跨隐私数据库加密数据等值连接共享协议

景旭[1,2]，李冰冰[2]，何东健[1]
(1. 西北农林科技大学机械与电子工程学院，712100，陕西杨凌；
2. 西北农林科技大学信息工程学院，712100，陕西杨凌)

摘要：针对管理型SaaS(software as a service)中两个租户公平共享隐私数据的问题，提出一种跨隐私数据库加密数据等值连接共享协议.在该协议中：两个租户通过服务提供方(service provider，SP)用可交换加密函数交换某共有属性的全集，实现属性值交集共享；SP用该属性值生成的密钥加密对应元组其他属性值后向对方租户分发；租户用交集生成的密钥解密；组合双方属性交集中等值元素的元组，实现两个租户通过不可信SP公平共享隐私数据.完备性、安全性证明和效率分析结果表明，在半诚实模型下，协议安全可证，满足最少必要信息共享条件，计算代价和通信代价分别是用AGRAWAL协议实现公平共享的57%和75%.

关键词：隐私数据库；加密数据；等值连接；信息共享；SaaS
中图分类号：TP309 **文献标志码**：A **文章编号**：0253-987X(2012)08-0037-06

A Protocol of Encrypted Data Equijoin Sharing Across Private Database

JING Xu[1,2]，LI Bingbing[2]，HE Dongjian[1]
(1. College of Mechanical and Electronic Engineering，Northwest A&F University，Yangling，Shaanxi 712100，China；
2. College of Information Engineering，Northwest A&F University，Yangling，Shaanxi 712100，China)

Abstract：A protocol of encrypted data equijoin sharing across private database is proposed for the problem of fair sharing the private data between two tenants in management-type software as a service. The protocol realizes the sharing of the intersection on a mutual attribute with the help of service provider and using a commutative encryption function to exchange the universal set of the attribute between two tenants. The service provider encrypts other attributes using the key that is generated by the value of attribute，and then sends them to counterpart. Each tenant decrypts them with the key that is generated by intersection. Then the equijoin of private data is shared fairly by assembling both parties' tuples of equivalent elements in the intersection. Analyzing results for completeness，security and efficiency in a semi-honest model show that the protocol is proved safely，and meets the minimal necessary information sharing，and that the computation and the communication costs are 57% and 75% of those in utilizing the fair sharing information by AGRAWAL's protocol.

Keywords：private database；encrypted data；equijoin；sharing information；software as a service

管理型SaaS(software as a service，微软称为line-of-business service[1])中，服务提供方(service provider，SP)提供基于互联网的软件交付和服务，租户(tenant，企业或组织)以"使用而不拥有"[2]的租赁方式消费和利用.为获得规模效益，SP以多租户模式提供服务托管[3].多个租户共享一个应用和数据库实例，但各租户能独立配置应用，好像在独立的环境中运行，以达到软硬件资源共享[4].这些租户

收稿日期：2011-12-19. 作者简介：景旭(1971—)，男，副教授；何东健(通信作者)，男，教授，博士生导师. 基金项目："十二五"国家科技支撑计划资助项目(2011BAD21B05)；中央高校基本科研业务费专项资金资助项目(QN2011036)。
网络出版时间：2012-05-15 网络出版地址：http://www.cnki.net/kcms/detail/61.1069.T.20120515.1824.001.html

一般为业务同行,有共享数据的需求,租用同一 SP 服务为此提供了方便. 租户间数据存储在 SP 数据库中,逻辑上分属于不同的数据库. 租户间通过 SP 的数据共享实质上是跨数据库的数据共享. 等值连接是表间属性相等的广义笛卡尔积,体现了元组间的跨表关联. 共享跨数据库的等值连接具有重要的实用价值. 由于租户不完全信任 SP,一般将隐私数据加密存储,所以无法直接通过跨数据库的等值连接操作实现两个租户间隐私数据库的等值连接共享. 在不泄露其他数据的前提下,通过不完全可信的 SP,提供两个租户公平共享跨隐私数据库加密数据等值连接服务,是 SP 提供高质量服务的必然要求之一.

两个租户通过 SP 的跨隐私数据库等值连接共享可看作是通过第三方的跨隐私数据库等值连接共享. Agrawal 等[5-6]提出一种跨隐私数据库等值连接共享协议(简称 AGRAWAL 协议),是本研究的主要参照. 但它本质上只是一种查询协议,只有通信的一方获得了双方共有信息,不是公平的共享协议,为实现双方公平共享,需执行 AGRAWAL 协议两次;信息经过 Hash 变换后再加密,即使信息拥有者也不能由解密得到原始信息,只能通过密文从备份信息中搜索,而租户不可能租用 SP 服务,同时再备份隐私数据明文. Ma 等[7]提出一种隐私保护的连接查询方案,需要一个可信任的第三方,而管理型 SaaS 中很难找到租户都信任的第三方. Siegenthaler 等[8]提出不同所有者分布式数据库敏感信息共享方案,数据以明文存储,和数据库所有者一致,但租户隐私数据以密文存储,数据库由 SP 所有. Carbunar 等[9]提出了一种能执行二进制 Join 操作的机制,但只适用于单一租户的表间连接,不适用于两个租户间信息共享.

