文章编号:1009-3087(2014)03-0095-07

跨加密关系数据库等值连接大小共享协议

景 旭,李书琴*,谭戈旭
(西北农林科技大学 信息工程学院,陕西 杨凌 712100)

摘 要:针对管理型SaaS中2个租户需要公平共享等值连接大小的问题,提出了等值连接大小的组合公式和跨加密关系数据库等值连接大小共享协议。在该协议中,由一个共享属性及对应元组其它属性的哈希值构成2元组,SP将其全集发给对方租户;对2元组中的共享属性值再次加密后,租户发送给对方;租户分别计算共享属性值的密文交集以及元素对应各租户的元组数;利用等值连接大小的组合公式,2个租户公平共享了等值连接大小。完整性和安全性证明表明,协议在半诚实模型下安全可证,满足最少必要信息共享条件。效率分析表明,协议计算代价和通信代价仅是通过Agrawal协议实现公平共享的50%和约68%。原型测试可看出,协议基本能满足管理型SaaS的需求。

关键词:关系数据库;加密数据;等值连接大小;共享协议;SaaS
中图分类号:TP309.2 **文献标志码**:A

A Protocol of Equijoin Size Sharing Across Encrypted Relational Database

JING Xu, LI Shuqin, TAN Gexu*
(College of Info. Eng., Northwest A&F Univ., Yangling 712100, China)

Abstract: In order to solve the problem that two tenants required to fair share equijoin size in management-type SaaS (software as a service), both a combinational formula of equijoin size and a share protocol of encrypted data equijoin size across private database were proposed. The universal set of 2-tuple was sent to the other tenant by service provider, which was constituted by a sharing attribute value and a hash values of others attributes to corresponding tuple. After the sharing attribute value of 2-tuple was re-encrypted, tenant sent them to the other. Both the ciphertext intersection of sharing attribute and the number of each element corresponds to tuples were calculated by tenant independently. By the combinational formula of equijoin size, the equijion size was fairly shared between tenants. The integrity and security showed that the share protocol of equijoin size was perfect and safe in the semi-honest model. By the efficiency analysis, the computation and communication costs were 50% and about 68% of those in utilizing the fair sharing information by Agrawal's protocol. It met the basic needs of management-type SaaS by prototype testing.

Key words: relational database; encrypted data; equijoin size; sharing protocol; software as a service

管理型SaaS[1](software as a service)主要面向租户(tenant,企业或组织)的业务需求,采用多租户(multi-tenant)模式[2]处理结构化数据。管理结构化数据的有效方式是关系数据库。主流关系云数据库系统有Amazon RDS (Amazon relational database service)、Amazon EC2和SQL Azure[3],有利于管理型SaaS的水平扩展,所以作为主要研究对象。

服务提供方(service provider,SP)的租户多为同行,有共享业务数据如等值连接大小等的需求。例如,基于基因技术研究肉牛性状改良的租户R和S,在不同条件下,独立得到性状A、B、C和D、E、F的多组参数。当可比较性状A和D相等时,在不泄漏其它信息条件下,希望借鉴对方的试验参数,通过B、C和E、F的组合数校验自己的研究成果。等值连接(equijoin)[4]是表(关系)间等值属性的广义笛卡尔积。等值连接大小(equijoin size)[4]是等值连

收稿日期:2013-07-24
基金项目:国家科技支撑计划资助项目(2013BAD15B02);中央高校基本科研业务费资助项目(QN2011036);科技支疆计划项目(2013AB016)
作者简介:景 旭(1971—),男,副教授,博士. 研究方向:信息系统安全. E-mail:jingxu@nwsuaf.edu.cn.
*通信联系人 E-mail:lsq_cie@nwsuaf.edu.cn

接的元组个数。租户 R 和 S 通过计算性状 A 和 D 上的等值连接大小,可共享 B、C 和 E、F 的组合数。由于租户不完全信任 SP,存储在 SP 的隐私数据是加密的,所以 R 和 S 无法直接计算等值连接大小。在多租户模式下,租户数据存储于 SP 数据库,共享一个 license,但逻辑上却分属不同的数据库,所以租户间通过 SP 共享数据实质上是跨数据库的。通过不完全可信的 SP,提供 2 个租户公平共享跨加密数据库等值连接大小,是发挥管理型 SaaS 信息集中优势,向租户提供高质量服务的必然要求之一。

