ISSN 2047-3338

Web Acceleration for Wireless Networks

Muhammad Wasim
University of Engineering and Technology, Lahore
cmwasim@gmail.com

***Abstract*—With the widespread use of tablets and smartphones the demand for data services is increasing exponentially. WiMAX and LTE networks provide high speed data services to the end users but it is no match to the end user requirements. The users browse multimedia content websites or social media services and watch videos on wireless networks. The wireless devices of today consume high bandwidth. For example, the iPhone can use 24 times the spectrum of a traditional cell phone, while the iPad uses 122 times as much. This explosive growth of wireless device users has increased bandwidth requirements and enterprises are evaluating methods to make the available infrastructure more efficient. One such method is to use optimization techniques at TCP & HTTP level. In this study, we compare and evaluate HTTP and TCP optimization techniques. These techniques decrease bandwidth congestion, make the existing infrastructure more efficient, and improve the end-user experience at the same time.**

***Index Terms*—Web Acceleration, HTTP Optimization and Wireless Network Optimization**

I. INTRODUCTION

THE field of telecommunication has gained much consumer attention in past years. After Internet, mobile communications have been the major driving force in telecom services demand. The ever increasing demand for data over mobile Internet (after voice services) has led to huge interest in mobile Internet data applications. Wireless devices such as Customer Premises Equipments (CPEs), tablets, and smart phones have emerged as a popular way to stay in touch and view information online. The information provided online today is media enriched i.e. it's common for websites to have images and videos. Mobile devices consume a lot of bandwidth space and hence its essential for network service providers to keep up with the infrastructure requirements. It is not easy for service providers to extend existing infrastructure and so they are only limited with some software option to effectively use their network infrastructure.

The degradation of user experience in mobile network data services, due to limited bandwidth and varying bit error rate of channels have acquired much interest in performance evaluation of these networks [1], [2]. To improve the end user experience many techniques have been used such as Wan Optimization, Proxy caching, bundling, or Content Distribution Networks (CDN) [3]. Many techniques are emerging with time and these techniques are already playing an important role in mobile Internet. Applying such techniques greatly improve Quality of Service (QoS) and hence improve end user experience. In mobile networks, acceleration for servers is very important because mobile users utilize wireless Internet to get their resources, so the performance problems associated with it must be controlled.

The acceleration software can reside on the networking hardware as in CHARTS [4] or it can be implemented as a proxy server. Various web resources such as web pages, css, html, images and video can be processed to accelerate the performance of the web. The focus of research in wireless networks has been on two aspects i.e. TCP based acceleration and HTTP based acceleration.

TCP based acceleration includes increased initial window (RFC3390), TCP multiplexing, window scaling, TCP Offload Engine congestion avoidance, and fast retransmit etc. Other techniques that have been used include selective acknowledgments, explicit notification, and recovery are discussed in [5]. HTTP Optimization techniques include persistent connection, pipeling, compression, pre-fetching, and cache proxy. Other less popular techniques are code optimization and bundling.

Other methods to optimize traffic are by image compression or video transcoding. The compression of images implies minimizing the size in bytes of an image but keeping the quality of the image to an acceptable level. It's obvious that the compressed images will take less space on the hard disk and will be transmitted fast over the wireless network. Various algorithms have been developed over time for efficient image compression. Three very common image formats used on the Internet today are the PNG, JPEG, and GIF. PNG is used because of its relatively small size and support of alpha channel. The photographs are usually uploaded in JPEG format and line art or basic animation is done in GIF. Bitmap

images are not commonly used because of their heavy file size. Fractals and wavelets can also be used for image compression. Applying compression techniques, while forwarding image resources to wireless networks, can greatly decrease the network traffic and improve end user experience.

Finally, the ads can be blocked on websites. Blocking ads is much like blocking SPAM: the more aggressive we are about it, the more likely we are to block things that were not intended and the more likely that some things may not work as intended. So there is a trade off here. Ad filtering algorithms can be used to remove the advertised content from the web page. The advertisements can be in the form of animations, ad text or pictures. Various techniques for ad filtering are used such as blacklists, white-lists etc. Regular expression filters can also be used to filter advertised content. Although advertisements are a major source of revenue for any website, the benefits to remove such ads include cleaner looking web pages free from advertisements and decreased traffic over the wireless link.

The focus of this study is to survey TCP Optimization and HTTP acceleration techniques. In next section, a brief review of the literature is presented. Later, in section III trade-offs in various techniques are discussed. Finally, the paper concludes with challenges and opportunities in this area.

