

Performance Evaluation on WAP and Internet Protocol over 3G Wireless Networks

Hidetoshi Ueno, Norihiro Ishikawa, Hideharu Suzuki, Hiromitsu Sumino, and Osamu Takahashi

NTT DoCoMo, Multimedia Laboratories
3-5, Hikari-no-oka, Yokosuka, Kanagawa, 239-8536, Japan
{hueno, ishikawa, hideharu, sumino,
osamu}@mml.yrp.nttdocomo.co.jp

Abstract. This research analyses the performance of WAP 1.x in a comparison to the Internet protocol. We implement a WAP client and a WAP gateway based on WAP version 1.1 and assess the response time by comparing to that of HTTP and TCP. We use a W-CDMA simulator to evaluate its performance in high-speed wireless networks such as 2.5G and 3G. The results shows that both protocols have comparable performance (i.e. response time) except when transmitting large content sets (e.g. multimedia data files), in which case the performance of HTTP/TCP is better than that of WAP 1.x. We also evaluate WAP specific functions such as the binary encoding of WAP headers and contents. While binary encoding is effective for small content sets, its effectiveness and performance are questionable for large content sets. Finally, we propose a mobile Internet architecture that is suitable for 2.5G and 3G wireless networks based on the evaluation and our experience with the i-mode service. Our architecture consists of wireless optimized TCP, TLS, HTTP and XHTML.

1 Introduction

Services that will access the Internet from handheld devices such as weather forecasts, news, and mobile banking are attracting people's attention. Handheld devices tend to have many restrictions such as have less powerful CPUs, less memory, and smaller displays. Wireless networks also suffer from higher error rates, lower bandwidth, higher latency, and unexpected circuit failures. Since it was considered that the protocol used in the Internet might not be suitable for wireless environments, the Wireless Application Protocol (WAP) Forum developed the WAP version 1.x (WAP 1.x) protocol [1].

WAP 1.x is designed for various kinds of wireless network bearers (e.g. GSM and CDMA) [1]. However, it is unclear what sort of networks can take full advantage of WAP 1.x, and its performance of has not been fully evaluated. Thus, we developed a WAP client and a WAP gateway based on the WAP specifications [1], and then evaluated WAP performance by using a Wideband Code Division Multiple Access (W-CDMA) simulator. We compared the WAP 1.x protocol to the Internet protocol (HTTP/TCP). Finally, we created a mobile Internet architecture that is suitable for

next generation (2.5G) and third generation (3G) wireless networks since services over the International Mobile Telecommunication 2000 (IMT-2000) networks [2] have been started in Japan.

The WAP 1.x protocol is overviewed in section 2, and its implementation in the WAP 1.x test-bed system is described in section 3. Our evaluation of WML 1.x binary encoding is given in section 4 and is compared to WAP 1.x and HTTP/TCP in section 5. In section 6, we propose a mobile Internet architecture for high-speed wireless networks such as 2.5G and 3G.

2 WAP 1.x Overview

WAP defines an architecture and protocols with the goal of providing special functionalities such as telephony, push delivery, and suspend & resume. The following summarizes the WAP architecture and protocols.

The WAP architecture consists of WAP client, WAP gateway and origin server. It is based on the Internet World Wide Web (WWW) model with few enhancements. The WAP protocols are used between the WAP client and the WAP gateway, and the Internet protocols (i.e., HTTP and TCP) are used between the WAP gateway and the origin server. Optimizations and extensions have been made in order to satisfy the requirements of the wireless environment.

The WAP protocols consist of Wireless Datagram Protocol (WDP), Wireless Transport Layer Security (WTLS), Wireless Transaction Protocol (WTP), Wireless Session Protocol (WSP) and Wireless Application Environment (WAE) (Figure 1.).