Agrawal 等[4]提出了一种跨隐私数据库等值连接大小共享协议(简称 Agrawal 协议),没有给出详细流程,而且要求掌握对方属性值的分布,当属性值均不相同时,可能泄露交集,实质仅实现了查询功能。景旭等[1]提出一种跨隐私数据库加密数据等值连接共享协议,如果直接用于共享等值连接大小,会泄露交集和等值连接。Liang 等[5]提出了一种等值连接的分布式计算协议,但它是非对称的,一方利用盲签名保护隐私,另一方利用哈希函数,双方计算负荷也是非对称的。Yuan 等[6]基于可交换加密函数(combination commutative encryption)和茫然传输协议(oblivious transfer protocol)提出了隐私数据库安全查询协议,但不能实现等值连接大小共享。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 中很难找到一个所有租户和 SP 都信任的第三方。李凌[8]研究了云计算服务中数据的若干问题,但主要从数据信息应用的角度研究了基于身份信息的加密机制。

针对 2 个租户需通过 SP 共享等值连接大小的问题,根据等值连接[9]的概念,基于数学归纳法,作者研究等值连接大小的组合计算公式;基于安全多方计算[10-12]的思想,作者提出了一种跨加密关系数据库等值连接大小共享协议(equijoin size sharing across encrypted relational database protocol, ESSAERDP)。在最少必要信息共享条件下,ESSAERDP 协议解决了 2 个租户通过不完全可信的 SP,实现跨加密关系数据库隐私数据等值连接大小共享。本研究对通过不完可信的第三方共享跨加密关系数据库的等值连接大小具有重要的理论价值,有助于推动 SaaS 走向成熟,可应用于其它云服务。

1 基本模型

与文献[1]相同模型下,假设共享等值连接大小的发起方为租户 R、接受方为租户 S,服务提供方为 P,且彼此间存在安全的信道。R 和 S 的密钥为 e_R 和 e_S。密钥为 e,加密消息 x 的可交换加密函数为 $f_e(x)$[4]。R 和 S 的加密隐私数据逻辑在数据库 D_R、D_S 的表 T_R、T_S 存储。T_R、T_S 的共有属性为 A(实质上,只要求 T_R、T_S 的 2 个属性值可比较即可,为便于描述,本文假设为共有属性),记为 $T_R.A$、$T_S.A$。$T_R.A$、$T_S.A$ 的一个明文记为 v_R、v_S,全集记为 V_R、V_S。v_R、v_S 的密文分别为 $y_R = f_{e_R}(v_R)$、$y_S = f_{e_S}(v_S)$,全集记为 Y_R、Y_S。T_R、T_S 的一个元组记为 b_R、b_S,全集为 B_R、B_S。除属性 A 外,b_R、b_S 其它属性的明文为 $ext(v_R)_R$、$ext(v_S)_S$。V_R 和 V_S 的交集为 $V = V_R \cap V_S = \{v \mid v \in V_R \wedge v \in V_S\}$。$|O|$ 表示集合 O 的元素个数。共享等值连接大小的模型如图 1 所示。

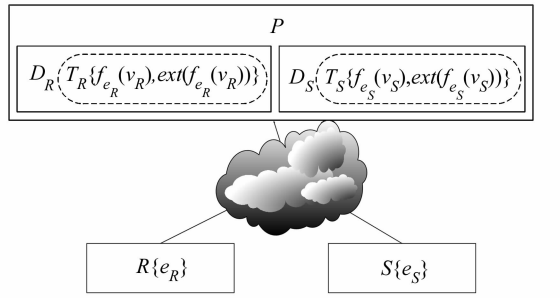

图 1 等值连接大小共享模型图

Fig. 1 Equijoin size sharing model

2 ESSAERDP 协议

2.1 等值连接大小

定理 1 设 T_R、T_S 在 A 上交集 V 大小为 n_V,V 中第 i 个元素对应 b_R、b_S 的个数分别为 n_{R_i}、n_{S_i},则 T_R、T_S 在 A 上等值连接大小为 $|T_{R_{T_R.A=T_S.A}\infty}T_S| = n_{R\infty S} = \sum_{i=1}^{n_V} n_{R_i} \times n_{S_i}$。

证明:利用数学归纳法证明。

1)当 T_R、T_S 在 A 上交集 V 大小为 1 时,即只有一个交集元素,该元素对应 b_R、b_S 的个数分别为 n_{R_1}、n_{S_1},根据等值连接[9]的定义,T_R、T_S 在 A 上等值连接是 A 值相等 b_R、b_S 的组合,则等值连接大小为 $|T_{R_{T_R.A=T_S.A}\infty}T_S| = n_{R_1} \times n_{S_1}$。

2)假设 T_R、T_S 在 A 上交集 V 大小为 n_V,V 中第 i 个元素对应 b_R、b_S 的个数分别为 n_{R_i}、n_{S_i},同理,根据等值连接[9]的定义,等值连接大小为 $|T_{R_{T_R.A=T_S.A}\infty}T_S| = n_{T_{R_{T_R.A=T_S.A}\infty}T_S} = \sum_{i=1}^{n_V} n_{R_i} \times n_{S_i}$。

