

## 基于聚类分析的客体聚合信息级别推演方法

曹利峰\* 陈性元 杜学绘 夏春涛  
(解放军信息工程大学电子技术学院 郑州 450004)

**摘要:** 多级客体关系的复杂性,使得等级化网络存在着客体聚合引起信息泄露的问题。针对这一问题,该文提出了基于聚类分析的客体资源聚合信息级别的推演方法,首先依据属性重要程度,对客体属性进行约简,形成属性矢量;然后通过形式概念分析,计算概念引力,对同一安全域内的客体资源进行相似性分析,实现客体资源聚类;最后,依据属性或属性子集级别模糊集可能性测度,推演出由同类客体推导出更高级别信息的可能性。通过该方法,能够有效地制定等级化网络区域边界访问控制策略,控制主体对同一类客体的受限访问,从而降低信息系统失泄密的风险。

**关键词:** 信息安全; 等级保护; 客体聚合; 概念分析; 级别推演

**中图分类号:** TP393.08

**文献标识码:** A

**文章编号:** 1009-5896(2012)06-1432-06

**DOI:** 10.3724/SP.J.1146.2011.01170

## A Level Inference Method for Aggregated Information of Objects Based on Clustering Analysis

Cao Li-feng Chen Xing-yuan Du Xue-hui Xia Chun-tao

(Institute of Electronic Technology, the PLA Information Engineering University, Zhengzhou 450004, China)

**Abstract:** The relations among objects with secure level are very complex, which leads to the problems of security in multi-level network, such as information leakage by object aggregation. This paper puts forward a level inference method for aggregated information of objects based on clustering analysis. This method makes the reduction of attributes by the importance degree of attribute in one object, and attribute vector is formed. Then, according to formal concept analysis, this method accomplishes comparability analysis of objects in the same secure domain by gravity among concepts so that objects can be aggregated. Finally, according to probability estimate of fuzzy set about secure level of attributes or sets of attribute, probability of higher level information inferred by aggregation of similar objects is computed, which may effectively establish access control policy in multi-level network, and accomplish restricted access of congeneric objects in order to reduce the risk of information system.

**Key words:** Information security; Classified security protection; Object aggregation; Concept analysis; Level inference

### 1 引言

信息系统进行分级保护后,系统间无疑会构建起严格的保护和隔离壁垒,形成新的数据孤岛,这与信息化建设中信息交换的实际需要相违背。因此,在不破坏原有信息系统的正常运行和等级化安全特性的前提下,继续保持信息系统间的互联互通是必须要解决的关键问题。

实现不同等级信息系统间的互联互通,必须对主体的访问进行有效控制。经典的多级安全模型BLP(Bell LaPadula)<sup>[1]</sup>,通过为主客体分配安全标记

来实施强制访问控制,尽管达到了多级信息系统间访问控制的目的,但是在等级化网络中,由于安全级别(密级)的存在,使得客体之间的关系变得异常复杂,更易于导致信息的泄漏。

例如,  $o_1, o_2, o_3, o_4, o_5$  为秘密级相似客体,当  $o_1, o_2, o_3, o_4$  信息融合在一起后,推导出的信息级别为机密,  $o_1, o_2, o_3, o_4, o_5$  信息融合后,同样推导出的信息级别也为机密。倘若秘密级主体  $s$  访问了  $o_1, o_2, o_3, o_4$  或  $o_1, o_2, o_3, o_4, o_5$ , 则  $s$  能够根据这些客体信息推导出更高级别的信息,从而造成失泄密。

由此可见,在等级化网络中,主体对客体的访问不再仅仅是遵从BLP模型的简单安全特性<sup>[2]</sup>,也应该考虑客体间的关系,控制主体对相似客体的访问数量,降低系统失泄密的风险。

