普适计算认证过程基于粗糙集的隐私保护策略①

张燕武　王国军　燕　锋　(中南大学 信息科学与工程学院 湖南 长沙 410083)

摘　要: 在普适计算环境中，用户要获得需要的服务，需要向对应的服务提供商提供一定的认证信息，而这些认证信息中往往包含有用户不希望泄漏的隐私信息。为了对这些隐私信息进行保护，本文提出了认证过程中基于粗糙集的隐私保护策略：用户将认证信息扩展成粗糙集提供给服务提供商；服务提供商根据策略从粗糙集中提取用户的真实认证信息对用户请求进行认证。该策略充分利用了粗糙集合的不确定性，能够有效地防止用户隐私泄漏。

关键词: 普适计算; 认证; 粗糙集; 隐私保护; 不确定性

Privacy Preserving Based on Rough Set Theory in Authentication for Pervasive Computing

ZHANG Yan-Wu, WANG Guo-Jun, YAN Feng

(School of Information Science and Engineering, Central South University, Changsha 410083, China)

Abstract: In the pervasive computing environment, the users have to provide certain authentication information to the services provider to get the services they need, but the authentication information often contains the privacy which the users do not want to be leaked. In order to protect this privacy, this paper proposed a privacy preserving policy based on rough set theory in authentication for pervasive computing: the users expand the authentication information to a rough set and then provide the rough set to the services provider, the services provider extracts the users' true authentication information from the rough set according to the pervasive preserving policy and then authenticate the users' service request. The privacy preserving policy makes use of the uncertainty of rough set and can effectively prevent the leakage of user's privacy.

Keywords: pervasive computing; authentication; rough set; privacy preserving; uncertainty

1　引言

在普适计算的环境中，人们可以随时随地、透明地享受数字化的服务[1]。用户享受服务的前提是向服务提供商提供对应的认证信息，服务提供商根据自己的认证策略来决定用户对该服务的访问权限。然而，这些认证信息中往往包含用户不希望泄露的隐私信息。例如：某病人通过自己的设备向医生提供病史信息时，他当然不希望这些信息被周围的人窃取，甚至被恶意地利用。在传统的计算模式下，用户可以通过加密手段来保护自己的隐私信息。然而，普适计算环境中的设备通常将嵌入式处理器和各种功能模块集成在一起，集成后体积小，并且集计算、通信、传感、供电等功能于一体[2]。这就决定了普适计算环境中的设备大多是资源有限的轻型设备，大量的加密算法由于复杂度高而不能在普适计算系统中大规模地使用。本文将粗糙集理论应用到用户的隐私保护中来，利用粗糙集合的不确定性，能够有效地防止用户隐私泄漏。

本文余下的部分组织如下：第二节介绍了粗糙集以及相关理论知识；第三节提出了基于粗糙集的隐私保护策略；第四节对本文提出的隐私保护策略进行了

① 基金项目:国家自然科学基金重大研究计划项目(90718034);国家自然科学基金面上项目(60773013)
收稿时间:2010-03-02;收到修改稿时间:2010-03-26

分析；最后总结全文。

2 相关概念与结论

本节介绍本文将用到的经典集合论中一些重要的概念和结论，这些概念和结论也是粗糙集的理论基础。

2.1 等价关系与等价类

定义 1. (等价关系)

如果集合 A 上的关系 R 满足：1)自反性，即 A 中的任一元素和自身具有关系 R；2)对称性，即 $\forall x,y\in A$，若 xRy，则必有 yRx；3)传递性，即 $\forall x,y,z\in A$，若 xRy 且 yRz，则必有 xRz；则 R 是集合 A 上的等价关系。

定义 2. (等价类)

设 R 是非空有限集合 U 上的等价关系，$\forall x\in U$，定义 $[x]_R=\{y\mid yRx\wedge y\in U\}$ 为元素 x 关于 R 的等价类。

2.2 集合的划分

定义 3. (集合的划分)