3)当 T_R、T_S 在 A 上交集 V 大小为 $n_V + 1$ 时,前

n_V 元素对应的等值连接大小为 $\sum_{i=1}^{n_V} n_{R_i} \times n_{S_i}$,第 n_V+1 个元素对应等值连接大小为 $n_{R_{n_V+1}} \times n_{S_{n_V+1}}$,则有 $|T_{R_{T_R.A=T_S.A}^{\infty} T_S}| = n_{T_{R_{T_R.A=T_S.A}^{\infty} T_S}} = \sum_{i=1}^{n_V} n_{R_i} \times n_{S_i} + n_{R_{n_V+1}} \times n_{S_{n_V+1}} = \sum_{i=1}^{n_V+1} n_{R_i} \times n_{S_i}$。

根据数学归纳法,T_R、T_S 在 A 上等值连接大小为 $|T_{R_{T_R.A=T_S.A}^{\infty} T_S}| = n_{T_{R_{T_R.A=T_S.A}^{\infty} T_S}} = \sum_{i=1}^{n_V} n_{R_i} \times n_{S_i}$。

2.2　协议描述

在不泄露各自私有信息的前提下,一组互不信任的参与方可通过安全多方计算[10-12]实现多方合作计算。本研究中,R、S 和 P 互不完全信任,R、S 通过 P 共享等值连接大小,属于安全三方计算。

最理想的隐私共享是仅共享如等值连接大小等授权敏感信息,但多数情况下无法达到。Agrawal 等[4]提出最少必要信息共享(minimal necessary information sharing),按照必须知道(need-to-know)的原则,允许最少量必要信息的泄漏。本研究的最少必要信息共享条件是:R 仅获得 $|V_S|$、$|V_S \cap V_R|$ 和 $|T_{R_{T_R.A=T_S.A}^{\infty} T_S}|$,$S$ 仅获得 $|V_R|$、$|V_S \cap V_R|$ 和 $|T_{R_{T_R.A=T_S.A}^{\infty} T_S}|$,$P$ 仅获得 $|V_R|$、$|V_S|$。

本研究采用理想的哈希函数 $h(x)$[4],即对于 $v \in V$ 均可计算 $h(v)$;对于独立的 $x \in_r Dom\ F$ 均可以选择 $x = h(v)$;相对于 $|V_S \cup V_R|$,假设 $|Dom\ F|$ 足够大,$h(v)$ 碰撞的概率小于指数级。

基于安全多方计算思想[10-12],在最少必要信息共享条件下,作者提出的 ESSAERDP 协议,首先,SP 计算 2 个租户表中某共有属性对应元组其它属性的哈希值,将它们构成的 2 元组全集分发给对方租户;其次,租户再次加密从 SP 收到的 2 元组全集的共有属性后,互发给对方;第三,2 个租户独立计算该属性值的密文交集以及每个元素对应各租户的元组数;最后,根据定理 1 计算等值连接大小,实现共享。协议流程如图 2 所示,详细过程如下:

① R 拟与 S 共享表 T_R、T_S 关于属性 A 的等值连接大小,R 向 S 请求通过 P 访问 Y_S,同时,向 S 颁发通过 P 访问 Y_R 的一次性授权证书。

② 若 S 同意共享请求,则向 R 颁发通过 P 访问 Y_S 的一次性授权证书。

③R、S 向 P 出示一次性授权证书,请求 Y_S、Y_R。

④P 通过检验 R 提交的授权证书后,针对每个

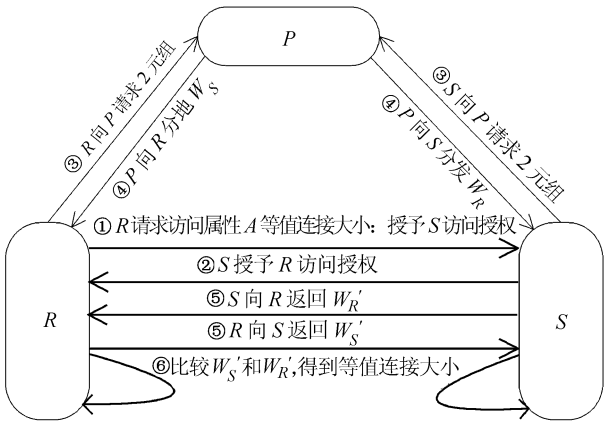

图 2　ESSAERDP 协议流程图

Fig. 2　ESSAERDP protocol flowchart

$y_S \in Y_S = \{y_S \mid y_S = f_{e_S}(v_S), v_S \in V_S\}$ 生成一个 2 元组 $w_S = \langle f_{e_S}(v_S), h(ext(f_{e_S}(v_S))) \rangle$,将 2 元组 w_S 的字典序(lexicographically)[13] 全集 W_S 分发至 R。