2011-11-10 收到, 2012-01-19 改回

国家 973 计划前期研究专项(2011CB311801)和河南省杰出科技创新人才计划(114200510001h)资助课题

\*通信作者: 曹利峰 caolf302@sina.com

针对这方面的研究，在文献[3,4]中有所提及，但对此都还未进行深入研究。文献[3]描述了客体之间可能存在的关系，文献[4]提出了基于风险的访问控制模型，在进一步展望中提到了访问控制中应该考虑客体之间的关系；文献[5,6]着重对多级关系数据库的推理问题展开研究，但是其方法并不适用于网络相似客体的聚合问题。而国内在这方面的研究较少。因此，本文在分析客体之间关系的基础上，提出了基于聚类分析的客体聚合信息级别推演方法，该方法通过形式概念分析，对同一安全域内的客体进行了相似性分析，从而实现客体资源的聚类，并依据属性或属性子集级别模糊集可能性测度，推演出由同类客体推导出更高级别信息的可能性，最后根据聚合信息的可能性级别，制定相应的访问控制策略，以有效降低多级网络的泄密风险。

## 2 基于概念相似的客体聚类

为解决同类客体信息聚合而引起的泄密问题，本文首先对客体进行了相似性分析，提出了基于概念相似的递归客体聚类算法，其思路为：对客体进行属性约简，去除不必要、冗余的属性，提取客体概念，通过概念引力，对客体进行相似性分析，并依据聚类因子与噪声因子对聚类进行合理的调整。

### 2.1 客体属性约简

令  $O = \{o_1, \dots, o_n\}$  为客体集合， $A = \{\alpha_1, \dots, \alpha_n\}$  为属性集合； $SL = \{C_1, \dots, C_k\}$  为安全级别集合。一个客体可表示为  $O(\alpha_1, C_1, \alpha_2, C_2, \dots, \alpha_k, C_k)$ 。

**定义 1** (属性重要度) 给定一个客体  $O = (A_o, C_o)$ ，其中  $A_o \subseteq A$ ， $C_o \subseteq SL$ ，对于  $\forall B \subseteq A_o$ ，若  $\forall \alpha \in B$ ，则定义  $\text{sig}(\alpha, B, A_o) = H(B) - H(B - \{\alpha\})$  为属性  $\alpha$  对属性集  $B$  的重要度。 $H(B)$  为  $B$  的信息熵。

在上述定义的基础上，本文给出了基于属性重要度的客体属性约简算法(Algorithm of Attribute Reduction Based on Significance, AARBS)，即在保持客体特性不变的情况下，删除不必要、冗余的属性，以保证客体聚类的有效性。算法如表 1 所示。

该算法主要包括 3 个关键部分：第 1 部分计算属性集  $B$  中某一个属性对属性集  $B$  的重要度；第 2 部分判断该单属性对属性集  $B$  来说是否重要，若重要度小于阈值  $\eta$  时，将该属性从属性集中去除；第 3 部分判断属性集  $B$  和原始的属性集  $A$  的信息熵是否相同，若不同，则说明该属性的去除将影响到客体  $O$  的特性。算法计算复杂度为  $O(n)$ 。

### 2.2 概念与概念核

$A$  表示某一安全域内的属性空间，由单属性  $\alpha_i$  所组成； $N$  表示  $A$  的属性值空间，属性值  $v_i \in N$ 。

表 1 客体属性约简算法

```

input A
Output B(B is a reduction of A)
(1) B=A; n=|A|; j=0;
(2) for i=1 to n do
(3) begin
(4)    $\alpha_i \in B$ ; sig =  $H(B) - H(B - \{\alpha_i\})$ ;
(5)   if sig <  $\eta$  then
(6)      $B = B - \{\alpha_i\}$ ; j = i;
(7)   if  $H(B) <> H(A)$  then  $B = B + \{\alpha_j\}$ ;
(8) endfor
(9) return B;

```

**定义 2** (属性矩阵) 令  $o_i \in O$ ，客体  $o_i$  的属性用  $\omega = \langle \alpha_{i1}, \alpha_{i2}, \dots, \alpha_{ik} \rangle$  表示；客体  $o_i$  的属性值用  $v = \langle v_{i1}, v_{i2}, \dots, v_{ik} \rangle$  表示；客体  $o_i$  的属性安全级别用  $\pi = \langle C_{i1}, C_{i2}, \dots, C_{ik} \rangle$ ；客体  $o_i$  的属性概率用  $\rho = \langle p_{i1}, p_{i2}, \dots, p_{ik} \rangle$  表示，即  $\langle \alpha_{ij}, v_{ij}, C_{ij} \rangle$  在  $o_i$  中出现的频率；存在属性矩阵为

$$M = \omega \times v \times \rho \times \pi = \begin{bmatrix} \alpha_{i1} & \alpha_{i2} & \dots & \alpha_{ik} \\ v_{i1} & v_{i2} & \dots & v_{ik} \\ p_{i1} & p_{i2} & \dots & p_{ik} \\ C_{i1} & C_{i2} & \dots & C_{ik} \end{bmatrix}$$

