What is LOOPS?

Localized Optimizations over Path Segments

IETF 105, 2019-07-22

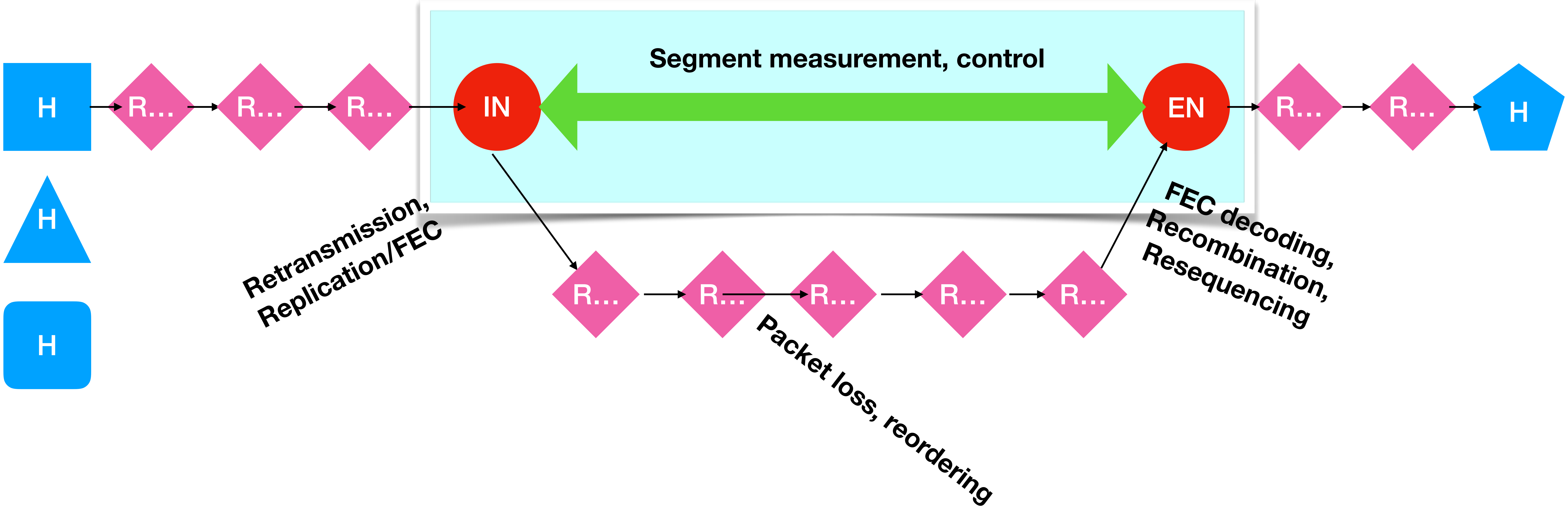
Packet loss

Packet loss

Packet loss

Tail loss!

LOOPS Opportunity



Recover Packets Locally

Reduce end-to-end packet loss

Recover locally, **where needed**, with low latency

In the **network**

Host participation not required

Don’t look
Don’t touch

Works with any kind of IP packets

How to recover?

- **Retransmission**
  - Reverse information needed: ACK/NACK ⬋ **Piggyback (Tunnel), separate Packets**
  - Forward information: sequence numbering (if needed) ⬋ **Tunnel**
- **Forward Error Correction** (redundancy)
  - Can use dynamic selection of block size/rate: measurement input
  - “Retransmission” also possible by adding FEC

- Aim for low setup overhead
- Keep most setup out of protocol (“controller model”)

How not to blow up the Internet

- Concealing losses removes important congestion signal
  - End-hosts would ramp up to higher rates, increase congestion

- Need **congestion feedback**
  - Preferred: ECN
  - Fallback: Selective dropping (selective recovery, actually)
- Host transport protocol improvements will help improve LOOPS performance, but are not prerequisite to obtaining benefit

Elements of LOOPS

- Information model for local **recovery**: in-network retransmission/FEC
    - Can be encapsulated in a variety of formats; define some of those
- Local **measurement**: e.g. segment forward delay/variation
    - To set recovery parameters
    - To determine if loss was caused by congestion
- Congestion **feedback**:
  ECN (or drops) to inform end hosts about congestion loss

Freezer (not in scope today)

- Multipath
- Measurement along string of LOOPS pairs (“almost e2e”)
- MTU handling, fragmentation, aggregation,
  header compression
- Selection of one or more specific tunnel encapsulation or measurement formats
  (beyond “sketches” showing it can be made to work)

Documents out there

- "LOOPS (Localized Optimizations on Path Segments) Problem Statement and Opportunities for Network-Assisted Performance Enhancement" <draft-li-tsvwg-loops-problem-opportunities>

Background (not discussed today, but worth looking at):