同理,P 检验通过 S 提交的授权证书后,针对每个 $y_R \in Y_R = \{y_R \mid y_R = f_{e_R}(v_R), v_R \in V_R\}$,$P$ 生成 $w_R = \langle f_{e_R}(v_R), h(ext(f_{e_R} v_R)) \rangle$,将 2 元组 w_R 的字典序[13] 全集 W_R 分发给 S。

⑤收到 W_S 后,R 用 e_R 计算 $w_S' = \langle f_{e_R}(f_{e_S}(v_S)), h(ext(f_{e_S}(v_S))) \rangle$,将 2 元组 w_S' 的字典序[13] 全集 W_S' 返回 S。

同理,收到 W_R 后,S 用 e_S 计算 $w_R' = \langle f_{e_S}(f_{e_R}(v_R)), h(ext(f_{e_R}(v_R))) \rangle$,将 2 元组 w_R' 的字典序[13] 全集 W_R' 返回 R。

⑥R、S 独立比较 W_S' 和 W_R' 2 元组元素 w_S' 和 w_R' 的第 1 项,根据可交换加密函数[4] 的可交换性知,当 $f_{e_R}(f_{e_S}(v_S)) = f_{e_S}(f_{e_R}(v_R))$ 时,虽然 R 和 S 均不能获知 v_S 和 v_R 的具体值,但可得知 $v_S = v_R$,记 $V = \{v_i \mid v_i = v_R = v_S, v_i \in V_R \cap V_S, 0 \le i \le l\}$。对应于 $v_i \in V$,$h(ext(f_{e_S}(v_S)))$ 的个数记为 n_{S_i},$h(ext(f_{e_S}(v_R)))$ 的个数记为 n_{R_i}。

根据定理 1,R、S 独立计算 T_R、T_S 在属性 A 上的等值连接大小 $|T_{R_{T_R.A=T_S.A}^{\infty} T_S}| = \sum_{i=0}^{l} (n_{Si} n_{Ri})$,$(0 \le i \le l)$,实现了等值连接大小的共享。

2.3　完备性证明

定理 2　半诚实模型下,在 ESSAERDP 协议中,R 精确获知 $|V_S|$、$|V_S \cap V_R|$ 和 $|T_{R_{T_R.A=T_S.A}^{\infty} T_S}|$,$S$ 精确获知 $|V_R|$、$|V_S \cap V_R|$ 和 $|T_{R_{T_R.A=T_S.A}^{\infty} T_S}|$,$P$ 精确获知 $|V_R|$、$|V_S|$。

证明:假设 R、S 和 P 都是“半诚实的[14]”。下面在半诚实模型[14]下分析 R 的完备性。

1) 通过比较 W_S' 和 W_R' 2元组 w_S' 和 w_R' 的第1项,根据可交换加密函数[4]的双射性和可交换性知,$|V_S \cap V_R| = |\{v_R = v_S \mid f_{e_S}(f_{e_R}(v_R)) = f_{e_R}(f_{e_S}(v_S))\}|$,虽然 R 不能通过解密得到 v_R 和 v_S,但可判断 $f_{e_S}(f_{e_R}(v_R)) = f_{e_R}(f_{e_S}(v_S))$ 是否成立。

2) 由于 W_S' 和 W_R' 中的2元组元素 w_S' 和 w_R' 按字典序[13]排列,所以根据 $f_{e_R}(v_R)$、$f_{e_S}(v_S)$ 的顺序推测 v_R、v_S 的值,在计算上是不可行的。

3) W_S' 和 W_R' 2元组元素 w_S' 和 w_R' 的第2项是 A 属性值对应元组其它属性值的哈希值,根据哈希函数[15]的性质知,由其推测 A 的值及对应元组其它属性值,在计算上是不可行的。

4) 对于 $v_i \in V$,不同 $h(ext(f_{e_S}(v_S)))$ 的个数记为 n_{S_i},不同 $h(ext(f_{e_R}(v_R)))$ 的个数记为 n_{R_i},由定理1知,表 T_R、T_S 属性 A 等值连接大小为 $|T_{R_{T_R.A=T_S.A}\infty T_S}| = \sum_{i=0}^{l}(n_{Si}\, n_{Ri})$,$(0 \le i \le l)$,即 R、S 精确获知 $|T_{R_{T_R.A=T_S.A}\infty T_S}|$。

