分布式信息集成系统安全性研究

Security of Distributed Information Integration System

丁立波[1]　胡金伟[2]　王宏志[3]

(上海工业自动化仪表研究院[1],上海　200233;华东电力设计院[2],上海　200063;哈尔滨工业大学[3],黑龙江 哈尔滨　150001)

摘　要:为保证分布式环境下信息集成系统的安全,从授权机制、通信安全保障和 Wrapper 安全性保证三个方面对信息集成系统的安全性策略进行了研究,提出了面向模式的授权机制。对全局模式和数据源局部模式进行授权,并针对通信中的不安全因素,提出了加密和数字签名的保障方法。同时设计了分布式环境中 Wrapper 的安全性策略,以保证分布式环境下信息集成系统的安全,避免信息集成系统受到攻击。

关键词:信息集成系统　分布式环境　授权机制　安全保障　数字签名

中图分类号:TP309.2　　**文献标志码**:A

Abstract: In order to ensure the security of information integration system under distributed environment, the security strategies of information integration system are researched from three aspects, i. e. the authorization mechanism, guarantee of the communication security and Wrapper pledge; and the authorization mechanism based on scheme is proposed. The global scheme and data source local scheme are authorized, and aiming at the unsafe factors, the methods encryption and digital signature are put forward. In addition, the Wrapper safety strategy in distributed environment is designed to ensure the security of information integration system under distributed environment to avoid possible attacks.

Keywords: Information integration system　Distributed environment　Authorization mechanism　Security guarantee　Digital signature

0　引言

使用统一接口对分布、自治和异种的数据源进行查询是信息集成系统的工作[1-4]。为了保证有效服务,安全性是系统设计必须考虑的问题。文献[4]中讨论了基于 Mediator 的信息集成系统的安全性问题,但该文仅面向关系数据库,没有考虑分布式系统的安全性。面向 Web 的分布式环境信息集成为信息安全提出了新问题。若网络不安全,则可能造成信息在传输过程中丢失,且有的用户需要系统对查询本身保密,还有的用户需要返回数据的保密。因此,如何在不安全的网络环境下保证信息集成安全是有待解决的问题。

针对分布式信息集成系统中的安全性问题,本文定义了系统的授权机制,并提出在用户有需求的情况下保障通信安全的技术。

1　授权机制

在基于 Mediator 的信息集成系统中,存在一个面向用户的全局模式,用户的查询就是定义在这个全局模式下的。全局模式可以描述为树型结构,定义为全局模式树;对于结构化和半结构化数据,每一个局部数据源(包括关系数据库、OR 数据库、OO 数据库和半结构化数据库)的模式也可以表示成为树型结构;局部模式可以通过映射函数映射到全局模式树的一个子图上。

定义全局模式上的授权方式,即:global_authorization(T,i),表示第 i 级用户有访问全局模式树的子图 T 的权限。若有 global_authorization(T_i,i)、global_authorization(T_j,j),且 $i<j$,则有 $T_j \subseteq T_i$,即 T_j 是 T_i 的子图。

结构化和半结构化的数据源也有类似的定义,数据源 i 的 j 级别用户授权方式定义为 local_authorization(T,j),表示第 j 级用户有访问局部模式树 T 的权限。同理有 local_authorization(T_r,r)、local_authorization(T_s,s),且 $r<s$,则有 $T_s \subseteq T_r$。

对于同时有读和写授权的数据源,这一授权仅限于读授权。全局授权 G 和局部授权 L_i 存在着对应关系。这种对应关系体现为函数 $f_i: G \rightarrow 2^L$,且授权等级对应函数 f 是单调不递减函数,亦即全局模式有授权 m 个授权等级 $\{g_0, g_1, g_2, \cdots, g_{m-1}\}$,局部模式有 n 个授权等级 $\{l_0, l_1, l_2, \cdots, l_{n-1}\}$。在映射中,当 $i<j$ 时,不存在 $l_i \in f(g_r)$,$l_j \in f(g_s)$,且 $r>s$。

