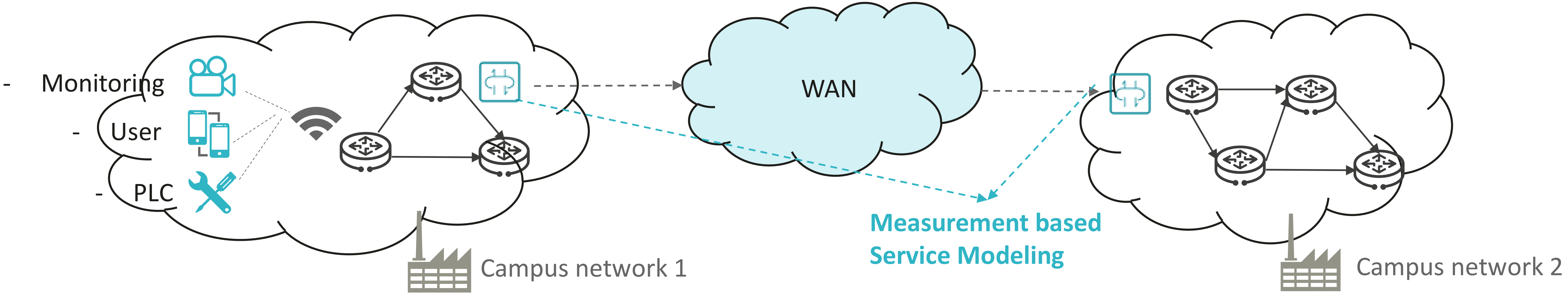
Controller function

- NC based model for delay guaranteed configuration

- **Challenge:**
  - Online traffic fitting and recognizing → Ingress edge
  - Dynamic Delay constrained routing & resource allocation : fast and incremental support, mobility support →Controller

Using NC to support a high quality IP network

Inter-carrier network



Goal: Different SLA are required across two campus networks, connected by a WAN. The service model is hardly obtained , neither is e2e delay bound.

- Idea 1: Modeling Service from measurement
  - Correction for theoretical service model
  - Obtain the service capacity (bound) for unknown device / scheduling
- Idea 2: Adaptive Solution to ensure SLA
  - Adaptively adjust the transmitting,
  - Choosing different tunnels for load balance

Summary

Towards better model of network to guarantee performance:

- Traffic fitting
- Service model
- Network analysis
- SLA guaranteed solutions for dynamic traffic

Welcome cooperation in research, projects, standards

Thank you.

把数字世界带入每个人、每个家庭、
每个组织，构建万物互联的智能世界。

Bring digital to every person, home and organization for a fully connected, intelligent world.

HUAWEI