综上可知,R 精确获知 $|V_S \cap V_R|$ 和 $|T_{R_{T_R.A=T_S.A}\infty T_S}|$。

同理,S 精确获知 $|V_S \cap V_R|$ 和 $|T_{R_{T_R.A=T_S.A}\infty T_S}|$。

在协议流程中,R 获得 Y_S,S 获得 Y_R,P 获得 Y_R、Y_S,因为 $|V_R| = |Y_R|$、$|V_S| = |Y_S|$,所以 R、S 和 P 精确获知 $|V_S|$、$|V_R|$。

2.4 安全性证明

定理3[4] 当 $m,n \in Z, 0 \le m \le n, i: x_i, z_i \in_r Dom\ F, e \in_r Dom\ F$ 时,如式(1)所示 $2\times n$ 元组 D_n^n 和 D_m^n 分布在计算上是不可区分的。

$$\begin{cases} D_n^n = \begin{pmatrix} x_1 & \cdots & x_m & x_{m+1} & \cdots & x_n \\ f_e(x_1) & \cdots & f_e(x_m) & f_e(x_{m+1}) & \cdots & f_e(x_n) \end{pmatrix}, \\ D_m^n = \begin{pmatrix} x_1 & \cdots & x_m & x_{m+1} & \cdots & x_n \\ f_e(x_1) & \cdots & f_e(x_m) & z_{m+1} & \cdots & z_n \end{pmatrix} \end{cases} \tag{1}$$

定理4[16] 假设分布 E 和 J、J 和 K 均是计算不可区分的,则分布 E 和 K 是计算不可区分的。

定理5 半诚实模型下,ESSAERDP协议满足最少必要信息共享条件,R 仅获知 $|V_S|$、$|V_S \cap V_R|$ 和 $|T_{R_{T_R.A=T_S.A}\infty T_S}|$,$S$ 仅获知 $|V_R|$、$|V_S \cap V_R|$ 和 $|T_{R_{T_R.A=T_S.A}\infty T_S}|$,$P$ 仅获知 $|V_R|$、$|V_S|$。

证明:在半诚实模型[14]下,假设攻击者模拟 R 攻击 ESSAERDP 协议。令 $V_S = \{v_1, \cdots, v_t, v_{t+1}, \cdots, v_m\}$,$V_R = \{v_{t+1}, \cdots, v_m, v_{m+1}, \cdots, v_n\}$,$t = |V_S - V_R|$,$m = |V_S|$ 和 $n = |V_R \cup V_S|$,则攻击者拥有的信息包括 $h(x)$、V_R、$|V_S \cap V_R|$、$|V_S|$、$|T_{R_{T_R.A=T_S.A}\infty T_S}|$ 和 e_R,希望得到的信息包括 $V_S - V_R$ 和 $\{ext(v_S) \mid v_S \in V_S - V_R\}$。攻击者模拟 R 的过程如下:

1) 在协议的第5)步中,模拟攻击者随机生成 n 个 $y_i, \zeta_i \in_r Dom\ F$ $(i = 1, \cdots, m)$,模拟生成所有 $v \in V_S \cup V_R$ 的 $f_{e_S}(v_S)$ 和 $h(ext(f_{e_S}(v_S)))$。由于没有模拟产生 e_S,所以不能确定 y_i, ζ_i 代替 $f_{e_S}(v_S)$ 和 $ext(f_{e_S}(v_S))$ 的值。模拟者通过 $\{y_{t+1}, \cdots, y_n\}$ 用 e_R 生成 $Z_R = \{f_{e_R}(y_{t+1}), \cdots, f_{e_R}(y_n)\}$,模拟生成 $\zeta_i = h(ext(f_{e_S}(v_i)))$,$i = t+1, \cdots, n, v_i \in V_S - V_R$。

2) 根据定理3可知,式(2)分布 D_0^n 和 D_n^n 在计算上是不可区分(indistinguishability)[4]的。

$$\begin{cases} D_0^n = \begin{pmatrix} x_1 & \cdots & x_n \\ y_1 & \cdots & y_n \\ \zeta_1 & \cdots & \zeta_n \end{pmatrix}, \forall i: x_i, y_i, \zeta_i \in_r Dom\ F; \\ D_n^n = \begin{pmatrix} x_1 & \cdots & x_n \\ y_1 & \cdots & y_n \\ \zeta_1 & \cdots & \zeta_n \end{pmatrix}, \forall i: x_i \in_r Dom\ F, y_i = f_{e_S}(x_i), \\ e_S \in_r KeyF, \zeta_i = h(ext(f_{e_S}(v_S))) \end{cases} \tag{2}$$