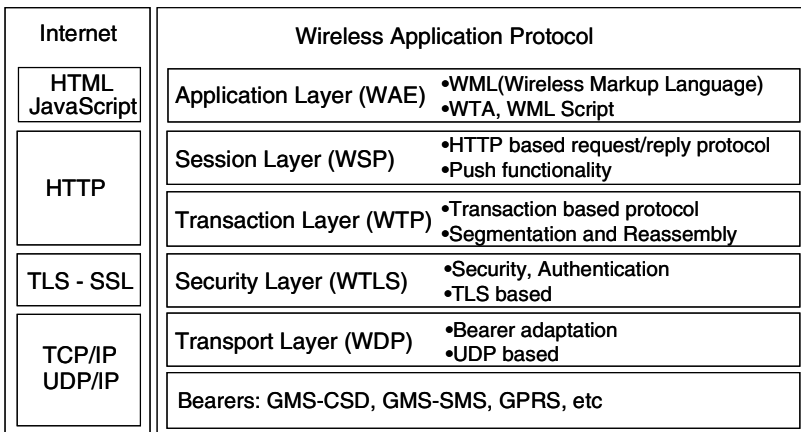


Fig. 1. WAP protocol overview and its comparison to the Internet protocol

WDP provides functions for bearer adaptation, which absorbs the differences of lower wireless network protocols. When the bearer network supports IP, UDP is used for WDP. WTLS is developed based on TLS and provides the means for supporting security functions such as authentication and confidentiality. WTP provides transaction type communication and has three transaction types (i.e. class 0, 1, and 2).

Class 2 transaction in particular realizes reliable communication by supporting packet retransmission. WTP supports segmentation and reassembly (SAR), which provides the means for transmitting large content sets whose size exceeds one Maximum Transfer Unit (MTU). WSP provides session management functions including session initiation and session suspend and resume. WSP provides header compact encoding and push capability in addition to the basic functionality of HTTP.

WAE is a general term of application environments in WAP, and consists of several components. WAP defines the Wireless Markup Language (WML) [1] as the markup language, and WML Script as the scripting language, and Wireless Telephony Application (WTA) as the telephony-related application. WML is based on Extensible Markup Language (XML) and defines its own tags, and has no compatibility with Hyper Text Markup Language (HTML). WML uses the binary representation format [1] in order to reduce content transfer volume. WML content is encoded into binary representation format at the WAP gateway to the wireless network.

2.1 WAP Related Works

Since there are a lot of Internet contents written in HTML, WAP clients must be able to access HTML contents. To do so the WAP gateway needs to support content conversion from HTML to WML. Reference [3] investigated the problems associated with the conversion from HTML to WML. It proposed some techniques for converting HTML to WML but several problems remained. The overhead of content conversion is also an issue because poor gateway scalability becomes a serious handicap when the number of subscribers increases. Although it is very important to investigate and consider WAP gateway scalability, research has been insufficient to date.

As for evaluating the WAP protocol, one paper evaluated the WTP class two protocols by implementing the WAP protocol stack [4]. It pointed out some inconsistencies in the WAP specifications but didn't investigate WTP performance over wireless networks.

Reference [5] analyzed the network traces generated by a mobile browser application. The research observed daily and weekly cycles, and found some evidence of self-similarity in the network traffic produced by the application. The research also compared and contrasted the mobile browser traffic characteristics with the results for WWW traffic published in the literature. The results of this research are very significant for the design of wireless networks. However, since the network characteristics of 3G networks are quite different from those of 1G and 2G networks, in-depth research of 3G network traffic characteristics is needed.

While 3G commercial services started in Japan in October 2001, no research has examined WAP 1.x over 3G networks.

3 Implementation of WAP 1.x Client and Gateway

We developed a WAP client and a WAP gateway based on WAP version 1.1 (WAP 1.1) specifications [1], and simulated a high-speed wireless environment by using a hardware-based W-CDMA emulator. The W-CDMA emulator allows several of the

parameters related to the wireless network (e.g. bearer speed, error rate and maximum number of retransmission) to be set up. The parameters used, see Table 1, are based on the FOMA implementation¹.