II. LITERATURE SURVEY

Many research efforts have been made to optimize both the TCP layer and HTTP optimization over wireless link. The most significant benefit can be achieved by using compression. Compression can provide up to 75% byte savings for different text files including html, CSS, and JavaScript [1]. Other techniques include code optimization, Caching, and image compression etc.

Barron and David [6] proposed a solution, WebExpress, for optimizing web browsing in wireless networks. Issues such as connection overhead for each individual page resource, redundant capability of transmission, and verbosity of HTTP protocol were identified. They used caching to cache the objects being called again and again, differencing to load only changed parts of the page, protocol reduction to reduce acknowledgments and header reduction to overcome the above mentioned issues. They introduced Server Side Intercept (SCI) and Client Side Intercept (CSI) or local web proxy to accomplish high performance. They found an average reduction of 85% in bytes and 64% reduction in response time.

Jiang Song [7] proposed architecture for wireless links. They used content compression and a modified version of TCP called BoostedTCP. The solution could work with both using a local proxy on the client side and without using the client side technology. For compression, GZip tool was used which is popular on Internet and most browsers support it. By avoiding Slow Start, Congestion because of packet loss, 3 way handshakes and changing the window size, the researchers were able to reduce the download time to 50% or more than the actual time required to download the same web page.

Qin Xu [8] has presented many acceleration technologies for data services. Some of these techniques use HTTP/1.1 features which are usually turned off by web servers such as persistent connections and pipeling, compression, caching, and load balancing. The researcher also evaluated HTTP optimization techniques [8]. Typical schemes of deployment of accelerator include symmetrical and asymmetrical methods. They deployed OSN, a product of Bytemobile, to achieve the above mentioned optimizations and found obvious improvements in user experience.

Paul Davern et al. [9] worked on split HTTP Performance Enhancing Proxy (HTTPEP) for accelerating web browsing on high latency satellite networks. They deployed a remote site and ground site proxy and persistent connections between these proxies. Using compression and bundling of web resources, they achieved a performance improvement of about 27%. They did not use caching in their architecture which could further improve the performance of the system.

Xiaohui Chen et al. [10] have extensively studied approaches to reduce web latency in mobile networks. He has summarized the literature based on two categories i.e., HTTP Optimization and TCP acceleration. The author has also discussed the challenges such as utilization of link information, TCP modification for asymmetry, and proxies with concurrent connections. Following are some other optimization techniques that have been discussed in literature for HTTP optimization [10] (Table I).

TABLE I: HTTP OPTIMIZATION TECHNIQUES

	Bandwidth Usage Reduction	Delays Reduction	Speedup Mobile User Experience
Persistent Connection	No	Yes	Yes
Pre-fetching	No	Yes	No
Compression	Yes	Yes	Obvious
L4 Switch	No	Yes	No
L7 Switch	No	Yes	No
Pipeling	No	Yes	No
Cache Proxy	Yes	Obvious	Obvious

There are many TCP optimization techniques that have been used to optimize wireless networks [10]. Table II shows some of these optimization techniques.

TABLE II: TCP OPTIMIZATION SKILLS

	Bandwidth Usage Reduction	Delays Reduction	Speed up Mobile UE
Window Scaling	No	Obvious	No
Congestion Avoidance	No	Obvious	Obvious
Explicit Notification	No	Yes	Obvious
Selective Acknowledgment	Yes	Yes	Obvious
Layer 2 Enhancement	Yes	Yes	Obvious

III. DISCUSSION AND REVIEW

We now review and discuss the techniques that we have covered so far. This is a comprehensive overview of all the popular web acceleration techniques.

A) Persistent Connection & Pipling

With the introduction of HTTP/1.1, it provides facility of setting up a persistent connection between an HTTP client and server. A client sends a request to server and after the server gives a response to client, the connection remains open for the next request and response. The connection is closed after all requests of that client are entertained. Requests from a client are pipelined, i.e. a client doesn't need to wait for a response of one request before it can send another request. In this way, server also sends quick response to clients. The server works efficiently because it already has the requests from a client and no time is wasted waiting for a new request. One drawback of keeping the connection open is that it may cause the server resources come to an end very soon. An intermediate proxy server can be used to persist open connections between client and server. This way, the web server will be free and proxy server will utilize the resources in local premises.

