WAP may Stumble over the Gateway

(Security in WAP-based Mobile Commerce)

Niels Christian Juul and Niels Jørgensen

***Abstract*—The key design idea underlying the Wireless Application Protocol (WAP) is to use a gateway at the intersection of the wireless mobile network and the traditional, wired network. The WAP gateway forwards web content to the mobile phone in a way intended to accommodate the limited bandwidth of the mobile network and the mobile phone's limited processing capability. However, the gateway introduces a security hole which may render WAP unsuitable for m-commerce and other security-sensitive transactions and services on the emerging mobile Internet.**

The paper explains the security hole and the gateway-based design that has led to it, including the technical and business considerations underlying the design. A number of ways to correct the situation are discussed, including a complete re-design of WAP as proposed for the future version 2.0 of the protocol.

***Index Terms*—WAP, gateway, Internet, end-to-end security, protocols, mobile commerce.**

I. Introduction

WHEN a customer places an order with an e-merchant, sensitive information is exchanged with the merchant, typically including credit card number, delivery address, etc. If there is a risk that the privacy of this data will be violated anywhere in between the parties, the customer is not likely to engage in this form of e-commerce, and as a consequence the e-merchant is not likely to invest in the technology either.

If the customer uses a mobile phone and the *Wireless Application Protocol (WAP)* [1], the privacy of the data is in fact not guaranteed. Even when encryption is used in accordance with WAP's security protocols [2-4], the WAP gateway constitutes a security hole since, inside the gateway, the data is transmitted in its original, un-encrypted form. What WAP fails to provide is end-to-end security, which is defined as a secure communication channel between the two parties on top of a potentially insecure network. In WAP, there is a point (the gateway) between the two end points (the customer and the merchant) where the data may be compromised. This critique pertains to version 1.2.1 of the WAP standard, from June 2000.

The WAP gateway is a piece of software. Typically, it runs on a computer in a building under the control of the mobile service provider, MSP. Figure 1 illustrates the role of the mobile service provider as an intermediary in a WAP-based e-commerce transaction. Specifically, the security weakness of WAP discussed in this paper means that all data exchanged may be available to people with privileged access to the WAP gateway. for example, a system administrator of the machine that the WAP gateway is running on. Thus, the privacy of the data depends on such things as the internal security policy of the mobile service provider, the methods used to grant computer access to technical staff, etc.

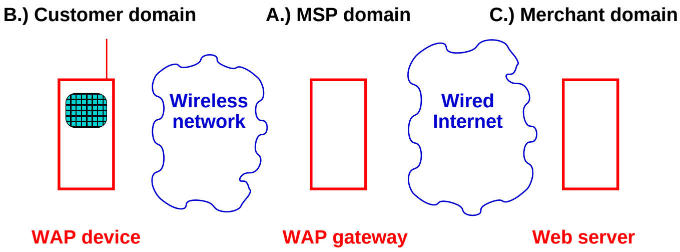


Fig. 1. The three parties involved in a generic m-commerce transaction using WAP: (A). The mobile telephone company (referred to as the mobile service provider, MSP) uses a WAP gateway to connect between the wired and wireless Internet. (B). The customer uses a WAP-enabled mobile phone. (C). The e-merchant's web site is connected to the fixed Internet. The networks illustrated are simplifications; the first is a mostly wireless, mobile access network, whereas the second is a fixed, mostly wired Internet.

The existence of the security hole sketched above is commonly recognized (see e.g. [5]). In this paper we attempt to analyze various strategies to avoiding the security hole, taking into account the technical and business rationales that led to the gateway-based design and the associated security weakness, in particular the technical rationale which is based on the limitations of the wireless network and the mobile phones.

There are three major remedies to WAP's breakage of end-to-end security:

(1) Putting the gateway inside the "vault". The WAP gateway can be placed at the web server end of the connection, that is, inside the same security zone. When residing inside the local network of the merchant the gateway is protected against the outside world in a similar fashion to the way the web server is protected.

We argue that this solution to the security problem conflicts with one of the fundamental purposes of the WAP gateway, namely to convert between two distinct protocol suites, one for the wireless network (WAP) and one for the traditional wired network (the Internet protocol suite, including HTTP and TCP). One crucial point of difference between these two stacks is the protocols used for transport. Assuming that the wired and wireless networks are sufficiently different to justify the use of the different transport protocols, the gateway should be placed at the intersection of the two networks, so that transportation over both networks uses the appropriate protocols.

(2) Application level security on top of WAP. This amounts to introducing security at a software layer above WAP, and considering WAP merely as a potentially insecure communication means. Instead of using WAP's security features, security is taken care of by means of dedicated software running at the two "ends", the mobile phone and the e-merchant's web server.