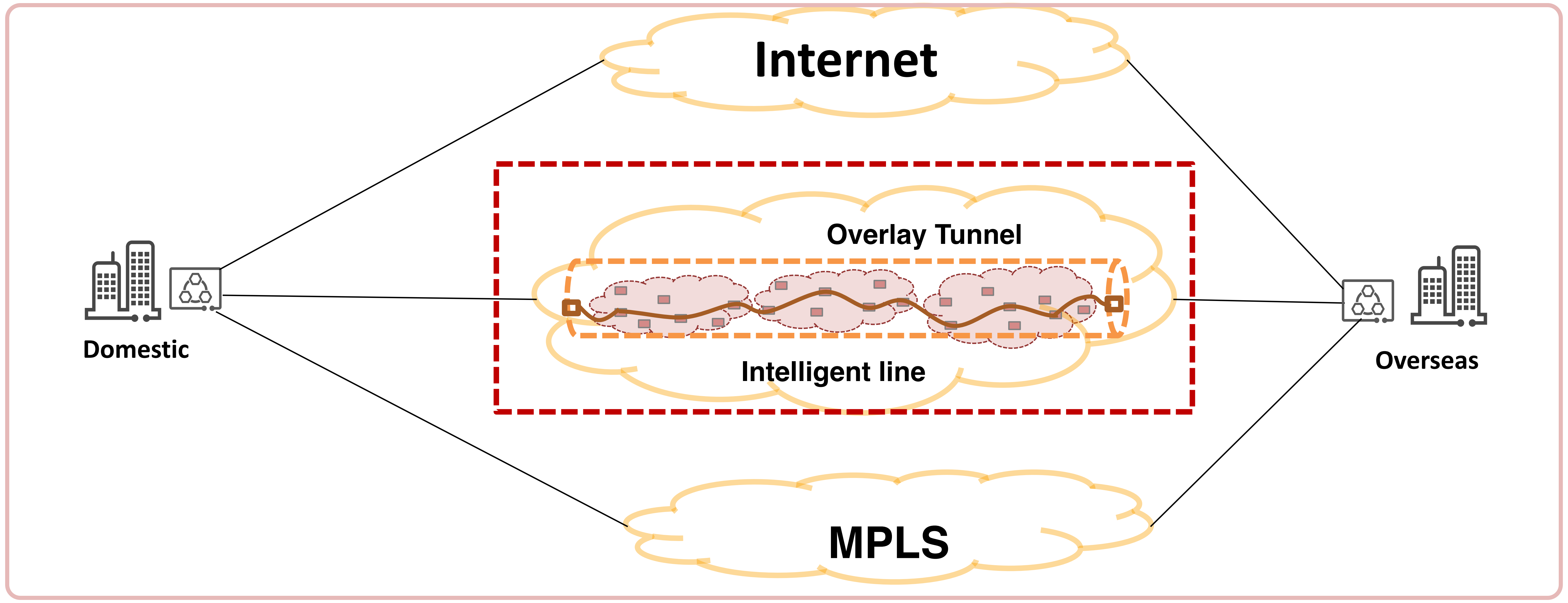
- "LOOPS Generic Information Set” <draft-welzl-loops-gen-info>
- Charter proposal for a LOOPS WG <https://github.com/loops-wg/charter>
- LOOPS mailing list loops@ietf.org

LOOPS Problem & Opportunities

draft-li-tsvwg-loops-problem-opportunities

Yizhou Li
Xingwang Zhou
Mohamed Boucadair
Jianglong Wang

Usage Scenario & Motivations



- Default path does not always give the best latency
- Cloud-Internet Overlay Network (CION): Build a better WAN path via overlay nodes in different geographic sites in multiple clouds

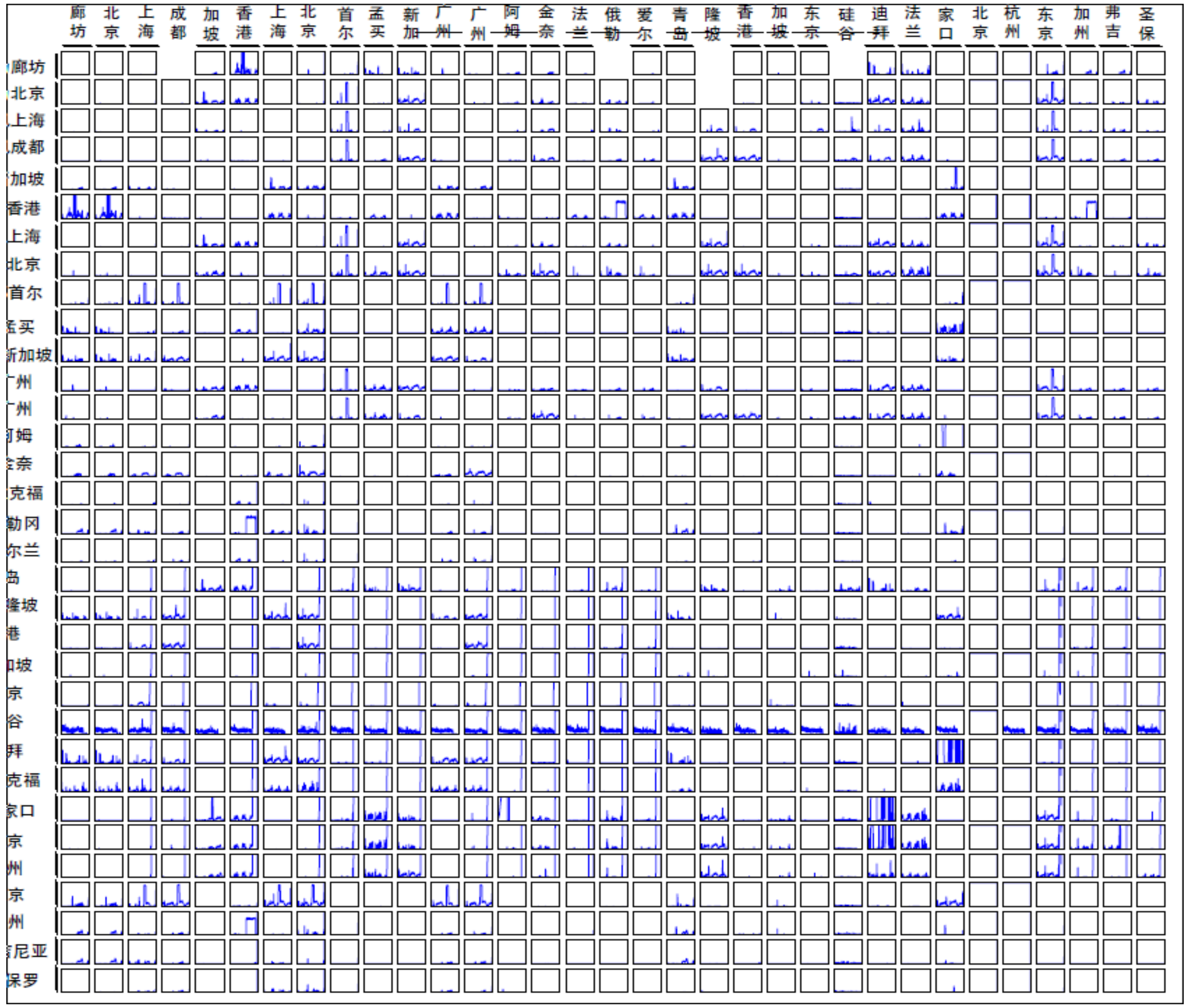
- Experiments based on 37 cloud routers globally: 71% chance of finding a better overlay path
- Problems: loss still exists in a selected path

