Study on Weight-value-Extended MLS and Its Application for Ontology modeling

Meng YuLong, Yin GuiSheng, Wang HuiQiang, Xu Dong

(Department of Computer Science and Technology, Harbin Engineering University, Harbin 150001, China)

Abstract: There exists a variety of heterogeneous data that need to be integrated within a secure domain. In order to solve the issues of the security and integrity of data more flexibly in the process of integration, ontology and multilevel security (MLS) is introduced. The application and flexibility in aspects of authentic subject access is poor. An Weight-value-Extended MLS（WEMLS）was proposed. A concept of "Authentic Weight value" was defined，then authorization mechanisms for processes and the mechanism of calculation of trust value were established, so the authentic processes can access object more flexibly. Finally, WEMLS has been verified by applying to a security data integration model based on ontology.

Key words: Multilevel security; Security data; Authentic Weight value; Ontology

扩展权值 MLS 及其在本体建模中的应用研究

孟宇龙，印桂生，王慧强，徐东

（哈尔滨工程大学计算机科学技术学院 黑龙江哈尔滨 150001）

mengyulong@hrbeu.edu.cn

【摘要】为更灵活解决异构安全数据集成中数据安全性和完整性问题，考虑引入本体和多级安全策略(MLS)的思想。针对 MLS 在可信主体访问方面的应用性和灵活性较差问题。本文提出一种扩展权值 MLS 策略（WEMLS）。定义了一个可信权值的概念，进而建立进程授权机制和信任计算机制，使可信进程能够更灵活地访问客体。最后，将之应用于一个基于本体的安全数据集成模型中加以验证。

【关键词】多级安全策略；安全数据；可信权值；本体

1 引言

近些年随着相关部门信息化工作的深入，安全数据的生成方式、存放方式和方法日益多样化，造成了大量多源异构安全数据源的存在。这些异构安全数据源使得安全信息共享和安全业务协同处理出现了严重的阻塞，对系统本身的安全检测、分析、度量、预防和处理带来极大的难度，造成了安全服务和安全系统之间的集成困难。在软件体系结构层次上，描述安全模型的工作正在逐步展开[1]。

Bell-LaPadula(BLP)保密性模型是第一个能够提供分级别数据机密性保障的安全策略模型(多级安全)[2][3]，由于 BLP 模型存在不保护信息的完整性和可用性，不涉及访问控制等缺点，1977 年 Biba 模型[4]作为 BLP 模型的补充而提出。目前，MLS(multilevel security)策略被广泛应用于众多的安全系统中。在 MLS 中任何主体都是不可信任的，因此 MLS 规定一个用户不能同时处理多种密级的信息。虽然一些改进方案陆续被提出，但这些模型对安全数据集成过程中安全数据访问提取的应用性无法解决。

基于以上问题，本文从分析概念、系统、数据与本体的关系及安全数据特点入手，在 MLS 策略基础上提出了一个扩展权值 MLS 策略(Weight-value-Extended MLS，WEMLS）并将之应用到安全数据集成的本体建模中，提出了一种为异构安全数据提供的本体建模策略。通过这种策略，对于减少安全数据的使用和管理的复杂性以及方便安全系统之间的集成和管理具有一定的作用。

2 扩展权值的多级安全策略（WEMLS）

MLS 的目的是防止高密级的信息泄漏给低密级的主体。但是，从数据处理的角度来看，一些数据处理程序常常必须同时处理多种安全等级的信息。同时，同一信息在不同主体中可能体现不同的安全级。在依据本体进行集成过程中应该允许本体元数据抽取程序访问它，如何在访问、抽取元数据的过程中保证信息