属性矩阵  $M$  等价于  $\{\beta_1, \beta_2, \dots, \beta_k\}$ ，其中  $\beta_j = \langle \alpha_{ij}, v_{ij}, p_{ij}, C_{ij} \rangle$  ( $j = 1, 2, \dots, k$ ) 为  $M$  的列，表示客体  $o_i$  的单个属性向量。

**定义 3** (概念<sup>[7,8]</sup>) 令  $\beta$  为所有属性矩阵  $M$  列向量的集合，给定一个三元组  $\langle O, \beta, R \rangle$ ，其中， $R$  为一个二元关系， $R \subseteq O \times \beta$ ，表示某一客体  $o_i$  拥有属性  $\beta_j$ 。令  $X \subseteq O$ ， $X' = \{\zeta \in \beta \mid \forall x \in X : (x, \zeta) \in R\}$ ；令  $Y \subseteq \beta$ ， $Y' = \{x \in O \mid \forall \zeta \in Y : (x, \zeta) \in R\}$ ；若  $X' = Y$ ， $Y' = X$ ，则称二元对  $\langle X, Y \rangle$  为属性空间  $A$  上的一个概念，记作  $\text{concept}(X)$ 。

**定理 1** 任何一个客体  $o_i$ ，均存在一个概念  $\langle \{o_i\}, \{\beta_j\} \rangle$  与之对应。

**证明** 令  $o_i$  的属性矩阵  $M = \{\beta_1, \beta_2, \dots, \beta_k\}$ ，令  $\{o_i\} = X \in O$ ，由于存在二元关系  $o_i R \beta_j$ ，故有  $X' = \{\beta_j\}$ ；令  $\{\beta_j\} = Y \subseteq \beta$ ，对于  $\forall \beta_j$ ，均存在二元关系  $o_i R \beta_j$ ，故有  $Y' = \{o_i\}$ 。可见  $X' = Y$ ， $Y' = X$ 。

**定义 4** (概念核) 令  $X = \{o_i, o_j, \dots, o_k\} \subseteq O$ ，且  $X$  中的客体相似，它们共同的必要属性项所组成的集合为  $\{\beta_l\} \subseteq \beta$ ，令  $Y = \{\beta_l\}$ ，则称  $\langle X, Y \rangle$  为  $o_i, o_j, \dots, o_k$  的概念核，记  $\text{Ccore}(X)$ 。

**定理 2** 给定某一安全域内的客体集合  $O$ ， $o_1, o_2 \in O$ ，若  $o_1, o_2$  相似，则  $o_1, o_2$  必有相似的概念，反之亦然。

**定理 3** 概念核  $\text{Ccore}(X)$  是属性集  $A$  的一个概念，且  $\text{Ccore}(X) = \bigcap \text{concept}(o_i)$ ， $o_i \in X$ 。

2.3 概念引力与聚类评估因子

为有效实现客体的聚类, 本文引入了引力场理论<sup>[9]</sup>, 通过计算概念之间的引力, 发现相似概念。引力越大, 相似性越高。

**定义 5 (概念引力)** 给定  $o_i, o_j \in O$ , 两客体的概念质量分别记为  $m_i, m_j$ , 概念距离记为  $r_{ij}$ , 概念引力定义为  $g(o_i, o_j) = G(m_i m_j / r_{ij})$ , 其中  $G = 1$ , 为引力常量;  $m_i = \sum_{\alpha_k \in A(o_i)} H(\alpha_k)$ , 指的是概念中属性信息熵和。

$$r_{ij} = \sqrt{\sum_{\alpha_s \in A(o_i), \alpha_t \in A(o_j)} (\alpha_s, \alpha_t)^2}$$

