

文章编号:1001-9081(2009)12-3191-03

基于否定选择的身份认证技术

唐俊^{1,2}

(1. 同济大学 软件学院, 上海 200092; 2. 湖南城建职业技术学院 信息工程系, 湖南 湘潭 411101)

(xttangjun@163.com)

摘要:为了提高用户身份认证的有效性,给出了一个集成肯定认证机制和否定认证机制的双层认证模型。首先,基于人体免疫系统T细胞识别自体和非自体的原理,设计了基于否定选择的身份认证机制;接着研究了否定认证机制的关键技术;最后给出双层认证模型的实现细节及性能分析。仿真实验表明,该身份认证模型能够承受各种口令攻击,有效过滤非法用户的登录请求,具有较好的鲁棒性和可用性。

关键词:身份认证;免疫系统;否定选择算法;自体;非自体

中图分类号: TP18;TP309.2 文献标志码:A

Identity authentication based on negative selection algorithm

TANG Jun^{1,2}

(1. School of Software Engineering, Tongji University, Shanghai 200092, China;

2. Department of Information Engineering, Hunan Urban Construction College, Xiangtan Hunan 411101, China)

Abstract: To improve the validity of user identity authentication mechanism, a two-double authentication model integrating positive authentication mechanism and negative authentication mechanism was given. Firstly, inspired from the principle of immune cell identifying self and non-self, an identity authentication mechanism based on negative selection was designed; Secondly, key technologies of negative authentication mechanism were researched, and implementation details of model were given in the end. Simulation tests show that the identity authentication model can stand with password attacks, filtrate out invalid login requests available, and have advantages of good robustness and reliability.

Key words: identity authentication; immune system; negative selection algorithm; self; non-self

0 引言

基于用户名和口令的用户身份认证机制,被网络信息系统广泛采用,它在保护网络系统安全性和用户隐私信息方面担负着重要的角色。该认证机制认为只有合法的用户才能通过认证登录系统,然而事实上并非如此,大量的安全事件表明攻击者通过一些技术手段可以方便地获得非认证访问,窃取网络系统信息^[1-2]。基于口令的用户身份认证机制自身存在的隐患(如弱的口令加密算法、较短的用户口令、口令加密算法自身的缺陷),以及目前出现的先进的攻击方式(如暴力攻击、字典攻击、分布式攻击、基于规则的攻击和混合攻击)和攻击工具(如 THC Hydra、Rainbow Crack、Brutus、L0phcrack 等),使得口令攻击者有很多机会实施攻击并获得成功。

目前,出现了许多加强认证机制安全性和鲁棒性的技术手段(如用户标识码技术、认证令牌技术、口令管理技术、一次性口令技术、哈希认证协议、图口令技术等^[3-5]),这些技术在一定程度上提高了用户身份认证机制的性能。但是认证机制仍然采用的是肯定选择机制,即当一个用户试图登录到一个系统时,若输入的用户名和密码与系统保存的用户信息匹配成功的话,认证通过。认证过程中包含 2 个主要步骤:用户身份标识(identification),需要系统调出相应的用户信息数据库;用户身份验证(Verification),验证登录者宣传的身份是合法的。从上面的认证过程可看出,在认证过程中用户信息数据库是暴露在攻击进程面前的,攻击进程可以读取用户信息数据库中的用户信息,甚至通过恶意代码把新的用户信息写入该文件。这种基于肯定选择认证机制的验证信息暴露缺

陷,给网络系统的用户身份认证带来了很多安全隐患。而基于否定选择的身份认证(即不读取用户信息数据库情况下,识别出非法用户)尚未见报道,基于此,受人体免疫系统识别自体和非自体原理的启发,本文提出了一种基于否定选择的身份认证机制。

1 否定选择认证机制

免疫细胞识别抗原的能力是完备的,可随机形成大量的抗原受体。抗原分为人体自身内的分子(自体)和外部入侵分子(非自体)两类。免疫细胞对自己分子(自体)不应答现象被称为免疫耐受。否定选择是使免疫系统具有免疫耐受功能的重要机制,其过程是:未成熟的 T 细胞在胸腺中发育期间,若与自体发生应答,则死亡;成熟后的 T 细胞分布在人体内循环系统中,只与非自体结合,对自体耐受^[6]。受人体免疫耐受和免疫识别原理的启发,文献[7]开发了否定选择算法(Negative Select Algorithm, NSA)。该算法模拟了 T 细胞在胸腺中的难受过程和识别非自体过程。非自体集是自体集的补集。在 T 细胞识别非自体过程中,自体信息是隐藏的(即成熟后的 T 细胞与自体不匹配),T 细胞仅对非自体(否定的信息)识别。在用户认证过程尽量不直接访问后端的用户信息数据库,才能保证认证过程本身的安全。否定选择机制可以利用非法用户信息(非自体)进行否定身份认证。