基金项目：国家自然科学基金资助项目（90718003）；国家"863"计划重点基金资助项目（2007AA012401）

的机密性和完整性是数据集成中一个需要考虑的问题。

为解决实际应用的问题，本文考虑允许将某些违反 MLS 策略的安全主体进程视为可信主体，使其获得能够通过访问控制机制的特权。将可信进程纳入到安全策略中来。

借鉴 BLP 和 Biba 模型定义术语和规则，下面给出 WEMLS 模型的相关定义和规则：

定义 2.1 在系统 F 中，$P(p)$ 为机密等级分类集合。$I(i)$ 为完整权限集合。$S(s)$ 为主体集合。$S_t(s_t)$ 为信任主体集合，$S_t \subseteq S$。$O(o)$ 为客体集合。$E(S,<r,w>,O)$ 为操作集合，r 表示只读，w 表示只写，E^e 表示向上进行操作，E_e 表示向下进行操作。

定义 2.2 信任域集合 $TDom$ 表示可信主体进程允许活动的域值。表示为 $TDom=\{rsDom, wsDom, esDom\}$，$rsDom$ 表示读信息安全域，$wsDom$ 表示写信息安全域，$esDom$ 表示结果信息安全域。

定义 2.3 在系统 F 中，$T(d,t)$ 为当前信任权限计算函数，$d \in TDom$。在 $TDom$ 内，$T^{(d,t)}$ 和 $T_{(d,t)}$ 表示在 d 内提高或降低 t 的信任权值。

定义 2.4 在信任域内，$R=(d,r)$ 为其上主体进程或安全数据的集合，$d \in TDom$，$r \in \{S,O\}$。对于 $\forall R_i \in R$（$R_i \subseteq d_1 \times d_2 \times \cdots \times d_n \times r_1 \times r_2 \times \cdots \times r_n$），有且只有 $R_i^P \subseteq R$ 且 $R_i^I \subseteq R$（$i=1,2,\cdots,n$）使得 R_i^P 与 R_i^I 中的元素在 P 和 I 上实现一一映射，则 S 安全性和完整性的信任权值 K 相等，记为 $P(s)=I(s)$，如果是满射，则记为 $P(s) \geq I(s)$。否则记为 $P(s) \leq I(s)$。

定义 2.5 $K(k)$ 为初始信任权值计算函数。该信任权值用来判定系统安全性和完整性的重要程度，在系统中允许动态调节，具体形式可由系统选择，多为线性函数。

系统 F 在 WEMLS 下的主体访问规则如下：

若 $s \in S_t$：

$$(1)\quad (T(s) \leq T(o) \wedge I(s) \geq P(s) \wedge I(s) \geq K(s)) \wedge T^{(d,s)} \rightarrow E^{(s,r,o)}$$

$$(2)\quad (T(s) \geq T(o) \wedge P(s) \geq I(s) \wedge P(s) \geq K(s)) \wedge T_{(d,s)} \rightarrow E_{(s,r,o)}$$

$$(3)\quad (T(s) \geq T(o) \wedge P(s) \geq I(s) \wedge P(s) \geq K(s)) \wedge T^{(d,s)} \rightarrow E^{(s,w,o)}$$

$$(4)\quad (T(s) \leq T(o) \wedge I(s) \geq P(s) \wedge I(s) \geq K(s)) \wedge T_{(d,s)} \rightarrow E_{(s,w.o)}$$

规则（1）（2）表示对于可变信任主体，当系统对完整性需求大于（小于）安全性需求且其完整性信任权值大于 K 时，为保证完整性（安全性）的实施，主体可适当提高（降低）自身权限，系统此时允许主体向上（下）读的请求。规则（3）（4）同理。若 $s \notin S_t$ 则依据其对安全性或完整性的重要程度，选择 BLP 模型或 Biba 模型。

以上规则定义的前提是该主体进程为可变更权限的可信主体，该主体由系统状态检测模块监测下更改自身权限，同时要求其具有保证系统完整性和安全性不被破坏的能力。

3 概念、本体和数据集成

3.1 概念化与本体

任何一个系统、领域都有其内在的、固有的概念规则及关联，这种通过对概念、术语及其相互关系的规范化描述，勾画出某一领域的基本知识体系和描述语言称之为本体。本体是为说明某语言词汇表的内在意义而设计的一套逻辑公理。本体是语言相关的，概念化则是语言无关的。任何通过概念化抽取的信息均可通过本体进行描述，包括集成。

