Reducing Internet Latency: a survey of techniques and their merits

Bob Briscoe, Anna Brunstrom, Andreas Petlund, David Hayes, David Ros, Ing-Jyh Tsang, Stein Gjessing, Gorry Fairhurst, Carsten Griwodz, Michael Welzl
summary

• industry roadmap of techniques
• gain vs pain
  – latency reduction against deployability

• “A Survey of Latency Reducing Techniques and their Merits”
  – 322 references
  – available via http://riteproject.eu/publications/

• evolved from BT roadmap work, but repurposed
  – a company tries to prioritise the quick wins
  – an industry also needs to identify hard problems being avoided
roadmap around body of survey I

Sources of delay and techniques for reducing latency

Structural delays § II
  - Sub-optimal routes/paths § II-A
  - Name resolution § II-B
  - Content placement § II-C
  - Network proxies and caches § II-C1
  - Client caches § II-C2
  - Prediction and latency-hiding § II-C3
  - Structured peer-to-peer § II-D1
  - Cloud server placement § II-D2
  - Cloud cache placement § II-D3
  - Virtualizing chains of network functions § II-D4

Interaction between endpoints § III
  - Transport Initialization § III-A
  - Secure session initialization § III-B
  - Building encryption into TCP § III-B2
  - Fast opening of TCP connections § III-A3
  - Application pipelining § III-A4
  - Path MTU discovery § III-A5
  - Faster transport security negotiation § III-B1
  - Bootstrapping security from the DNS § III-B3
  - Application tolerance to loss § III-C1
  - Reduce packet loss detection times § III-C2
  - Combining redundancy and retransmission § III-C3
  - Explicit congestion notification § III-C4

Packet loss recovery delays § III-C

Signal propagation delay § IV-A
  - Straighter cable paths § IV-A1
  - Higher signal velocity § IV-A2

Medium acquisition delays § IV-B
  - Higher velocity with straighter routes § IV-A3

Serialization delay § IV-C
Fig. 1. Techniques for reducing latency organized by sources of delay.
case (1a): small (20kB) flow over WAN

- QoS
- microwave
- TFO
- AQM
- DNS pre-fetch
- RTO-restart
- IW10
- sender only

For example...
- all at once
- both ends & network
- both ends
- network only
- network only

Reduction in completion time:
- 0%
- 50%
- 100%
case (1b): small (20kB) flow over LAN

for example...

- QS
- hollow fibre
- Straighter links
- CDN
- AQM
- RTO-restart
- TFO
- IW10
- DNS pre-fetch
case (2a): large flow over WAN

- Data pre-fetch
- CDN

For example... all at once, both ends & network, network only.
case (2b): large flow over LAN

Data pre-fetch

For example...
- all at once
- both ends & network
- network only
- sender only