While technically possible, this approach of neglecting the existing security features of WAP (and in fact also those of

Niels Christian Juul and Niels Jørgensen are with the Department of Computer Science, Roskilde University, PO Box 260, DK-4000 Roskilde, Denmark. Fax +45 4674 3072.

Niels Christian Juul may be reached by e-mail: ncjuul@acm.org, URL: http://www.ruc.dk/~ncjuul, or phone +45 4674 3860.

Niels Jørgensen may be reached by e-mail: nielsj@ruc.dk, URL: http://www.ruc.dk/~nielsj, or phone +45 4674 3702.

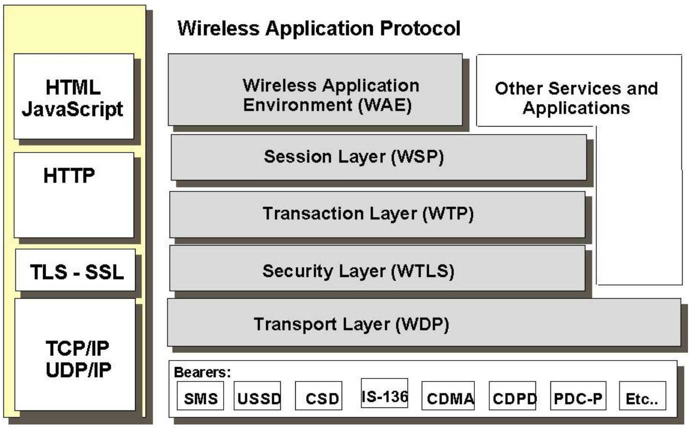


Fig. 2. The WAP vs the Internet Protocol Stack. (Source: WAP Forum)

the Internet protocols) neutralizes most of the optimization provided by the WAP gateway. This includes loosing the benefits of the data compression taking place in the gateway to accommodate the limited bandwidth of the wireless network. Moreover, the approach may be difficult to standardize, a significant obstacle to its widespread acceptance.

(3) The third and last approach is to re-design the WAP protocol to not use a gateway, and employ the existing Internet standards, including the transport protocol (TCP), for the entire wired and wireless part of a connection. By definition, this solves the security problem introduced by the gateway.

This change of design has been proposed by the WAP Forum for the future version 2.0 of the WAP protocol. It constitutes a fundamental change of design which does away not only with the security problem, but also the optimization for the wireless network, made possible by the gateway, and the gateway's potential for integration with other mobile telephone services. In addition, the change creates compatibility problems.

The remainder of the paper is organized as follows:

An overview of WAP is given in Section II, followed by an explanation of the gateway security hole in Section III. The security solution based on putting the WAP gateway inside the vault of the e-merchant is discussed in Section IV, while applying end-to-end security at the application level is discussed in Section V. Section VI concludes and discusses the fundamental re-design of WAP proposed by the WAP Forum.

II. The gateway-based design of WAP

The Wireless Application Protocol (WAP) [6, 1, 7][1] is a suite of evolving[2] standards for browsing the web with a thin client browser, e.g. a mobile phone. The standard describes a full suite, sometimes referred to as a "stack" of protocols, basically in compliance with the ISO-OSI model for network protocols [8]. The protocol stack of WAP is compared with the Internet protocol stack on Figure 2. Bear in mind that, although illustrated side by side, each protocol layer does not communicate across the two stacks. The obligation of the gateway is to utilize the two stacks and converts between the two protocols HTTP and WSP at the top. This section introduces the WAP standard with particular emphasis on the WAP gateway.

```
<wml>
<!-- A deck with two cards -->
<card id="cc1" title="CCPayment">
  <p align="center"><b><big><big>
    Payment
  </big></big></b></p>
  <p>
    Enter your CreditCard #
  </p>
  <p>
    <input name="cc" title="number"
     format="NNNNNNNNNNNNNNNN"/>
  </p>
  <p><anchor title="Next"><go href="#cc2"/>
    Next
  </anchor></p>
</card>

<card id="cc2" title="CCVerify">
  <p align="center"><b><big><big>
  <p align="center">
    Please verify your credit card number is:
  <b></b></p>
</card>
</wml>
```

Fig. 3. A WML code fragment for a credit card payment. The fragment is a deck with two cards. The first card asks for the credit card number, and stores it as the value of the variable "cc". When the user clicks "Next", the second card will be displayed, asking the user to confirm the number. (Among the things left out is code in the second card for comparison of the two numbers, and for moving on to further cards e.g. for sending the credit card number.)

