Protocol Replacement Proxy for 2.5 and 3G Mobile Internet

Victor Khashchanskiy
First Hop Ltd., Finland

Andrei Kustov
First Hop Ltd., Finland

Jia Lang
Nice Business Solutions Finland, Finland

INTRODUCTION

Providing mobile Internet access in GPRS and UMTS networks is not an easy task. The main problem is in rather challenging network conditions (Inamura, Montenegro, Ludwig, Gurtov, & Khafizov, 2003). Latency in these networks could be an order of magnitude higher than in wired networks, with round-trip time (RTT) reaching up to one second. Moreover, there occur delay spikes in the network, when latency can exceed average RTT several times (Gurtov, 2004). Furthermore, in wireless networks, the risk of experiencing packet losses is considerably higher in comparison to that in wired networks. This is because packets can easily be lost due to corruption, either during deep fading leading to burst losses, or cell re-selections, resulting in a link black-out condition. Such characteristics of wireless cellular networks significantly affect performance of the principal Internet protocol—TCP—as it was designed to work in conditions of low-latency reliable networks.

TCP assumes that all segment losses indicate congestion as they are traditionally (i.e., in wired links) caused by buffer overflows in routers. If losses are detected, in addition to retransmitting the lost packet, TCP adjusts the values of sending window size and retransmission timeout (RTO) in order to slow down transmission. Packet losses or even long enough delays can lead to TCP timeouts. However, in both cases the timeouts are not caused by congestion, hence the basic working assumption for TCP is incorrect, and consequently the countermeasures are also not optimal. TCP flow control is achieved by complex mechanisms trying to probe for a data rate as high as possible but backing off as soon as congestion occurs. A TCP sender adapts its use of bandwidth based on feedback from the receiver. The high latency characteristic of cellular networks implies that TCP adaptation is correspondingly slower than in wired networks with shorter delays. Similarly, delayed acknowledgements exacerbate the perceived latency on the link.

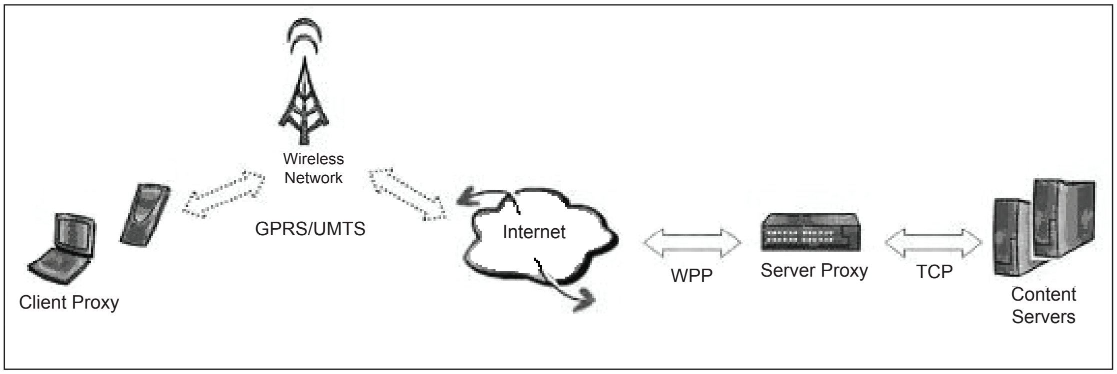
The central performance issues of TCP in wireless cellular networks lie in the inability to correctly detect the nature of the error, and so it is incapable of responding in an appropriate manner (Tsaoussidis & Matta, 2001). In addition, the protocol lacks efficient monitoring of the network conditions, rapid window size readjusting in response to changes in these conditions. Thereby overall performance of TCP is degraded through additional retransmission and wasted opportunities in maintaining the communication pipe full. There have been numerous attempts to improve TCP in wireless environments, for example, Chandran, Raghunathan, Venkatesan, and Prakash (2001), Kim, Toh, and Choi (2000) and Liu and Singh (2001). A standard implementation (Wireless Profiled TCP) is defined by WAP forum, which comprises state-of-the-art works on the subject, and implementation of TCP stack in modern operation systems supports them by default (Macdonald & Barkley, 2000).

As an addition to TCP optimization, the performance enhancing proxy (PEP) concept was developed (Border, Kojo, Griner, Montenegro, & Shelby, 2001) to further improve performance of Internet applications over wireless links. Different types of PEPs are used in different environments to overcome different link characteristics, which affect the performance.

Kustov et al. (2002) proposed an idea for raising efficiency of data transfer over high-latency low-bandwidth links by combining application-level PEP and TCP replacement with a lightweight protocol stack. This approach is called protocol replacement proxy (PRP).

As shown on Figure 1, data transfer over wireless link is performed using WAP peer protocol (WPP). At both ends of the link client and server proxies perform protocol translation between TCP and WPP. For example, when a mobile user downloads a Web page from a content server, the client proxy accepts a TCP connection from the browser, passes the request to the server proxy, which, in turn, establishes a TCP connection to the content server. The content server response is delivered to the browser in the same way.

