Internet 网络 HTTP

计算机科学 2000Vol. 27№. 7

22-23,42

HTTP 的安全机制

On the Security Mechanism in HTTP

张信明 陈国良 TP393.408

（国家高性能计算中心（合肥） 中国科技大学计算机科学与技术系 合肥 230027）

Abstract This paper analyses the security mechanism in HTTP. It includes basic authentication and its flaws in HTTP/1.0, digest access authentication in HTTP/1.1, proxy authentication, privacy, and cookie's influence on privacy. For those in e-commerce, S-HTTP is also briefly discussed.

Keywords HTTP/1.0, HTTP/1.1, Security mechanism

一、引言

随着 Internet 越来越广泛地深入到世界的各个角落，越来越多的人在使用和浏览 Web。然而作为 Web 基础的 HTTP 安全性尚不尽如人意，商业界、政府部门亦或普通百姓有理由为其机密、敏感信息可能遭泄漏、窃取、串改而感到担忧，因此全面认识 HTTP 的安全机制，对于 Web 相关的各方，无论是一般用户，还是应用程序设计者，或者服务提供者，甚至协议制定者，都非常重要。在 Web 服务器诸多安全措施中，对用户的身份进行鉴别是最简单也是使用最广泛的方法。本文详细介绍了 HTTP/1.0 基本认证机制、HTTP/1.1 摘要访问认证机制、代理认证机制。对于用户而言如若因访问 Web 而泄漏隐私、机密显然是可怕的，为此本文介绍了 HTTP 是如何保护用户隐私的，并剖析了 Cookie 对隐私、安全的影响。

二、基本认证机制

HTTP/1.0 提供了基本认证机制以询问/回答方式来实施访问控制。源服务器用标明认证模式(scheme)、领域(realm)的 WWW-Authenticate 头标对需要认证的请求进行响应。其中领域定义了一个特别的保护空间，该空间仅是资源的一种概念性的划分，每个划分具有独自的认证模式与认证数据库，例如可将一资源划分为普通用户和网络管理员领域，两者具有不同的权限及认证模式。认证步骤是：

(1)客户向服务器发送不带 Authentication 头标的请求；

(2)服务器返回带有 401(Unauthorized)状态码及 WWW-Authenticate 头标的响应，头标包含所需认证类型及参数的询问；

(3)客户(用户 agent)针对特定领域查询用户的用户名、口令等证件，并将这些证件包含在 Authorization 头标中，然后重复原来的请求。

源服务器在认证客户证件是可接受的就响应相应的内容，接着客户可用相同的证件在同一领域内的同一服务器上请求其它资源，这样便可省去多余的询问/回答开销。在客户/服务器认证过程中，任何代理或网关都必须是透明的。

基本认证机制的一个严重缺陷是证件中的用户名、口令没有加密，易于被偷窥。证件也没有有效期限制，别有用心的人可以从容地收集之，并在以后使用。

三、摘要访问认证机制

摘要访问(Digest access)认证机制仍然使用基本认证的基本框架，但克服了许多缺陷。摘要访问认证的消息流、头标与基本认证的消息流、头标相同，但使用了"摘要"模式。在摘要访问认证中，服务器的询问使用了一些包括批次(nonce)的信息。为成功回答，客户必须计算用户名、口令、批次、HTTP 请求方法以及被请求 URI 的校验和(缺省算法为 MD5)。该机制不仅仅对口令加密，而且给定的响应仅与唯一的资源、方法相匹配。进行网上偷窥的攻击者只能重放请求。不过这种攻击方法是无效的，因为重放请求的响应他已经看到过了。与基本认证不同的是，在摘要访问认证机制中已获得的证件并不能访问其它资源。与基本认证相同的是，客户可对同一领域提出进一步的请求，可包括用适宜的请求方法及 URI 计算过的摘要证件。不过源服务

张信明 博士生，副教授，研究方向为计算机网络与网络计算。陈国良 教授，博士生导师。

器的批次值可能是有有效期限制的，批次过期后服务器就拒绝证件并提供一新的批次，此时客户无须再向用户要求用户名、口令即可重新计算证件。

摘要访问认证的新特性是：除了可以直接认证外，通过 Authentication-Info 头标还提供了另外两种特性：支持第三方认证服务器及有限的消息完整特性。

四、代理认证

代理认证机制使得客户在被允许接触源服务器前首先向代理进行认证，代理服务器只对正确认证的客户提供服务，这就阻止了未经认证的用户从一个不设防的代理处窃取带宽。为支持代理认证，HTTP/1.1 引入了 Proxy-Authenticate 和 Proxy-Authorization 头标。这两个头标除了不是端对端而是逐跳的外，所起的作用与 HTTP/1.0 的 WWW-Authenticate 及 Authorization 头标相同，代理认证既可使用摘要访问认证机制也可使用基本认证机制，但前者更好些。