为评估聚类效果, 本文给出了两个聚类评估因子, 分别为聚类因子(Aggregation Gene, AG)和噪声因子(Noise Gene, NG)。

**定义 6 (聚类因子)** 令  $g_{in}$  为类内容体聚合力,  $g_{out}$  为各类之间的平均引力。用  $Z_i$  表示某一个类,  $q$  为聚类的数量。则有  $g_{in}^{Z_k} = (1/|Z_k|) \cdot \sum_{o_i, o_j \in Z_k} g(o_i, o_j)$ ,  $g_{out} = g(Z_i, Z_j)$ 。其中  $g(Z_i, Z_j)$  的计算依据  $Z_i, Z_j$  的概念核中的属性。  $\bar{g}_{in} = \frac{1}{q} \sum_{k=1}^q g_{in}^{Z_k}$ , 表示类内容体聚合力均值。  $\bar{g}_{out} = \frac{1}{q} \cdot \sum_{Z_i, Z_j \in Z} g(Z_i, Z_j)$ , 表示类间引力均值。聚类因子定义为  $AG = \frac{\bar{g}_{in}}{\bar{g}_{in} + \bar{g}_{out}}$ 。

**定义 7 (噪声因子)** 假定有  $t$  个噪声点(未被分类的客体), 则噪声  $o_i$  对类的引力为  $g_{noise}^{o_i} = (1/q) \cdot \sum_{k=1}^q g(o_i, Z_k)$ , 其中  $g(o_i, Z_k)$  的计算依据  $o_i$  的概念、 $Z_k$  的概念核中的属性。  $\bar{g}_{noise} = \frac{1}{t} \sum_{i=1}^t g_{noise}^{o_i}$ , 即噪声对类的引力均值。噪声因子  $NG = \bar{g}_{in} / (\bar{g}_{in} + \bar{g}_{noise})$ 。

可见, AG, NG 越接近1, 聚类效果越好。

2.4 基于概念相似的递归客体聚类算法

在此基础上, 本文提出了基于概念相似的递归客体聚类算法 (Recursive Clustering Algorithm Based on similar Concept, RCABC)。如表 2 所示。

该算法主要分为 3 大部分, 第 1 部分为客体属性约简(1-5 行), 并形成客体概念; 第 2 部分为客体概念相似性分析, 并形成类(6-19 行); 第 3 部分为依据类内与类间聚合因素, 评估聚类的好坏, 若达不到系统要求, 则调整相似度阈值  $\delta$ , 再次聚类, 直到满足需求(20-40 行)。时间复杂度为  $o(n \lg n)$ 。

表 2 递归客体聚类算法

---

Input: $O, A$ (objects set; Attribute set)
Output: $Z$ (cluster set)
(1) for $i=1$ to $ O $ do
(2) begin
(3) $A = A(o_i); B_i = AARBS(A);$
(4) $M_i = \maxtrix(B_i); \text{concept}(o_i) \leftarrow M_i;$
(5) endfor
(6) $k=1; Z_1 = \{o_1\};$
(7) for $i=1$ to $ O $ do
(8) begin
(9) if $i < > 1$ then
(10) if $o_i$ is not yet assigned to a cluster then
(11) for $j=i+1$ to $ O $ do
(12) begin
(13) $g = g(o_i, o_j);$
(14) if $g \geq \delta$ then
(15) if $o_j$ is not yet assigned to a cluster then
(16) $Z_k \leftarrow o_j;$
(17) endfor
(18) $k=k+1;$
(19) endfor
(20) $g_{in} = 0; g_{out} = 0; g_{noise} = 0; t=0;$
(21) for $i=1$ to $k$ do
(22) begin
(23) $g_{in} = g_{in} + g_{in}^{Z_i};$
(24) endfor
(25) for $i=1$ to $k$ do
(26) begin
(27) for $j=i+1$ to $k$ do
(28) begin
(29) $g_{out} = g_{out} + g(Z_i, Z_j);$
(30) endfor
(31) endfor
(32) for $i=1$ to $ O $ do
(33) begin
(34) if $o_j$ is not yet assigned to a cluster then
(35) compute $g_{noise}^{o_i}; t=t+1;$
(36) $g_{noise} = g_{noise} + g_{noise}^{o_i};$
(37) endfor
(38) compute AG, NG;
(39) if $AG < \theta; NG < \omega$ then $\delta \leftarrow \delta'; \text{goto } 8;$
(40) else output $Z;$