本文设计的否定选择认证机制描述如下:

1) 自体和非自体。

抗原空间记为 U , 自体集 S 定义为合法用户集, $S \subset U$; 非自体集 T 定义为非法用户集, $T \subset U$ 。满足 $S \cap T = \emptyset$, $S \cup$

$T = U$ 。例如,若取用户帐号 6 位字符,口令 6 位字符,用字符的 ASCII 码对应的二进制表示字符,则用户信息可表示为长度为 96 位的二进制串, U 为长度为 96 位的所有二进制串集。

2) 检测器(检测细胞)。

检测器(检测细胞) b 为长度为 l 的二进制串,检测器长度是可以变化的。它融合了人体免疫系统中 B 细胞、T 细胞和抗体的性质,用于检测和识别非法用户。

3) r 连续位匹配。

设检测器 $b = c_1c_2\cdots c_l, c_i \in \{0, 1\}$, l 为检测器的长度; 抗原(用户信息) $u = d_1d_2\cdots d_m, u \in U, M$ 为抗原长度。检测器与抗原(用户信息)的匹配采用 r 连续位匹配,如式(1)所示。

$$f_r(b, u) = \begin{cases} 1, & \exists i \exists j, b_k = u_k \\ & \wedge k = i, \dots, j \\ & \wedge j - i \geq r \\ & \wedge i, j, k \in \mathbb{N} \\ 0, & \text{其他} \end{cases} \quad (1)$$

4) 否定认证过程。

当系统收到一个用户登录请求时,把该请求转换为一个抗原(长度为 96 位的二进制串),提交给检测器集(抗原提呈)进行 r 连续位匹配,若检测器集中的某个检测器与该抗原匹配成功,说明该抗原为非自体,即该请求登录的用户为非法用户,可采用蜜罐技术,将该非法用户引入陷阱区。

通过否定选择认证的用户登录请求,被送往正常的登录验证进程,进行肯定选择认证。否定选择过滤了大量的非法登录请求,屏蔽了恶意口令攻击(如暴力口令攻击),而且由于在否定选择认证中,不对用户信息数据库操作,攻击者没机会读或写用户信息数据库,保证了认证过程的安全性。

2 否定选择认证机制的关键技术

2.1 用户信息映射

在应用系统中,用户登录信息是经过加密处理后保存在用户信息数据库中的。常采用哈希函数作为用户登录信息加密方式,处理过程如图 1 所示。

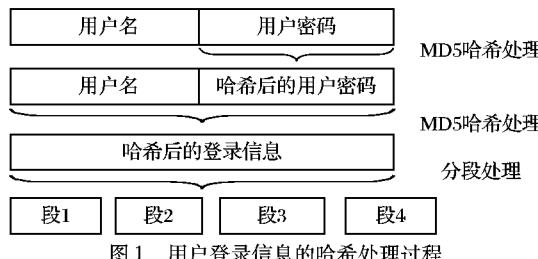


图 1 用户登录信息的哈希处理过程

明码用户信息和哈希处理过的用户信息都是采用字符串表示的(问题空间),而检测器对自体耐受及对非自体结合过程都是基于二进制字符串(编码空间)进行的,因此必须完成从问题空间到编码空间的映射。

这种映射工作发生在两种情形:

1) 检测器构造。随机产生的二进制串形式的检测器要经过对自体耐受后才能成为成熟检测器。此时,自体集必须完成从问题空间到编码空间映射。

2) 检测器识别。应用系统的客户端收到一个用户登录请求后,为了预防网络截获攻击,该请求经过加密处理后才送往服务器端,服务器端必须完成从问题空间到编码空间的映射后,驻留在服务器端的检测器才能识别非主体。

一般地,加密和解密可采用同样的 MD5 哈希算法^[8]。为

了提高映射速率,本文开发了一个映射函数(map function),负责完成从问题空间(长为 12 位字符串空间)到编码空间(长为 96 位二进制串空间)的映射。

2.2 检测器集构造

检测器集的构造是否定选择算法的关键,好的检测器集可以降低误检(错误肯定)率和漏检(错误否定)率。检测器集 B 的构造应满足式(2)~(4):

$$\forall b \in B, \forall s \in S, f_r(b, s) = 0 \quad (2)$$

$$\forall t \in T, \exists b \in B \wedge f_r(b, t) > \omega \quad (3)$$

$$\neg \exists B_1 \wedge T_{B_1} = T_B \wedge |B_1| < |B| \quad (4)$$

其中: ω 为漏检率阈值($0 < \omega < 0.04$), T_B 表示检测器集 B 覆盖的非自体空间。式(2)要求检测器集对自体集不覆盖(即检测器对自体难受),式(3)要求检测器集对非自体集全覆盖(能检测出所有的非自体),式(4)要求检测器集应是满足需求的规模最小的检测器集。