3) 令函数 $Q(M) = \langle h, e_R, Y_S, Z_R, \zeta_S \rangle$,其中,随机密钥 e_R,$Y_S = \{y_1, \cdots, y_n\}$,$\{f_{e_R}(y_{t+1}), \cdots, f_{e_R}(y_n)\}$,$\zeta_S = \{h(ext(f_{e_S}(v_{t+1}))), \cdots, h(ext(f_{e_S}(v_n)))\}$。

若 M 相当于分布 D_0^n,则模拟 R 的数据为 $Q(M)$;若 M 相当于分布 D_n^n,则 $Q(M)$ 可看作 R 的真实数据。

4) 由于分布 D_0^n 和 D_n^n 在计算上是不可区分的,多项式时间内可计算函数 $Q(M)$,根据定理4有,分布 $Q(D_0^n)$ 和 $Q(D_n^n)$ 在计算上不可区分[4]的。

综上可知,在半诚实模型下,R 仅获知 $|V_S|$、$|V_S \cap V_R|$ 和 $|T_{R_{T_R.A=T_S.A}\infty T_S}|$。

同理可得,S 仅获知 $|V_R|$、$|V_S \cap V_R|$ 和 $|T_{R_{T_R.A=T_S.A}\infty T_S}|$。

5) 由于 R、S 的隐私数据以密文托管于 P,所以 P 无法获得数据内容,仅能获知 $|V_R|$、$|V_S|$。

在基本模型图1中,假设 R、S 和 P 彼此间有安全信道,所以在传输过程中隐私数据是保密的。

因此,在半诚实模型下,ESSAERDP 协议满足

最少必要信息共享条件,R 仅获知 $|V_S|$、$|V_S \cap V_R|$ 和 $|T_R \underset{T_R.A=T_S.A}{\infty} T_S|$,$S$ 仅获知 $|V_R|$、$|V_S \cap V_R|$ 和 $|T_R \underset{T_R.A=T_S.A}{\infty} T_S|$,$P$ 仅获知 $|V_R|$、$|V_S|$。

2.5　效率分析

设在文献[4]相同条件下,$Dom\ F$ 的密文字长为 k bit,R、S 运行 $f_e(x)$ 的计算代价为 C_e,P 运行 $h(x)$ 的计算代价为 C_h,$|ext(v_S)|$ 表示 T_S 的属性数减 1,$|ext(v_R)|$ 表示 T_R 的属性数减 1。在理想条件下,分析等值连接大小共享协议的计算代价和通讯代价,如在协议第 4) 步中,R、S 分别在 P 存储 Y_R、Y_S,忽略执行加密的代价。

1)计算代价

$2C_e(|V_R|+|V_S|)+C_h(|ext(v_S)|+|ext(v_R)|)$。

由于 $C_e \gg C_h$,所以本协议的计算代价近似于 $2C_e(|V_R|+|V_S|)$。

在与文献[4]相同的条件下,Agrawal 协议的计算代价为 $2C_e(|V_R|+|V_S|)$,其实现公平共享的计算代价为 $4C_e(|V_R|+|V_S|)$,所以本协议的计算代价仅是通过 Agrawal 协议实现公平共享的 50%。

2)通信代价

$(2(|V_R|+|V_S|)+|h(ext(v_S))|+|h(ext(v_R))|)k$ bit。

由于 $h(x)$[17] 的通信代价很少,所以本协议的通信代价近似于 $2(|V_R|+|V_S|)k$ bit。

在与文献[4]相同的条件下,Agrawal 协议的通信代价为 $(|V_R|+2|V_S|)k$ bit,其实现公平共享的通信代价为 $2(|V_R|+2|V_S|)k$ bit,所以本协议的通信代价是通过 Agrawal 协议实现公平共享的 $(|V_R|+|V_S|)/(|V_R|+2|V_S|)$。当 $|V_R|$ 和 $|V_S|$ 相近,即 $|V_R| \approx |V_S|$ 时,$(|V_R|+|V_S|)/(|V_R|+2|V_S|) \approx 68\%$。

3　测试与分析

为了验证 ESSAERDP 协议的有效性,将基于 JSP + Tomcat + MySQL 平台研发的肉牛性状改良管理系统部署 P 服务器上。

服务器的性能:CPU Intel(R) Core i7-2600 3.4 GHz;内存 DDR3 8192Mbytes 802.7 MHz;操作系统 Windows 7 64 位;Java 平台 JDK 1.6.0_22。

选用 2 台 PC 机作为 R、S 的测试终端。测试终端的性能:CPU Pentium E5200 2.5 GHz;内存 2G 2.5 GHz。通过 100 Mbps 局域网组成图 1 所示的逻辑结构,忽略网络速度对协议的影响。

