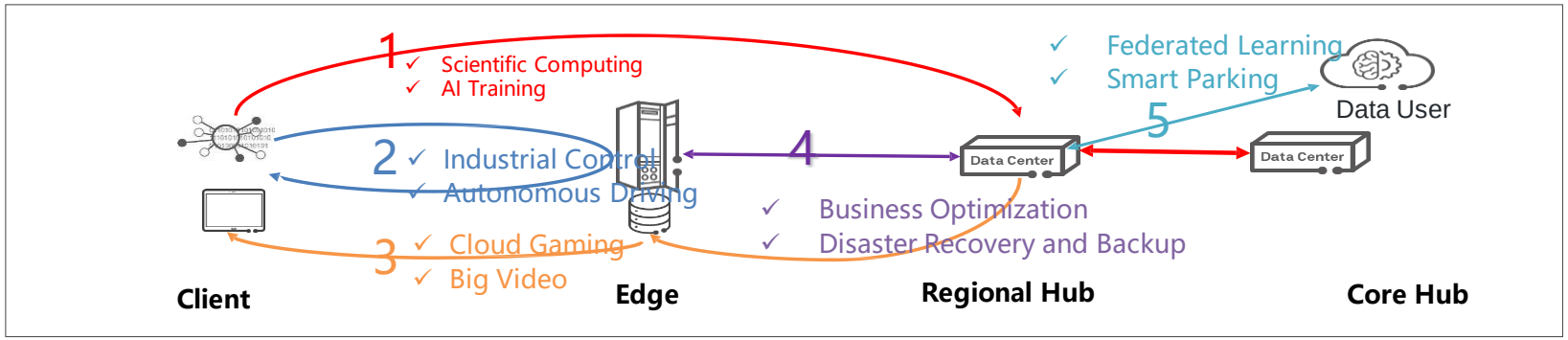
RDMA proxy Inter-DC Over WAN Using Gateway

Rubing Liu (H3C presenter)