Negative impacts of packet loss in long haul network

- Tail loss or short flows:
  - TCP retransmission may take an additional e2e RTT and kick out of slow-start
  - Need to wait for timeout in tail loss
  - Long flow completion time, especially for short flows
- Packet loss in real time streams:
  - Playout buffer grows with retransmissions
- Packet loss in large flows:
  - TCP sender reduces sending rate even when the loss is not caused by persistent congestion ➔ Throughput degradation

- In summary, e2e retransmission takes time in long haul network

Further analysis on packet loss

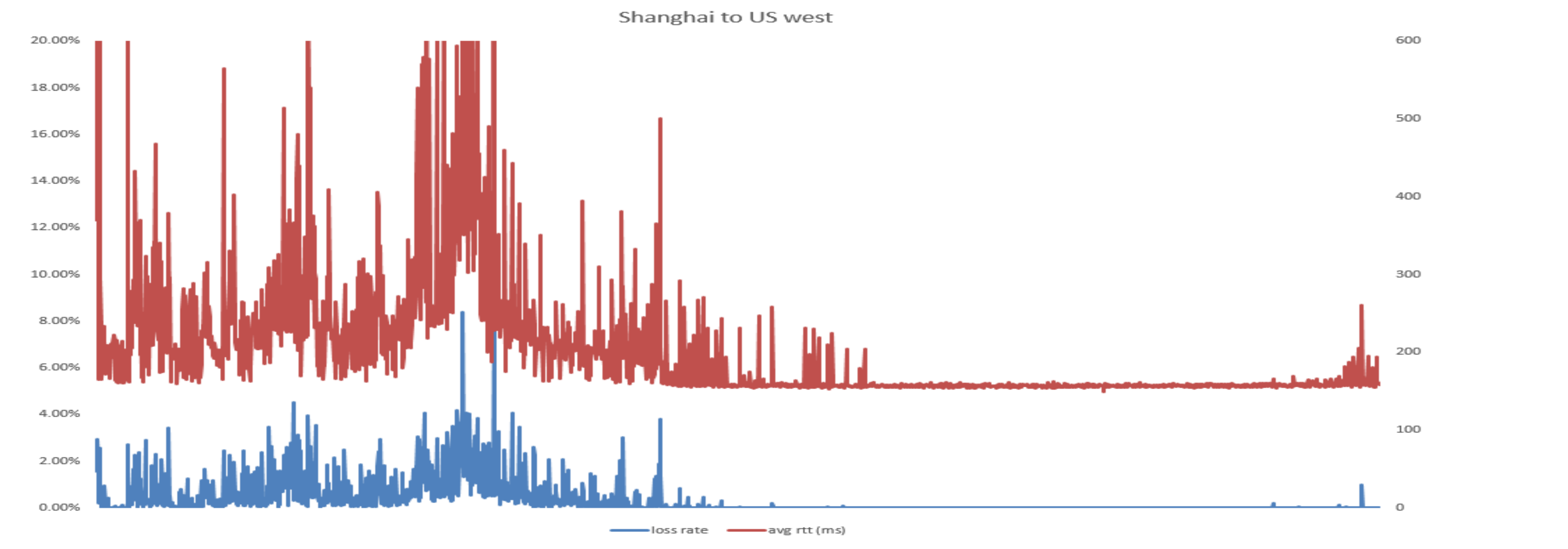


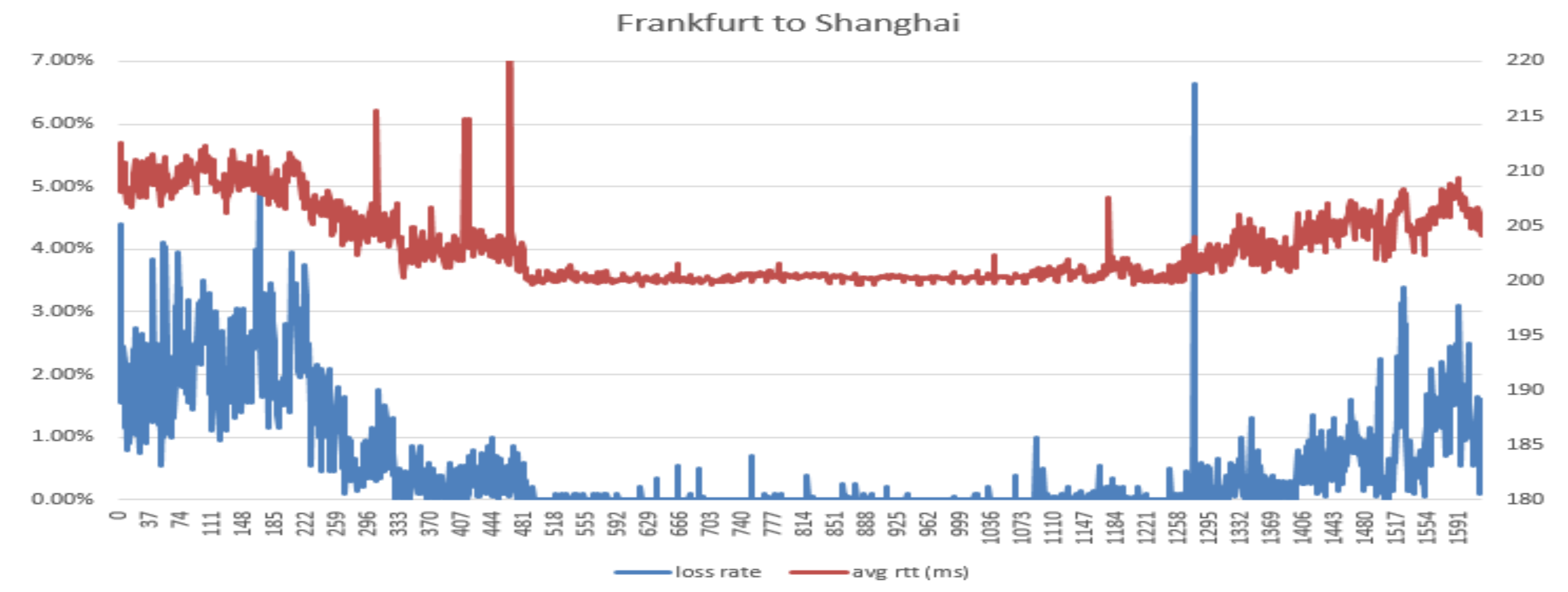
- Loss over path segments has different characteristics and may vary over time
- Loss over a specific segment may affect end to end path loss rate significantly

New Opportunities for Solving the Problems

- Overlay nodes partition the whole path into shorter segments. per-segment operation enables:
  - quicker recovery from loss
  - adaptive recovery
- Overlay nodes have computing and memory resources, capable of providing
  - loss detection & recovery
  - measurement
  - ECN marking

Tests show cause of packet loss can be deduced by measurement in some cases





- In some cases, delay and packet loss rate changes have strong correlation, and:
  - high delay means congestive loss,
  - otherwise non-congestive loss.

Summary

- Introduction of overlay nodes allow to improve handling loss over specific tunnel based path segments
- Mechanisms need to be defined to achieve local recovery while minimizing undesired side effects

- Deployment: Binding to existing overlay tunnel encapsulations, do not invent a new encapsulation