肉牛性状改良管理系统是一个 web 系统,用 web 系统扩展为多租户 SaaS 原理[18],可以方便地将其扩展为多租户[19]管理型 SaaS。在测试原型中,分别用 R、S 的基因性状表模拟协议中的 T_R、T_S 表,表中有 8 个属性,加密属性值采用 MySQL 中的 char 类型存储,采用 JDK 中的 java.math.BigInteger 类处理。可交换加密函数用 $f_e(x) = x^e \bmod p$[4]。哈希函数用 Sha-512,用 javax.xml.crypto.dsig.DigestMethod 接口实现。时间计数器用 System.nanoTime 方法实现。在 T_R、T_S 表的元组数分别为 1 000、5 000、10 000、15 000 和 20 000 条,对应的等值连接大小分别为 6 283、199 645、582 301、1 309 894 和 2 401 958 个的条件下,当可交换加密函数的密钥长度为 512、1 024、2 048 和 3 072 bit 时,取相同条件下执行 50 次的均值作为测试 ESSAERDP 协议的执行代价,测试结果如表 1 所示。

表 1　ESSAERDP 协议的执行代价

Tab. 1　ESSAERDP's Performance

密钥长度/bit	执行代价/ms				
	1 000	5 000	10 000	15 000	20 000
512	5 071	27 718	45 324	69 024	88 190
1 024	11 504	59 059	97 269	148 113	195 867
2 048	32 741	187 340	312 775	474 231	650 225
3 072	60 799	363 593	60 8794	917 293	1 228 070

执行代价与元组数、密钥长度及三者关系分别如图 3、4 所示。

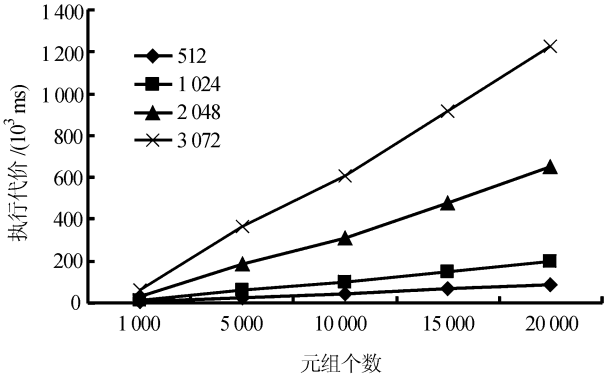

图 3　元组数与执行代价关系图

Fig. 3　Relation between tuple number and performance

原型测试可以看出:

1)由图 3 可看出,当密钥长度相等时,执行代价和共享表中的元组数成线性增长;密钥越长,增长率越高。

当前,SET(secure electronic transaction)协议中建议 RSA 密钥长度为 1 024 bit。当密钥长度为

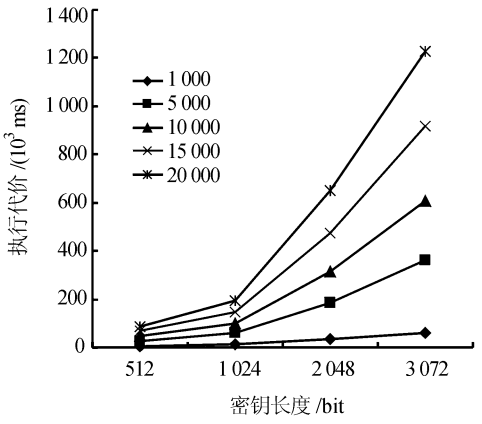

图4 密钥长度与执行代价关系图

Fig. 4 Relation between key size and performance

1 024 bit 时,1 000 条元组的执行代价为 11 504 ms,20 000 条元组的执行代价为 195 867 ms,在租户忍受范围内,所以本协议基本能满足管理型 SaaS 的需求。

2)由图 4 可看出,当共享表中的元组数相等时,执行代价和密钥长度成线性增长;当密钥长度小于 1 024 bit 时,随着密钥长度的增加,执行代价增加不明显,增长率较平稳;当密钥长度大于 1 024 bit 时,随着密钥长度的增加,执行代价增加非常明显,增长率激升。

为了解决密钥长度增加引起执行代价激增的问题,租户可将隐私数据分级存储在不同的表中。由于高安全强度的元组数较少,密钥长度对执行代价影响不显著,可采用更长的密钥(如 2 048 bit,甚至 3 072 bit)以达到更高的安全强度,由此引起的执行代价增加不明显,仍能满足管理型 SaaS 的需求。