A. *The technical rationale for the WAP gateway*

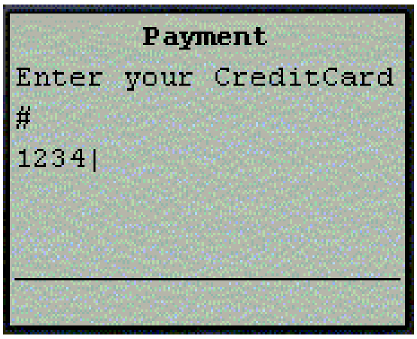
In addition to mobile phones - our prototype example of a WAP device - the WAP standard aims at other hand-held digital devices such as pagers, two-way radios, smart-phones, etc. These devices are characterized by their limited capacities of: processing power, storage capacity, input/output (key-pad and display), and power (battery capacity).

WAP was designed to work not only with GSM but most other digital wireless telephone networks. Some bearers are illustrated on Figure 2. Compared to the well-known Internet, mobile wireless networks are characterized by: limited communication capacity (bandwidth), higher latencies, higher variation in packet-loss (jitter), and variation in long-term connectivity/availability (on/off).

WAP's only requirement pertaining to a document to be delivered from a web server to a mobile phone is that the document is written in WML (Wireless Markup Language). WML is WAP's replacement of HTML and designed specifically for the small and diverging displays of the mobile phone. A fragment of a WML document that could be part of an e-merchant's software for mobile commerce is shown in Figure 3. The WML document asks the user to provide a credit card number, using elements of WML that correspond to a so-called form in HTML. A WML-document is shown on the display of the WAP

[1] The reader is referred to the web site of the WAP Forum (`www.wapforum.org`) for the full list of standard documents, both approved and proposed standards.

[2] The standards discussed here are the latest approved set of WAP standards, version 1.2.1 of June 2000. Whenever the emerging proposals for the coming version 2.0 WAP standard is referred to, an explicit notion of version and current approval status of the document is noted in the text.



Payment

Enter your CreditCard

#

1234|

Fig. 4. A WML browser showing the first card of Figure 3. The user has typed in 1234. If the credit card number is sent to the e-merchant using WAP's security protocol it will be compromised at the WAP gateway.

device using a WML browser, WAP's equivalent of an ordinary (HTML) browser. Figure 4 indicates how a mobile phone may display part of the WML-code of Figure 3.

A WML document is ordered and sent in a request-reply cycle in which the roles of the WML browser, the gateway, and the web server are as follows (the reader may wish to refer again to Figure 1):

The request:

The user of the mobile phone orders information by clicking on a link which is shown on the display by the WML browser. This action spawns a so-called WSP (Wireless Session Protocol) request sent from the phone to the WAP gateway. The WSP request parallels the HTTP request sent when an ordinary browser requests a document from a web server. In the gateway, the WSP request is converted to a standard HTTP request which is send to the web server.

The reply:

The web server's response to the request is the action of sending the WML document to the WAP gateway. After some processing in the gateway (see below), the modified document is sent to the mobile phone.

In order for the WAP gateway to help dealing with the low-capacity wireless **network connection**, it has the following functions:

1. **Switching between transport protocols**

   The gateway in it's communication with the mobile phone uses a transport protocol (WDP, for Wireless Datagram Protocol[3]), which is of a so-called connection-less and unreliable nature. In contrast, the transport protocol used by HTTP on the ordinary Internet is the connection-oriented and reliable protocol, TCP. In TCP, there is additional communication to ensure that all data actually gets through, and if not, it is re-transmitted. Omitting this in the wireless part of the connection incurs a risk of losing data but saves a significant amount of bandwidth.

2. **Compression**

   The gateway compresses the WML document before sending it to the mobile phone. The gateway applies so-called loss-less data compression, which means that the same information can be sent in a format that occupies fewer bites. In comparison, compression of text documents is not used by default on the ordinary Internet.

In addition, to accommodate the limited capacity of the **mobile phone**, the gateway also has the following functions:

3. **Compilation**

   If a WML document contains embedded source code, the gateway compiles such code into a so-called bytecode format, something that relieves the mobile phone of the task of parsing the code. WML documents may contain code written in the scripting language WML-script, which is similar to the JavaScript language which may be embedded inside HTML documents. Such code is executed inside the browser. For example, it may be used for validation, say, checking that a credit card number contains a certain number of digits and allowing the user to correct it before sending the information to the web server.

4. **Decompression**

   The gateway reads and interprets the original WSP request, which is in a compact form that cannot be understood by the web server. The web server understands lengthy HTTP requests in which various commands are given as ordinary text rather than by numerical codes. Also, the gateway translates a symbolic Internet address into an IP number. The gateway does this in the usual way by communicating with a DNS (Domain Name Service) server. Relieving the mobile phone from the use of lengthy HTTP-style commands and performing DNS look-ups also reduces the amount of data that must be sent from the mobile phone.