给定一个集合 A 和 A 的非空子集簇 $p=\{A_1,A_2,\mathbf{L}\ A_m\}$，如果满足：1) $A_i\neq\varnothing$，2) $A=\bigcup_{i=1}^{m}A_i$，3) $A_i\cap A_j=\varnothing$，$i\neq j$，$1\le i,j\le m$；则称子集簇 p 是集合 U 的一种划分，划分的元素 A_i 称为划分 p 的块。有了上述的定义，下面给出一个很有用的定理。

2.3 集合等价关系与划分一一对应定理

定理 1. 非空有限集合 A 上的等价关系与集合 A 的划分是一一对应的[3]。

证明：

充分性：设给定一个非空有限集合 A，R 是 A 上的一个等价关系，由等价类的性质可知 $p=\{[x]_R\mid\forall x\in A\}$ 是 A 的一个子集簇。1)由 $x\in[x]_R$ 得知 $[x]_R\neq\varnothing$；2)若 $y\in[x]_R$ 由关系 R 的对称性可知 $[x]_R=[y]_R$，若 $y\notin[x]_R$ 则 $[x]_R\cap[y]_R=\varnothing$；3)显然 $\bigcup_{\forall x\in A}[x]_R=A$。由定义 3 可知由等价关系 R 导出的集合 p 是集合一个划分。

必要性:设给定非空有限集合 A 以及 A 的一个划分 $p=\{A_1,A_2,\mathbf{L}\ A_m\}$，设由该划分导出的关系为：$R=\{(x,y)\mid\exists k,k\le m,x,y\in A_k\}$ 显然 R 满足自反性和对称性。下面证明其传递性:设 $\forall x,y,z\in A$，$xRy\Leftrightarrow x,y\in A_k$，且 $yRz\Leftrightarrow y,z\in A_l$，如果 $A_k\neq A_l\Rightarrow A_k\cap A_l\neq\varnothing,k\neq l$，则这与划分的定义矛盾，因此有 xRz。由此可知 R 具有自反性、对称性以及传递性，根据等价关系的定义可知由划分 p 导出的关系 R 是 A 上的等价关系。这样定理就得到了证明。

2.4 粗糙集介绍

1982 年，波兰华沙理工大学的科学家帕拉克(Z. Pawlak)[3]提出了粗糙集理论。粗糙集理论具有很强的定性分析能力，能够有效地表达不确定的或不精确的知识，成为一种处理不确定性问题的新的数学工具。粗糙集理论是经典集合论的推广，在经典集合论中，一个对象 x 是否隶属于某个集合 A 必须是“非此即彼”的，要么 $x\in A$，要么 $x\notin A$，二者必居其一。而粗糙集理论推广了经典集合论的对象与集合的关系[4,5]，一个对象 x 是否隶属与某个集合 A 是“亦此亦彼”的，对象 x 肯定属于集合 A，记为 $x\in A$，或对象 x 肯定不属于 A，记为 $x\notin A$，或对象 x 属于 A 的可能程度为 $m_A^R(x)$，$m_A^R(x)\in$ (0,1).即论域 U 的任意一个子集都一一对应唯一的值域为[0,1]的一个特征函数，即 $\forall A\subset U$，

$$A\leftrightarrow m_A^R(c_A):U\leftrightarrow[0,1],x\ \mathbf{a}\begin{cases}1,x\in A,\\ r,0<r=m_A^R(x)<1,\\ 0,x\notin A.\end{cases}$$

其中 $m_A^R(x)=\left|[x]_R\ \mathbf{I}\ A\right|/\left|[x]_R\right|$ 称为 x 对于 A 的粗糙隶属度，$|\bullet|$ 表示集合的基数(集合元素的个数)。

2.5 粗糙集的定义

在介绍粗糙集定义之前，我们先介绍粗糙集理论中两个非常重要的概念：上近似和下近似[3]。我们称一个论域 U 和 U 上的一簇等价关系 S 的二元组合 $K=(U,S)$ 为论域 U 上的一个知识库。给定知识库 $K=(U,S)$，则 $\forall X\subseteq U$ 和论域 U 上的一个等价关系 R，X 关于等价关系 R 的下近似和上近似分别定义为：