Satellite Use Cases for LOOPS

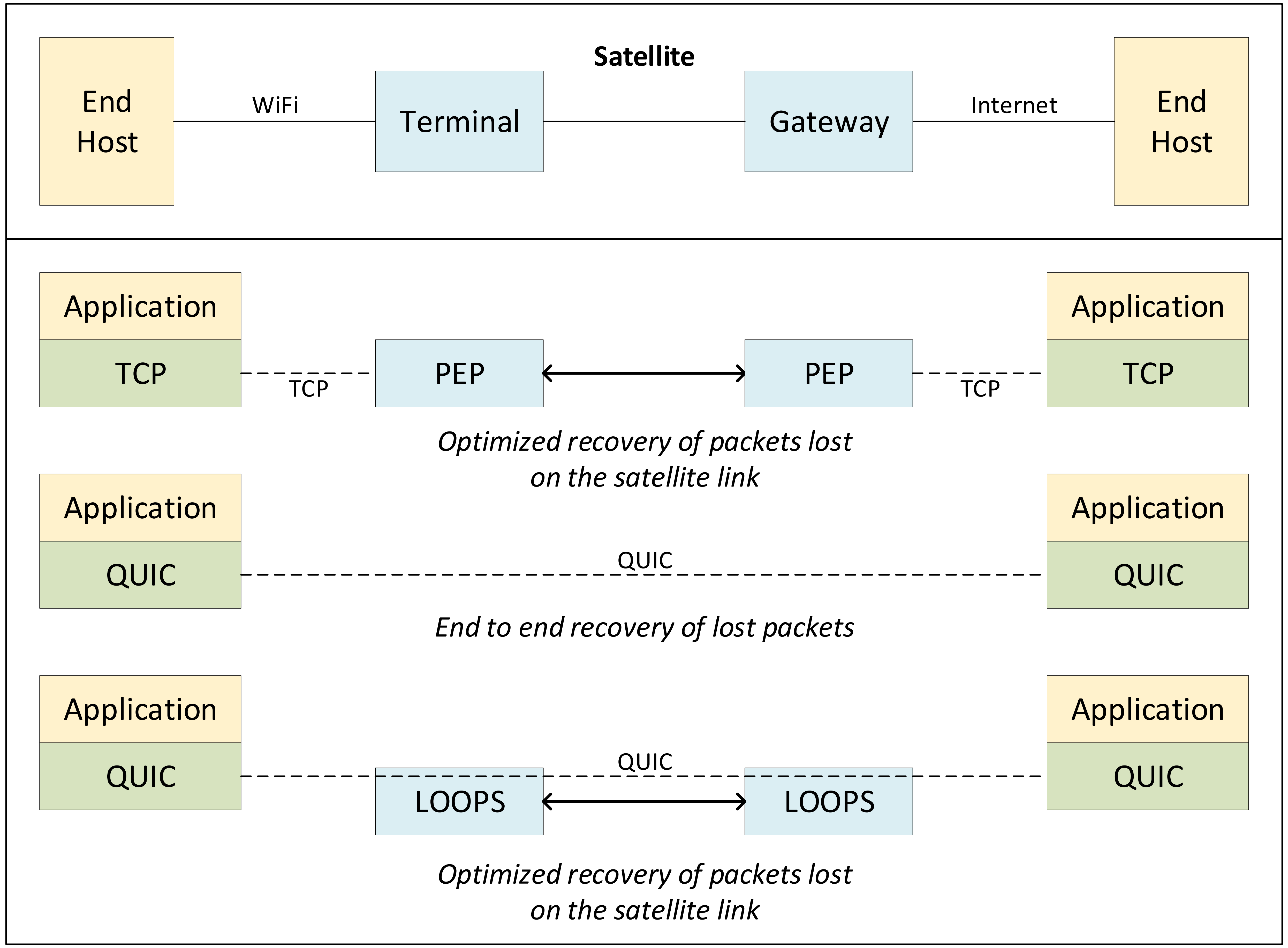
IETF 105 LOOPS