Also, local validation of input inside the mobile phone (3) and avoiding DNS look-ups from the mobile phone (4) serve to eliminate request/reply message cycles between the mobile phone and the Internet. Thus, these features of the gateway also reduce consumption of the scarce network bandwidth of the wireless connection.

While the gateway is essential for most of the techniques employed by WAP to accommodate the limitations of the mobile phone and the wireless network, other means such as various features of WML are independent of the gateway-based design. This includes the splitting of a WML document into "cards" (see Figures 3 and 4). The user browses from one card to the next by clicking on links, as when requesting a new document, which can be used to guide the mobile phone user through a sequence of steps in an e-commerce transaction. The advantage is that all cards within the same document (or "deck") are transferred in the same request/reply cycle. Without the division of decks into cards, the customer would not be able to move on to the next step immediately, but would have to wait while it is being retrieved in the next request/reply cycle across the high-latency wireless network.

B. The business rationale for the WAP gateway

In addition to the technical rationale, there is also a business rationale underlying WAP's gateway-based design. In particular, the gateway may open new business opportunities for the

[3] For the purposes of this discussion WDP is identical to the Internet's UDP transport protocol

mobile service provider. The gateway's impact on business, and usability in a wider sense, may be summarized as follows from the point of view of the mobile service provider, the mobile customer, and the e-merchant.

A. The Mobile service provider (MSP) may use the gateway to tie its mobile customers to a portal-like access point to the Internet. Such portals are well-known on the fixed Internet, where they are provided to private/home-based users of the Internet by Internet Service Providers such as America On-line and Tiscali/World On-line. There are, however, two important aspects where a mobile portal may differ from an ordinary Internet-portal: First, WAP services may be combined with the MSP's basic phone services. (For the integration into WAP of basic phone services there is a dedicated protocol, Wireless Telephony Application Interface (WTAI) [9].) Second, the MSP of any given customer has a special, privileged position in the competition with other service providers available to the customer, because the MSP has access to privileged *location* information about the mobile customer, i.e. the exact cell in which the mobile phone is located. The customer's own MSP may utilize this information as part of location dependent WAP services, e.g. ordering take-away meals from nearby restaurants.

B. The mobile WAP user has a dual interest: On the one hand, she would presumably like to have access to those of her own MSP's services that integrate WAP and ordinary mobile telephone service, as made possible by accessing the Internet via the MSP's gateway. On the other hand, she is clearly interested in the cheapest possible, universal access to all WAP-based services on the Internet.

C. The typical e-merchant has one obvious interest in relation to WAP which we want to emphasize: the interest of minimizing the cost - in terms of hardware, software, maintenance etc. - of extending e-commerce to WAP. For most e-merchants, WAP-based e-commerce is just an additional channel to be added to e-commerce via the web. The chief advantage of the gateway is that the e-merchant can open this channel merely by hosting a new type of documents - WML documents - because the adaptation to the special requirements of the wireless network and devices are taken care of by the gateway.

As an aside, we note that the development of the WAP standard has been initiated, not by representatives of any of the above parties, but by producers of hardware and software for mobile devices and networks: Nokia, Motorola, Ericsson, Phone.com (previously Unwired Planted), and others. The main interests of these producers may be identified as selling hardware and software to the mobile service providers (and the companies that own the mobile networks), and mobile phones and other devices to the users. In turn, the prospects for this is of course highly depending on a successful integration with the Internet, including e-commerce.

III. The security hole at the WAP gateway

In general, the mobile customer and the e-merchant involved in an m-commerce transaction want (or should want) to ensure:

Confidentiality: messages are kept secret.

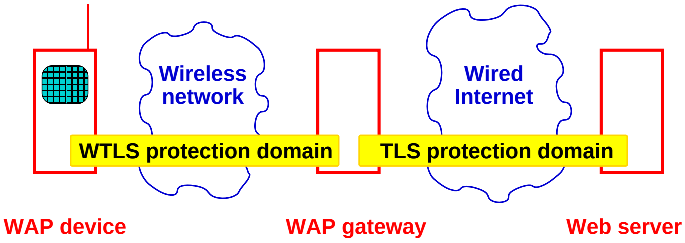


Fig. 5. Security zones using standard security services (WTLS and TSL)

Authentication: each party knows who the other party is.

Message integrity: messages are passed unaltered from sender to receiver.

Prevent replay attack: any unauthorized re-sending of messages is detected and rejected.