---

3 相似客体聚合信息级别的推演

在依据上述算法对客体进行有效分类之后, 本文通过属性级别模糊集可能性测度<sup>[10-12]</sup>对同类客体推导出更高安全级别信息的可能性进行了评估。其基本思想为: 在空间  $A \times N$  上定义属性安全级别模糊集  $F_c = \{f_{c_1}, f_{c_2}, \dots, f_{c_k}\}$ , 其中  $c_1 < c_2 < \dots < c_k$ , 依据同类客体属性集子集在  $F_c$  上的可能性分布, 模

糊推演出同类客体推导更高安全级别信息的可能性，并给出相似客体访问的阈值。

### 3.1 属性级别模糊集可能性测度

为模糊推演出多个同类客体是否能够推导高级别信息的可能性，本文引入了属性子集在不同安全级别上的可能性测度。

**定义 8** 设  $U$  为一个属性域， $U = A \times N$ 。  $F_c = \{f_{c_1}, f_{c_2}, \dots, f_{c_k}\}$  为论域  $U$  上的模糊集， $f_{c_i}$  代表不同的安全级别属性子集集合，令  $X$  为  $U$  上取值的变量，与  $X$  有关的可能性分布，记为  $\Pi_x$ ， $\Pi_x$  的可能性分布函数用  $\pi_x$ ，并在数值上定义等于  $F_c$  的隶属度，即  $\forall u \in U, \pi_x(u) = F_c(u)$ 。

**定义 9** 设  $B$  为安全级别  $C(f_{c_i})$  的属性域  $U$  上的模糊集， $\pi_x$  是与变量  $X$  关联的可能性分布， $X$  在  $U$  中取值，则  $B$  的可能测度定义为

$$\text{Poss}(X \text{ is } B) \cong \bigvee_{u \in U} (B(u) \wedge \pi_x(u))$$

其中  $B(u)$  为  $B$  的隶属函数； $\pi_x(u)$  为与  $X$  有关的可能性分布函数。

### 3.2 相似客体聚合信息级别推测公式

**定义 10** 相似客体推导出更高安全级别信息的可能性为相似客体属性集合  $\Omega$  ( $\Omega$  属性元素允许重复，重复的在子集划分时可视为一个元素) 的各个子集在高于相似客体最大安全级别的模糊集上的可能性测度。

$$\text{Pro} = \text{Poss}(X \text{ is } B \mid X \subseteq \Omega, B \in F_c,$$

$$C(B) > \max(C(O_1), C(O_2), \dots, C(O_n)))$$

该公式表示了由相似客体聚合信息推导出更高安全级别信息的可能性， $X$  为属性子集。

**定义 11** 若  $\text{Pro} \geq \tau$ ，且  $f_{c_i}(Y) = \text{pro}$ ，则认为由相似客体聚合信息(由属性子集  $Y$  组成)能够推出安全级别为  $C_i$  信息。令  $Y[O] = \{\xi_1, \xi_2, \dots, \xi_k\}$  表示能够获取属性子集  $Y$  的相似客体子集的集合， $\xi_i$  为客体子集。令  $\psi = \min\{|\xi_1|, |\xi_2|, \dots, |\xi_k|\}$ ，则称  $\psi$  为相似客体被读取的阈值。

### 3.3 相似客体聚合信息级别推演算法

据此，本文给出了相似客体聚合信息推导更高级别信息可能性推演算法 LIAAI(Level Inference Algorithm of Aggregated Information)。如表 3 所示。