4 结 论

针对管理型 SaaS 中 2 个租户通过 SP 共享跨加密隐私关系库等值连接大小的问题,本研究利用数学归纳法,证明了等值连接大小的组合计算公式;基于安全多方计算思想,利用可交换加密函数和哈希函数,提出了一种跨加密关系数据库等值连接大小共享协议。完备性和安全性证明表明,在半诚实模型下,本协议安全可证,满足最少必要信息共享条件。从效率分析可看出,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 50% 和约 68%。原型测试表明,本协议基本能满足管理型 SaaS 的需求。本研究对于不完全可信双方通过不完可信的第三方实现隐私共享具有重要的理论价值,有助于推动 SaaS 模式走向成熟,可推广应用于其它云服务。

参考文献:

[1] Jing Xu, LI Bingbing, He Dongjian. A protocol of encrypted data equijoin sharing across private database[J]. Journal of Xi'An Jiaotong University, 2012, 46(8): 37 – 42. [景旭,李冰冰,何东健. 跨隐私数据库加密数据等值连接共享协议[J]. 西安交通大学学报,2012,46(8):37 – 42.]

[2] Sengupta B, Roychoudhury A. Engineering multi-tenant software-as-a-service systems[C]//Proceedings of International Conference on Software Engineering. New York: ACM, 2011: 15 – 21.

[3] Lin Ziyu, Lai Yongxuan, Lin Chen, et al. Research on cloud databases[J]. Journal of Software, 2012, 23(5): 1148 – 1166. [林子雨,赖永炫,林琛 等. 云数据库研究[J]. 软件学报,2012,23(5):1148 – 1166.]

[4] Agrawal R, Evfimievski A, Srikant R. Information Integration across autonomous enterprises. United States: US 0065910 A1[P]. 2008.

[5] Liang Gang, Chawathe S S. Privacy-preserving inter-database operations[C]//Proceedings of Intelligence and Security Informatics Second Symposium on Intelligence and Security Informatics. Berlin: Springer, 2004: LNCS 3073: 66 – 82.

[6] Yuan Xianping, Zhong Hong, Huang Hongsheng, et al. Efficient query protocol for database's privacy[C]//Proceedings of 2nd International Conference on Theoretical and Mathematical Foundations of Computer Science. Berlin: Springer, 2011, 164: 109 – 115.

[7] Ma Sha, Yang Bo, Li Kangshun, et al. A privacy-preserving join on outsourced database[C]//Proceedings of 14th International Conference on Information Security. Berlin: Springer, 2011, 7001: 278 – 292.

[8] Li Ling. Research on some issues of data security in cloud computing services[D]. He Fei: University of Science and Technology of China, 2012. [李凌. 云计算服务中数据安全的若干问题研究[D]. 合肥:中国科学技术大学,2012.]

[9] Huang K T, Davenport W B Jr. A model for equi-join query processing in distributed relational databases[R]. The

Laboratory for Information and Decision Systems Technical Report Series 1211,1981.

[10]Tang Chunming,Shi Guihua,Yao Zhengan. Secure multi-party computation protocol for sequencing problem[J]. Scientia Sinica Informationis,2011,41(7):789 - 797.[唐春明,石桂花,姚正安. 排序问题的安全多方计算协议[J]. 中国科学:信息科学,2011,41(7):789 - 797.]

[11]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.[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.]

[12]Zhang Yu,Hu Jie. A trust model of UC secure computation[J]. Journal of Schuan University:Engineering Science Edition,2012,44(3):106 - 11.[张妤,胡杰. UC 安全计算的一种信任模型[J]. 四川大学学报:工程科学版,2012,44(3):106 - 11.]

[13]Wiedermann I. The complexity of lexicographic sorting and searching[J]. Aplikace Matematiky,1981,26(6):432 - 436.

[14]Fletcher C,van Dijk M,Devadas S. A secure processor architecture for encrypted computation on untrusted programs[C]//Proceedings of 7th ACM Workshop on Scalable Trusted Computing. New York:ACM,2012:3 - 8.

[15]Agarwal S,Rungta A,Padmavathy R,et al. An improved fast and secure hash algorithm[J]. Journal of Information Processing Systems,2012,8(1):119 - 132.

[16]Rogaway P. The round complexity of secure protocols[D]. Boston:Massachusetts Institute of Technology,1991.

[17]Agarwal S,Rungta A,Padmavathy R,et al. An improved fast and secure hash Algorithm[J]. Journal of Information Processing Systems,2012,8(1):119 - 132

[18]叶伟. 互联网时代的革命软件——SaaS 架构设计[M]. 北京:电子工业出版社,2009.

[19]Chong F,Carraro G. Architecture strategies for catching the long tail [R/OL]. http://msdn. microsoft. com/en-us/library/aa479069. aspx,2011.

(编辑　杨　蓓)