本文针对管理型 SaaS 中两个租户通过 SP 共享跨隐私数据库加密数据等值连接的问题,基于安全多方计算[10-11]的思想,提出一种跨隐私数据库加密数据等值连接共享协议(encrypted data equijoin sharing across private database protocol,EDESAP-DP). 在最少必要信息共享(minimal necessary information sharing)[5]条件下,将解决两个租户通过不完全可信的 SP 实现跨隐私数据库加密数据等值连接共享问题. 本研究对半诚实模型下通过不完全可信的第三方,实现跨隐私数据库数据共享具有重要的理论价值,有利于推动 SaaS 模式走向成熟,也可应用于其他云服务.

1 基本模型

信息共享的发起方租户为 R、接受方租户为 S,服务提供方为 P,P、R 和 S 两两之间都有保密信道如 SSL/IPSec. 基本模型如图 1 所示.

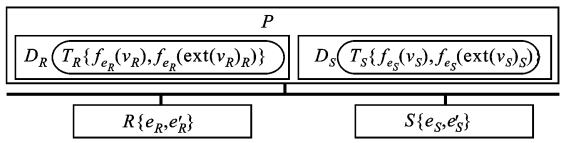

图 1 基本模型图

R 密钥为 e_R、e'_R,S 密钥为 e_S、e'_S,R、S 向 P 安全分发 e'_R、e'_S. 在 e_R、e_S 作用下,R、S 用可交换加密函数 f_e 加密隐私数据,逻辑存储于隐私数据库 D_R、D_S 的表 T_R、T_S 中. 设 T_R、T_S 共有属性 A,记为 $T_R.A$、$T_S.A$;T_R、T_S 的某元组为 b_R、b_S,全集为 B_R、B_S;$T_R.A$、$T_S.A$ 的某明文值为 v_R、v_S,全集为 V_R、V_S;v_R、v_S 的加密值为 y_R、y_S,加密值全集为 $Y_R=\{y_R\mid y_R=f_{e_R}(v_R),v_R\in V_R\}$、$Y_S=\{y_S\mid y_S=f_{e_S}(v_S),v_S\in V_S\}$. 除 A 外,b_R、b_S 的其他属性明文为 $\mathrm{ext}(v_R)_R$、$\mathrm{ext}(v_S)_S$. V_{ext}是除属性 A 外元组其他属性值的全集;C_{ext}是 V_{ext}加密值全集,$\boldsymbol{D}_{\mathrm{ext}}$是 C_{ext}上的一个分布;集合 M 元素个数记为 $|M|$.

V_R 和 V_S 的交集为 $V=V_R\cap V_S=\{v\mid v\in V_R\wedge v\in V_S\}$,$T_R$、$T_S$ 在 A 上的等值连接为 $J=R\infty S=\{(v,\mathrm{ext}(v)_R,\mathrm{ext}(v)_S)\mid v\in V,\ b_R.A=v\wedge b_S.A=v\}$.

R、S 希望共享 D_R、D_S 中 T_R、T_S 属性 A 的等值连接. R、S 的隐私数据加密存储在 P 数据库,不可能达到只共享 J 的理想状态. 按照必要知道原则(need-to-know basis),本文研究最少必要信息共享的条件是 R 只获得 $|V_S|$、$V_S\cap V_R$ 和 $\{\mathrm{ext}(v)_S\mid v\in V_S\cap V_R\}$,$S$ 只获得 $|V_R|$、$V_S\cap V_R$ 和 $\{\mathrm{ext}(v)_R\mid v\in V_S\cap V_R\}$,$P$ 仅获得 $|V_R|$、$|V_S|$,无其他信息泄露.

2 相关定义

定义 1[5] 计算不可区分(indistinguishability):设 k bit 有限域 $\Omega_k\subseteq\{0,1\}^k$,分布 $\boldsymbol{D}_1=\boldsymbol{D}_1(\Omega_k)$和 $\boldsymbol{D}_2=\boldsymbol{D}_2(\Omega_k)$,$x\in\Omega_k$ 上多项式时间算法(polynomial-time algorithm, PPT)$A_k(x)$返回值是 True 或 False,如果对于任意 $A_k(x)$、多项式 $p(k)$和所有足够大的数 k,定义在随机变量 $x\in\Omega_k$ 上,$\boldsymbol{D}_1$ 和 $\boldsymbol{D}_2$ 分布总满足式(1),则称 $\boldsymbol{D}_1$ 和 $\boldsymbol{D}_2$ 分布是计算不可区分的(简称

不可区分).

$$P_r[A_k(x) \mid x \sim \boldsymbol{D}_1] - P_r[A_k(x) \mid x \sim \boldsymbol{D}_2] < 1/p(k) \quad (1)$$

式中:$x \sim D$ 表示 x 服从 D 分布;$P_r[A_k(x)]$表示 $A_k(x)$返回 True 的概率.