Non-repudiation: neither parties can later reject that the exchange took place.

All these issues must be addressed in a secure system. Indeed, it is the ambition of the WAP Forum to develop WAP into a standard that covers all relevant aspects of security, and many steps have been taking already with version 1.2.1.

For security, WAP provides a secure protocol for data transport: WTLS, Wireless Transport Layer Security [2]. WTLS also contains features for authentication of both parties, as well as for non-repudiation using message digests and digital signatures. Authenticating of the user of a GSM phone may utilize the phone's SIM card [4]. Finally, the definition of WML-Script includes a specification of a function library called a “crypto package” [3]. As of WAP version 1.2.1, this package is extremely small and specifies only a single function, but assuming the package is extended in the future it may be useful for developing secure applications on top of WAP, in the style discussed in Section V.

In the sequel, we merely discuss to what extent WAP achieves message confidentiality.

A. WAP's two-stage security model

Consider a WAP-based m-commerce transaction in which a document, say `www.weSellIt.com/orderNow.wml`, must be transferred to the mobile phone in a secure manner.

The basic security feature of WAP provides secrecy in the two half parts of the path that connects the WAP client and the web server: the (presumed) wireless path between the WAP client and WAP gateway, and the (presumed) wired path between the gateway and the web server (Figure 5).

In the middle point constituted by the gateway, incoming data is decrypted; then while the data is in its original, un-encrypted form, it is subjected to some processing; and finally it is encrypted again before it is sent off along the other path. The processing done on the un-encrypted data corresponds to the protocol layers above the security layers on Figure 2 in the gateway.

For encryption over the wired path, WAP simply relies on the Transport Layer Security (TLS) protocol [10], a widely used Internet standard. TLS is standardized by the Internet Engineering Task Force, and is also known under the name Secure Socket Layer (SSL) given to it by Netscape, the company that developed the standard originally.

For encryption over the wireless path, WAP uses WTLS ([2]) which is in essence an adoption for wireless communication of the TLS protocol. The changes to TLS embodied in WTLS do not weaken security.

WAP's topmost protocol, the Wireless Session Protocol (WSP)[11], initiates and links the two secure halves of the connection. The major steps in this are as follows: The user of the WAP client selects the URL `shttp://www.weSellIt.com/orderNow.wml`. This tells the WSP layer at the WAP client to initiate the setting up of a WTLS connection to the WAP gateway, and then pass a WSP request (the equivalent of an HTTP request) over the connection for the particular file. Thus prompted by the WAP client, the WSP layer at the gateway will initiate a TLS connection to the web server, in a way completely similar to setting up a connection between an ordinary web client and the server.

This combination of WTLS and TLS provides secrecy (indeed, also integrity) over both halves of the WAP client / web server connection. The crucial weakness is, of course, that all data transferred between the WAP client and the web server is decrypted at the WAP gateway, i.e., all data such as credit card numbers, etc. exists as free text in the memory of the gateway. Technical solutions, such as various programming techniques applied to the software that implements the WAP gateway, can make it somewhat difficult to get access to the data, but not impossible. Organizational solutions, such as tightening the security policy of the organization that hosts the WAP gateway, may limit access to the gateway and it's data; but from the point of view of the two "end users" it is unsatisfactory that the privacy of their data is not under their own direct control.

B. *Encryption vs. the functions of the gateway*

The breakage of end-to-end security at the WAP gateway is not an accidental error. Rather it is merely a disadvantage of the gateway-based design which the designers felt was outweighed by its advantages. To argue this, before we discuss (in Sections and IV and V below) how the gateway-based design can be circumvented, let us consider again the technical rationale underlying that design. The question is, which of the functions listed in Section II could still be fulfilled by the gateway had the WAP standard prescribed a different approach to confidentiality, one that attained end-to-end security as provided by TLS on the ordinary web ?

We believe the answer is that it would probably be possible to use the gateway as a point of switching between the two transport protocols TCP and WDP (cf. function 1), but that it would not be possible for the gateway to employ compression of WML or compilation of WML-Script, nor for the mobile phone to use the bandwidth-saving WSP requests (cf. functions 2, 3, and 4).

To introduce end-to-end encryption while retaining the gateway's function as the point of switching between the transport protocols would be possible by letting the gateway act as a proxy for the WAP client at the level of TCP, similarly to the way it already acts as a proxy at the level of HTTP. A deeper discussion of this strategy is beyond the scope of this paper.

