Chapter 6

HTTP\textsuperscript{v}:
Verifiable HTTP across an Untrusted Channel

Subrata Acharya
Towson University, USA

ABSTRACT

There is a need to be able to verify plaintext HTTP content transfers. Common sense dictates authentication and sensitive content should always be protected by SSL/HTTPS, but there is still great exploitation potential in the modification of static content in transit. Pre-computed signatures and client-side verification offers integrity protection of HTTP content in applications where SSL is not feasible. In this chapter, the authors demonstrate a mechanism by which a Web browser or other HTTP client can verify that content transmitted over an untrusted channel has not been modified. Verifiable HTTP is not intended to replace SSL. Rather, it is intended to be used in applications where SSL is not feasible, specifically, when serving high-volume static content and/or content from non-secure sources such as Content Distribution Networks. Finally, the authors find content verification is effective with server-side overhead similar to SSL. With future optimization such as native browser support, content verification could achieve comparable client-side efficiency.

1. INTRODUCTION

Recent events indicate malicious modification of static HTTP content does occur. For example, the Tunisian government successfully compromised the Facebook, Yahoo, and Gmail accounts of protesters by injecting JavaScript into the front pages of these services Ragen (2011). Even though the authentication was sent over HTTPS and therefore not susceptible to eavesdropping, the injected script sent a copy of the username and password to a government-controlled destination. This attack could have been prevented if the browser could have detected the page was modified in transit before enabling active content such as JavaScript.

Less intrusive modification of pages is also possible: for example, a government controlled Internet Service Providers (ISPs) may censor information by on-the-fly rewriting of HTML containing certain keywords. While such censorship can easily be detected by comparing content from an uncensored source, or by word-of-mouth,
this takes time and a way to access the content in an uncensored fashion. If HTTP content can be verified, the user can immediately know something is amiss.

Censorship and spying issues aside, ISPs have in the past been known to modify HTTP traffic in-transit for revenue purposes (Odvarko, 2008). Two competing technologies, Phorm and NebuAd, both have the ability to track browsing behavior and add or replace advertisement content on webpages with “targeted” advertising designed to generate revenue stream for these ISPs Topolski (2008). This type of activity is not necessarily harmful to end-users, but deprives websites of advertisement revenue. To this effect, Verifiable HTTP could inform sites of the occurrence of this event.

1.1 Dynamic vs. Static Content

For the purposes of this research work, we will place web content into two broad categories: dynamic and static. Dynamic content is generated specifically by the server for the client, and will vary between clients and between sessions. Static content is generated once and does not change. In a website with mostly or entirely public content, much of the content is likely to be static. Examples of such a site would be a website for a restaurant, featuring menus, daily and weekly specials, hours and directions. This information changes infrequently, at most once per day, and often not at all. Even a website with majority user-generated content, such as a blog, forum or social networking, has many static elements.

A user’s home page on any generic user-generated site pulls content from many sources, from dynamic to entirely static. The main content of the page is entirely dynamic, generated specifically by or for the users and changes whenever there is any activity on the site. The profile pictures (representing friends, forum posters, blog commenters, etc.) displayed will vary based on the generated page content, but the individual pictures will only change when a user changes his or her picture. If there is advertising on the site, the ads displayed will change with each page fetched, but all users will get the ads currently in rotation, and the content of each ad does not change. The site’s logo, graphics, navigation menus plus the non-displayed elements such as style-sheets and scripts are the same for all users and do not change. If a visitor to the site is not logged in, they will be directed to a public page asking them to log in or create an account. This serves to illustrate even sites with dynamic content have static pages or page content.

1.2 Static Content vs. Dynamic Delivery Systems

Many sites optimize the serving of static content. This may range from using separate web servers for static content to using the services of a Content Distribution Network (CDN). A CDN is a service that delivers content from a source more optimal to serve multiple clients. This may mean the CDN source is closer to the client in terms of network topology (and therefore often geography), or has higher bandwidth. CDNs often serve multiple customers from the same set of distributed physical resources. In addition to content provider a system, many organizations and ISPs operate caching proxies to locally mirror frequently accessed static content (Davidson, 2008).

1.3 SSL Overview

HTTPS, or HTTP over SSL/TLS, is a mechanism to provide end to end security of the HTTP session. SSL (renamed TLS in more recent versions) is a general purpose protocol for encapsulating application-layer data to provide confidentiality, integrity, and authenticity services (Magnini, 2005). Encryption and message authentication codes provide the former two services, while public-key cryptography and certificate verification provider the latter.