July 22, 2019

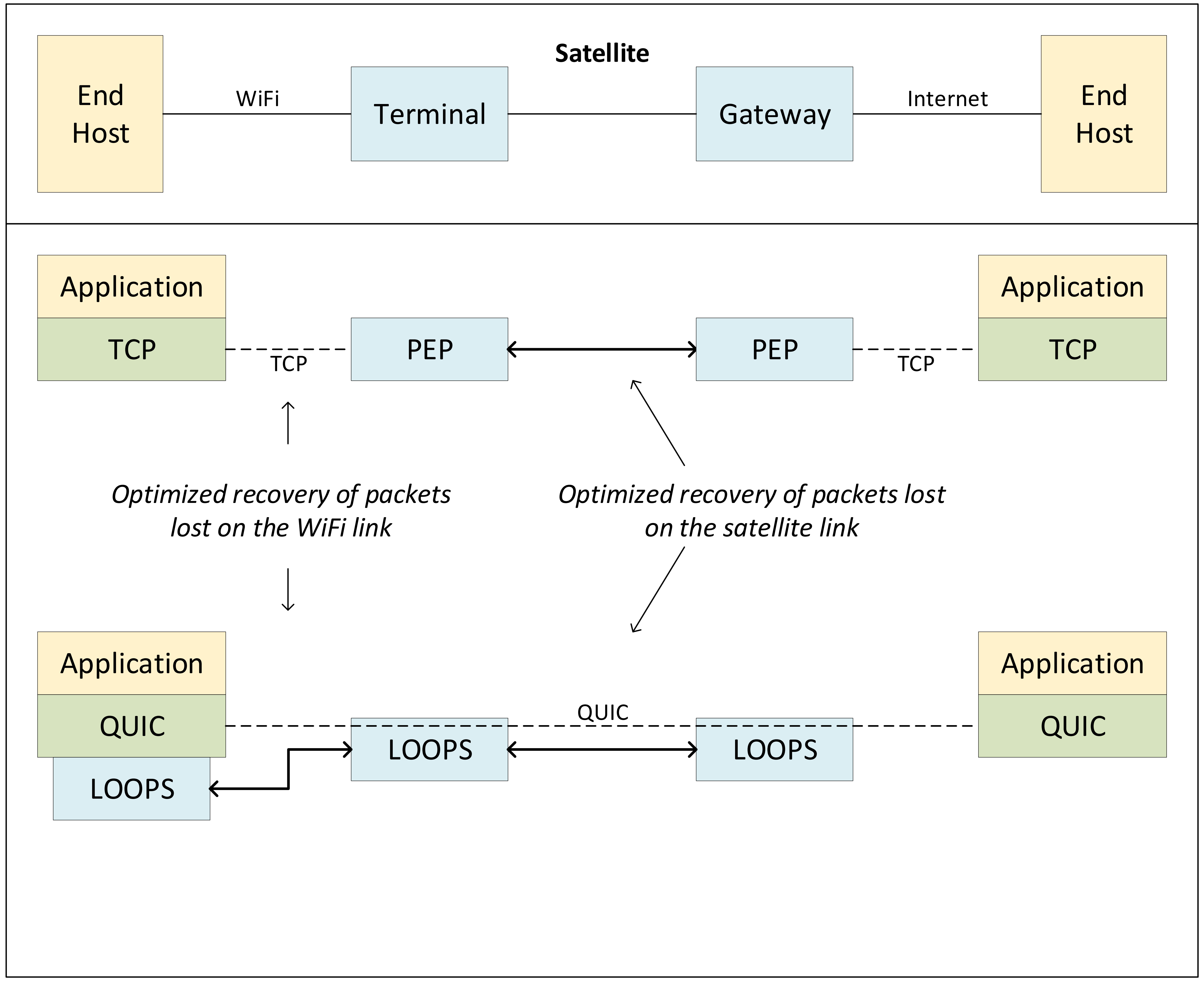
John Border (Hughes)