The remaining functions of the gateway cannot be achieved if end-to-end encryption is employed. Clearly, if code written in WML-Script is encrypted, it cannot be compiled, unless the gateway can first decrypt the script, thus breaking the end-to-end encryption. For the same reason, it would not be possible for the gateway to understand (decompress) the encrypted WSP requests, so the mobile phone would have to use the much higher latency-incurring and more bandwidth-consuming HTTP requests. Finally, the gateway's compression of ordinary WML documents is ruled out: If the gateway cannot recognize, for example the WML tags `<card>` and `</card>`, it cannot compress them (by replacing them with numbers that consume fewer bytes). In general, the characters of (well) encrypted text are randomly distributed, i.e. with no apparent patterns, and therefore such data cannot be compressed. (For a discussion on encryption versus compression, see [12])

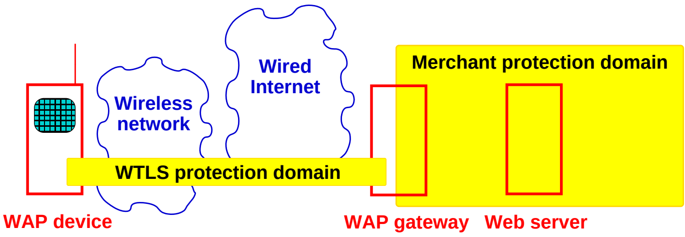


Fig. 6. Security zones by replacement of the WAP gateway. The customer may utilizes a dial-up point at the MSP and continue across the Internet, or utilizes the Merchants own dial-up point bypassing the Internet. The secure channel is protected by WTLS in both cases.

Thus modifying the WAP standard to provide full end-to-end encryption as in TLS conflicts with the compilation, decompression, and compression functions of the gateway that serve to limit the amount of data that has to be transmitted over the wireless network, although, on the other hand, it does not prohibit the use of the fast WDP protocol on that part of the connection.

IV. Moving the WAP gateway to the web server

The first way that one can use the existing WAP standard, but escape the breakage of end-to-end security caused by the gateway, is to move the gateway to the web server end point as show on Figure 6. The security chain is broken in the gateway, but this would only be harmful if the domain of the e-merchant was already insecure, and so does not decrease security.

A. *The Mobile Service Provider*

For the Mobile Service Provider the main problem is a possible loss of business opportunities, e.g. "locking in" the customer to the provider's m-portal. There are a number of possible WAP-services that the MSP can offer only if it hosts the gateway, this being the MSP's only way of ensuring that all data used in those services stays within the mobile network that the MSP is in control of. For instance, location dependent services utilize the MSP's access to information from its own network about the exact location (cell) of the mobile phone. There may be strong business and security reasons for the MSP to not want such data to be available outside of the network. In the case of location information, the data is clearly sensitive, at the same time, the data may have a high business value, exactly because it is a prerequisite for certain services.

B. The User

The user of the WAP client may witness degraded performance: The WAP protocols, which are tailored for the characteristics of wireless networks, are now used for the entire transport of the web content to the WAP client, instead of merely for the wireless part as intended. This may incur increased delays if there is congestion in the wired network.

The end-user interface becomes less friendly, because the user of the WAP client will be forced to swap between gateways during Internet browsing. For example, the user that wants to buy from two distinct web sites needs to use the WAP gateways of those sites (given that both transactions are secured by the method discussed in this section).

Swapping gateways raises two problems when changing the gateway profile of the phone. The gateway profile includes a dial-up phone number, the IP number of the gateway, etc. First, in many current implementations of WAP, the user of a WAP device has to go through a cumbersome procedure that involves clicking through several menus, the typing of an Internet address in the form of an IP number, etc. Future WAP implementations may provide features for easy switch of gateway, however it seems that the user should, at least, be asked to confirm a gateway switch. Second, different gateways may have slightly different properties themselves. Although standardized by the WAP Forum, current gateways perform differently on the same task.

The option of the mobile service provider running a kind of default gateway - one that the user may switch back to, after using a particular e-merchant's gateway - is not as simple as it may appear at first. A default gateway either requires that the client must explicitly “log out”, or alternatively, a “time out” mechanism must be included in the WAP device, letting it switch back automatically.

C. The Merchant

The e-merchant is burdened with a complex piece of additional software (the gateway) that must be acquired and maintained. For a WAP gateway, maintenance work includes, for example, ensuring that the security-related software in the gateway is up-to-date, and updating the gateway to new versions of the WAP protocols. Furthermore, location based services, micropayment, etc. may not be available to the m-commerce application neither from the MSP-detached gateway, nor from the short-cut MSP directly.

V. Independent end-to-end security

An alternative method of achieving end-to-end security is for the WAP client and web server to negotiate and apply appropriate security measures independently of WAP. Since this can be done with the gateway in place at the wired/wireless cross over, it does not introduce the possible disadvantages discussed in the previous section.