Note that WAP 1.1 is not the newest version of the WAP 1.x series, but there is no measurable difference as regards protocol performance.

Table 1. This table shows principal parameters set in the W-CDMA bearer simulator. In W-CDMA, one to twelve PDUs (i.e. Radio Link Control frames) constitute one Forward Error Correction (FEC) frame [6]. The actual size of the FEC frame depends on the link conditions and bandwidth allocation. Since the error rate value is based on the typical average value on our experimental 3G system, it captures wireless-specific characteristics (e.g. fading behavior). The error rate is per FEC frame.

Parameter	Value
Bearer Speed	(Downlink) 384 and 64 Kbps, (Uplink) 64 Kbps
Layer 2	Radio Link Control (RLC) Protocol [6]
Bearer MTU	1500 bytes
Error Rate	5% (per FEC frame)

3.1 WAP Test-Bed System Overview

The WAP 1.1 test-bed system is shown in Figure 2.

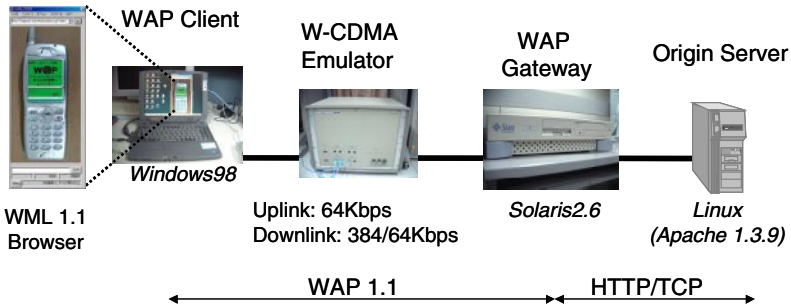


Fig. 2. This is the WAP 1.1 test-bed system. The W-CDMA emulator is set between the WAP client and the WAP gateway.

- 1) WAP client

We wrote the WAP client in C++ for the Windows 98 platform. It consists of WML 1.1 browser, WAP protocol module and so on. The WML 1.1 browser can emulate a mobile client and has the capability of displaying WML contents.
- 2) WAP gateway

We wrote the WAP gateway in C language for Solaris 2.6. It consists of WAP and Internet protocol modules, gateway application module, and so on. The

¹ FOMA is the first 3G commercial service; it started in October 2001. For further information, please see (<http://www.nttdocomo.co.jp/english/>).

gateway application module realizes WML content parsing and binary encoding if the content is WML, and forwards the encoded data to the WAP client.

3) Origin Server

We used Apache 1.3.9 as the WWW server application. It also supports the Common Gateway Interface (CGI) function so that it is possible to create dynamic WML content.

3.2 WAP Applications

Table 2 lists the applications we implemented in the test-bed system. GET and POST are used for Web browsing. As for the push application, the WAP gateway becomes a push server and pushes contents to the WAP client by using the WSP push or confirmed push method. We also developed an e-mail application by using CGI and an e-mail receiving application by using WSP push.

Table 2. Applications developed for the test-bed system

Service	Content	WSP Function
Browsing	WML	GET, POST method
Push	GIF, Text	WSP Confirmed/Unconfirmed Push
E-mail	Text	(Send) POST method (Receive) WSP Confirmed/Unconfirmed Push

4 Evaluation of WAP 1.x Binary Encoding

WAP 1.x defines two types of binary encoding functions: WML binary encoding and WSP header compact encoding. Since these functions are peculiar to WAP 1.x, we evaluated both of them.

4.1 Evaluation of WML Binary Encoding

The WAP Gateway encodes WML contents (WML tags as well as control codes) after XML parsing. Since the effectiveness of the binary encoding depends on the content, we evaluated the encoding rate by using several typical examples of WML contents as shown in Figure 3. The contents used essential tags such as “wml”, “card” and “p” and text data in addition to line break tags (i.e., “br”). We evaluated the binary encoding of WML contents using several content set sizes (i.e., 500, 1000, 1400, 20K, 100K, 360K bytes) by changing the size of the data part. The evaluation focused on the following two factors.