本文借鉴了实值否定选择算法^[9] 中检测器集构造思路。根据认证应用中非自体空间的实际情况,采用了变长的检测器,修改后的检测器生成算法如下:

$Self$: 自体集合(合法用户信息的集合)。

T_{\max} : 检测器的最大数目。

t : 当前已经生成的检测器数目。

D : 检测器集合。

l : 事先设定的检测器间的相似度阈值。

x : 随机生成的候选检测器。

第 1 步 $D \leftarrow \emptyset, t \leftarrow 0$ 。

第 2 步 随机生成一个候选检测器 x 。

第 3 步 对每一个 $d_i \in D$, 计算 x 与 d_i 之间的欧氏距离 w_1 。

第 4 步 如果 $w_1 > l$, 则转到第 2 步。

第 5 步 对每一个 $s_i \in Self$, 计算 x 与 s_i 之间的欧氏距离 w_2 。

第 6 步 如果 $w_2 > l$, 则转到第 2 步。

第 7 步 $D \leftarrow D \cup \{d_i\}; t \leftarrow t + 1$ 。

第 8 步 如果 $t \geq T_{\max}$, 则转第 9 步; 否则转第 2 步。

第 9 步 结束。

2.3 基于危险理论的非自体识别

Matzinge 在研究了人体免疫系统否定选择机制后,提出了危险理论模型^[10],他认为抗原提呈细胞(Antigen Presentation Cells, APCs)不全是以自体-非自体为识别依据的,外界危险信号存在与否及信号强度也影响着 APCs 的工作效能。在用户身份认证过程中,若一旦用户信息与某个检测器匹配,就认为是非法用户,这种判断过于武断。事实上,合法用户偶尔输错口令是常有的事情。因此,本文借鉴了危险理论,把在一个连续的时间段内(如 30 s)同一个用户登录请求的用户信息与检测器匹配成功的次数看作危险信号,只有当危险信号强度超出阈值(例如连续输入错误的用户名和口令),才认为是非法用户,从而使认证具有较好的智能性。

3 双层认证模型

本文设计的双层认证模型如图 2 所示。

否定认证能有效地过滤掉非授权用户的登录请求,但是对于合法用户的登录请求没有做任何处理,为此,必须将否定认证机制和肯定认证机制有机地集成在一起,利用否定认证

机制过滤非授权用户的登录请求,然后利用肯定认证机制验证合法用户的权限,最终完成用户的身份认证。

下面对双层认证模型的性能进行分析:

设用户口令信息(用户名称,密码)长度为 l , 口令取值于 128 位字符空间, 系统中现有合法用户数目为 t , 设黑客一次口令攻击能成功获得一个合法用户信息的概率为 α , 则 $\alpha \approx t/(128^l)$ 。理论上黑客经过 $(128^l)/t$ 次尝试就能成功地获取一个合法用户信息。

再设检测器漏检率为 β , 一般地 $0 < \beta < 0.05$; 则黑客一次访问通过否定认证的概率为 β , 再次通过肯定认证的概率仍为 α , 通过两次认证的概率为 $\alpha\beta$ 。

这样,理论上黑客经过 $(128^l)/(t\beta)$ 次尝试才能成功地获取一个合法用户信息。由于 $t\beta$ 的值较小, 使得 $(128^l)/(t\beta)$ 数值极大, 攻击难度极大, 即攻击成本极大。按照网络安全理论, 若攻击的成本大于预期获取的收益, 则认为系统是安全的。

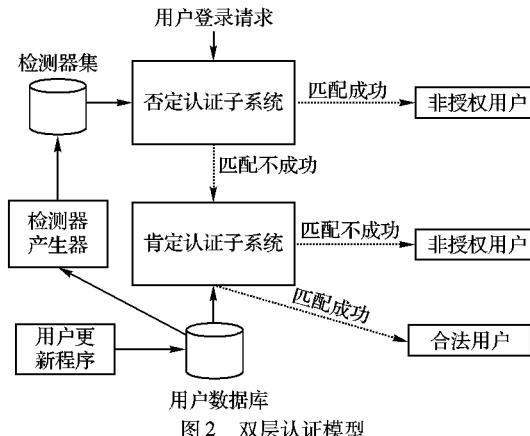


图 2 双层认证模型

4 仿真实验

本文设计的否定选择认证系统, 其认证过程是在传统的肯定认证之前进行的, 因此它可以方便地部署, 而不需要对现有的认证系统进行修改。为了验证本文设计的身份认证系统的效能, 在 C++ 环境下, 编程实现了否定选择算法和用户信息映射函数, 并设计了用户信息数据库进行测试。