这个授权映射和全局模式到局部模式的模式映射之间可能产生冲突,存在的情况是可能存在全局模式中的某个结点 n_g 和它对应的数据源 i 的本地模式对应

修改稿收到日期:2010-05-07。

第一作者丁立波,男,1977 年生,2000 年毕业于东南大学热能自动化专业,获学士学位,工程师;主要从事自动化系统集成的设计及应用。

结点 n_l，其中，n_g对应最高的安全性级别是 j，n_l对应最高的安全性级别是 k，而 $f(g_j) > l_{n_l}$。在这种情况下，以数据源的保密级别优先为原则。执行查询时的验证授权过程如下。

设一个具有授权级别 i 的用户查询 q，涉及到的全局模式为 g_q，求得 $g_q' = g_q \cap T_i$，其中 T_i表示 i 对应的模式，g_q'对应到数据源 w 的局部模式上涉及到的模式为 l_q，j 为 i 对应到 w 上的授权级别。在数据源上，j 级别对应的模式为 T_{l_j}，则可以在数据源上查询涉及的模式为 $l_q' = l_q \cap T_{l_j}$。

2　通信数据安全保障

考虑到网络通信的不安全因素，为了保障用户查询和返回数据的安全，在用户提出需求的情况下，系统可以对网络上传输的数据进行加密。

首先，用户可以在查询语句中定义是否查询需要加密或者返回的结果需要加密。定义语法如下：①ENCRYPT(<查询语句>)，查询语句需要加密；②WITH RESULT ENCRYPT(<查询语句>)，结果需要加密；③ENCRYPT(<查询语句> WITH RESULT ENCRYPT)，查询和结果都需要加密。查询和结果采用的加密方法是不同的。

在信息集成系统中，假定每一个在用户端的 Wrapper 都可能被攻击，系统中 Mediator 和每一个 Wrapper w_i存在着一个密钥 k_i，则用户提出的查询在 Mediator 中经过查询改写和转换后，变换成为面向每个数据源的查询。如果用户在查询语句中要求对查询进行加密，则在将这个查询发给 Wrapper 的过程中需要进行加密；对于 Wrapper w_i的查询，经过 k_i加密后发送给 w_i。

对 Web 进行查询的情况则更加复杂，因为涉及到一个作为面向搜索引擎和潜在数据源进行查询改写以及对它们返回的数据进行信息提取的查询 proxy。因此，在每一个查询传送的阶段都需要进行数据加密。一方面 Mediator 要保存和查询 proxy p_i之间的密钥 k_{p_i}；另一方面查询 proxy p_i需要保存它及其所管辖的 Wrapper w_j之间的密钥 k_{ij}，利用 k_{ij}对 p_i和 w_j的查询传递进行加密。

对于返回信息的加密，考虑到具有相同授权级别的不同用户之间可能互相窃取数据，因此，我们采取 Diffie-Hellman 算法[7]，即由 Mediator 和 proxy/Wrapper 或者 proxy 和 Wrapper 产生密钥。Wrapper 和 Mediator 存有双方协定的大素数 n 和 g。密钥产生过程如下。

① Wrapper 选择一个大的随机整数 x，计算 $X = g^x \bmod n$，将 X 发送给 Mediator；

② Mediator 选择一个大的随机整数 y，计算 $Y = g^y \bmod n$，将 Y 发送给 Wrapper；

③ Wrapper 计算 $k = Y^x \bmod n$；

④ Mediator 计算 $k' = X^y \bmod n$。

k 和 k'同时作为加密用的密钥，这样，即使 Mediator 和 Wrapper 协商密钥的过程被窃听，窃听者也没有办法得到密钥。

由于数据源是自治的，因此可能存在数据源的欺诈行为，即存在其他方面的数据源伪造数据对系统进行破坏，如存在一个破坏者伪造大量非法的数据返回给 Mediator，造成 Mediator 的负载过大，从而达到其破坏目的。同时也存在第三方伪装 Mediator，从 Wrapper 中窃取数据。对于这样的潜在不安全因素，我们采取数字签名的方法来解决，即采用 DSA(digital signature algorithm)[6]。由于每个 Wrapper w_i都存在着唯一和 Mediator 约定的密钥 k_i，这个 k_i是其他 Wrapper 不能伪造的，因此，对于每个 Wrapper，这个密钥是安全的。