- Compression rate comparison between WML binary encoding and gzip
- Time taken for WML compression, which includes XML parsing time.

The result of the evaluation is shown in Figure 4. In the evaluation, we used RXP beta 15 as the XML parser.

```

<?xml version="1.0"?>
<!DOCTYPE wml PUBLIC "-//WAPFORUM//DTD WML 1.1//EN"
"http://www.wapforum.org/DTD/wml_1.1.xml">
<wml>
<card id="card2" ontimer="./auto500-2.wml#card1">
<timer value="50"/>
<p mode="wrap">
*Test of 1400 Octet Contents*<br/>
***deck1***<br/>
WAP is a protocol that is designed for wireless
environment.<br/>
WAP is a protocol to access to the Internet from
cellular phones, PDAs and so on. <br/>
(The rest is omitted.) </p>
</card>
</wml>

```

Fig. 3. Example of WML content. It uses essential tags such as “wml”, “card” and “p” and text data in addition to line break tags (i.e., “br”).

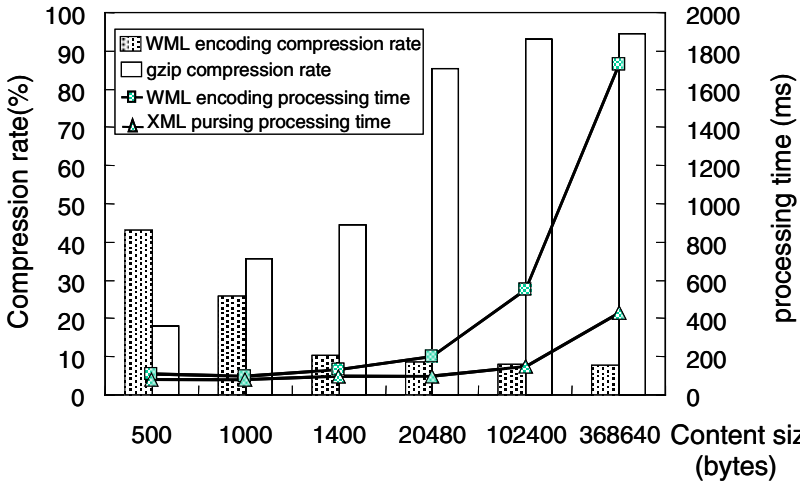


Fig. 4. The bars show WML encoding compression rate and gzip compression rate. The lines show WML encoding processing time and XML parsing time. This paper defines the compression rate as the ratio of the original WML content size to the reduced size of compressed content.

WML binary encoding is effective only if the content set size is small. The reason for this is that the XML and DOCTYPE declarations, which offer very high compression rates (i.e. 115 octets → 3 octets), are a significant fraction of the data only if the data set is small.

As for the WML encoding processing time, both WML encoding and XML parsing time increase rapidly with content set size (e.g., 1730 ms for 360 Kbytes). For a 500 bytes content set, 72.7% of the WML binary encoding processing time is occupied by XML parse time while the remainder is for the actual encoding. For a 368 Kbytes

content set, however, the XML parse time occupies 24.9% of the WML binary encoding processing time. Therefore, actual binary encoding takes longer than XML parsing if the content set size is large.

WML binary encoding is not so effective if high-speed networks are used. For example, a 1400 bytes content set is binary encoded to 1117 bytes. The transmission delay of the size reduction (i.e., 283 bytes) is about 6 ms at 384 Kbps, and 236 ms at 9600 bps. Because it takes approximately 130 ms to encoding a 1400 bytes WML data set, encoding is only effective if the bearer network speed is low, i.e. 9600 bps.

