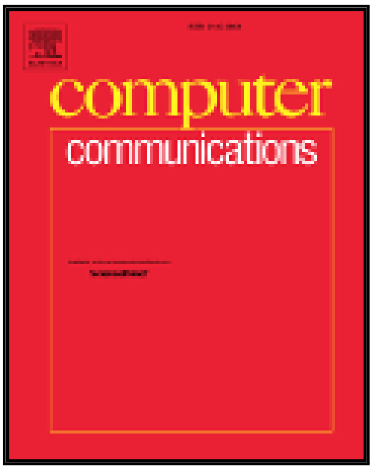

Using Multiple Paths in SCTP to Reduce Latency for Signaling Traffic

Johan Eklund, Karl-Johan Grinnemo, Anna Brunstrom

PII: S0140-3664(17)31087-3
DOI: 10.1016/j.comcom.2018.07.016
Reference: COMCOM 5735

To appear in: *Computer Communications*

Received date: 12 October 2017
Revised date: 5 June 2018
Accepted date: 7 July 2018

Using Multiple Paths in SCTP to Reduce Latency for Signaling Traffic

Johan Eklund, Karl-Johan Grinnemo, Anna Brunstrom

Department of Mathematics and Computer Science, Karlstad University, Karlstad, SWEDEN
Email: {johaeklu, karlgrin, annab}@kau.se

Abstract

The increase in traffic volumes as well as the heterogeneity in network infrastructure in the upcoming 5G cellular networks will lead to a dramatic increase in volumes of control traffic, i.e., signaling traffic, in the networks. Moreover, the increasing number of low-power devices with an on-off behavior to save energy will generate extra control traffic. These increased traffic volumes for signaling traffic, often generated as bursts of messages, will challenge the signaling application timing requirements on transmission. One of the major transport protocols deployed for signaling traffic in cellular networks is the Stream Control Transmission Protocol (SCTP), with support for multiple paths as well as for independent data flows. This paper evaluates transmission over several paths in SCTP to keep the latency low despite increasing traffic volumes. We explore different transmission strategies and find that concurrent multipath transfer over several paths will significantly reduce latency for transmission over network paths with the same or similar delay. Still, over heterogeneous paths, careful, continuous sender scheduling is crucial to keep latency low. To this end, we design and evaluate a sender scheduler that considers path characteristics as well as queuing status and data flows of different priority to make scheduling decisions. Our results indicate that by careful dynamic sender scheduling, concurrent multipath transfer could lead to reduced latency for signaling traffic irrespective of path or traffic characteristics.

Keywords: SCTP, Multipath, Scheduling, Latency

1. Introduction

The promises of increased capacity and reduced latency in the upcoming 5G cellular networks are envisioned to make cellular connectivity more attractive. The evolution of 4G to 5G is mainly driven by three incentives [1]: massive broadband that delivers gigabytes of bandwidth – on demand – to mobile devices; massive machine-to-machine communication that offers billions of sensors and machines interconnectivity; and critical machine-type communication which enables remote control over robots and autonomous driving. Already today the number of low-power mobile devices such as smart phones and tablets, connecting to the Internet using cellular networks is increasing. The increase of such devices and sensors is expected to explode by the materialization of Internet of Things (IoT) [2]. To reduce the energy consumption, some of these low-power devices will tear down connections to the network after the completion of a data-transfer, only to re-establish the connection when a new transfer request is issued. Moreover, future cellular networks is expected to encompass a variety of networks, so-called HetNets [3], where smart client devices will be equipped with multiple radio interfaces capable of leveraging many different radio access technologies. Furthermore, the 5G mobile network architecture is envisioned to include both physical and virtual network functions, as well as edge- and central-cloud deployments.

In cellular networks, the control information needed for the supervision, maintenance and operation of the network is exchanged among the network elements by signaling messages. The scenarios mentioned above will lead to a dramatic increase in signaling-traffic volumes in the cellular core networks [4].

Signaling-traffic characteristics differ significantly from the characteristics of bulk traffic, as signaling traffic is often generated as bursts of data-limited, independent data flows [5]. Moreover, dif-