Figure 1. PEP approach with transport protocol replacement



WPP stack uses UDP as a transport protocol and utilizes WTP and WDP layers of standard WAP protocol stack, with proprietary WPP layer added on top of WTP layer. This scheme makes use of both application-specific optimization techniques (e.g., for HTTP metadata, applets removal, lossy image compression), and transport protocol optimization, which is achieved by reducing overhead and adapting it to wireless link characteristics.

PROBLEM DEFINITION

The concept of PRP, proposed by Kustov et al. (2002) was experimentally studied in Kustov and Lang (2005) for its practical realization (Nokia Wireless Accelerator). The study focused on mobile Internet experience for the two most commonly used Internet protocols—FTP and HTTP—for network technologies ranging from GPRS to UMTS. Application-specific optimization depends strongly on the transferred data; for example, for office documents and HTTP browsing, data reduction was in the range of 60-90%. Transport-level optimization was found to reach 35% in the case of GPRS, which was the main target for PRP.

At the same time, it was found that the wireless accelerator yields negative time savings in the 3G network. Taking into account the fact that compression was applied to the downloaded data, all this indicated inefficient bandwidth utilization of WPP stack compared to that of wireless profiled TCP. The reason was found to be static protocol sending window, decreasing efficiency of WPP at higher data rates. Initially optimized for slower GPRS/EDGE, the same window only allowed partial bandwidth utilization in 3G.

For PRP to work properly in 3G networks as well, WPP protocol stack needs improvements.

DYNAMIC WPP STACK ADAPTATION

In this article we propose the solution to the problem—correction of the WPP protocol stack, allowing the transport layer to adapt in real-time to available bandwidth.

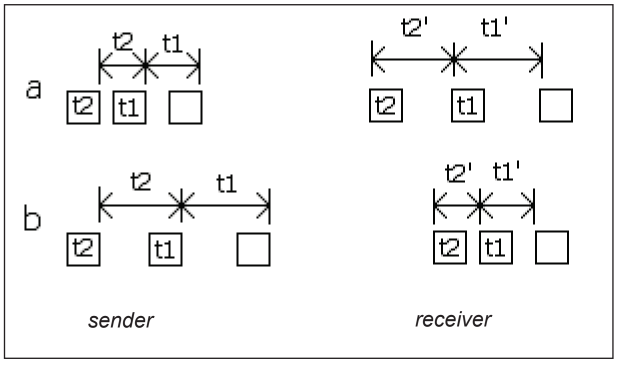
The optimal size of sending window is equal to bandwidth-delay product. As this value can not be measured directly, we have developed the method of adaptation based on adjusting sending window size according to bandwidth utilization feedback. For that, all packets are supplied before sending with time intervals placed in transport information items (TPI) of WTP packet headers. At the receiving side, time intervals between the moments of arrival of two consequent packets are compared to the intervals between the packets departure.

Figure 2 explains the idea in detail. Time interval **t1** between sending of the first and the second packets is conveyed from the sender to the receiver within the second packet. At the receiving side, the interval between times of arrival **t1'** is measured and then the difference (**t1'** – **t1**) is sent as feedback to the sender within WTP ACK packet.

If the difference is positive like in the case (a), this means that packets are sent faster than it is possible to deliver with currently available network bandwidth. In the long run it might result in packet losses and thus retransmissions. To avoid this, sending window size has to be decreased in order to restrict the data-sending rate.

If the difference is negative like in the case (b), the network is capable to deliver packets faster than they are currently sent, so available bandwidth is not utilized fully and sending window size has to be increased.

Figure 2. Window adaptation algorithm for WPP protocol stack improvement



2 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/protocol-replacement-proxy-mobile-internet/17175

Related Content

The Wi-INET Model for Achieving M-Health Success
N. Wickramasingheand S. Goldberg (2007). *Encyclopedia of Mobile Computing and Commerce (pp. 1004-1010).*
www.irma-international.org/chapter/inet-model-achieving-health-success/17210

Mobile Multimedia Streaming Using Secure Multipath in Wireless Ad Hoc Networks
Lei Chenand Chung-wei Lee (2010). *International Journal of Handheld Computing Research (pp. 54-74).*
www.irma-international.org/article/mobile-multimedia-streaming-using-secure/43604

An Overview of the Capabilities and Limitations of Smartphone Sensors
Avi Klausner, Ari Trachtenberg, David Starobinskiand Mark Horenstein (2013). *International Journal of Handheld Computing Research (pp. 69-80).*
www.irma-international.org/article/an-overview-of-the-capabilities-and-limitations-of-smartphone-sensors/79960

Fall Behavior Recognition Based on Deep Learning and Image Processing
He Xu, Leixian Shen, Qingyun Zhangand Guoxu Cao (2018). *International Journal of Mobile Computing and Multimedia Communications (pp. 1-15).*
www.irma-international.org/article/fall-behavior-recognition-based-on-deep-learning-and-image-processing/214040

Scenario-Based Cluster Formation and Management in Mobile Ad Hoc Networks
V. S. Anithaand M. P. Sebastian (2009). *International Journal of Mobile Computing and Multimedia Communications (pp. 1-15).*
www.irma-international.org/article/scenario-based-cluster-formation-management/4060