不存在明显可区分概率空间 $\boldsymbol{D}_1$ 和 $\boldsymbol{D}_2$ 的 $A_k(x)$,即任何 PPT 成功区分 $\boldsymbol{D}_1$ 和 $\boldsymbol{D}_2$ 的概率总是可忽略的.

定义 2[5]　可交换加密函数:在有限域上,多项式时间内可计算且满足性质①~④的函数 f:Key $F \times$ Dom $F \rightarrow$Dom F. 令 $f_e(x) \equiv f(e,x), \in_r$ 表示在有限域中均匀地随机取值. 可交换加密函数性质:①可交换性,即对于所有的 e、e',有 $f_e \cdot f_{e'} = f_{e'} \cdot f_e$ 成立;②双射性,即每一个 f_e:Dom $F \rightarrow$Dom F 是双射的;③可计算性,即给定 e,反函数 f^{-1}在多项式时间内是可计算的;④计算不可区分性,即对于 $e \in_r$Key F 和 $x,y,z \in_r$Dom F,$\langle x, f_e(x), y, f_e(y)\rangle$和$\langle x, f_e(x), y, z\rangle$是计算不可区分的. 也就是说当 e 未知时,对于给定的随机值 x 和密文 $f_e(x)$及另一个随机值 y,多项式时间内不可能区分 $f_e(y)$和随机值 z.

定义 3[12]　半诚实(semi-honest/passive)模型:参与方严格执行协议规程,不会中途强行退出或恶意掺入虚假数据,但在协议执行过程中,参与方可能会保留所有能搜集到的信息,以期望在协议结束后推断出另一参与方的输入信息.

本文假设协议参与方都是"半诚实的".

定义 4[5]　加密函数 K:$K(k,\text{ext}(v))$表示在密钥 k 作用下,加密 ext(v)的函数. K:Dom $F \times V_{\text{ext}} \rightarrow C_{\text{ext}}$ 的性质为:①可计算性,即给定 k 时,能有效解密 $K_k(x) \equiv K(k,x)$;②完善保密性,即当 $k \in_r$Key F 时,对于任何ext(v),K_k(ext(v))的值和 C_{ext}上固定的分布 $\boldsymbol{D}_{\text{ext}}$是不可区分的.

3　EDESAPDP 协议

3.1　协议描述

本文提出的 EDESAPDP 中,两个租户通过 SP 利用可交换加密函数交换某共有属性的全集,实现属性值交集的共享;SP 用该属性值生成的密钥加密对应元组其他属性值后向对方租户分发;租户利用交集生成的密钥解密,得到对方与交集对应的元组密文;通过组合自己和对方属性值交集中等值元素对应的元组,实现等值连接共享. 基于 UML 时序图的协议流程如图 2 所示,协议具体过程如下.

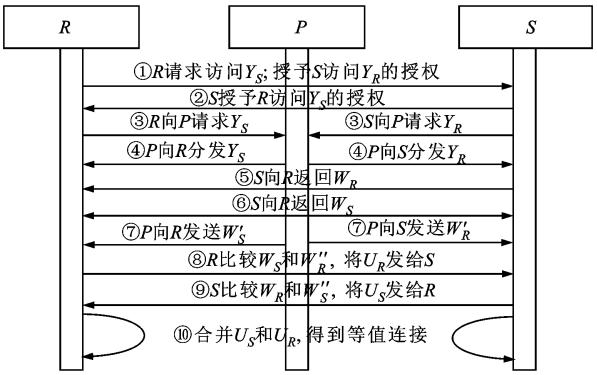

图 2　EDESAPDP 协议流程图

步骤 1　R 希望与 S 共享 T_R、T_S 在 A 上的等值连接,R 向 S 发起通过 P 访问 Y_S 的请求,同时,授予 S 通过 P 访问 Y_R 的授权.

步骤 2　如果 S 接受 R 的请求,授予 R 访问 Y_S 的授权.

步骤 3　R、S 分别向 P 出示 S、R 的授权,请求获得 Y_S、Y_R.

步骤 4　P 验证 R、S 授权后分发 Y_S、Y_R.

步骤 5　对 $y_R \in Y_R$,S 计算 $f_{e_S}(y_R)$、$f_{e'_S}(f_{e_S}(y_R))$,得到 3 元组 $w_R = \langle y_R, f_{e_S}(y_R), f_{e'_S}(f_{e_S}(y_R))\rangle = \langle f_{e_R}(v_R), f_{e_S}(f_{e_R}(v_R)), f_{e'_S}(f_{e_S}(f_{e_R}(v_R)))\rangle$,将 w_R 的全集 W_R 发送给 R.

步骤 6　对 $y_S \in Y_S$,R 计算 $f_{e_R}(y_S)$、$f_{e'_R}(f_{e_R}(y_S))$,得到 3 元组 $w_S = \langle y_S, f_{e_R}(y_S), f_{e'_R}(f_{e_R}(y_S))\rangle = \langle f_{\S}(v_S), f_{e_R}(f_{e_S}(v_S)), f_{e'_R}(f_{e_R}(f_{e_S}(v_S)))\rangle$,将 w_S 的全集 W_S 发送给 S.