IETF 121 11/7/2024, rev 1.0

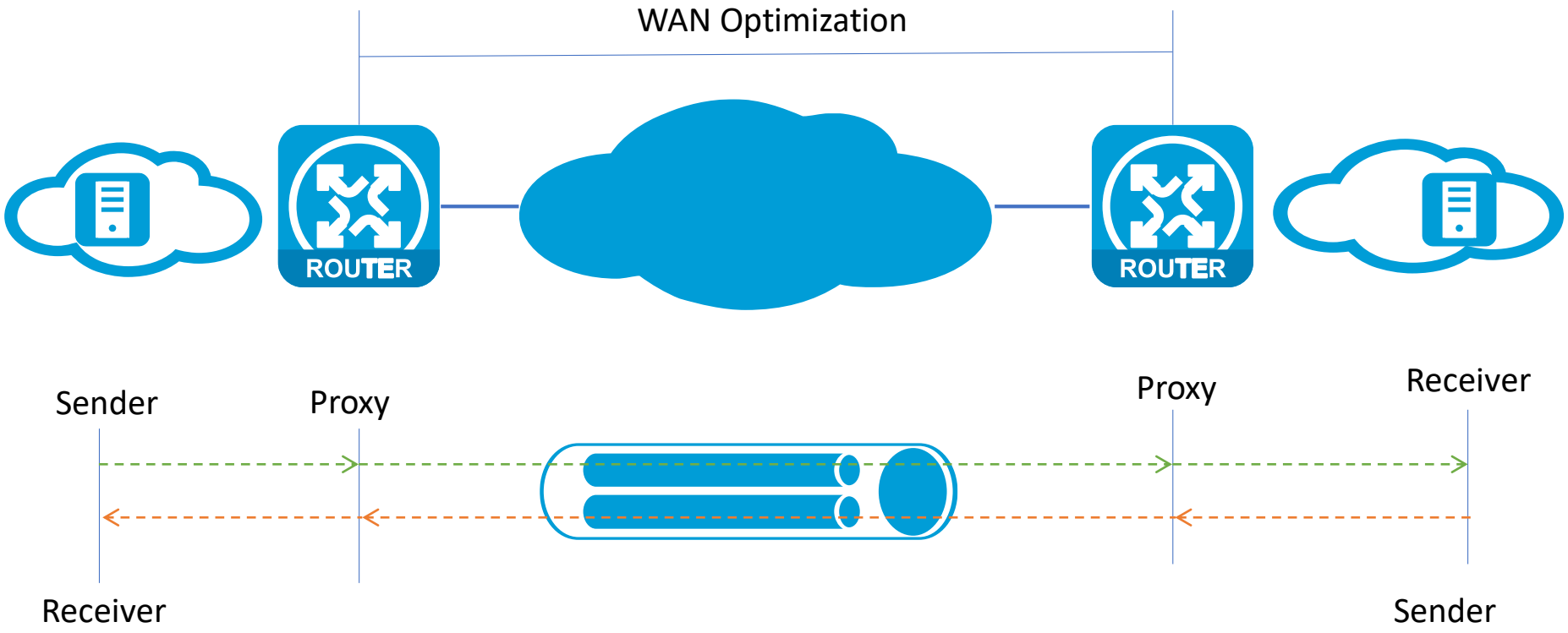
Issues and Innovations

Issues and Innovations



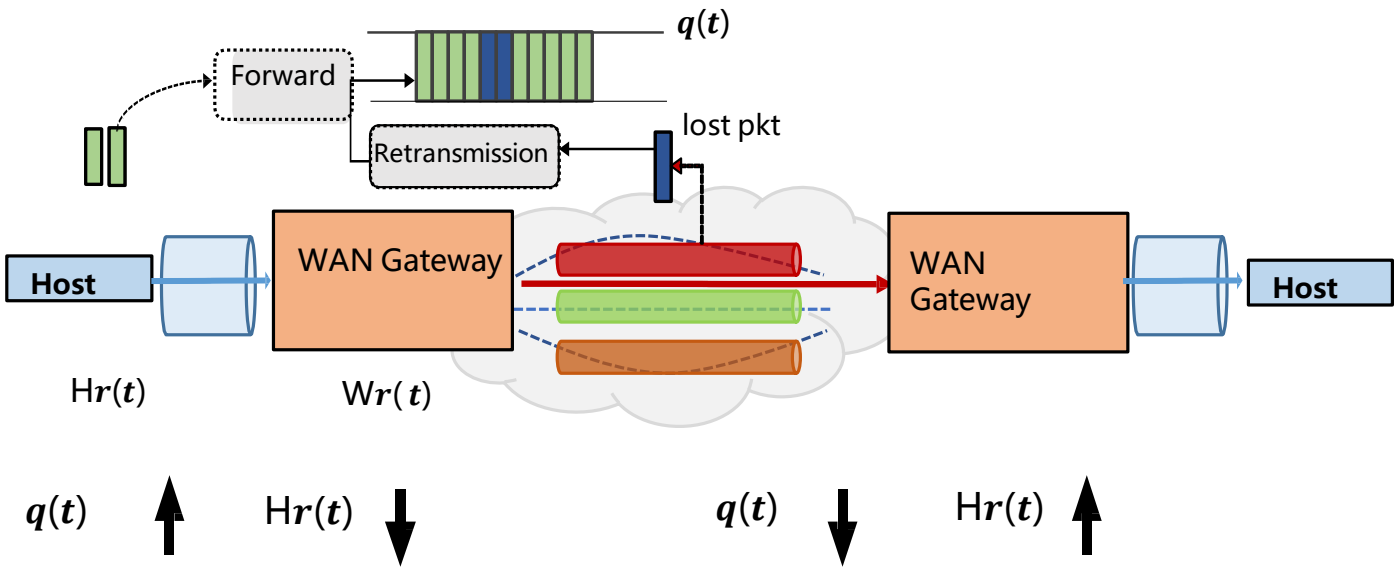
scenarios	data flow		network requirements
computing power scheduling and large-scale data transmission	1	Users upload data to the data center, utilizing a computing power network for ultra-large data transfers.	Ultra-wideband and low latency
digital production	2	Industry users collect data, employ edge computing, and execute real-time closed-loop processes at the endpoint.	Agile access and determinism
user gaming/entertainment	3	Data is distributed from the data center to the edge and then further delivered to the user.	Service awareness and determinism
large ISP business optimization	4	Different services of an application are scheduled between the edge and central data center based on real-time requirements to alleviate pressure on the edge.	Elastic, flexible, and agile interaction between edge and cloud
data transactions	5	Provide data and computing power services to data users, enabling data to be "usable but not ownable“.	Determinism, Security, and Trustworthiness

A Gateway approach



We suggest setting up special routers at each end of the WAN to help RDMA connections. These routers use fake ACK messages to cut down on wait times and choose the best congestion control methods to adjust the buffer size automatically. This plan should boost RDMA performance, make better use of the network, and allow for faster data transfer.