代理认证过程是：客户首先向服务器发送不带 Proxy-Authorization 头标的请求，代理服务器收到请求后并不转送，而是用 407(Proxy Authentication Required)状态码及包含询问的 Proxy-Authenticate 头标响应客户，此时客户必须首先增加代理服务器所需证件的 Proxy-Authentication 头标，然后重复原先的请求，代理认证成功后，客户在后续的请求中不需再被询问而直接向代理发送相同的 Proxy-Authorization 头标。

五、保护用户的隐私

资源的 URI 常常代表了一些用户视为隐私的信息，比如用户可能并不希望他们访问过某些站点之事被弄得天下皆知。

Referer 头标的作用是：客户将 Referer 头标发送给服务器，实际上向服务器提供了从前一个对象到当前对象的 URI。采用 Referer 头标，服务器能够生成资源的返回链接表，以便进一步进行研究、注册、优化缓存等。

HTTP/1.1 为克服 Referer 头标副作用做了努力，为防止意外侵犯隐私，HTTP/1.1 规范不遗余力地阻止不当发送 Referer 头标。例如，当一用户键入 URI 时，应用程序不应发送描述当前对象的 Referer 头标，客户也不应在相应对象已被安全传送后用一不安全的请求发送 Referer 头标。用户提交 HTML 表单时，浏览器通常将其编码后再发送给服务器。基于 GET 的表单会引起表单参数出现在请求-URI 中，许多代理服务器登记这些请求-URI。为防止泄漏 URI 中的诸如口令或信用卡号等敏感信息，HTTP/1.1 规范强力地阻止使用基于 GET 表单提交以上数据，而基于 POST 的表单可防止表单参数出现在请求-URI 中，进而被不恰当地登记。

六、Cookie 对隐私及安全的影响

HTTP 请求是无状态的，服务器对各个请求的处理是相互独立的。然而对于 Web 应用，状态有时是有用处的，例如在购物应用程序中随着一系列 HTTP 请求购物单亦发生变化，程序自然就要记录这种变化。使用 CGI 环境变量 PATH_INFO、QUERY_STRING 和表单的隐含文本框可以保存客户端用户的状态信息，但是它们的功能是非常有限的，至少当浏览器关闭以后这些被保存的状态信息才会消失。另一个更有力的保存客户端用户状态信息的方法是 HTTP Cookie。

Netscape 在其 1.1 版浏览器里引入了 Cookie 作为状态管理机制，后来 IETF 在 RF2109 中对 Cookie 进行了标准化。

基本 Cookie 机制：源服务器在其响应中向客户发送任一状态信息，客户负责保存该信息并在下一请求中将其返回给源服务器。通常将这些存储在客户端主机硬盘、内存中的状态信息称为 Cookie。RFC2109 及 Netscape 最初规范对上述基本模型作了放松，这就是 Cookie 可返回给一组相关服务器中的任一个，而不是特定的一个，规范同时也对 Cookie 可以返回给服务器的那些 URI 作了限制。服务器可给 Cookie 赋予一个有效期，有效期结束就不再使用。

由于 Cookie 的内容是任意的，故而其中可能会包含与应用有关的敏感信息，例如信用卡号、用户名、口令或其他个人信息等。应用程序若在未加密的连接中发送敏感信息，就易被偷窥。存于客户端的 Cookie 亦可能会将敏感信息泄漏给该客户的其他用户(或入侵者)。

在保护隐私方面 RFC2109 是有争议的。最大的争议与"未核实事务(unverifiable transactions)"及"第三方 Cookie(third-party cookies)"有关，考虑下述情形：

(1)用户访问 http://www.example1.com/home.html。

(2)返回的页面包含一指向广告 http://ad.example.com/adv1.gif 的 IMG(图像)tag。

(3)用户浏览器自动请求上述图像，响应中包含来自 ad.example.com 的 Cookie。

(4)用户访问 http://www.example2.com/home.html。

(5)返回的页面包含一指向广告 http://ad.example.com/adv2.gif 的 IMG(图像)tag。

(下转第 42 页)

表2 几种相关反馈方法在 Infolite 上的部分测试结果

	初始检索	Rocchio	Ide dec-hi	Adaptive
人民日报(3200篇)	0.3547	0.4922 +38.79%	0.5933 +67.26%	0.6215 +75.21%
联合早报(1200篇)	0.1984	0.2857 +44.00%	0.3066 +54.53%	0.2913 +46.77%
其它报刊(1600篇)	0.2159	0.3517 +62.87%	0.3804 +76.19%	0.4136 +91.57%
平均(6000篇)	0.2563	0.3763 +48.82%	0.4268 +66.52%	0.4421 +72.49%

结论 多年的研究和实验结果表明，相关反馈技术是一种非常有效的提高系统性能的方法。本文通过三种方法的简要介绍和实验比较，证明对于中文信息检索系统，适应性查询构造方法优于 dec-hi 和 Rocchio 方法。建议在实际系统中采用这种方法。