步骤 7　$y_S \in Y_S$,P 用 e'_S生成 $k(v_S) = f_{e'_S}(f_{e_S}(v_S))$,作为定义 4 中加密 $f_{e_S}(\text{ext}(v_S)_S)$的密钥;用定义 4 的 K 生成 2 元组 $w'_S = \langle f_{e_S}(v_S), c(v_S)\rangle = \langle f_{e_S}(v_S), K(k(v_S), f_{e_S}(\text{ext}(v_S)_S))\rangle$,向 R 发送 w'_S全集 W'_S.

同理,对于 $y_R \in Y_R$,P 生成 $w'_R = \langle f_{e_R}(v_R), c(v_R)\rangle = \langle f_{e_R}(v_R), K(k(v_R), f_{e_R}(\text{ext}(v_R)_R))\rangle$,向 S 发送 w'_R全集 W'_R.

步骤 8　收到 W_R 和 W'_S后,首先,R 用 e_R 解密 $w_R \in W_R$,得 $w''_R = f^{-1}_{e_R}(w_R) = \langle v_R, f_{e_S}(v_R), f_{e'_S}(f_{e_S}(v_R))\rangle$;其次,在 W'_S和 W''_R中,如果 2 元组w'_S第 1 分量和 3 元组 w''_R第 2 分量相等,则 w''_R的第 1 分量 v 属于 $V_R \cap V_S$;再次,用与 v 对应 3 元组 w''_R的第 3 分量作为密钥,解密对应 2 元组 w'_S的第 2 分量,得到 b_S 其他属性的密文 $f_{e_S}(\text{ext}(v)_S)$;然后,通过 v 检索 T_R 得到 b_R 的 ext(v)$_R$;最后,将 $U_R = \{\langle v, f_{e_S}(v), \text{ext}(v)_R\rangle \mid v \in V_R \cap V_S\}$发给 S.

步骤 9 S 接收到 W_S 和 W'_R 后，执行类似步骤 8 的操作，将 $U_S=\{\langle v, f_{e_R}(v), \mathrm{ext}(v)_S\rangle | v\in V_R\cap V_S\}$ 发给 R.

步骤 10 R、S 分别利用 U_S、U_R 中 $\langle v, f_{e_R}(v)\rangle$、$\langle v, f_{e_S}(v)\rangle$ 验证 U_S、U_R 的完整性，通过 v 检索 T_R、T_S 得 b_R、b_S 的 $\mathrm{ext}(v)_R$、$\mathrm{ext}(v)_S$，将 $\mathrm{ext}(v)_R$、$\mathrm{ext}(v)_S$ 组合，得等值连接 $J=R\infty S=\{(v, \mathrm{ext}(v)_R, \mathrm{ext}(v)_S) | v\in V, b_R.A=v\wedge b_S.A=v\}$.

3.2 完备性证明

定理 1 在 EDESAPDP 中，R 准确获得 $|V_S|$、$V_S\cap V_R$ 和 $\{\mathrm{ext}(v)_S | v\in V_S\cap V_R\}$，$S$ 准确获得 $|V_R|$、$V_S\cap V_R$ 和 $\{\mathrm{ext}(v)_R | v\in V_S\cap V_R\}$，$P$ 准确获得 $|V_R|$、$|V_S|$.

证明 由定义 2 知 f_e 具有可交换性和双射性，在协议步骤 5、6 中，R 获得 w_R、w_S，比较它们的第 2 个分量有 $v\in V_R\cap V_S$ iff $f_{e_S}(f_{e_R}(v_R))=f_{e_R}(f_{e_S}(v_S))\wedge v_R\in V_R\wedge v_S\in V_S$，即 R 准确获得 $V_S\cap V_R$. 同理，S 准确获得 $V_S\cap V_R$.

在协议步骤 7 中，R 收到所有的 $\mathrm{ext}(v_S)_S$，但是在密钥 $k(v_S)=f_{e'_S}(f_{e_S}(v_S))$ 作用下用定义 4 的加密函数 K 加密. 在协议步骤 8 中，通过比较 2 元组 w'_S 第 1 分量和 3 元组 w''_R 第 2 分量得到 $V_S\cap V_R$；用与 v 对应 3 元组 w''_R 的第 3 分量作为密钥，解密对应 2 元组 w'_S 的第 2 分量，得到 b_S 其他属性的密文 $f_{e_S}(\mathrm{ext}(v)_S)$，但无法解密 $\{f_{e_S}(\mathrm{ext}(v_S)_S | v_S\notin V_S\cap V_R\wedge v_S\in V_S)\}$. 唯有 R 知道 e_R，所以唯有 R 能生成 $f_{e_R}(v)$. 在协议步骤 10 中，唯有 R 能利用 U_S 中的 $\langle v, f_{e_R}(v)\rangle$ 验证 U_S 的完整性，准确获得 $\{\mathrm{ext}(v)_S | v\in V_S\cap V_R\}$. 同理，$S$ 准确获得 $\{\mathrm{ext}(v)_R | v\in V_S\cap V_R\}$.