John.Border@Hughes.com

Satellite Use Case – Near Term



Satellite Use Case – Longer Term



SRv6-based Enterprise WAN Connection Use Case for LOOPS

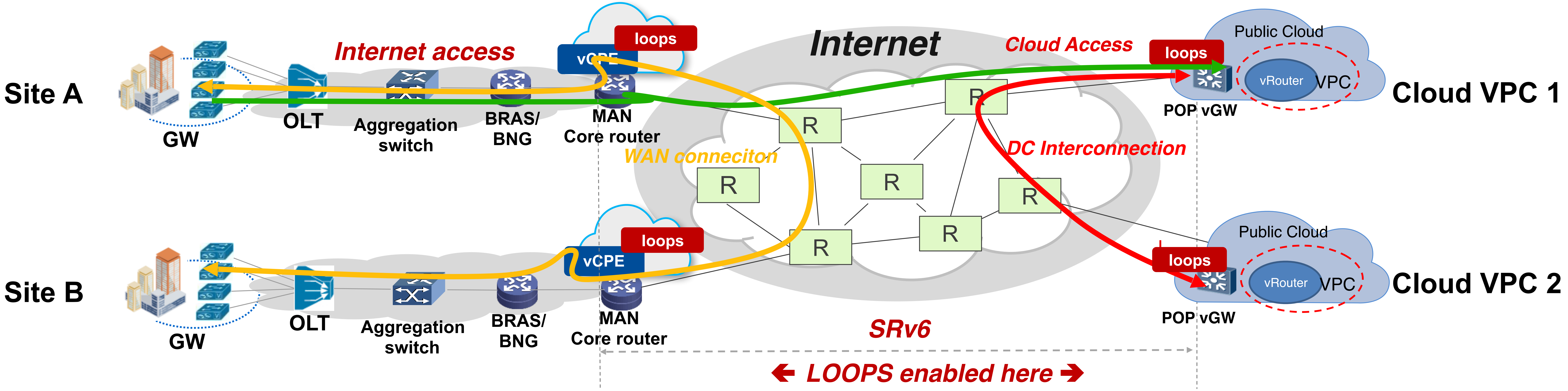
Jianglong Wang, Bo Lei, Cong Li（China Telecom）

Montreal@IETF105
2019.07

Background

- ChinaNet, the largest public INTERNET backbone network in China, covering 300+ MANs. It mainly provides broadband Internet and E-Cloud access. The entire network devices are IPv6-enabled, we plan to upgrade to support SRv6 to provide the path selection to meet SLA.
- Enterprises usually require network connections between the branch offices or between branch offices and cloud data center over geographic distances.
  - Enterprise branch WAN connection via Internet
  - Enterprise Cloud Access
- The traffic on path segments over Internet is subject to loss due to the best effort networks, relying on the endpoint for packet loss recovery. There is an urgent need here to do network optimization to provide high-quality Internet for specific service that needed the reliability of data transfer (e.g. video conference application in SDWAN over Internet).

Use case



- The enterprise accesses the Internet backbone network through vCPE (NFV virtual node) for WAN connection, and vCPE connects to a cloud based PoP which further directs traffic to vPC for Cloud access.
- vCPE to vCPE (●), vCPE to PoP (●), POP to POP (●) can be over a long distance, which could be divided into multiple sub-path based on SRv6 segments, some of which have packet loss.
- LOOPS can be enabled for segments of this sub-path to solve packet loss and provide data for path selection.
- Deploying SRv6 with LOOPS, we can provide high-quality Internet connection in terms of loss rate.

Next Step

- LOOPS can be applied to Enterprise WAN connection scenarios
- SRv6 could be a specific encapsulation protocol for LOOPS;
  LOOPS + SRv6 could be considered in the following work
- Welcome more comments and discussion if you are interested in this topic

LOOPS Generic Information Set

draft-welzl-loops-gen-info-00
LOOPS BoF
IETF 105 - Montreal

Michael Welzl, U. Oslo
Carsten Bormann, ed., U. Bremen TZI

Why look at this draft now?