该算法首先通过相似客体形成属性集  $\Omega$ ，将  $\Omega$  划分为  $2^n - n$  ( $n$  不包括重复的属性) 个属性子集，并根据相似客体中最大安全级别  $L$ ，分别计算属性子集在各个安全级别  $c_i$  ( $c_i > L$ ) 模糊集上的可能性  $\text{pro}_{c_i}$ ，若有相似客体聚合信息推导  $c_i$  级别信息的可能性非常大，则根据属性子集获取相应的相似客体

表 3 相似客体聚合信息推导更高级别信息可能性推演算法

```

Input:  $O$  (similar objects set)
Output:  $\text{pro}; \psi$ 
(1)  $\Omega \leftarrow O$  ;// form attribute subset
(2)  $\Omega = \{A_1, A_2, \dots, A_k\}$  ;
(3)  $L = \max(C(o_i) \mid o_i \in O)$  ;// maximum level
(4) for each  $c_i \in C$  do;
(5)   begin
(6)     if  $c_i > L$  then{
(7)        $B = f_{c_i}$  ;//fuzzy set of  $c_i$ 
(8)        $\text{pro}_{c_i} = \text{Poss}(X \text{ is } B \mid B = f_{c_i}, X \in \Omega)$  ;
(9)       if  $\text{pro}_{c_i} \geq \tau_{c_i}$  then {
(10)        Attr  $\leftarrow \{A_i \mid p(A_i \mid C_i) = \text{pro}_{c_i}\}$  ;
(11)        get  $\xi_i$  which can achieve Attr;
(12)         $\psi = \min(|\xi_1|, |\xi_2|, \dots, |\xi_k|)$  ;
(13)        output( $\text{pro}_{c_i}, \psi$ ) ;
(14)      }else{ output( $\text{pro}_{c_i}, \infty$ ) }//no exist
(15)    }
(16)  endfor

```

子集，最终得到相似客体在级别  $c_i$  上的最大访问数量。算法的时间复杂度为  $o(k \cdot n \lg n)$ ， $k$  为安全级别的个数， $n$  为客体的数量。

## 4 仿真实验分析

为了说明本文所提出的推演方法的执行效果，我们采用人工合成文件和实际文件作为仿真实验客体，人工合成文件的目的是有两种，一种是为了产生与实际文件内容相近的文件，一种是将实际文件分割为不同的相近文件，实际文件来源于本单位科研性文档。实验的环境为 Visual C++ 6.0, Intel® Pentium® 4 CPU 3.00 GHz, 2 G 内存，软件环境是 Microsoft Windows Server 2003 sp2。

在实验中，我们将客体划分为公开、内文、秘密、机密和绝密 5 个安全等级，并依据客体特征提取客体属性，对客体属性、属性值进行量化，构造了属性及属性子集在各个安全级别上的可能性分布。实验评估的指标为算法执行效率、聚类质量与高级别信息推导准确率等。

针对上述选取的客体文件，我们对该方法中的聚类算法以及推演算法进行了性能评估。设定客体个数  $n$  分为 400, 800, 1000, 1200, 1600, 1800, 2500 个，客体相似度阈值  $\delta = 1$ ，聚类因子阈值  $\theta = 0.75$ ，噪声因子阈值  $\omega = 0.8$ ，推出高级别信息可能性阈值  $\tau = 0.75, 0.7, 0.65, 0.6$ 。通过实验，客体相似度阈值  $\delta$  不断变化，满足聚类因子、噪声因子后，最终  $\delta = 2.95$ 。如表 4 所示。

在实验中， $\delta$  的初始值为 1，依据聚类因子、噪

表4  $\delta$  迭代变化表

次数	$\delta$	$\theta = 0.75$	$\omega = 0.8$
1	1	<	<
2	2	<	<
3	3	>	<
4	2.5	<	<
5	2.75	<	<
6	2.95	>	>

声因子进行调整。表4中，前3次迭代 $\delta$ 值的变化幅度较大(为1)，当 $\delta=3$ 时，聚类因子 $AG > \theta$ ，而噪声因子 $NG < \omega$ ，说明此时尽管分类比较合理，但是噪声较多，将会影响聚类的质量，影响算法的准确率。本文采用中间取值法，最终得到客体相似度阈值 $\delta=2.95$ 时，聚类质量较好。