概念化(conceptualization)作为知识形式化表达的基础，是所关心领域中的对象、概念和其它实体，以及它们之间的关系。与一般逻辑理论不同的是：本体的描述、使用首先进行概念化，而且这里的概念化是：领域中实体、对象、概念、事物、事件等等所关心的客体及其它们之间的关系。

数据作为信息的重要载体，针对某些特殊的“安全数据”，其保密性、可用性、可控性和完整性等安全问题均可通过概念化进行本体进行数据集成。

3.2 数据集成

目前，比较常见的数据集成方法主要有联邦数据库法[5]、中介器法[6]、数据仓库法[7]和基于语义的方法。基本上，有三种比较典型的本体集成建模方法。它们针对不同种类的本体，互相间不能整合一致，也没有遵循统一的开发模式和协议。

Uschold 和 King 方法[8]，该方法包括：明确构造

978-1-935068-01-3 © 2009 SciRes.

本体的目的；构造本体，包括概念提取与定义、编码和集成已有的本体；

Methontology 方法[9]，该方法由是结合了 Uschold 的骨架法和 Gomez Perez 方法后，提出的一种更为通用的本体建设方法，其目的是在知识层上构造本体。

Gruninger 和 Fox 方法[10]，该方法基本过程是先对所要描述的领域给出非形式化的规范说明，然后在此基础上给出形式描述。

目前最有影响的本体构建方法是 Gruber 于 1995 年提出的 5 条准则[11]：明确性和客观性、完全性、一致性、最大单调可扩展性和最小承诺。

以上这些方法都是具体领域本体开发过程中总结出来的，因此应用领域有限，方法细节比较粗，多数理解困难，而且相关技术比较少，没有对于安全数据方面的阐述，导致构建本体的过程中存在的问题较多，有着一定的局限性。

4 引入了 WEMLS 的安全数据集成

4.1 本体分析和概念关系

安全数据集成本体模型建立的关键是根据安全数据的特征及其关系，进行安全数据访问控制技术和策略的分类，首先要抽象出其关键概念、属性，用合理的方式表达属性之间的关系，定义类及类层次关系。

定义 4.1 域关系：域是一个集合，设 $D_1, D_2, \cdots, D_n$ 是 n 个域。其上的笛卡尔乘积的任意子集合称为定义在域上的关系。

定义 4.2 关系等价：$R=(P,E)$，R 为关系集合，P 为 R 中的属性集，E 为 R 中的元组集，O 为 R 在语义视图中的映射语义集合，o 为 O 中的原子关系。对于 $\forall r \in R$ $r \subseteq P_1 \times P_2 \times \cdots \times P_n \times E_1 \times E_2 \times \cdots \times E_n$，有且只有 $o_r^i \subseteq O$ $i=1,2,\cdots,n$，使得 r 与 o_r^i 中的元素能够一一映射，则 R 与 O 等价。

对安全数据来说，其一定是来自该安全领域的概念化抽象并为该领域服务。对于系统 S，考虑其在 D 上的关系 $Rel=(A_1 \times A_2 \times \cdots \times A_n)$，其中 $A_i \in Dv_i$，A_i 为属性，Dv_i 为 A_i 的值域。若 S 在 D 上有 n 种关系：$Rel_1, Rel_2, \cdots Rel_n$。对于 Rel_i，可以抽取其属性为本体词汇形成本体元结构词表 St_i。