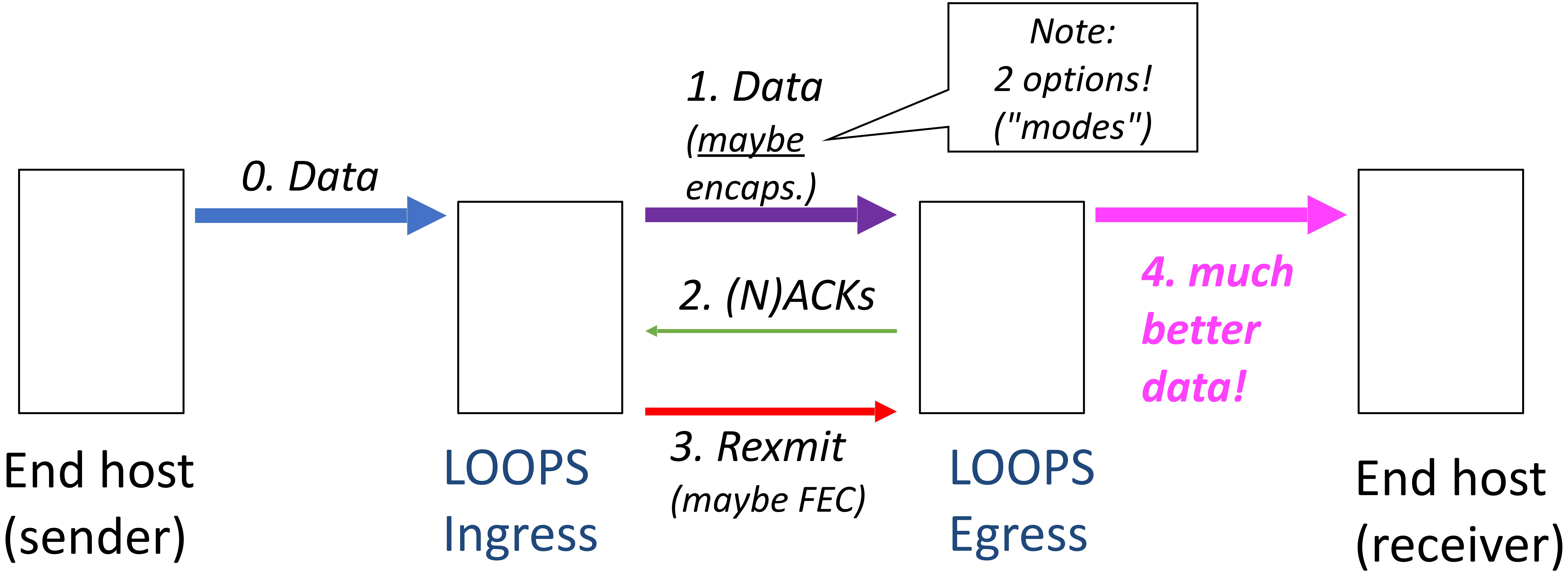
- "The present document is a strawman for the set of information that would be interchanged in a LOOPS protocol, without already defining a specific data packet format."

→ an overview of how a LOOPS protocol could work

… as an existence proof, and to aid visualization

We are not picking alternatives today.

Context



0. *Data*

1. *Data* (*maybe* *encaps.*)

Note:
2 options!
("modes")

2. *(N)ACKs*

3. *Rexmit* *(maybe FEC)*

4. much better data!

Problems to address

- From previous slide:
  - Loss detection/retransmission
  - FEC control

- Also: detect congestion on ingress-egress segment
  - Measurement / congestion detection
  - Congestion signaling to end hosts if congestion was detected

- Next: some concrete ideas on how to deal with these problems
  - Just a strawman (two, actually)!

Tunnel mode

1. Ingress forwards

- Encapsulate: not tied to any specific overlay protocol
  - Document contains sketches of bindings to GUE and Geneve
- We don't try to understand data after the IP header
  - Hence, need to add a Packet Sequence Number (PSN)
- Some more information added
  - Tunnel type
  - "ACK desirable" flag (asks for feedback block 1, next slide)
  - Anything needed by FEC

2. Egress answers

- Block 1 (optional, only upon "ACK desirable")
  - PSN being acknowledged
  - Absolute time of reception for the packet acked (PSN)

- Block 2 (optional)
  - an ACK bitmap (based on PSN)
  - a delta indicating the end PSN of the bitmap

- Can be interspersed and repeated
- Can be piggybacked on a reverse direction data packet or sent separately
- Usually aggregated in some useful form

3. Ingress retransmits

- Detects need for rexmit via NACK or RTO
  - Make decision based on congestion

  → Use ECN if possible

  → Calculate latency variation from timestamps in feedback blocks 1

- ... Or, rather than "just rexmit", send FEC repair packets

4. Egress forwards

- De-FEC

- Inform end hosts about congestion <u>if needed</u>
  - Might be able to distinguish "real" congestion from, e.g., corruption loss
  - ECN much preferred as a signal!

