Content-based control of HTTPs mail for implementation of IT-convergence security environment

Young Ran Hong · Dongsoo Kim

Received: 30 September 2011 / Accepted: 22 May 2012 / Published online: 5 June 2012
© Springer Science+Business Media, LLC 2012

Abstract Convergence among industries has made network data loss prevention (DLP) more important, in that in the IT-convergence environment, web mail is utilized as a means of information delivery. The web mail used in various industries has two facets: convenience for information transfer, and vulnerability to leakage of confidential information. Monitoring blocking and logging of web mail are a few of the major security methods that have been employed for prevention of such leakage. However, application of HTTP over SSL (HTTPs) to web mail systems such as Gmail has revealed the limitations of existing web mail security methods in controlling web mail. Most importantly, the existing method cannot control encrypted contents of web mail. In this paper, we propose a method that controls HTTPs web mail contents by using a proxy server and distributing the secure socket layer (SSL) certificate to user’s PC. The proxy server plays the Certificate Authority role between the users’ PCs and the web mail server, distributing its own SSL certificates to the users’ PCs. The SSL certificate is the key to encryption and decryption of HTTPs web mail contents. Using the protocol derived in the present study, HTTPs web mail contents can be controlled as an effect of content-awareness. Network DLP is effected by monitoring, blocking and logging suspicious mail contents using HTTPs.

Keywords HTTPs · SSL · Web-mail security · Proxy server · Content-aware DLP · SSL certificate

Introduction

In the IT-convergence environment, HTTP web mail has been the preferred method of information delivery owing to two major advantages: convenience, and real-time delivery. These web mail features, however, can be major sources of information leakage in an organization. Nowadays, security is considered as an essential part of various services in many applications, for example, smart grid (Ling and Masao 2011; Hsu et al. 2011). Indeed, the convergence phenomenon in various networks has posed a significant challenge to enterprises’ security systems planning (Kryvinska et al. 2011). Xie et al. (2011) emphasized the importance of secure communication in integrated heterogeneous wireless networks. Despite the limited resources of wireless networks and their devices, this issue looms larger than ever (Imani et al. 2010).

There are two major network threats to organizational security. One is outside intrusion for the purposes of stealing information or disrupting information systems. Anti-virus, the IDS (Intrusion Detection System) and firewalls represent the main technologies that have been developed to respond to this danger. The other major threat is data loss or information leakage from inside to outside via the network. According to the IT-convergence trend in industries, organizations have taken advantage of networks’ availability and convenience for information transfer. This means, however, that there can be more channels opened to internal information leakage (Oppliger et al. 2006). In response, governments have introduced rules and regulations and instituted technological standards for prevention of data loss. Certainly, the necessity of network security, including web mail security, has grown. Organizations themselves have sought to deal with such issues by establishing, often for specific compliance purposes, security policies on web mail.
From the organizational point of view, there are two important rationales for web mail security. First, it contributes to data loss prevention (DLP) by blocking mail content that includes confidential information. Second, it can provide for accountability in the form of logs that, after an incidence of information leakage occurs, can audit the following factors: Who sent this? Who received this? When did it happen? What was transferred through the web mail?

Such mail-content blocking and logging by analyzing the plain text of web mails of the HTTP type has been used to satisfy the two requirements of web mail security (Xia et al. 2005).

Unfortunately, the recent adoption of HTTPs (HTTP over SSL) web mail using secure socket layer (SSL) has revealed the limitations of the existing web mail security methods. For example, Google’s Gmail is a typical HTTPs web mail service. Because it encrypts contents and transfers mail on SSL, it has the merit of being strong against outside attack (Fossi et al. 2010). But at the same time, this advantage is a weak point in web mail security. For an organization’s web mail security, all mail contents should be analyzed for blocking of forbidden contents or for logging full text. However, Xia and Brustonlini (2005) showed that because HTTPs web mail is encrypted, it cannot be analyzed and controlled based on its contents. This encryption technology has been proposed and applied because cryptography has functions for keeping the security properties (Andreeva et al. 2010).

In this paper, we suggest that HTTPs web mail should be analyzed and controlled on the enterprise level by means of the HTTPs proxy server. Our proposed method takes advantage of the vulnerability to MITM (man-in-the-middle) attack in SSL. Generally, SSL MITM employs an improper proxy server acting as a CA (Certificate Authority) in order to hook the ID and password of users. But this vulnerability, as in the present reeach, can be applied to the HTTPs web mail system to decrypt encrypted mail contents for content analysis and, thereby, secure HTTPs web mails. That is, the HTTPs proxy server in an organization plays a CA role in the SSL network and distributes SSL certificates. During this process, the key is exchanged between the PC client and the HTTPs proxy server instead of the web mail server located in a mail company like Google. The SSL certificate from the HTTPs proxy server is activated as a symmetric key in the HTTPs web mail contents. Researchers studying MITM have explored ways in which IDs and passwords can be protected from hooking. However, this research shows that the SSL certificates distributed from the HTTPs proxy server can be used as a symmetric key for the encryption and decryption of mail contents. It shows how the enhanced security in the web mail aspect is implemented by content-based control of HTTPs web mail.

The rest of the paper is organized as follows. “Problem definition of web mail security” section defines the problem of web mail security and presents related research. “Proposed system architecture” section describes the system architecture and methodology introduced herein to solve that problem. “Implementation results” section discusses the results of system implementation. Finally, “Conclusions” summarizes conclusions and looks ahead to future work.

### Problem definition of web mail security

#### E-Discovery

The representative means of web mail security compliance is e-Discovery (Electronic Discovery), a digital proof discovery of an organization. E-Discovery was launched in the US in December, 2006. In the case of a lawsuit, a company should discover the ESI (Electronically Stored Information) as a digital proof discovery. Electronic information differs from the paper variety in that there can be considerably more quantity, for instance in the form of PC word documents, e-mail, instant messenger contents, accounting data, CAD/CAM files, and indeed any kind of data that is stored electronically. If the company fails to submit e-mail data due to file erasure or damage, or does not maintain logs, it is likely to lose its case due to proof failure. This means that all e-mail contents transferred in an organization should be stored in a log format in order to maintain e-Discovery compliance (Ward et al. 2009).

Figure 1 illustrates the general process of e-Discovery. Box A represents the request process, and box B shows the main role of web mail security (Ward et al. 2009).