In order to achieve end-to-end security between the WAP client and the web server as shown in Figure 7, application level encryption software must be available to both the WAP device (at the WAE level) and the web server (in the HTTP implementation).

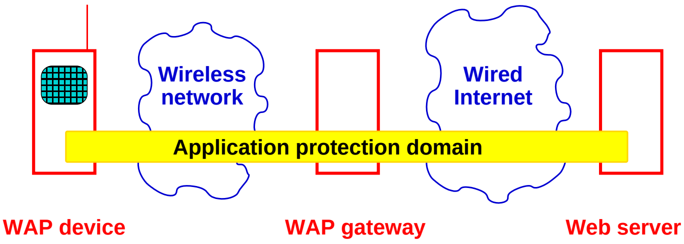


Fig. 7. Security zones using application level encryption

The easiest way to implement this form of security is for the e-merchant to provide WML-Scripts that are executed on the WAP client using the crypto library [3] of WML-Script. In order for this kind of application-level security to be consistent with WAP's basic philosophy, only small fragments of the data should be encrypted, not the data as a whole. This is because for the WAP gateway to perform compression, decompression, and compilation it must have access to the WML tags and WSP/HTTP commands in their original, un-encrypted form as discussed in Section II.

If this approach is feasible, it has the additional advantage of relieving the WAP client of the task of encryption/decryption the entire bulk data, as required when WTLS is used. This is in line with the conjecture about protocol design made in [13]: that an end-to-end function must be placed at a level where end-to-end control is available, i.e., at the application level.

There are two problems with the approach:

The first is that the (single) function currently specified in the crypto library is simply not powerful enough to implement the required functionality on the client side. However, it is likely that the crypto library will be extended in future versions of WAP.

The second problem is that the secrecy attained depends on a (WML-Script) program provided by the individual e-merchant. Users of the ordinary web are already reluctant to run foreign code on their computers. Indeed, to run inside a web browser a JavaScript provided by some web server can by harmful. It requires even more confidence in the provider of foreign code to trust that the code is capable of establishing a fully secure communication channel. In contrast, for the customer engaging in e-commerce over the ordinary Internet, it is possible to verify that STL-based end-to-end encryption is used and to obtain, if desired, information about exactly what kind of security is attained by that approach. Thus, for application-level security on top of WAP to become useful, it may be necessary to develop some further means to enable the customer to verify that an acceptable level of security is in place - either in the form of a standard, some kind of “branding”, or otherwise.

VI. Conclusion

The WAP Forum has recently released its proposal for the next major version (2.0) of the WAP standard [7, 14]. The proposal includes turning the standard in to what is really a set of alternative standards, including one alternative that corresponds to the current version 1.2.1 of WAP discussed in this paper, and a new, alternative approach that does not use a gateway at all.

Discarding the WAP gateway is the third and final remedy to the security hole associated with it. It would make it possible to attain the same high level of security for an m-commerce transaction as that of an e-commerce transaction on the ordinary web using full end-to-end encryption. Indeed, for WAP to discard the WAP gateway would turn the (fully) WAP-enabled mobile phone into an ordinary Internet device. The fact that the WAP Forum proposes this strategy (at least as one alternative) seems to confirm the critique proposed in this paper. There are inherent difficulties associated with the two other approaches that both build on the current WAP standard: to place the WAP gateway at the web server end of the connection, or to use application level security on top of WAP.

Discarding the WAP gateway implies, of course, a loss of the optimization it provides. Even when more powerful mobile phones are developed, and higher bandwidths are attained in future wireless networks, high latency will remain, and so does the relevance, in particular, of the idea underlying WAP's WSP protocol that all data should be obtained by the mobile phone in a single request/reply cycle over the wireless part of the network connection.

The WAP Forum has not released the considerations that actually led to the recent re-design that does away with the WAP gateway. Similarly, although the security hole associated with the gateway is commonly recognized, the WAP Forum did not release its previous deliberations as to why it felt the gateway's advantages would outweigh its disadvantages.

The introduction and subsequent deroute of the WAP gateway - whether preconceived or not - has increased the complexity of the WAP standard, comprising variations both with and without the gateway. The WAP standard has become more difficult to standardize, and difficult to implement for providers of software for mobile phones, e-merchants and other content providers.

It is interesting though, that once again has the old Internet technology beaten a vendor provided network solution. One may hope that the significance of openly discussed standardization will be learned also by the vendors behind the semi-clossed WAP Forum.

