Bidirectional Cross-layer QoE Optimization

Srikanth Sundaresan
International Computer Science Institute
Berkeley CA

M. Zubair Shafiq
The University of Iowa
Iowa City IA

Abstract
Modern day applications are complex and have diverse Quality of Experience (QoE) requirements. There is also a growing need to maintain tight control over application performance and minimize QoE impairments. Unfortunately, QoE impairments can be caused by a variety of reasons, including unexpected interactions between various layers, and a lack of direct communication between network elements. In this memo, we argue that applications need to interact more closely with the underlying network and network elements to minimize QoE impairments. We propose a bidirectional cross-layer information exchange mechanism to help both applications and the network to to work in harmony and deliver better QoE to end-users.

The case for bidirectional, cross-layer communication
Applications today are very complex and diverse, and consequently they have a wide range of requirements from the underlying network for optimum performance. It is not possible (beyond rudimentary mechanisms) in the modern Internet for applications to communicate with lower layers of the stack (and vice-versa) to influence performance, and therefore applications have tended to handle performance in the application layer itself. This also profoundly limits the Quality of Experience (QoE) that applications are able to deliver to end-users. This is because QoE impairments can be caused by a wide range of reasons including but not limited to *(a)* unexpected interactions between network protocols operating at different layers and *(b)* lack of cooperation between end-points (e.g., client and server) and network elements (e.g., routers, middleboxes). While there are good reasons for the current protocol stack design which makes it difficult for different layers to communicate, we argue that the increasing complexity of modern applications means that it is harder to optimize their performance purely from the application layer, or from the individual stack elements.

We present a case for a bidirectional, cross-layer information exchange mechanism for QoE optimization. There do exist some current techniques and proposals for cross-layer communication. For example, there have been proposals to pass link layer estimation of access link capacity in mobile networks to the transport layer to adapt its sending rate. We advocate a more general bidirectional information exchange mechanism between end-points and network elements. To extend the previous example, not only should base stations in a cellular network be able to inform end-points about imminent congestion at the last mile so that the transport layer can adapt sending rates, the end-points should also be able to pass information to base stations (and other intermediate network elements) about application priority and status so that

they can adaptively allocate resources for better global application performance. We expand on these ideas below and also discuss existing efforts that share similar goals.
A large body of prior work has focused on cross-layer optimizations to improve TCP performance. ECN [4], a network and transport layer construct that is used to control transport layer sending rates, is the most common example. Intermediate routers that support ECN set a flag indicating that a link is congested, which leads to the sender side halving its congestion window. Middleboxes today implement many types of performance enhancing proxies, but their utility is increasingly limited by the trend towards encrypted traffic. ECN has seen large scale real world deployment and is effective, but is at best rudimentary. Other work has proposed that lower-layers feed information back to TCP, particularly in wireless networks [1,2]. Such information exchange, even though it goes against the building principle of the Internet, is important today. Indeed, we argue that applications and protocols have become so complex that these efforts do not go far enough. A canonical example supporting our claim is the well understood difficulty that TCP has in differentiating random wireless loss from congestion loss. Recent work has also analyzed the complex and unintended interactions between TCP congestion control at the transport layer and video bitrate adaption at the application layer [3].

In our body of work, we plan to expand the scope of cross-layer interaction to not just network and link-layer information up to the transport layer, but also to to the application layer, and also in the *reverse* direction; from the application down to the network. We believe that this approach will improve applications' ability to control the QoE that they are able to offer significantly.

We illustrate the power of bidirectional cross-layer interaction with a simple, but hypothetical example for QoE optimization. Suppose that the application, a variant of DASH (Dynamic Adaptive Streaming over HTTP), is able to get information from the network, be it a cellular base-station, or a wireless access point, about the state of the link, it can drastically improve the feedback control delay (which currently happens in the application layer between the client and the server based on buffer status and chunk arrival rate). The server can use this information to decide on the best bitrate (at the application layer) and sending rate (at the transport layer). Now, suppose that the server can also tell the network about its own state: if the server could request the base station to prioritize sending its packets if the client buffer is running low, or deprioritize the next few chunks if the buffer is full, it would ideally result in a network that would perform better for all parties concerned. Such an approach obviously raises questions about implementation and incentives, which we hope to tackle as part of our ongoing work.

Takeaway
Recent developments in the Internet ecosystem such as (1) the rise of video as the dominant source of Internet traffic, (2) the emergence of a few large video content providers (Netflix, YouTube), (3) the ever increasing traffic pressure on spectrum-constrained mobile networks and peering interconnection points, and (4) the widespread deployment of encryption (HTTPS, HTTP/2) have led application developers, content providers and network operators to look for innovative, generalized solutions to address the challenge of providing good QoE to end-users. In this memo, we present a case for a bidirectional, cross-layer information exchange

mechanism. We argue that bottom-up *and* top-down cross-layer information exchange can result in tangible performance benefit due to applications (and network elements) having access to more relevant information quicker. Our approach goes against the basic tenet of layered protocol design, but the growing complexity and requirements of modern applications mean that this or similar approaches are inevitable.

References

1. Jiantao Wang, Lun Li, Steven H. Low, John C. Doyle, “Cross-layer optimization in TCP/IP networks”, IEEE/ACM Transactions on Networking, 13(3), pp. 582-595, 2005.

2. Feng Lu, Hao Du, Ankur Jain, Geoffrey M. Voelker, Alex C. Snoeren, Andreas Terzis, “CQIC: Revisiting Cross-Layer Congestion Control for Cellular Networks”, ACM HotMobile, 2015.

3. Te-Yuan Huang, Nikhil Handigol, Brandon Heller, Nick McKeown, Ramesh Johari, “Confused, Timid, and Unstable: Picking a Video Streaming Rate is Hard”, ACM Internet Measurement Conference (IMC), 2012.

4. K.K. Ramakrishnan, S. Floyd, D. Black, “The Addition of Explicit Congestion Notification (ECN) to IP”, RFC 3168, September 2001.