$\underline{R}(X)=\{x\mid(\forall x\in U)\wedge([x]_R\subseteq X)\}$，

$\overline{R}(X)=\{x\mid(\forall x\in U)\wedge([x]_R\cap X\neq\varnothing)\}$。

定义 4. (粗糙集)

对于上述的 $\underline{R}(x)$ 和 $\overline{R}(x)$，若 $\underline{R}(x)\neq\overline{R}(x)$，则称集合 X 是关于论域 U 的相对于 R 的 R 粗糙集。

集合的 X 上、下近似如图 1 显示。从定义以及图 1 可以看出，下近似是由那些根据等价关系 R 判断肯定属于集合 X 的论域 U 中的元素组成的集合，如图 1 中绿色网格表示的部分；而上近似是由那些根据 R 判断肯定属于或者可能属于集合 X 的论域 U 中的元素组成的集合。如图 1 中黄色网格和绿色网格的并集。集合 $\overline{R}(x)-\underline{R}(x)$ 为 X 的 R 边界域，边界域是由那些根据知识 R 既不能判断肯定属于 X 又不能判断肯定不

属于 X 的论域 U 中的元素组成的集合。如图 1 中黄色网格表示的部分。正是因为有了边界域的存在，粗糙集合才有了不确定性。

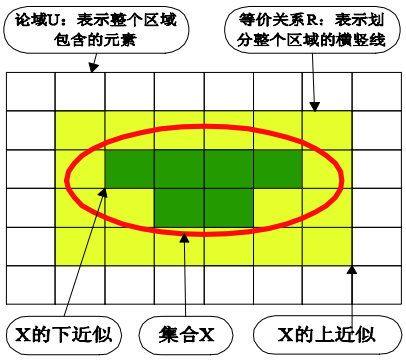

图 1　集合 X 的上、下近似示意图

3　基于粗糙集的隐私保护策略

以上述粗糙集理论为基础，本节提出了普适计算环境认证过程中的隐私保护策略。在由服务提供商、用户组成的普适计算环境中，用户 User 需要服务提供商 SP 提供的某种服务，而又不希望自己的隐私泄漏给周围环境中的其他用户。基于粗糙集的隐私保护策略的具体过程如图 2 所示：

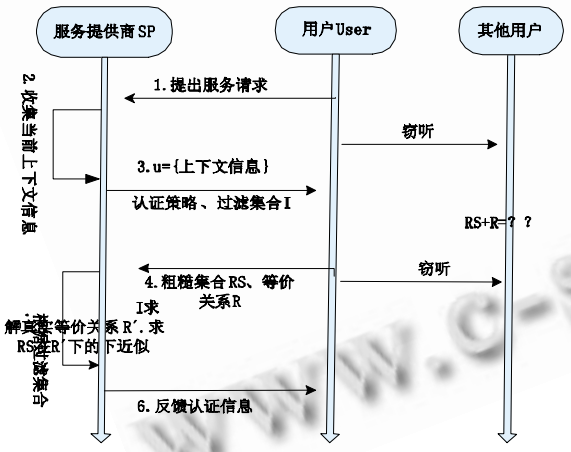

图 2　基于粗糙集隐私保护策略示意图

步骤 1：处于普适计算环境下的用户 User 感知到服务提供商 SP 提供的服务类型，并根据自身的需要向 SP 提出享受某种服务的请求。

步骤 2：服务提供商 SP 收到 User 的请求后，收集当前环境的上下文信息，这些上下文信息包括服务提供商 SP 能感知到的用户的基本属性，同时保存用户 User 的 ID。