We conclude that WML binary encoding offers some benefits but not for high-speed networks and/or large content sets. Thus it appears that WML binary encoding is not so effective since 3G networks offer high speeds and will be used to send large content sets such as multimedia data files.

4.2 WSP Header Compact Encoding

One unique function of WSP is WSP header compact encoding, which aims to reduce the size of the WSP header. Our test bed system took less than 10 ms to compress (185 bytes → 50 bytes) the WSP header on a connect message. Once again, compression is not effective if the network's speed is high. For example, if the network speed is 384 Kbps, the savings of 135 bytes is equivalent to about 3 ms. Obviously it is better to send the message without using WSP header encoding. A similar argument can be made for large data sets. If the network speed is 9600 bps then the transmission delay is approximately 112 ms, so WSP header encoding is effective. Given that 3G offers high speeds and will handle large data sets, WSP header encoding is problematic.

5 Comparing WAP 1.x to HTTP/TCP

Since WAP defines many specifications that vary from protocols to application environments, it is difficult to evaluate the overall performance of the WAP protocol. As the first step, we evaluated WAP functions (WAP 1.1) and their performance.

5.1 Functional Comparison between WAP 1.x and HTTP/TCP

While TCP is a byte-stream connection-oriented transport protocol, WTP provides transaction-type message transmission so that WAP 1.x can provide reliability in case of using WTP since WDP doesn't provide any reliability. While 3-way handshaking is necessary to establish a TCP connection, WSP establishes a session by using WSP connect and WSP connect reply messages. WTP can optionally send asynchronous transaction requests so that the WTP initiator can send subsequent packets without waiting to receive acknowledgements of the previous packets. This asynchronous transaction request function enables the WAP 1.x user to shorten the total communication time.

An example of the communication sequence of each protocol is shown in Figure 5. This example shows the case where the client receives one content set in one packet. Both protocols consist of:

- WSP session/TCP connection setup phase,
- Content pull phase and
- WSP session/TCP connection closure phase.

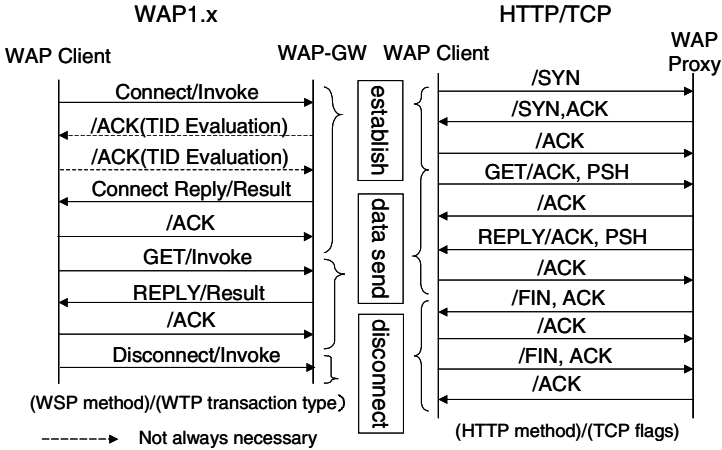


Fig. 5. Communication sequences of WAP 1.x and HTTP/TCP. The left and right figures are independent.

Since TCP offers the HTTP persistent connection defined in [7], setting up a new connection is not necessary if another connection is already established. WAP 1.x provides the same function by WSP.

5.2 Performance Evaluation Parameters

We choose three content set sizes (1K, 10K, 100K bytes). The 1 Kbytes set represents the small graphic elements common in web pages, and it is possible to send the set in one packet if the bearer Maximum Transfer Unit (MTU) is 1500 bytes. Since interactive users may see a page in segments and spend some time reading parts of it and clicking to see the next part of the page, it is the case for browsing such kind of small data. The 100 Kbytes set represents multimedia contents such as MPEG4 streaming data, and Java content. This kind of application is actually provided in Japan by the FOMA service. In case of WAP 1.x, large content sets that exceed one MTU can be conveyed if the SAR functionality of WTP is supported in WAP 1.x.