为验证 RCABC 算法聚类质量的好坏，本文依据聚类算法得出的分类与实验数据标准分类之间的差距占客体总数的百分比来进行考核。其中，实验数据标准分类由人工分析仿真实验客体获得，而差距为每个类内不同元素个数与噪声异同个数之和。

RCABC 算法相对于层次聚类算法 BIRCH 和基于密度的聚类算法 DBSCAN 的聚类质量对比如图1所示。从图1可看出，在相同客体数量情况下，RCABC 算法的聚类质量优于 BIRCH 算法和 DBSCAN 算法。

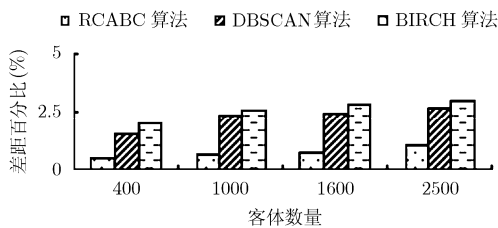


图1 RCABC 算法与其他算法聚类质量对比

通过实验，针对不同数量的客体，本文对 RCABC 算法和 LIAAI 算法进行了执行效率评估。对于 RCABC 相似客体聚类算法来说，执行的效率与客体数量、属性空间大小有关，也与相似阈值 $\delta$ 的设定有关，因为 $\delta$ 值设定是否合理决定着算法迭代的次数；对于 LIAAI 相似客体聚合信息级别可能性推演算法来说，执行效率与客体数量、安全级别数量有关，因为需要找出获取高级别属性子集的所有客体子集。两个算法的执行效率如图2所示。

LIAAI算法推演的准确率，与RCABC算法聚类质量、高级别信息可能性阈值 $\tau$ 有关。为验证LIAAI算法推演的合理性，本文依据LIAAI算法推演的准

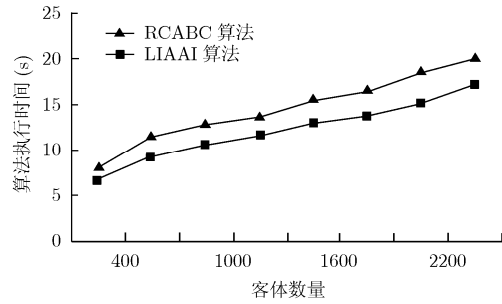


图2 RCABC 和 LIAAI 算法执行效率

确率(VR)、误差率(ER)和错误率(FR)进行衡量。准确率VR是指LIAAI算法推导出的客体聚类问题数占实验数据标准客体聚类问题总数的百分比。误差率ER是指LIAAI算法未推导出的客体聚类问题数占实验数据标准客体聚类问题总数的百分比， $ER=1-VR$ 。错误率FR是指未包含在实验数据标准客体聚类问题中的LIAAI算法推导出的客体聚类问题数占实验数据标准客体聚类问题总数的百分比。

在 $\delta=2.95$ ， $\tau=0.75, 0.7, 0.65, 0.6$ 下，算法推演的准确率、错误率如表5所示。

表5 LIAAI 算法推演的 VR, FR(%)

$\tau$	$n$					
	400		1600		2500	
	VR	FR	VR	FR	VR	FR
0.75	79.27	1.33	81.25	1.62	77.34	1.05
0.70	90.62	1.63	88.65	1.76	85.68	1.05
0.65	91.22	2.71	90.56	2.15	90.76	1.18
0.60	91.87	3.26	91.77	2.26	92.28	1.36

由于RCABC算法存在着聚类质量评估因子AG和NG，AG, NG越接近1则聚类质量越好，且本文通过聚类与噪声因子阈值调整客体相似度阈值 $\delta$ ，能够有效地保证聚类的质量。而阈值 $\tau$ 设置的越合理，挖掘出相似客体聚类问题的准确率越高。若 $\tau$ 越高，则放行的客体过多，准确率降低，从而导致泄密；若 $\tau$ 越低，则使得冗余的客体较多，错误率高，将会影响访问控制系统的执行效率。因此，如何能够自动挖掘出合理的阈值是提高算法执行效率、准确率的关键，这是下一步研究的难点。

