Journal of Network and Computer Applications 75 (2016) 236–258

Contents lists available at ScienceDirect

Journal of Network and Computer Applications

journal homepage: www.elsevier.com/locate/jnca

Review

The past, present, and future of transport-layer multipath

Sana Habib [a,*], Junaid Qadir [b], Anwaar Ali [b], Durdana Habib [c], Ming Li [d], Arjuna Sathiaseelan [e]

[a] Electrical Engineering Department of School of Electrical Engineering and Computer Sciences (SEECS), National University of Sciences and Technology (NUST), Islamabad, Pakistan
[b] Information Technology University-Punjab, Lahore, Pakistan
[c] National University of Computer and Emerging Sciences, Islamabad, Pakistan
[d] Department of Computer Science, Aalto University, Finland
[e] Computer Laboratory, University of Cambridge, UK

A R T I C L E   I N F O

Article history:
Received 27 May 2016
Received in revised form
13 August 2016
Accepted 9 September 2016
Available online 10 September 2016

Keywords:
Multipath
Flow control
Congestion control
Flow scheduling
Fairness
Stability
Responsiveness

A B S T R A C T

Multipathing in communication networks is gaining momentum due to its attractive features of increased reliability, throughput, fault tolerance, and load balancing capabilities. In particular, wireless environments and datacenters are envisioned to become largely dependent on the power of multipathing for seamless handovers, virtual machine (VM) migration and in general, pooling less proficient resources together for achieving overall high proficiency. The transport layer, with its knowledge about end-to-end path characteristics, is well placed to enhance performance through better utilization of multiple paths. Realizing the importance of transport-layer multipath, this paper investigates the modernization of traditional connection establishment, flow control, sequence number splitting, acknowledgement, and flow scheduling mechanisms for use with multiple paths.

Multipath rate control defines the fundamental feature of the transport layer. Thus, we study the restructuring of classical utility maximization framework and analyze its stability and convergence. We also examine the topic of multipath congestion control in the light of TCP fairness. To the best of our knowledge, this is the first in-depth survey paper that has chronicled the evolution of the transport layer of the Internet from the traditional single-path TCP to the recent development of the modern multipath TCP (MPTCP) protocol. Along with describing the history of this evolution, we also highlight in this paper the remaining challenges and research issues.

Contents

* Corresponding author.
E-mail addresses: 14mseeshabib@seecs.edu.pk (S. Habib), junaid.qadir@itu.edu.pk (J. Qadir), anwaar.ali@itu.edu.pk (A. Ali), durdana.habib@nu.edu.pk (D. Habib), ming.li@aalto.fi (M. Li), arjuna.sathiaseelan@cl.cam.ac.uk (A. Sathiaseelan).

http://dx.doi.org/10.1016/j.jnca.2016.09.005
1084-8045/

1. Introduction

"The best way to predict the future is to create it."— Abraham Lincoln

Traditionally, transport layer uses a single path for end-to-end communication between applications. However, single-path transport protocols are not able to keep up with the growing bandwidth as well as reliability and fault tolerance demands of multimedia applications and critical businesses (such as communication between government agencies and e-commerce). This deficiency of the single-path protocols has led to the increasing trend of multipathing in the Internet.

Multipathing elegantly solves the deficiencies of single-path protocols by employing *inverse multiplexing* of resources to send packets over a set of paths rather than a single-path. This striping of data across multiple paths provides better reliability, fault tolerance, and increased throughput (Ye et al., 2008; Kim and Shin, 2005; Paasch et al., 2013). Thus, applications can reap the benefits of resource pooling and diversity provided by multiple paths to achieve desired quality of service (QoS) as well as improved user's quality of experience (QoE). Instead of discussing multipathing at different protocol stack layers in terms of performance enhancements (Qadir et al., 2015; Singh et al., 2015; Li et al., 2016); this paper focuses on the use of multiple paths at the transport layer.

Multipath transport layer is well suited to devices equipped with heterogeneous access technologies (such as a mobile device equipped with both Wi-Fi and 3G) (Paasch et al., 2012). Use of multiple paths by mobile devices not only improves reliability by providing the ability to seamlessly shift from one technology to another (that can act as the backup interface in the case of outage or congestion) but users can also benefit from cheap prices (e.g., when both 3G and Wi-Fi are available, the latter can be used being less expensive). It can also become possible for mobile nodes in close proximity to pool their low bandwidths in order to support high bandwidth applications (Kim and Shin, 2005). Recent implementation of MPTCP by Apple iOS7 (https://support.apple.com/en-us/HT201373) has further encouraged the multipath trend at the transport layer.

Datacenters in particular represent an important use case of transport-layer multipathing. Also, multipath transport layer enables many topologies in datacenter that could not be realized with single-path TCP. For example, GRIN (Agache and Raiciu, 2012) uses MPTCP to efficiently utilize datacenter network by making minor topology changes. Raiciu et al. (2011a) showed that Amazon EC2 achieves three times throughput in comparison to single-path by exploiting path diversity.

To further emphasize on the importance of multipath at transport layer, we next describe some of the benefits in detail.

1.1. Merits of multipathing

Four major benefits obtained by using multiple paths are mentioned below.

(1) *Load balancing*: Load balancing is traditionally performed at the network layer, wherein a network operator can reroute traffic in order to avoid congested hotspots. However, traffic engineering at the network layer can be unstable and may lead to oscillations (aBT Innovate, 2009). Let us elaborate this argument with an example. Consider that a source–destination pair is using a path; say A, for communication and it is initially congested. Later, a better path B becomes available. Now, it is desirable to balance traffic between paths A and B. It is possible for the network layer to naively shift the bulk of the traffic to path B, thereby moving the congestion to path B rather than load balancing. With path B now becoming congested, the network layer will then force the traffic back to path A, leading to oscillations and instability. Transport layer, on the other hand, gradually increases congestion window size on path B over a period of several round-trip time (RTTs) and balances traffic between A and B in a more stable fashion.

Thus, use of multiple paths at the transport layer can enable better load balancing properties due to the availability of fine-grained information about the network's end-to-end characteristics such as available bandwidth, RTT, etc. This compelling observation was made by the research of Kelly and Voice (2005). In particular, the multipath transport protocols are designed to move traffic from more congested to less congested paths. With this shifting, the loss rates on less congested path increase and on more congested path decrease; the overall result is that the loss rates across the network tend to equalize (Dong et al., 2007b).

(2) *Resource pooling*: The idea of pooling resources together into one single resource, which has the aggregate abilities of all the resources, has been widely used in the Internet (Wischik et al., 2008). Specifically, pooling of multiple paths at the transport layer provides better aggregate path characteristics (e.g., bandwidth, delay, and RTT) than each individual path. Resource pooling allows to efficiently using network resources by dynamically allocating resources to meet the traffic surges. Once the traffic transaction is complete, the resources go back in the pool. Resource pooling enables the Internet to have higher reliability and robustness than the individual links and routers.

In single-path applications, if a path fails then alternate paths can be used. However, the failure of primary path causes a temporary interruption in application until an alternate path is established (e.g., transient problems on a radio interface). Pooling of multiple paths, on the other hand, enables the network to transparently shift traffic from faulty paths to non-faulty paths without any interruption in the operation of an application.

(3) *Diversification*: Diversity is a technique that is frequently deployed in datacenters, wireless environments, and the Internet to improve performance. It has been shown in the literature that