P 只能接触到 Y_R、Y_S，因为 $|V_R|=|Y_R|$，$|V_S|=|Y_S|$，所以 R、S 和 P 准确获得 $|V_S|$、$|V_R|$.

3.3 安全性证明

定理 2[5] 当 $m\in Z, i: x_i\in{}_r\mathrm{Dom}\ F, z_m\in{}_r\mathrm{Dom}\ F, e\in{}_r\mathrm{Dom}\ F$ 时，$2\times m$ 元组 $\boldsymbol{D}_m$ 和 $\boldsymbol{D}_{m-1}$ 分布是计算不可区分的.

$$\boldsymbol{D}_m=\begin{pmatrix} x_1 & \cdots & x_{m-1} & x_m \\ f_e(x_1) & \cdots & f_e(x_{m-1}) & f_e(x_m)\end{pmatrix};$$

$$\boldsymbol{D}_{m-1}=\begin{pmatrix} x_1 & \cdots & x_{m-1} & x_m \\ f_e(x_1) & \cdots & f_e(x_{m-1}) & z_m\end{pmatrix} \tag{2}$$

定理 3[5] 当 $m,n\in Z, 0\leqslant m\leqslant n, i: x_i, z_i\in{}_r\mathrm{Dom}\ F, e\in{}_r\mathrm{Dom}\ F$ 时，$2\times n$ 元组 $\boldsymbol{D}_n^n$ 和 $\boldsymbol{D}_m^n$ 分布是计算不可区分的.

$$\boldsymbol{D}_n^n=\begin{pmatrix} x_1 & \cdots & x_m & x_{m+1} & \cdots & x_n \\ f_e(x_1) & \cdots & f_e(x_m) & f_e(x_{m+1}) & \cdots & f_e(x_n)\end{pmatrix};$$

$$\boldsymbol{D}_m^n=\begin{pmatrix} x_1 & \cdots & x_m & x_{m+1} & \cdots & x_n \\ f_e(x_1) & \cdots & f_e(x_m) & z_{m+1} & \cdots & z_n\end{pmatrix} \tag{3}$$

定理 4[13] 假设 R 和 S 是计算不可区分的分布，S 和 T 是计算不可区分的分布，那么 R 和 T 是计算不可区分的分布.

定理 5 当 n 为任意自然数，$i: x_i, y_i, z_i\in{}_r\mathrm{Dom}\ F$ 和 $e, e'\in{}_r\mathrm{Dom}\ F$ 上时，$3\times n$ 元组 $\boldsymbol{D}_1$ 和 $\boldsymbol{D}_2$ 分布是计算不可区分的.

$$\boldsymbol{D}_1=\begin{pmatrix} x_1 & \cdots & x_n \\ f_e(x_1) & \cdots & f_e(x_n) \\ f_{e'}(f_e(x_1)) & \cdots & f_{e'}(f_e(x_n))\end{pmatrix};$$

$$\boldsymbol{D}_2=\begin{pmatrix} x_1 & \cdots & x_n \\ y_1 & \cdots & y_n \\ z_1 & \cdots & z_n\end{pmatrix} \tag{4}$$

证明 设 $3\times n$ 元组 $\boldsymbol{D}_3$，如式(5)所示，则有

$$\boldsymbol{D}_3=\begin{pmatrix} x_1 & \cdots & x_n \\ f_e(x_1) & \cdots & f_e(x_n) \\ z_1 & \cdots & z_n\end{pmatrix} \tag{5}$$

由于 $\boldsymbol{D}_1$ 和 $\boldsymbol{D}_3$ 的第 1、3 行分别组成 $2\times n$ 元组 $\boldsymbol{D}_n^n$ 和 $\boldsymbol{D}_0^n$，第 2 行相同，由定理 3 可知，$\boldsymbol{D}_n^n$ 和 $\boldsymbol{D}_0^n$ 分布是不可区分的，所以 $\boldsymbol{D}_1$ 和 $\boldsymbol{D}_3$ 分布是不可区分的.

$\boldsymbol{D}_2$ 和 $\boldsymbol{D}_3$ 的第 1、2 行分别组成 $2\times n$ 元组 $\boldsymbol{D}_n^n$ 和 $\boldsymbol{D}_0^n$，第 3 行相同，同理，$\boldsymbol{D}_2$ 和 $\boldsymbol{D}_3$ 分布是不可区分的.

由定理 4 知，$\boldsymbol{D}_1$ 和 $\boldsymbol{D}_2$ 分布是不可区分的.