4.2 建立本体框架

安全数据概念化描述后可生成本体词汇，即本体元结构词表，如图 1 所示。

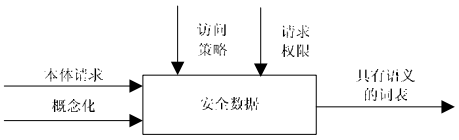

Figure 1. Ontology conceptualization of Security data

图 1 安全数据的本体概念化分析

在图 2 中，本体作为主体请求对安全数据进行概念化表示，依据安全访问策略可形成松散的具有语义的本体词表，该词表应该具有对数据的内核标记和内核封装，需要注意的是，安全数据的概念化要基于请求主体的安全级别，具体实现时，需要综合考虑各种因素，目前需要专家的参与。

4.3 WEMLS 下的基于本体的安全数据集成模型

对于异构安全数据集成访问的本体模型（SHDOM），引入 WEMLS 后，对其数据访问、抽取部分定义表示为 $S<UA, BC, SD, DIS>$，定义中的元素分别表示：进程授权模块、信任度计算模块、系统状态检测模块和信息抽取模块，如图 2 所示。

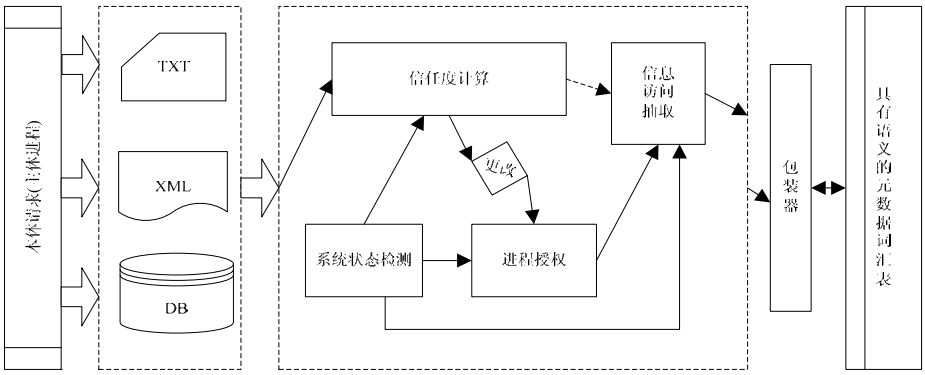

Figure 2. Security data integration model based on WEMLS

图 2 基于 WEMLS 的安全数据集成本体模型

对于任何主体进程来说，所有信息必须首先通过信任度计算模块。同时，系统状态检测模块对信任度计算和进程授权进行监控，未通过信任度计算的主体信息将遵循 BLP 和 Biba 进行数据访问，反之可遵循 WEMLS 策略进行数据访问。

例 1 此处给出高考报名系统考生体检部分的部分 UML 模型，如图 3 所示。

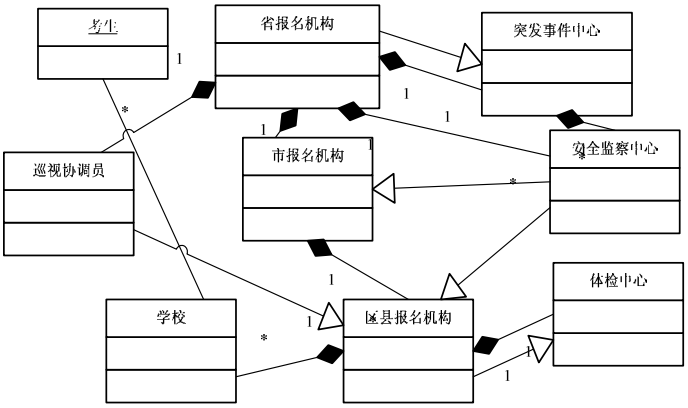

Figure3. Model of college entrance examination physical examination

图 3 高考体检报名系统部分 UML 模型

此系统存在 1 个信任主体 S_t（省招生办公室）和 2 个客体 O_1（区县报名机构）、O_2（本体词表），O_1 为源安全数据，初始值如表 1 所示。

Table 1. System initial value

表 1 系统初始值

	P	I	T	K
S_t	0.3	0.7	0.8	0.6
O_1	1	1	1	1
O_2	1	1	0.5	0.5

根据系统要求，假设系统分别由主体对客体发出操作请求如下：$request_1\ (s_t, r, o_1)$，$request_2\ (s_t, r, o_2)$ 和 $request_3\ (s_t, w, o_2)$。