Gateway consideration



The speed on the host side (Hr) needs to match the speed on the gateway side (Wr). Usually, the host side is faster, while the gateway side is slower. To address this, the gateway should apply backpressure to the host side to reduce its sending rate. This prevents wasting line and gateway processing resources.

Designing Buffer Size: Considering the buffer fluctuations caused by the entire congestion control process and the delay variations due to hardware.

Optimization aspect

- **Packet Buffering**

All sent packets are not lost; they accumulate in the device's buffer and are delivered through feedback and acknowledgment.

- **Congestion Notification**

Design a congestion metric based on waterline and buffer occupancy, which is sent back to the source via messages like CNP.

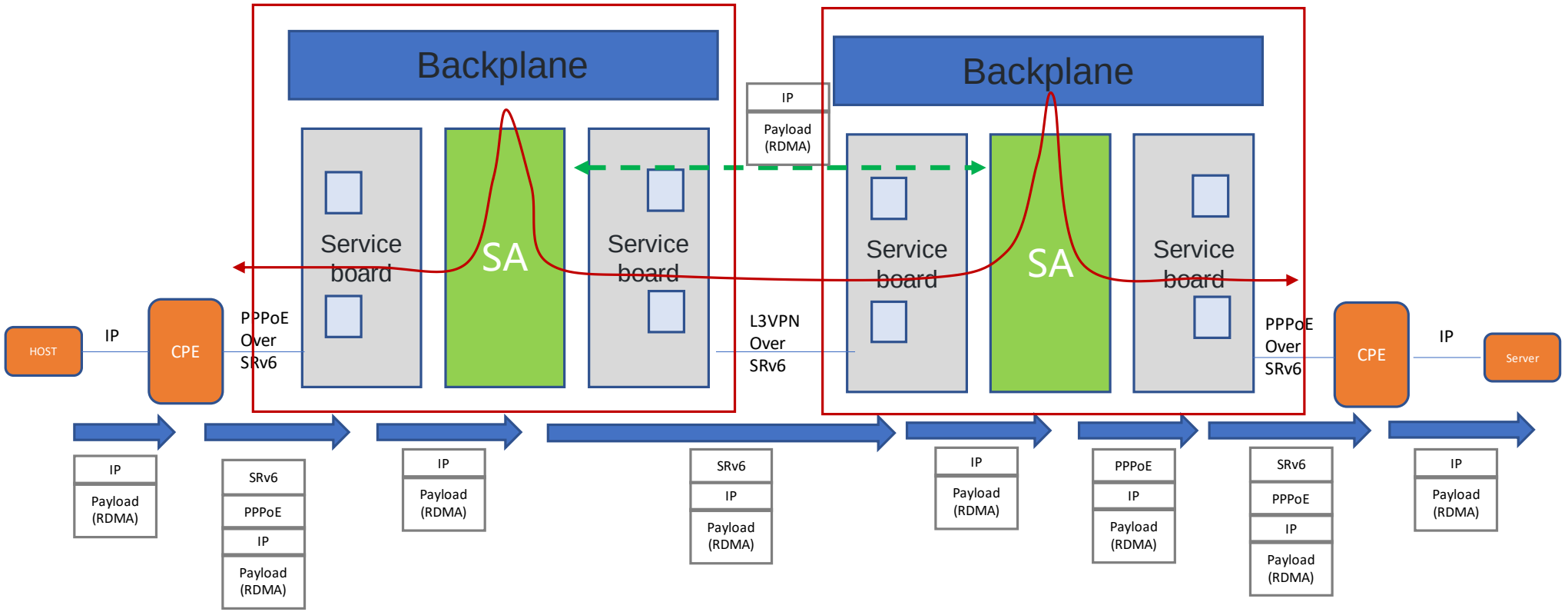
- **Selective Retransmission**

An end-to-end congestion control feedback mechanism enables the sender to adjust its transmission rate according to network conditions, such as slow start, congestion avoidance, selective retransmission, fast retransmission, and fast recovery.

