Low-Latency Communications and the Internet Architecture

draft-arkko-arch-low-latency-01.txt

Jari Arkko and Jeff Tantsura

IETF-99

Goals

- Trying to understand what the recent interest on low-latency / critical communications applications implies for the Internet

- New work at the IETF? Architectural evolution?

- Work in progress; please contribute!

Recent General Requirements for Low-Latency Communications

- Tactile Internet requires 1 ms reaction time
- Self-driving cars require 1ms latency
- Mission-critical 5G MTC requires low latency & high reliability and availability

Plenty of Wild Claims in This Space

We do not expect changes in the dynamics and economics of Internet evolution

But the World Does Care about Low-Latency

- Data centers distributed around the globe
- Including content served from operator premises
- Advanced optimisation techniques for connecting to data centers (DNS etc)
- Industry working HTTP2, QUIC, TLS.1 (0-RTT), L4S, DETNET, 802.1 TSN, 5G radios, …
- SDN and SFC replacing long chains of processing functions
- Industry working on ServiceWorker, AMP, …

Lets Recap To Be Clear

- Latency in L2 is being improved
- Latency in routing/forwarding is being improved
- Latency in transport is being improved
- Latency in security is being improved
- Latency in application protocols is being improved
- Network deployments are changing to take into account latency

And it is all part of our regular program anyway

All Done? Or Work Ahead?

- Not necessarily the big revolution some might claim; a lot of the tools are there
- Obviously much of this is work in progress
- But, more importantly, the Internet is changing and this may cause strain for the architecture

Architectural Thoughts 1

- Need to consider the system whenever thinking about this topic

- Trend of service placement in different locations: from global datacenters to more regional ones, co-operative solutions, edge computing

- Impacts on architectures that employ tunnelling

- There are and will be demands on cross-layer optimisation, is that a good thing?

Architectural Thoughts 2

- Tension between local networks (e.g., cars braking and informing nearby cars) and Internet-networking
- Designing applications entirely in their own silo vs. applications that also talk to peers in the Internet
- Tension between application/edge and network control of forwarding decisions (MPTCP vs. routing)
- Deployment story for new QoS or low-latency tech

Thank you

Questions and contributions welcome!