定理 6 m、n 和 t 为任意自然数，$c_i\in{}_rV_{\mathrm{ext}}$，形如 $\boldsymbol{D}'$ 的 $4\times n$ 元组 $\boldsymbol{D}'_1(\forall i: x_i\in{}_r\mathrm{Dom}\ F, y_i=f_e(x_i), z_i=f_{e'}(f_e(x_i)), \zeta_i=K(z_i, c_i), e, e'\in{}_r\mathrm{KeyF})$ 表示 R 真实数据分布，形如 $\boldsymbol{D}'$ 的 $4\times n$ 元组 $D'_2(\forall i: x_i, y_i, z_i\in{}_r\mathrm{Dom}\ F$，当 $i=1,\cdots,t$ 时，ζ_i 表示分布 $\boldsymbol{D}_{\mathrm{ext}}$ 上的独立随机数；当 $i=t+1,\cdots,m$ 时，$\zeta_i=K(z_i, c_i))$ 表示攻击者相应的模拟数据. 在 $\boldsymbol{D}'_1$ 和 $\boldsymbol{D}'_2$ 中，前 t 列表示 $V_S-(V_S\cap V_R)$，接着的 $m-t$ 列表示 $V_S\cap V_R$，最后的 $n-m$ 列表示 $V_R-(V_S\cap V_R)$，$\boldsymbol{D}'_1$ 和 $\boldsymbol{D}'_2$ 分布是计算不可区分的.

$$\boldsymbol{D}'=\begin{pmatrix} x_1 & \cdots & x_t & x_{t+1} & \cdots & x_m & x_{m+1} & \cdots & x_n \\ y_1 & \cdots & y_t & y_{t+1} & \cdots & y_m & y_{m+1} & \cdots & y_n \\ & & & z_{t+1} & \cdots & z_m & z_{m+1} & \cdots & z_n \\ \xi_1 & \cdots & \xi_t & \xi_{t+1} & \cdots & \xi_m & & & \end{pmatrix} \tag{6}$$

证明　令形如 $\boldsymbol{D}'$ 的 $\boldsymbol{D}'_3(\forall i: x_i, y_i, z_i \in_r \mathrm{Dom}\ F, \zeta_i = K(z_i, c_i))$,分析 $\boldsymbol{D}'_2$ 和 $\boldsymbol{D}'_3$ 可以看出,当 $i=1,\cdots,t$ 时,$\boldsymbol{D}'_2$ 中的 ζ_i 表示定义 4 中 $\boldsymbol{D}_{\mathrm{ext}}$ 分布上的独立随机数,$\boldsymbol{D}'_3$ 中的 $\zeta_i = K(z_i, c_i)$,其他部分相同. 当 $i=1,\cdots,t$ 时,z_i 不在 $\boldsymbol{D}'_2$ 和 $\boldsymbol{D}'_3$ 分布上,不影响 $\boldsymbol{D}'_2$ 和 $\boldsymbol{D}'_3$ 分布的独立性. 根据定义 4 的性质②可知,$\boldsymbol{D}'_2$ 和 $\boldsymbol{D}'_3$ 分布是不可区分的.

令多项式时间可计算函数 $Q(\boldsymbol{M})$ 表示由定理 5 中 $3\times n$ 元组 $\boldsymbol{M}$ 生成 $4\times n$ 元组 $\boldsymbol{M}'$ 的函数,其构造方法为:$\boldsymbol{M}'$ 和 $\boldsymbol{M}$ 前 3 行一致,但 $\boldsymbol{M}'$ 左边 $z_1,\cdots,z_t$ 空白;以 $\boldsymbol{M}$ 第 3 行对应的值作为参数 z_i 的值,由 $\zeta_i = K(z_i, c_i)$ 生成 $\boldsymbol{M}'$ 的第 4 行.

在 $Q(\boldsymbol{M})$ 作用下,如果 $\boldsymbol{M}$ 类似于定理 5 中 $\boldsymbol{D}_1$,那么 $\boldsymbol{M}'$ 就相当于定理 6 中 $\boldsymbol{D}'_1$;如果 $\boldsymbol{M}$ 类似于定理 5 中 $\boldsymbol{D}_2$,那么 $\boldsymbol{M}'$ 就相当于定理 6 中 $\boldsymbol{D}'_3$. 由定理 5 可知,$\boldsymbol{D}'_1$ 和 $\boldsymbol{D}'_3$ 分布是不可区分的.

由定理 4 知,$\boldsymbol{D}'_1$ 和 $\boldsymbol{D}'_2$ 分布是不可区分的.

定理 7　在半诚实模型下,EDESAPDP 中,R 仅获得 $|V_S|$、$V_S \cap V_R$ 和 $\{\mathrm{ext}(v)_S \mid v \in V_S \cap V_R\}$,$S$ 仅获得 $|V_R|$、$V_S \cap V_R$ 和 $\{\mathrm{ext}(v)_R \mid v \in V_S \cap V_R\}$,$P$ 仅获得 $|V_S|$、$|V_R|$.