系统中包括 1 024 位的素数 p、与 $(p-1)$ 互质的 160 位长的素数 q 以及 $g = h^{(p-1)/q} \bmod p$，其中 $h < (p-1)$，且为满足 $[h^{(p-1)/q} \bmod p] > 1$ 的任意数。

在 Mediator 和 Wrapper w_i之间的通信过程中，采用数字签名的生成和验证过程如下。

发送信息方产生一个小于 q 的随机数 k，则：

$$r = (g^k \bmod p) \bmod q$$

$$s = \{k^{-1}[H(m) + w_i r]\} \bmod q$$

式中：m 为发送信息方的标示；$H(m)$为单向散列函数；r 和 s 为数字签名，两者的和作为信息的头发送给信息的接收方。接收方按照以下方法对签名进行验证：

$$w = s^{-1} \bmod q$$

$$u_1 = [H(m)w] \bmod q$$

$$u_2 = (w_i w) \bmod q$$

$$v = [(g^{u1} y^{u2}) \bmod p] \bmod q$$

因此，如果 $v = r$，则签名有效。

3　Wrapper 安全保证

在分布式环境下，Wrapper 可能受到恶意数据源的攻击，特别是代码在数据源一方运行的 Wrapper 尤其容易受到攻击，其中的密钥或者代码如果被破解，将无法保证系统或者数据的安全性。

如果 Wrapper 代码运行在专用的主机上，那么维护主机的安全就是确保 Wrapper 的安全；如果 Wrapper 运行在数据源的主机或者第三方主机上，那么关于移动代理安全性的一些技术是可以采用的[9-11]。保证 Wrapper 安全的策略如下。

① 确认 Wrapper 运行环境，当 Wrapper 需要装载到某个非系统可控制主机时，需要事先确认主机的安

全性。通常情况下,主机的安全性是可以确定的,但并非绝对确定。

② 记录 Wrapper 运行日志,Wrapper 对和主机系统的每一次交互都进行记录,并将保存的记录及时发送给 Mediator;对于已遭受攻击的 Wrapper,由于它的交互记录被保存,当主机使用类似的手段攻击 Wrapper 时,就会采取相应的手段加以防范。

③ 对 Wrapper 中关键数据和算法进行加密,当系统中的关键数据和算法需要进行加密时,可以采用 CEF(cisco express forwarding)的方法进行加密[12]。当前 CEF 没有通用的解决方案,但是对于多项式和有理函数是完全可行的。可以储存一个利用加密的有理函数计算过的密钥进行加密,对需要保密的函数代码也可以用相关的方法加密。

4　结束语

安全问题是分布式环境下信息集成的重要问题。为了有效维护信息集成系统的安全,本文提出了一系列技术,从授权机制、通信安全保障和 Wrapper 安全性三个方面保证分布式信息集成系统的安全性。我们的安全性策略适用于分布式环境下多数据源的集成,对于分布式环境下信息集成系统稳定、高效的应用有着重大的意义和价值。

参考文献

[1] Hasselbring W. Information system integration[J]. Communication of the ACM,2000,43(6):33 - 38.

[2] Garcia-Molina H, Ullman J D, Widom J. Database system implementation[M]. USA:Prentice Hall,2000:254 - 272.

[3] Wirderhold G. Mediators in the architecture of future information systems[J]. Computer,1992,25(3):38 - 49.