步骤 3：服务提供商 SP 将收集的上下文信息集合 S={上下文信息}，用户 User 所申请服务的认证策略 $K=\{A_1,A_2,\mathbf{L}\,A_m\}$ 传输给请求服务的用户 User，K 中的元素代表用户申请的服务所对应的认证信息，比如用户的身份、位置信息等敏感信息，这些信息对用户来说属于隐私信息。同时服务提供商 SP 向用户提供一个集合 $I=\{i_1,i_2,\mathbf{L}\,i_n\}$，并满足 $I\cap S=\varnothing$。假设服务提供商 SP 能安全地将 I 传输给用户 User。

步骤 4：用户 User 收到 S，K，I 0 后开始整理认证数据。将认证信息 $k=\{a_1,a_2,\mathbf{L}\,a_m\}$，集合 I 以及冗余信息 $T=\{t_1,t_2,\mathbf{L}\,t_n\}$ 写成一个集合，记为 $RS=K\cup I\cup T$，且满足 $I\cap T=\varnothing$，然后在论域 $U=S\cup RS$ 下扩展成粗糙集，我们称这一过程为“粗糙化”。粗糙化的具体过程如下：

1)将论域 U 按下面的方式进行划分，

$$p=\{K\cup I,\{t_1,s_1\mathbf{L}\},\{t_2,s_2\mathbf{L}\},\mathbf{L},\{t_n,s_n\mathbf{L}\}\}$$

其中 $t_1,t_2\mathbf{L}\,t_n\in T$，$s_1,s_2\mathbf{L}\,s_m\in S$，且满足所有块中至少有一元素不属于冗余信息集合 T，这个条件是为了保证下一步中 $\underline{R}(RS)=K\cup I$。

2)由定理 1 可以求出与 p 唯一对应的等价关系 $R=R_0\cup R_1\cup\mathbf{L}\,R_n$，其中 R_i 为块 $\{t_i,s_i\mathbf{L}\}$ 对应的二元组的集合，即 $R_i=\{(x,y)\,|\,x,y\in\{t_i,s_i\mathbf{L}\}\}$，$i\neq 0$。$R_0$ 为集合 $K\cup I$ 中元素对应的二元组构成的集合。R_0 可以表示为 $R_0=\{(x,y)\,|\,x,y\in K\cup I\}$。例如：当 K ={a}，I ={b，c}时，R_0 ={(a,a),(a,b),(a,c),(b,a),(b,b),(b,c),(c, a),(c,b),(c,c)}。容易验证 $\underline{R}(RS)=\{x\,|\,(\forall x\in U)\wedge([x]_R\subseteq RS)\}=K\cup I$：

步骤 5：用户 User 将粗糙集 RS 以及等价关系 R 传给服务提供商 SP。

步骤 6：服务提供商收到 RS 与 R 后，对关系 R 进行处理，删除 R 中含有 I 中元素的二元组，得到新的关系 R'。求出 RS 在 R' 下的下近似：

$$\underline{R'}(RS)=\{x\,|\,(\forall x\in U)\wedge([x]_{R'}\subseteq RS)\}=K$$

K 即为用户的认证信息。服务提供商 SP 根据认证策略对用户 User 进行认证。对于前面的例子，此时 $\underline{R'}(RS)=K$ ={a}，即为用户提供的认证信息，服务提供商根据服务的认证策略判断{a}是否满足条件，从而决定用户 User 是否具有对所申请服务的访问权，将认证结果反馈给用户。

4　策略性能分析

1)策略开销方面

从策略可以看出用户所耗时间主要在于“粗糙化”

这一过程，其中划分 p 构造需要保证每一块中至少有一个元素不属于冗余集合 T ，这需要对块中元素和冗余集合 T 进行遍历。等价关系 $R = R_0 \cup R_1 \cup \mathbf{L}\ R_n$ 的构造过程所耗时间主要在与划分 p 中每一块对应的二元组集合的构造。服务提供商所耗时间主要在于求粗糙集 RS 在 R' 关系的下近似，主要是对集合的遍历运算。因此本文提出的基于粗糙集的隐私保护策略没有涉及复杂的数学运算。与加解密算法的算法复杂度相比，该策略更适合于普适计算。