We used the parameters recommended in the WAP specifications of WTP 1.x in the test-bed system. The W-CDMA emulator parameters are the same as those described in section 3.

Of the TCP optimization techniques specified in [8], we used Selective ACK (SACK) [9], increased TCP initial window [10] and large receiver's advertised window. The last technique enlarges the maximum window size that can be estimated

by the Bandwidth Delay Product (BDP) of the end-to-end path. The parameter of receiver's advertised window is one of the keys to improving transmission performance. For example, the IMT-2000 network can support up to 384 Kbps, and its Round Trip Time (RTT) of the end-to-end path varies from half a second to one second [8]. Therefore, the BDP of the end-to-end path over the IMT-2000 network can be very large. The sender cannot make full use of the network bandwidth if the receiver's advertised window size doesn't suit this BDP. In our early investigations, we found that the value of the receiver's window size should be set at appropriately 32 Kbytes (calculation based on BDP over FOMA) to maximize network performance. The RTT is estimated to be up to 667 ms if the window size is 32 Kbytes. We decide that this value is suitable considering the tradeoff between memory cost and expected performance gains. As for the WTP, we also used 32 Kbytes as the maximum group size parameter to be consistent with TCP.

5.3 Performance Comparison between WAP 1.1 and HTTP/TCP

We measured the response time as time between sending a request from the WAP client to receiving a response from the WAP gateway. We also measured the response time of HTTP/TCP. Note that neither of these response times includes connection setup time because we assume that the TCP persistent connection is used and a continuous WSP session is available.

The results of our measurement showed that the WAP 1.1 and HTTP/TCP have comparable performance (i.e. response time) except when transmitting larger contents (e.g. multimedia data types), in which case the performance of HTTP/TCP is better than that of WAP 1.x (Figure 6).

If we look at the result more in detail, we find that this is related to the difference in the flow control mechanisms of WTP and TCP.

- For 1 Kbytes set: HTTP/TCP is slightly better than WAP 1.1.

The reason of this is that the TCP slow start mechanism doesn't matter because the 1 Kbytes set can be sent in one packet. The difference between WAP 1.1 and HTTP/TCP is due to the overhead of WAP 1.1 protocol conversion at the WAP gateway (i.e., from WAP 1.1 protocol stack to Internet protocol stack). However, the impact on performance is insignificant.

- For 10 Kbytes set: WAP 1.1 is slightly better than HTTP/TCP

This is due to the difference in the flow control algorithms. In WTP, seven packets are used to send a 10 Kbytes set so the WAP gateway transmits these seven packets simultaneously within the maximum group size. However, the TCP slow start algorithm generally prevents full use being made of network resources, since TCP can send only two packets at the beginning of communication (assuming that the increased TCP initial window [10] is used). Therefore, the window size of TCP is not fully expanded when the sender finishes sending the 10 Kbytes set.

- For 100 Kbytes set: HTTP/TCP is better than WAP 1.1

If the packet size is as large as 100 Kbytes, TCP communication is finished after the TCP window size is fully expanded and adjusted appropriately based on underlying bearer's speed. In WAP 1.1, the sender has to wait until the sent packets are acknowledged. This incurs a slight waiting time because the WAP-GW needs to wait for a single ACK packet (Figure 7).