### 5 结束语

本文通过将同一安全域内客体进行相似性挖掘，依据属性特征、属性级别模糊集可能性测度对相似客体间聚合后能否得出更高密级信息的可能性进行了推演。该方面的研究将改变多级网络环境中

区域边界安全防护的基本原则, 使其拓宽到等级化客体间的关系上来, 以有效地控制多级网络中主体对客体的访问, 从而进一步降低失泄密的风险。

下一步我们将针对关联客体聚合信息推导更高级别信息的可能性进行深入研究, 进一步解决由于等级化客体关系的存在的泄密问题。

### 参考文献

- [1] Bell P D E and Padula L J L. Secure computer system: unified exposition and multics interpretation, ESD-TR-75-306[R]. MTR 2997 Rev.1, The MITRE Corporation, 1976.
- [2] 秦超, 陈钟, 段云所. 中国墙策略及其在多级安全中的应用[J]. 北京大学学报, 2002, 38(3): 369-374.  
Qin Chao, Chen Zhong, and Duan Yun-suo. Chinese wall policy and its extension in multilevel security system[J]. *Acta Scientiarum Naturalium Universitatis Pekinensis*, 2002, 38(3): 369-374.
- [3] Pernul G, Winiwarter W, and Tjoa A M. The entity-relationship model for multilevel security[C]. Proceedings of the 12th International Conference on the Entity-relationship Approach: Entity-relationship Approach, Arlington, Texas, USA, Dec. 15-17, 1994: 166-177.
- [4] Cheng P C, Rohatgi P, and Keser C. Fuzzy multi-level security: an experiment on quantified risk-adaptive access control[C]. IEEE Symposium on Security and Privacy. Oakland, CA, May 20-23, 2007: 222-230.
- [5] Emilin C and Swamynathan S. Reason based access control for privacy protection in object relational database systems [J]. *International Journal of Computer Theory and Engineering*, 2011, 3(1): 32-37.
- [6] Santosh Kumar Chauhan. A study of inference control techniques [D]. [Ph.D. dissertation], National Institute of Technology Rourkela, India, 2010.
- [7] 张明卫, 刘莹, 张斌, 等. 一种基于概念的数据聚类模型[J]. 软件学报, 2009, 20(9): 2387-2396.
- [8] 史金成, 胡学刚. 基于二部图的概念聚类研究[J]. 计算机工程与应用, 2010, 46(13): 132-134.  
Shi Jin-cheng and Hu Xue-gang. Research on conceptual clustering based on bipartite graph[J]. *Computer Engineering and Applications*, 2010, 46(13): 132-134.
- [9] 涂文燕, 李德毅, 王建民. 一种基于数据场的层次聚类方法[J]. 电子学报, 2006, 34(2): 258-262.  
Gan Wen-yan, Li De-yi, and Wang Jian-min. A hierarchical clustering method based on data fields [J]. *Acta Electronica Sinica*, 2006, 34(2): 258-262.
- [10] 罗承忠. 模糊集引论(下册)[M]. 北京: 北京师范大学出版社, 2007: 13-14.  
Luo Cheng-zhong. Fuzzy Set Theory (Volume II)[M]. Beijing: Beijing Normal University Press, 2007: 13-14.
- [11] Jan Noessner and Mathias Nipert. ELOG: a probabilistic reasoner for OWL EL[C]. RR'11 proceedings of the 5th International Conference on Web Reasoning and Rule Systems, Galway, Ireland, Aug. 2011: 281-286.
- [12] 刘雄, 卓雪君, 汤永利, 等. 一种基于通信信道容量的多级安全模型[J]. 电子学报, 2010, 38(10): 2460-2464.  
Liu Xiong, Zhuo Xue-jun, Tang Yong-li, et al. A multilevel security model based on communication channel capacity [J]. *Acta Electronica Sinica*, 2010, 38(10): 2460-2464.

曹利峰: 男, 1981年生, 讲师, 博士生, 研究方向为网络安全。  
陈性元: 男, 1963年生, 博士, 教授, 博士生导师, 研究方向为信息安全、分布式操作系统。  
杜学绘: 女, 1968年生, 博士, 教授, 硕士生导师, 研究方向为信息安全、算法分析。