- **Flow Control**

In long-distance scenarios, use mechanisms like sliding windows, ACKs, and CNP to backpressure the sender with an allowable range of data to be sent, based on RTT.

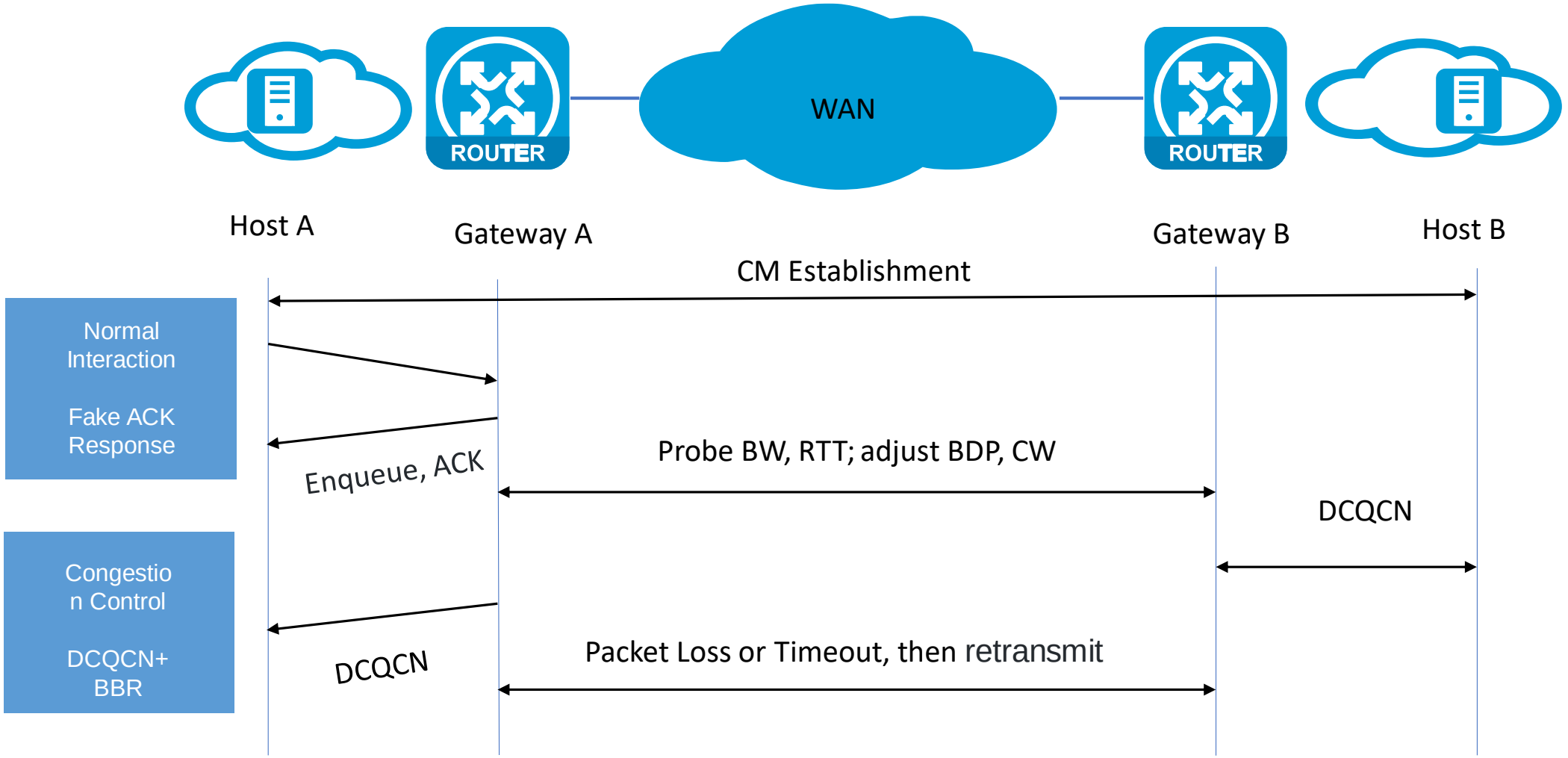
H3C Architecture



Develop a type of service card to improve the scalability of core routers. By adding new service cards into service gateway devices, we can identify applications and improve traffic Inter-DC over WAN.

We have already created a prototype using software, and in the future, we can use FPGA for implementation.

H3C Architecture



Thank you!