为了保证适应性查询构造算法的收敛性，应确保优先关系是弱线性的。这一点，我们可以通过一段检测程序来判断优先关系是否是弱线性的。

参 考 文 献

1 Maron M E, Kuhns J L. On Relevance, Probabilistic Indexing and Information Retrieval. Association for Computing Machinery, 1960, 7(3): 216～244

2 Rocchio J J Jr. Relevance Feedback in Information Retrieval. In Salton G, ed. The SMART Retrieval System-Experiments in Automatic Document Processing. Englewood Cliffs, NJ: Prentice-Hall, 1971: 313～323

3 Ide E. New Experiments in Relevance Feedback. In Salton G, ed. The SMART Retrieval System-Experiments in Automatic Document Processing. Englewood Cliffs, NJ: Prentice Hall, 1971: 337～354

4 Bollmann P, Wong S K M. Adaptive Linear Information Retrieval Models. In: Proc of the Tenth ACM SIGIR Conf on Research and Development in Information Retrieval. New Orleans, 1987: 157～163

5 Wong S K M, Yao Y Y. Query Formulation in Linear Retrieval Models. J of the American Society for Information Science, 1990, 41: 334～341

6 Wong S K M, Yao Y Y, Bollmann P. Linear Structure in Information Retrieval. In: Proc of the Eleventh ACM SIGIR Conf on Research and Development in Information Retrieval. Grenoble, France, 1988: 219～232

(上接第23页)

(6)用户浏览器自动请求上述图像，并在请求过程中将上一次接收到的 Cookie 发送给 ad.example.com。从 ad.example.com 的响应包含一个新的 Cookie。

关心隐私的人们担心：

(1)用户在第3步从一个他自己都不知道已访问过(一次未核实的事务)的站点 ad.example.com 收到了一个第三方 Cookie，

(2)第6步中在第二次图像请求里将第一次的 Cookie 返回给了 ad.example.com。

如果随着每一次图像请求都有一个 Referer 头标发送给 ad.example.com，那么 ad.example.com 就会根据用户访问过哪些站点积累出用户的概况。这里是 http://www.example1.com/home.html 和 http://www.example2.com/home.html。诸如此类的广告站点可以选择一些广告吸引用户，孤立地看获取用户概况的过程其危害程度相对说来并不严重，但若用户概况涉及的并不仅仅是一个体而是一实实在在的人时，其隐私性就显现出来了。比如用户在 www.example1.com 进行某种注册时，就会出现隐私被侵犯的情况。

RFC2109通过要求用户 agent 拒绝来自未核实事务响应的 Cookie 来限制 Cookie 可能的危害作用。RFC2109进一步规定用户可以配置 agent 接收缺省值为 not 的 Cookie。如此选择缺省值与业务模型依赖于 Cookie 的广告网络公司有关，这些公司在规范基本完成(1996年7月)到 RFC 出台(1997年2月)期间已经得到蓬勃发展。

七、S-HTTP

S-HTTP(Secure HTTP)技术是在电子商务活动中发展起来的。为了推广电子商务，美国成立了商业 Internet 联盟。在 Internet 上开展商务活动所必须解决的首要难题就是安全性问题，为此该联盟组织开发了在 HTTP 基础上扩充并增加安全功能的 S-HTTP。S-HTTP 可以采用多种方式封装信息。它的封装包括加密、签名和基于 MAC(Message Authentication Code)的认证。同时一个消息可以被反复封装加密。S-HTTP 还定义了头标以便来进行密钥传输，认证传输，并支持多种加密协议。

结束语 Web 已是 Internet 的主流业务，随着越来越多的个人、公司、网络以及国家连入 Internet，Web 的用户会变得更加庞大，其安全性就更显重要。本文主要介绍了作为 Web 基础的 HTTP 安全机制，它包括：HTTP/1.0基本认证机制及其缺陷；HTTP/1.1摘要访问控制机制；代理认证；如何保护用户的隐私；Cookie 对隐私、安全的影响以及用于电子商务的 S-HTTP。值得指出的是近些年 IETF 非常关注安全问题，目前 Internet 的安全体系结构日趋成熟，一个安全的 Internet(包括 Web)指日可待。

参 考 文 献

1 Krishnamurthy B, et al. Key differences between HTTP/1.0 and HTTP/1.1. Computer Networks, 1999, 31(11/16): 1737～1751

2 Stallings W. Data and Computer Communications. Fifth edition. 北京：清华大学出版社，Prentice Hall，1997

3 Tanenbaum A S. Computer Networks. Third edition. 北京：清华大学出版社，Prentice Hall，1997

4 裴有福. Web 技术大全. 北京：中国水利水电出版社，1998

5 蒋继洪，黄月江. 计算机系统、数据库系统和通信网络的安全与保密. 成都：电子科技大学出版社，1995