VII. Acknowledgment

An equipment grant from Siemens and a WAP/GSM-air-time grant from the National Danish Telecommunication Company, TDC are greatly acknowledged. We are in debt to Torben Lauritzen of Siemens for many fruitful discussions on WAP. This research was also partly funded by the IT-University of Copenhagen and The Research Center for Electronic Commerce at the Copenhagen Business School.

References

[1] WAP Forum, "WAP Architecture: Wireless Application Protocol Architecture Specification," Approved Specification (Part of the June 2000 (WAP 1.2.1) conformance release): WAP 100, WAP Forum Ltd., URL: `http://www.wapforum.org/`, Apr. 1998.

[2] WAP Forum, "WAP WTLS: Wireless Application Protocol Wireless Transport Layer Security Specification," Approved Specification (Part of the June 2000 (WAP 1.2.1) conformance release): WAP 199, WAP Forum Ltd., URL: `http://www.wapforum.org/`, Feb. 2000.

[3] WAP Forum, "WMLScript Crypto Library: Wireless Application Protocol WMLScript Crypto Library Specification," Approved Specification (Part of the June 2000 (WAP 1.2.1) conformance release): WAP 161, WAP Forum Ltd., URL: `http://www.wapforum.org/`, Nov. 1999.

[4] WAP Forum, "Wireless Application Protocol Identity Module Specification (WIM) part: Security," Approved Specification (Part of the June 2000 (WAP 1.2.1) conformance release): WAP 198, WAP Forum Ltd., URL: `http://www.wapforum.org/`, Feb. 2000.

[5] Sandeep Singhal, Thomas Bridgman, Lalitha Suryanarayana, Daniel Mauney, Jari Alvinen, David Bevis, Jim Chan, and Stefan Hild, *WAP - the Wireless Application Protocol Writing Applications for the Mobile Internet*, ACM Press, 2001.

[6] Wireless Application Protocol Forum Ltd., Ed., *Official Wireless Application Protocol: The Complete Standard with Searchable CD-ROM*, Wiley Computer Publishing, 2000.

[7] WAP Forum, "WAP Architecture: Wireless Application Protocol Architecture Specification," Proposed Version (Intended as part of WAP 2.0 release): WAP 210, WAP Forum Ltd., URL: `http://www.wapforum.org/`, Oct. 2000.

[8] Andrew S. Tanenbaum, *Computer Networks*, Prentice Hall, third edition, 1996.

[9] WAP Forum, "WAP WTAI : Wireless Application Protocol Wireless Telephony Application Interface Specification," Approved Specification (Part of the June 2000 (WAP 1.2.1) conformance release): WAP 170, WAP Forum Ltd., URL: `http://www.wapforum.org/`, July 2000.

[10] T. Dierks and C. Allen, "The TLS protocol version 1.0," RFC 2246, Internet, URL: `http://www.rfc-editor.org/rfc/rfc2246.txt`, Jan. 1999.

[11] WAP Forum, "WAP WSP: Wireless Application Protocol Wireless Session Protocol Specification," Approved Specification (Part of the June 2000 (WAP 1.2.1) conformance release): WAP 203, WAP Forum Ltd., URL: `http://www.wapforum.org/`, May 2000.

[12] Shmuel T. Klein, Abraham Bookstein, and Scott Deerwester, "Storing text retrieval systems on CD-ROM: compression and encryption considerations," *ACM Transactions on Information Systems*, vol. 7, no. 3, pp. 230–245, July 1989.

[13] J. H. Saltzer, D. P. Reed, and D. D. Clark, "End-To-End Arguments in System Design," *ACM Transactions on Computer Systems*, vol. 2, no. 4, pp. 277–288, Nov. 1984, Revised version of a paper from the Second International Conference on Distributed Computing Systems, Paris, France, April 8-10, 1981, pp. 509-512.

[14] WAP Forum, "WAP TLS Profile and Tunneling WAP-219-TLS: Wireless Application Protocol TLS Profile and Tunneling Specification," Proposed Version (Intended as part of WAP 2.0 release): WAP 219, WAP Forum Ltd., URL: `http://www.wapforum.org/`, Apr. 2001.

Niels Christian Juul was born in Copenhagen in 1955. He received his Master of Science in Computer Science and Mathematics (1988) and PhD in Computer Science (1993) from University of Copenhagen, Denmark. His has researched *distributed systems* at University of Copenhagen, University of California, Riverside, and Copenhagen Business School. He is currently Associate Professor at the Department of Computer Science at Roskilde University in Denmark. He is a member of ACM, IEEE Computer Society, USENIX, DISP, CPSR, and IFIP.

Niels Jørgensen is an Associate Professor at the Department of Computer Science at Roskilde University in Denmark. Besides distributed systems and security, his research interests include optimized compilation and open source software development.