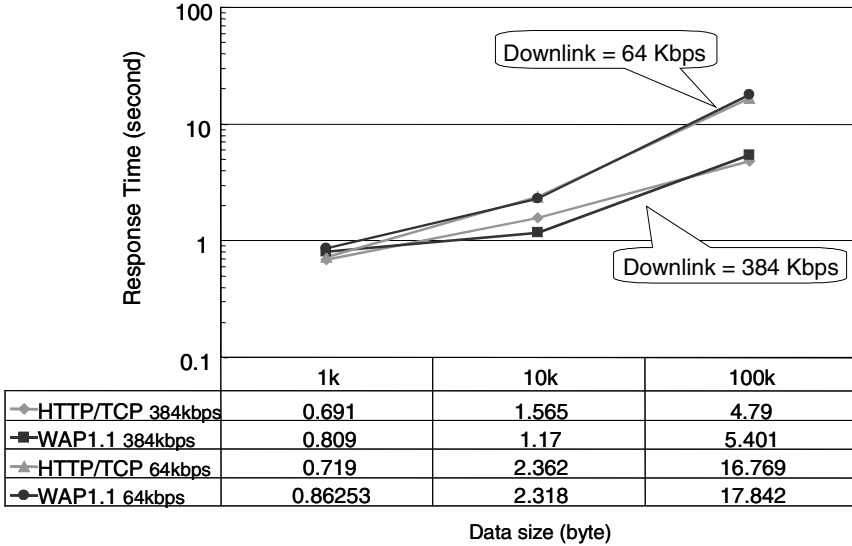


Fig. 6. The response times of WAP 1.1 and HTTP/TCP. Note that both axes have logarithmic scale.

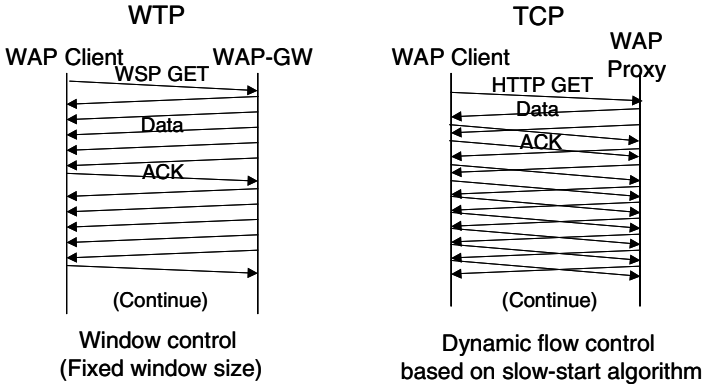


Fig. 7. In this example of WTP usage, there is waiting time after sending data packets until the ACK packet is received. The WAP-GW sends five packets continuously and waits for ACK packet reception since this example assumes that the receiver window size is 32 Kbytes (the number of packets is calculated by dividing the receiver window size by the MTU size). In this example of TCP usage, the ACK packet is piggybacked. The window size is not extended during at the beginning of communication because of the slow start mechanism.

From the results of our evaluation, we feel that TCP is applicable for high-speed wireless networks such as 3G since they are more likely to transmit large content sets such as multimedia data files.

6 Proposed Mobile Internet Architecture for 3G Wireless Networks

As mentioned in sections 4 and 5, WAP 1.x is appropriate for low-speed networks such as 1G and 2G since WAP 1.x has several encoding technologies to reduce the amount of data size and it also offers protocol optimisation. However, HTTP/TCP offers better performance over high-speed wireless networks such as 2.5G and 3G. Accordingly, our mobile Internet architecture, intended for high-speed wireless networks such as 2.5G and 3G (Figure 8), that adopts HTTP and Wireless Optimised TCP (W-TCP) [8], while TLS is used to achieve transport layer security. XHTML [11] is adopted as the markup language because it is expected to become the next generation standard markup language.

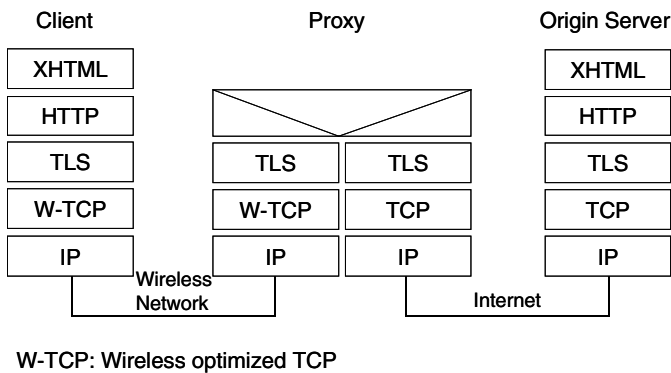


Fig. 8. Proposed mobile Internet architecture for 3G wireless networks. The proxy terminates the wireless-optimized TCP and acts as a proxy. TLS is used to achieve the end-to-end security needed to access the origin server directly through the proxy.