证明　半诚实模型下,设模拟 R 的攻击协议. 令 $V_S = \{v_1,\cdots,v_t,v_{t+1},\cdots,v_m\}$,$V_R = \{v_{t+1},\cdots,v_m,v_{m+1},\cdots v_n\}$,$t = |V_S \cap V_R|$,$m = |V_S|$ 和 $n = |V_S \cup V_R|$,则攻击者拥有 V_R、$V_S \cap V_R$、$|V_S|$、e_R、e'_R、$W = \{\langle v, \mathrm{ext}(v)_R, \mathrm{ext}(v)_S\rangle \mid v \in V_R \cap V_S\}$,希望获得 $V_S - V_R$ 和 $\{\mathrm{ext}(v)_S \mid v \notin V_S - V_R, v \in V_S\}$. 模拟攻击过程如下.

(1) 在协议步骤 5 中,攻击者生成 n 个 $y_i \in_r \mathrm{Dom}\ F (i=1,\cdots,n)$ 模拟 $f_{e_S}(y_R)$,n 个 $z_i \in_r \mathrm{Dom}\ F (i=1,\cdots,n)$ 模拟 $f_{e'_S}(y_R)$,用 e_R、e'_R 生成 3 元组 $\langle f_{e_R}(v_i), f_{e_R}(y_i), f_{e'_R}(f_{e_R}(z_i))\rangle (i=1,\cdots,n)$.

(2) 在协议步骤 7 中,模拟攻击过程如下:首先,对于 $\{v_i \mid v_i \in V_R \cap V_S, t+1 \leqslant i \leqslant m\}$,利用 $\zeta_i = K(z_i, \mathrm{ext}(v_i)_R)(t+1 \leqslant i \leqslant m)$ 得 2 元组 $\langle y_i, \zeta_i\rangle (t+1 \leqslant i \leqslant m)$;其次,产生 t 对 $\langle y_i, \zeta_i\rangle (i=1,\cdots,t)$,其中,$\zeta_i$ 是定义 4 中 C_{ext} 在 $\boldsymbol{D}_{\mathrm{ext}}$ 上的分布,y_i、ζ_i 是各自域上的随机值.

令 $x_i = f_{e_R}(v_i)$,则真实数据和定理 6 中的 $\boldsymbol{D}'_1$ 分布一致,攻击者模拟数据和定理 6 中 $\boldsymbol{D}'_2$ 的分布一致. $\boldsymbol{D}'_1$ 和真实数据、$\boldsymbol{D}'_2$ 和模拟数据的唯一区别是真实数据、模拟数据是用 f_{e_R} 加密的,故用 $\boldsymbol{D}'_1$ 和 $\boldsymbol{D}'_2$ 的区分关系代替真实数据和模拟数据的区分关系是有效的.

由定理 6 知 $\boldsymbol{D}'_1$ 和 $\boldsymbol{D}'_2$ 分布是不可区分的,则由定理 4 有,真实数据和模拟数据是不可区分的. 所以 R 仅获得 $|V_S|$、$V_S \cap V_R$ 和 $\{\mathrm{ext}(v)_S \mid v \in V_S \cap V_R\}$.

同理,模拟 S 的攻击者攻击过程相似,S 仅获得 $|V_R|$、$V_S \cap V_R$ 和 $\{\mathrm{ext}(v)_R \mid v \in V_S \cap V_R\}$.

R、S 的隐私数据以密文托管于 P,根据定义 4 性质②可知,P 无法解密,仅获得 $|V_S|$、$|V_R|$.

R、S 和 P 彼此间都有安全的信道,隐私数据在传输中安全.

3.4　效率分析

假设 Dom F 上加密字长为 k bit,R、S 执行 f_e 的计算代价为 C_e,P 执行 f_e 的计算代价为 C'_e,P 执行 K 的计算代价为 C_K,n 个密文集中检索代价为 $n\log n \cdot C_S$,$|\mathrm{ext}(v)_S|$ 表示 T_S 中属性个数 -1,$|\mathrm{ext}(v)_R|$ 表示 T_R 中属性个数 -1.

Agrawal 等[5]估算执行代价和通信代价的条件如下:计算 C_e(如 $x^y \bmod p$)的代价是 Pentium III 处理 1 024 bit 的时间为 0.02 s;通信带宽为 1.544 Mbit/s($\approx$5 Gbit/h);并行处理器个数为 10.

分析理想条件下的代价,例如在协议步骤 4 中,R、S 已将 Y_R、Y_S 存储在 P,不计算加密代价;在协议步骤 8~10 中,R、S 分别存储 Y_R、Y_S,不再计算传输代价.

(1) 计算代价:$4C_e(|V_R|+|V_S|)+2C_e|V_R \cap V_S|+C_K|V_R \cap V_S| \cdot |\mathrm{ext}(v)_S|+C_K|V_R \cap V_S| \cdot |\mathrm{ext}(v)_R|+2(|V_R|\log|V_R|+|V_S|\log|V_S|)C_S+2|V_R \cap V_S|\log|V_R \cap V_S|C_S+C'_e(|V_R|+|V_S|)+C'_K(|V_R|+|V_S|)$.

由于 P 采用大型的服务器,租户采用 PC 机,甚至资源受限型设备如 iPad 等,所以 $C_e \gg C'_e$. 假设 $nC_e \gg n\log n\, C_S$,$C_e \gg C_K$,$C_e \gg C'_K$,计算代价近似于 $4C_e(|V_R|+|V_S|)+2C_e|V_R \cap V_S|$.