2)可行性方面

该策略在用户方进行“粗糙化”将隐私信息扩展成粗糙集，然后将粗糙集 RS 与 R' 等价关系传给服务提供商。服务提供商收到用户传来的数据后，先对等价关系 R' 进行过滤得到新的等价关系 R' ，最后求出 RS 在新等价关系 R' 下的下近似即为用户的认证请求信息。用户对划分 p 以及等价关系 R 的构造，能够保证服务提供商根据用户提供的数据求出的下近似恰好为用户的认证请求信息。从数据的构造与传输过程中可以看出，该策略是可行的。

3)安全性方面

根据该策略，用户的认证数据中除了用户的隐私信息外，还有其它的冗余信息。即使传输过程中信息被他人窃听，他们获得的是一个粗糙集合。从粗糙集合中挖掘出真实的隐私信息这一过程在统计学上属于独立不重复试验，任何两次的猜测结果都是相互独立的。一次试验从粗糙集合中挖掘出真实数据的可能性非常小。假设用户认证信息集合 K 中元素个数为 m ，过滤集合 I 中元素个数为 n ，而扩展成的粗糙集合 RS 中元素个数为 N 。则分析下面两种情况的安全性:

a.如果用户传输给服务提供商的数据中粗糙集合 RS 被窃听者窃取，则此时窃听者一次试验从 RS 中获得用户隐私信息的概率为 $m!(N-m)!/N!$;

b.如果用户传输给服务提供商的数据中粗糙集合 RS 和等价关系 R 同时被窃听者窃取，窃听者通过求 RS 在关系 R 下的下近似得到含有冗余信息的集合，此时窃听者一次试验从该集合中获得用户隐私信息的概率为 $n!m!/(n+m)!$ 。从上述两种情况下的概率可以看出，随着 N 和 n 的增大，窃听者获得用户隐私信息的概率逐渐减小。当 N 和 n 足够大时，窃听者一次试验获得用户隐私信息可以看成是小概率事件。因此本文提出的策略对用户的隐私信息能够进行有效的保护。

4)适用性方面

在本文提出的基于粗糙集的隐私保护策略中，用户和服务提供商可以通过配置冗余信息集合 T 以及过滤集合 I 来获得不同等级的安全保护，用户能够根据当前上下文以及自身的实际情况灵活地设置安全等级，这是与普适计算的特点相吻合的。事实上 T 和 I 中元素个数越多，窃听者从窃听到的数据中获得用户隐私信息的概率越小，但这必须以增加传输的数据量为代价，因此在策略的实际实施过程中需要权衡安全性和通信开销的问题。

5 结论

在充满计算和通信能力的普适计算环境下随时随地无处不在的计算给人们的生活带来方便的同时，也给用户的隐私保护带来了极大的挑战[6]。本文提出的普适计算环境下基于粗糙集的隐私保护策略，旨在认证过程中防止用户的隐私信息泄露给除服务提供商之外的第三者或被恶意窃听者获取甚至加以利用。这在普适计算中有非常多的应用场景。理论分析表明该策略能够很有效地保护用户隐私。在以后的研究中，我们应该考虑在保证用户隐私不泄露给周围环境中的其他用户的同时，尽量降低用户隐私对服务提供商的透明度。

参考文献

1 徐光祐,史元春,谢伟凯.普适计算.计算机学报, 2003,26(9):1042－1050.

2 吴介,裘正定.普适计算环境中的安全机制.计算机安全, 2005(5):74－84.

3 苗夺谦,李道国.粗糙集理论、算法与应用.北京:清华大学出版社, 2008:10－30.

4 Pawlak Z. Rough sets. International Journal of Computer and Information Sciences, 1982,11:341－356.

5 Pawlak Z, Skowron A. Rough Sets and Conflict Analysis. E-Service Intelligence, Series on Computational Intelligence, Berlin, Heidelberg: Springer-Verlag, 2007:35－74.

6 Weiser M. The computer for the twenty-first century. Scientific American, 1991,265(3):94－104.