Since the WAP 1.x protocol is different from what is used in the Internet, it is theoretically unable to provide end-to-end security. If, however, we adopt HTTP/TCP, TLS can be used so the client can communicate with the origin server by using TLS tunneling through the WAP proxy.

One of the major reasons for the success of i-mode² is that it supports an HTML subset as the markup language so that it is compatible with the contents of the Internet. We think that it is easy to shift from HTML to XHTML because it is easy to create XHTML contents by anyone who has some skill in HTML content generation because XHTML is just a rewrite of HTML following the XML syntax. Moreover, XHTML is more flexible and extensible than HTML since it can be extended based on XML.

² For further information, please see (<http://www.nttdocomo.co.jp/english/>).

7 Conclusions

We investigated the performance of WAP 1.x by comparing it against the Internet protocol. Our result has shown that both protocols have comparable performance (i.e. response time) except when transmitting large content sets (e.g. multimedia data files), in which case HTTP/TCP offers better performance than WAP 1.x. We have also evaluated some WAP-specific functions such as the binary encoding of WAP headers and contents. Binary encoding is effective for small content sets, but its effectiveness is questionable for large content sets. Therefore, we concluded that the Internet standard protocol (i.e., HTTP/TCP) is suitable for high-speed wireless environments. We proposed a mobile Internet architecture for high-speed wireless networks such as 2.5G and 3G.

We have submitted our proposal to the WAP forum, which adopted it in August 2001 as the next version of WAP 1.x, called WAP version 2.0 (WAP 2.0) [1]. It was developed for high-speed wireless networks such as 2.5G and 3G, and its goal is to achieve Internet convergence.

Future research should look at fourth-generation (4G) mobile Internet services and protocols. Since 4G networks can support up to 100 Mbps, the services provided over such high-speed networks must be changed drastically. Considering the applicability of WAP 2.0 or other Internet-based protocols to the 4G wireless networks will be a future task and we plan to investigate the future mobile Internet architecture and applications.

References

- [1] WAP Forum Specifications, <http://www.wapforum.org/>.
- [2] P. Chaudhury, et al., "The 3GPP Proposal for IMT-2000," IEEE Communications Magazine, December 1999.
- [3] M. Metter, et al., "WAP enabling existing HTML applications," Proceedings First Australasian User Interface Conference, January 2000.
- [4] S. Gordon, et al., "Analyzing the WAP Class 2 Wireless Transaction Protocol Using Coloured Petri Nets," Proceedings of the 8th International Aerospace Congress incorporating the 12th National Space Engineering Symposium, September 1999.
- [5] Thomas Kunz, et al., "WAP traffic: Description and comparison to WWW traffic," Proceedings of the Third ACM International Workshop on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM 2000), August 2000.
- [6] 3rd Generation Partnership Project (3GPP) "RLC protocol specification," September 2001.
- [7] R. Fielding, et al., "Hypertext Transfer Protocol - HTTP/1.1," RFC 2616, June 1999.
- [8] H. Inamura, et al., "TCP over 2.5G and 3G Wireless networks," Internet-draft, February 2002.
- [9] S. Floyd, et al., "An Extension to the Selective Acknowledgement (SACK) Option for TCP," RFC 2883, July 2000.
- [10] M. Allman, et al., "Increased TCP's Initial Window," RFC 2414, September 1998.
- [11] World Wide Web Consortium (W3C), "XHTML 1.0: The Extensible HyperText Markup Language," W3C Recommendation, January 2000.