在与文献[5]相同的分析条件下,AGRAWAL 协议的计算代价为 $7\times10^4 C_e$,通过其实现双方公平共享的计算代价为 $14\times10^4 C_e$;本协议的计算代价为 $8\times10^4 C_e$,所以本协议的计算代价仅是通过 AGRAWAL 协议实现公平共享的 57%.

(2)通信代价:$3(|V_R|+|V_S|)k+|V_R \cap V_S| \cdot |\mathrm{ext}(v)_R|k+|V_R \cap V_S| \times |\mathrm{ext}(v)_S|k$.

在与文献[5]相同的分析条件下,AGRAWAL 协议的通信代价为 4×10^7 bit,通过其实现双方公平共享的通信代价为 8×10^7 bit;本协议的通信代价为 6×10^7 bit,所以本协议的通信代价仅是通过 AGRAWAL 协议实现公平共享的 75%.

4 结 论

针对管理型 SaaS 中两个租户通过 SP 共享跨隐私数据库等值连接的问题，本文提出一种跨隐私数据库加密数据等值连接共享协议. 在两个租户和 SP 都不完全可信的条件下，本协议可以实现租户间公平获得隐私数据的等值连接. 从完备性证明和安全分析结果可以看出，在半诚实模型下，本协议安全可证，满足公平共享和最少必要信息共享条件. 在相同分析条件下，本协议的计算代价和通信代价仅是通过 AGRAWAL 协议实现公平共享的 57% 和 75%. 本研究对于在半诚实模型下通过不完全可信的第三方实现隐私共享具有一定的理论价值，有利于推动 SaaS 模式走向成熟，可推广应用于其他云服务.

参考文献：

[1] CHONG F, CARRARO G. Architecture strategies for catching the long tail [R/OL]// (2006-04-01) [2011-06-10]. http://msdn.microsoft.com/en-us/library/aa479069.aspx.

[2] PAPAZOGLOU M P, TRAVERSO P, DUSTDAR S, et al. Service-oriented computing: state of the art and research challenges [J]. IEEE Computer, 2007, 40(11): 64-71.

[3] KWOK T, NGUYEN T, LAM L. A software as a service with multi-tenancy support for an electronic contract management application [C]// Proceedings of the 2008 IEEE International Conference on Services Computing. Washington DC, USA: IEEE Computer Society, 2008: 179-186.

[4] BEZEMER C, ZAIDAM A. Challenges of reengineering into multi-tenant SaaS applications, TUD-SERG-2010-012 [R]. Delft, Netherlands: Delft University of Technology. Software Engineering Research Group, Department of Software Technology, Faculty of Electrical Engineering, Mathematics and Computer Science, 2010.

[5] AGRAWAL R, EVFIMIEVSKI A, SRIKANT R. Information integration across autonomous enterprises: US, 2008/0065910 A1[P]. 2008-05-13.

[6] AGRAWAL R, EVFIMIEVSKI A, SRIKANT R. Information sharing across private databases[C]// Proceedings of the 2003 ACM SIGMOD International Conference on Management of Data. New York, NY, USA: ACM, 2003: 86-97.

[7] MA Sha, YANG Bo, LI Kangshun, et al. A privacy-preserving join on outsourced database [C]// Proceedings of the 14th International Conference on Information Security. Berlin, Germany: Springer-Verlag, 2011: 278-292.

[8] SIEGENTHALER M, BIRMAN K. Sharing private information across distributed databases[C]// Proceedings of the 8th IEEE International Symposium on Network Computing and Applications. Piscataway, NJ, USA: IEEE, 2009: 82-89.

[9] CARBUNAR B, SION R. Joining privately on outsourced data[C]// Proceedings of the 7th VLDB Conference on Secure Data Management. Berlin, Germany: Springer-Verlag, 2010: 70-86.

[10] LINDELL Y, PINKAS B. Secure multiparty computation for privacy-preserving data mining [J]. Journal of Privacy and Confidentiality, 2009, 1(1):59-98.

[11] 程柏良，曾国荪，揭安全. 基于安全多方计算的可信防共谋协议模型[J]. 通信学报，2011，32(8):23-30.
CHENG Bailiang, ZENG Guosun, JIE Anquan. Trusted coalition-proof protocol model based on secure multi-part computing [J]. Journal on Communications, 2011, 32(8): 23-30.

[12] 刘文，罗守山，王永滨. 安全两方向量优势统计协议及其应用[J]. 电子学报，2010，38(11)：2573-2577.
LIU Wen, LUO Shoushan, WANG Yongbin. Secure two-party vector dominance statistic protocol and its applications [J]. Acta Electronica Sinica, 2010, 38(11): 2573-2577.

[13] BEAVER D, MICALI S, ROGAWAY P. The round complexity of secure protocols [C]// Proceedings of the Twenty-Second Annual ACM Symposium on Theory of Computing. New York, NY, USA: ACM, 1991:503-513.

（编辑 武红江）