[4] Candan K S, Jajodia S, Subrahmanian V S. Secure mediated databases[C]// Proceedings of the Twelfth International Conference on Data Engineering, IEEE Computer Society, Washington DC, USA, 1996:28 - 37.

[5] Bray T, Paoli J, Sperberg-McQueen C M. Extensible markup language(XML)1.0[S]. 1998.

[6] Stalling W. 密码编码学与网络安全:原理与实践[M]. 杨明,译. 北京:电子工业出版社,2001: 351-357.

[7] Schneier B. Applied cryptography:protocols, algorithms and source code in C[M]. USA:Jon Wiley & Sons,Inc,1996:412 - 418.

[8] Kudo M, Hada S. XML document security based on provisional authorization[C]//Proceedings of the 7th ACM Conference on Computer and Communications Security, New York, NY, USA, 2000: 87 - 96.

[9] Ertaul L, Panda J. Mobile agent security[C]//Proceedings of the 2006 International Conference on Security & Management, Las Vegas, USA,2006:172 - 178.

[10] Chu Y H, Feigenbaum J, LaMacchia B, et al. Referee:trust management for Web applications[J]. Computer Networks and ISDN Systems,1997,29(8-13):953 - 964.

[11] 刘建勋,李仁发,张申生. 移动 Agent 的安全性问题探讨[J]. 小型微型计算机系统,2000,21(12):1316-1319.

[12] Sander T, Tschudin C. Protecting mobile agents against malicious hosts[C]//Proceedings of Mobile Agents and Security, Springer-Verlag, London, UK,1997:44-60.

(上接第 11 页)

EMHR 协议和 LEACH 协议,并能进行扩展而应用于不同规模的网络。

参考文献

[1] Akyildiz F I, Su W, Sankarosubramaniam Y, et al. Wireless sensor networks:a survey[J]. Computer Neworks, 2002, 38(4):393 - 422.

[2] Ilyas M, Mahgoub I. Handbook of sensor networks:compact wireless and wired sensing systems[M]. CRC Press:New York,2005:13 - 18.

[3] Heinzelman W R, Chandrakasan A, Balakrishnan H. Energy-efficient communication protocol for wireless micro-sensor networks[C]//Proceedings of the 33rd Hawaii International Conference on System Sciences, Maui, HI,2000:1 - 10.

[4] Manjeshwar A, Agarwal D P. TEEN:A routing protocol for enhanced efficiency in wireless sensor networks [C]//International Parallel and Distributed Processing Symposium IPDPS 2001, San Francisco, California, USA,2001:2009 - 2015.

[5] Manjeshwar A, Agrawal D P. APTEEN: A hybrid protocol for efficient routing and comprehensive information retrieval in wireless sensor networks[C]//International Parallel and Distributed Processing Symposium:IPDPS 2002, Fort Lauderdale, Florida,2002:195 - 202.

[6] Xu Y, Heidemann J, Estrin D. Geography informed energy conservation for Ad hoc routing[C]//Proceedings of the 7th Annual ACM/IEEE International Conference on Mobile Computing and Networking, Rome, Italy,2001:70 - 84.

[7] Younis O, Fahmy S. HEED: a hybrid, energy-efficient, distributed clustering approach for Ad hoc sensor networks [J]. IEEE Transactions on Mobile Computing,2004,3(4):366 - 379.

[8] Huang Wenwen, Peng Yali, Wen Jian, et al. An energy-efficient multi-hop hierarchical routing protocol for wireless sensor networks [C]// Proceedings of the 2009 International Conference on Networks Security, Wireless Communications and Trusted Computing, Washington DC, USA,2009:469 - 472.

[9] Heinzelman W B, Chandrakasan A P, Balakrishnan H. An application specific protocol architecture for wireless micro-sensor networks[J]. IEEE Transactions on Wireless Communications, 2002, 1(4):660 - 670.

[10] Hu Limin. Distributed code assignment for CDMA packet radio networks[J]. IEEE/ACM Transactions on Networking, 1993 (6): 668 - 677.