B) Code Optimization

The optimization of source code plays an important role to decrease the traffic over the wireless network. It's the principal of sending as little as possible to achieve the same objective. The comments, extra spacing and long variable names all contribute to big code files. The code optimization on the fly can greatly help reduce the traffic. Currently, many developers dont optimize the code and hence the websites online are not code optimized which increases the possibility to improve the performance of the network using this technique.

C) Compression

Many research papers propose to compress the web content. The compression of web content over HTTP was introduced in HTTP 1.1 but many web servers don't compress the content to save the processing time to deliver the web page. The proxy server which routes the traffic from wired network to wireless network can compress the content on the fly using standardized algorithms such as Gzip. It is a lossless compressed data format and has been used for decades for compression. The positive point about Gzip is that many major browsers now support the Gzip. This means that we don't use any separate software to decompress the content and the browser will itself decompress and show the web page to the end user. This greatly improves the performance of wireless network as the amount of data is decreased to a great extent.

D) Cache Proxy Server

Proxy server has been used for many years by service providers. Such servers are used for caching which is a well known concept in computer science. When the client needs to access the same resources again and again, it can be cached to reduce the traffic over the network.

E) Bundling

The 80% of the end-user response time is spent on the frontend. The web page consists of multiple resources and each resource is requested separately. This increases the number of connections and the overall loading time. Using persistent connections with pipling, a proxy server can request all the resources at once and then transmit them on the wireless medium. This greatly improves the efficiency of page load. This bundling mechanism is getting popularity over wireless networks.

F) TCP Optimization

The TCP is badly affected by various wireless characteristics. The bandwidth oscillation may trigger time-out based retransmission, also known as spurious timeout [11]. Moreover, other parameters that affect the wireless network performance are window size, slow start, advanced congestion control algorithms and modification in acknowledgment strategies.

IV. CONCLUSION

In this paper, we discussed various approaches and architectures that have been developed in last decade to improve the performance of wireless networks. As the demand for Internet access is increasing over wireless network, the acceleration of wireless asymmetrical networks needs deeper study. Further, it is evident that a mix of more than one optimization strategy can greatly improve the performance of the network.

REFERENCES

[1] M. C. Chan and R. Ramjee, "Improving tcp/ip performance over thirdgeneration wireless networks," Mobile Computing, IEEE Transactions on, vol. 7, no. 4, pp. 430–443, 2008.

[2] M. Fornasa, N. Zingirian, and M. Maresca, "Extensive gprs latency characterization in uplink packet transmission from moving vehicles," in Vehicular Technology Conference, 2008. VTC Spring 2008. IEEE, 2008, pp. 2562–2566.

[3] T. Grevers Jr, J. C. C. No et al., Application acceleration and wan optimization fundamentals. cisco Press, 2012.

[4] J. Lee, P. Sharma, J. Tourrilhes, R. McGeer, J. Brassil, and A. Bavier, "Network integrated transparent tcp accelerator," in Advanced Information Networking and Applications (AINA), 2010 24th IEEE International Conference on. IEEE, 2010, pp. 285–292.

[5] C. Burns and D. Newman, "Vendors choose to tout disparate application acceleration techniques," Network World, vol. 1, p. 16, 2006.

[6] B. C. Housel and D. B. Lindquist, "Webexpress: A system for optimizing web browsing in a wireless environment," in Proceedings of the 2nd annual international conference on Mobile computing and networking. ACM, 1996, pp. 108–116.

[7] J. Song and Y. Zhang, "Architecture of a web accelerator for wireless networks," in Proceedings of the thirtieth Australasian conference on Computer science-Volume 62. Australian Computer Society, Inc., 2007, pp. 125–129.

[8] Q. Xu, L. Wang, H. Wan, and Y. Zhong, "Acceleration for data service in mobile networks," in Global Mobile Congress 2009. IEEE, 2009, pp. 1–5.

[9] P. Davern, N. Nashid, A. Zahran, and C. J. Sreenan, "Http acceleration over high latency links," in New Technologies, Mobility and Security (NTMS), 2011 4th IFIP International Conference on. IEEE, 2011, pp. 1–5.

[10] X. Chen, W. Wang, and G. Wei, "Reducing web latency in mobile network," in Computer and Information Science, 2009. ICIS 2009. Eighth IEEE/ACIS International Conference on. IEEE, 2009, pp. 829–833.

[11] H. Balakrishnan, V. N. Padmanabhan, S. Seshan, and R. H. Katz, "A comparison of mechanisms for improving tcp performance over wireless links," Networking, IEEE/ACM Transactions on, vol. 5, no. 6, pp. 756–769, 1997.