以下操作中涉及信任权值计算函数部分采用线性指数函数表示。BC 模块对所有请求的信任权值进行评估。$request_1$ 中，$T(s) \le T(o)$，根据规则（1）和图 3 中的安全访问模型，BC 模块通过预先确定的线性指数函数进行 $T^{(d,s)}$ 操作，则三个请求均可被允许。如表 2 所示。

Table 2. Resulting data of system operation

表 2 执行操作后的结果

	$P(s_t)$	$I(s_t)$	$T(s_t)/T^{(d,s_t)}/T_{(d,s_t)}$	$K(s_t)$	*result*
$request_1$	0.3	0.7	0.8/0.9/-	0.6	y
$request_2$	0.3	0.7	0.8/-/-	0.6	y
$request_3$	0.3	0.7	0.8/-/0.5	0.6	y

主体在遵循 BLP 和 Biba 模型时的访问客体不会改变其可信度，在不同情况下根据对主客体要求访问控制的严格程度，则系统对可信主体请求的权值响应动态变化，显然，这种变化首先应该得到进程授权的允许。

5 结论

本文从异构安全数据视角出发，针对异构数据源集成的本体建模进行了分析。针对主体可能会访问不同级别的安全数据问题，考虑如何保证该安全数据的安全性、完整性和可信主体访问的灵活性，本文借鉴安全系统中的 MLS 策略思想，将之应用于安全数据集成，提出了一种更灵活的扩展权值 MLS 策略 WEMLS，并将其进行了概念化。与传统 MLS 相比，

978-1-935068-01-3 © 2009 SciRes.

该策略通过计算主体进程的信任度来动态调节其权值，对其进行量化，从而保证可信主体更灵活地完成对数据的访问。最后，将之应用于基于本体的安全数据集成建模并进行了验证。

References(参考文献)

[1] Mei Hong, Chang Jichuan, Yang Fuqing. Composing software components at architectural level [C].The Int'l Conf on Software Theory and Practice, Beijing, 2000.

[2] D Bell L LaPadula. Secure computer systems: Mathematical foundations and model[R]. Technical Report M74-244, Mitre Corp, 1973.

[3] D E Bell, L LaPadula. Secure Computer Systems: Mathematical Foundations. MTR 2547 Vol. I- III, Mitre Corporation, 1996-11.

[4] K. J. Biba. Integrity considerations for secure computer systems. USAF Electronic System Division, Hanscom Air Force Base, Tech. Rep. : ESD-TR-76-372 , 1977.

[5] Sheth A P, and Larson J A; Federated databases for managing distributed, heterogeneous, and autonomous databases [M]; Computing Surveys; 1990.

[6] Chawathe S, Garcia-Molina H, Hammer J, et al. The TSIMMIS Project: Integration of Heterogeneous Information Sources. In: 10th Meeting of the Information Processing Society of Japan (IPSJ), 1994.7~18

[7] Chaudhuri S, Dayal U. An overview of data warehousing and OLAP technology. SIGMOD record, 1997, 26(1):65~74

[8] Uschold M. Building Ontologies: Towards A Unified Methodology [J]. Expert System 96, Cambridge, 1996.

[9] Fernandez M, Gomez2 perez A, Juristo N. Methontology: from ontological art towards ontological engineering. AAA I-97 Spring Symposium on Ontological Engineering, Stanford University, March 24-26, 1997

[10] Gruniger M, Fox M S .Methodology for the design and evaluation of ontologies[C]. Workshop on Basic Ontol ogical Issues in Knowledge Sharing, JCA I- 95

[11] Gruber T R. Towards principles for the design of ontologies used for knowledge sharing. International Journal of Human-Computer Studies, 1995; (43): 907-928