Summary: information exchanged

- Forward: encapsulation, containing...
  - Packet Sequence Number (PSN)
  - Tunnel type
  - "ACK desirable" flag (asks for feedback block 1, next slide)
  - Anything needed by FEC
- Backward: optional blocks type 1 and 2...
  (can be piggybacked, aggregated, interspersed, repeated, ...)
  - Block 1 (optional, only upon "ACK desirable")
    - PSN being acknowledged
    - Absolute time of reception for the packet acked (PSN)
  - Block 2 (optional)
    - an ACK bitmap (based on PSN)
    - a delta indicating the end PSN of the bitmap

Transparent mode

A bit more radical... describing least intrusive method here:

"Never delay and don't even tunnel"

Just discussing rexmit; FEC could also be done

1. Ingress forwards

- Just forward
- Also, keep a copy of packets, with a hash for identification
  - From immutable header fields
  - May need to include data beyond IP

2. Egress answers

- ACK everything; no NACK possible
  - Same hash calculation as ingress
- ACK format similar to tunnel mode

3. Ingress retransmits

- Detects need for rexmit via RTO only
  - Decide based on congestion estimation as before

- Cost of hash collisions is low: misses retransmit opportunity.

4. Egress forwards

- That's all it does. There will be re-ordering.

Summary: information exchanged

- Forward: nothing extra

- Backward: roughly as before - optional blocks type 1...
  (can be piggybacked, aggregated, interspersed, repeated, ...)
  - Block 1
    (*limited in some way: was optional, only upon "ACK desirable" for tunnel mode, but egress doesn't get this information in transparent mode*)
    - PSN being acknowledged
    - Absolute time of reception for the packet acked (hash)
  - Block 2 (hmm)
    - an ACK Bloom filter?

Conclusion

- Spectrum of possibilities, from "full-fledged" to min-intrusive
  - Various trade-offs between these options
- In all cases: LOOPS can be very beneficial when the LOOPS segment RTT is much shorter than the e2e RTT
  - Wireless NICs use this fact
- Some packet drops really cause pain
  - LOOPS can help

Tail loss!

Clarifying questions?

(Don't forget to think "strawman".)

(F)AQ (1)

- So this is only about encrypted traffic?
  - Any traffic is welcome, we just don’t try to peek beyond L3 info
- So how do you know which packets are worth recovering?
  - Today we don’t. If more L3 marking becomes available, we’d use it.
- How do you transport your measurement-related information?
  - Forward info: In encapsulation extension (e.g., with sequence number). Reverse info: The same way we transport the ACK channel. Depends on encapsulation.

(F)AQ (2)

- How do you avoid spending more for LOOPS encapsulation than the performance enhancement is worth?
  - LOOPS will need some management that is weighing this (and doing the pair setup in the first place)
  - Can dynamically switch off and on (e.g., based on monitoring)
- How to relay congestion for non-ECN-capable transports?
  - Dropping. Or, actually, not even requesting a retransmission when a congestion event would be relayed anyway.

Related work in the IETF

- Note: Lots of related work outside the IETF, e.g., see Bob Briscoe's mail <https://mailarchive.ietf.org/arch/msg/loops/lsVH1PKFnfjs06MilrVH_iRyhao>; link layers, …
- Inside IETF/IRTF:
  - Measurement work: IOAM in-band, other IPPM active (TWAMP/OWAMP/STAMP), IPFIX serialization/transfer of measurements
  - Recovery: 6lo Fragment Recovery, e.g., between 6LN and 6LBR
  - NWCRG/TSVWG work on sliding window FEC
  - TCF (Tunnel Congestion Feedback)

IOAM and related IPPM work

- IOAM (In-Situ OAM) is used to collect “operational and telemetry information in the packet while the packet traverses a path between two points in the network”.
    - Multi-hop collection of traces (node data lists)
    - Might return measurement to third party
    - LOOPS looks much more like a classical transport protocol
- IOAM uses a “generic information model” approach, from which we can learn.
- Data formats and measurement methods from IPPM may be applicable.

6LoWPAN Fragment Recovery

- draft-ietf-6lo-fragment-recovery (WGLC passed in 6lo WG)
- 6LoWPAN has adaptation-layer fragmentation (~ 80 byte fragments)
- Adaptation Layer Fragments can be forwarded in a 6LoWPAN (draft-ietf-lwig-6lowpan-virtual-reassembly); packet loss multiplies…
- Fragment recovery is between fragmenter and reassembler
- Pacing (Inter-Frame Gap); congestion control is limited to radio mesh but mostly left as an exercise to the reader, as is congestion feedback

Sliding Window FEC

- Sliding windows fit quite well to LOOPS application
  (Can also use traditional block formats)

- Various drafts for FEC scheme and specific embeddings in NWCRG and TSVWG, e.g.,

  - "Sliding Window Random Linear Code (RLC) Forward Erasure Correction (FEC) Schemes for FECFRAME" <draft-ietf-tsvwg-rlc-fec-scheme-16.txt>

  - "Forward Error Correction (FEC) Framework Extension to Sliding Window Codes" <draft-ietf-tsvwg-fecframe-ext-08.txt>

Tunnel Congestion Feedback

- draft-ietf-tsvwg-tunnel-congestion-feedback-07:
  - feeding back inner tunnel congestion level,
  - from egress to ingress
- Using IPFIX as a transfer layer, defines IPFIX Information Elements
- Could be used as a generic information model/protocol for LOOPS (negotiation part to be handled by controller model)

Work that is out of scope

- Assumes host participation:
    - PANRG
    - Various proposals to improve QUIC or measure QUIC
    - Packet Loss Signaling for Encrypted Protocols
- For now:
    - Spin Bit (but could provide great market differentiation potential)

Technical Discussion