实验参数设置: 用户名和口令组成的用户信息字符串长度为 10~20 字符, 则对应的二进制表示的抗原(用户信息)字符串长度 l ($80 < l < 160$); 检测器字符串长度 m ($80 < m < 160$); 由于检测器集中各个检测器长度可能不相同, r 取为抗原字符串长度的 90%。考虑到认证系统的实际情况, 用户信息数据库存储了 10 000 个用户, 其中 50% 用于训练系统, 另 50% 用于测试系统对未知合法用户的识别情况。采用人工输入合法用户信息, 口令攻击工具自动产生非法用户信息, 合法与非法用户信息的比例为 1:100, 依次进行了 4 组实验, 每组实验分别做 100 次取平均结果, 依次测试平均漏检率和误检率。实验结果如表 1 所示。

表 1 不同参数下的实验结果

抗原长度 l	检测器长度 m	漏检率/%		误检率/%	
		平均	方差	平均	方差
固定: $l=120$	固定: $m=120$	2.3	5.3	2.1	2.5
固定: $l=120$	可变: $80 \leq m \leq 160$	1.6	3.4	1.1	1.7
可变: $80 \leq l \leq 160$	固定: $m=120$	3.2	7.7	2.5	4.2
可变: $80 \leq l \leq 160$	可变: $80 \leq m \leq 160$	0.9	1.4	0.8	1.6

表 1 表明, 无论抗原长度固定与否, 可变检测器的检测效果都优于固定长度检测器; 当抗原长度和检测器长度均可变时, 由于检测时可以根据当前抗原长度情况从检测器集中选择最优检测器进行检测, 此时平均漏检率和误检率达到最小值, 分别是 0.9% 和 0.8%。而在抗原长度可变检测器长度固定情况下, 检测效果最差, 其平均漏检率和误检率分别为 3.2% 和 2.5%, 这也比系统要求的漏检率(4%)偏小些。这说明, 设计的检测算法可以满足需求。实验是在模拟超恶劣环境下进行的, 如模拟网络瞬间遭到大量的口令攻击。若本文设计的认证算法工作在真实的网络环境中, 性能表现会比实验时要好一些。

5 结语

身份认证的作用是安全地标识一个系统用户。目前存在的认证系统是基于肯定配合法用户机制的, 认证过程必须读取用户数据库, 认证过程本身的脆弱性使得口令攻击工具有机会对用户数据库进行读写, 给认证安全带来隐患。本文设计的否定认证系统, 由于在认证过程中, 不需要对用户数据库读取, 可以过滤掉大量的口令攻击。经过否定认证后的用户, 再提交给传统的肯定认证系统, 这种双重认证机制, 可以有效地杜绝非法用户对系统的非授权访问, 提高了认证系统的鲁棒性。当然, 本认证系统仅保证非法用户无能力访问系统, 若黑客利用其他手段(如截获了管理员无意间泄露的用户信息)偷取的合法用户信息, 登录系统后实施内容攻击, 本认证系统对此无能为力。对此种攻击, 可以采取用户行为监控技术解决, 这不是本文的研究内容。

下一步要研究的工作有: 在系统动态添加或删除一个用户的情况下, 检测器集自适应更新问题; 多系统间身份认证传递问题(怎样把一个系统的认证结果传递给被信任的下一个系统)。

参考文献:

- [1] 田野, 张玉军, 刘莹, 等. 移动 IPv6 网络基于身份签名的快速认证方法[J]. 软件学报, 2006, 17(9): 1800~1888.
- [2] 田野, 张玉军, 张瀚文. 移动 IPv6 网络基于身份的层次化接入认证机制[J]. 计算机学报, 2007, 30(6): 905~915.
- [3] SMITH R E. Authentication: From passwords to public keys [M]. New York: Addison-Wesley, 2002.
- [4] THORPE J, van ORSCHOT P C. Towards secure design choices for implementing graphical passwords [C] // ACSAC: 20th Annual Computer Security Applications Conference. Tucson: ACM Press, 2004: 138~145.
- [5] BIRGET J C, HONG DA-WEI, MEMON N. Graphical passwords based on robust discrimination [J]. IEEE Transactions on Information Forensics and Security, 2006, 1(3): 395~399.
- [6] 朱锡华. 生命的卫士: 免疫系统[M]. 北京: 科学技术文献出版社, 2004: 198~213.
- [7] DASGUPTA D, FORREST S. Artificial immune systems and their applications [M]. Berlin: Springer-Verlag, 1999: 312~344.
- [8] 卢开澄. 计算机密码学[M]. 北京: 清华大学出版社, 1998: 200~213.
- [9] ZHOU JI, DASGUPTA D. Real-valued negative selection using variable-sized detectors [C] // GECCO: Genetic and Evolutionary Computation Conference. Washington: ACM Press, 2004: 120~132.
- [10] MATZINGER P. The danger model: A renewed sense of self [J]. Science, 2002, 296(5566): 301~305.