doi:10.3969/j.issn.1000-436x.2014.11.012

GUC 安全的关系联结算子保密计算协议

田园，孙荣辛，蔡悟洋
(大连理工大学 软件学院，辽宁 大连 116620)

摘 要：分布式数据库系统的关系算子的保密计算协议是多方保密计算(MPC)理论的重要应用领域之一，目前该方向的绝大部分工作主要针对如何构造查询类算子的保密计算协议，对如何构造数据生成类算子的保密计算协议则较少涉及。针对广泛应用的关系联结(join)算子，基于保密及匿名的身份基公钥加密(IBE)方案及其用户私钥盲生成协议给出联结算子的 2-方保密计算协议的一种通用的、不依赖于随机 oracle(即标准模型)的有效构造，并证明该构造具有 GUC(generlized universal composability)安全性。
关键词：多方保密计算；GUC 安全模型；分布式数据库；联结算子
中图分类号：TP309 **文献标识码**：A **文章编号**：1000-436X(2014)11-0107-10

GUC-secure protocol for private relational join operator computing

TIAN Yuan, SUN Rong-xin, CAI Wu-yang
(Software School, Dalian University of Technology, Dalian 116620, China)

Abstract: It is one of important applications of secure multiparty computation that privacy-preserving SQL computation in distributed relational database. There are only few works dealing with provably-secure privacy-preserving data manipulations in contrast with comparatively abundant works on privacy-preserving data-query in database, among which the join operator is the most powerful in generating new data (relation). By making use of anonymous IBE (identity-based encryption) scheme and its user private-keys blind generation techniques, a very general cryptographic protocol framework is proposed for secure 2-party join computation. This construction is provably GUC (generalized universally composable) secure in the standard model with acceptable efficiency.
Key words: secure multiparty computation; GUC security; distributed relational database; join

1 引言

关系数据库系统是当代最重要的应用软件之一，其中分布式关系数据库技术也日益成熟和广泛应用。关系数据库系统的核心是关系数据模型和作用于这类数据模型上的关系代数算子[1]，如带各种约束的查询算子(σ)、投影算子(π)、联结算子(join)及相关的集合算子。分布式数据库系统将关系算子的语义扩展到分布式关系模型并且已发展出较成熟的实现技术[2]。在分布式数据库系统中，数据分布于各个站点，于是自然地引出一类针对关系算子的多方保密计算(MPC)问题：在站点彼此不信任的情形下，如何构造实现这些运算的安全协议使得既能导出算子的输出，同时又保证协议各参与方的私有数据彼此保密？

尽管对一般性的 MPC 问题已经给出肯定的存在性结论和普适的求解方案，但由于效率的缘故，针对具体问题构造高效求解方案仍然具有重要的实用价值。对分布式数据库领域中关系算子的安全计算问题，应用工作者和密码学理论工作者都进行了探讨，前者的工作主要基于启发式分析，缺乏精确的安全理论基础，后者的工作则基于精确的安全

收稿日期：2014-06-05；**修回日期**：2014-09-10
基金项目：国家自然科学基金资助项目(61370144)
Foundation Item: The National Natural Science Foundation of China (61370144)

模型并强调严格的安全证明，这也是本文采用的研究方法。目前为止所解决的各类问题大部分属于保密查询的协议构造(相当于构造σ和π算子的保密计算协议)[3~7]以及保密的集合运算协议(其应用不完全限于数据库系统)[8,9]、从关系数据模型导出语义决策树[10]等。就目前所知，除联结算子之外对其他重要的关系算子都基本给出了保密计算协议，这些方案的安全语义几乎均为 standalone 意义下的安全性[11]，尚未有满足 UC 安全[12]特别是最近完善的 GUC 安全性[13]，并且绝大多数工作致力于 2-方保密计算(除文献[9]中的部分方案之外)。此外，现有方案的计算和通信效率也仍有进一步改进的必要。

本文致力于构造联结算子的保密计算协议。与大多数工作相似，致力于 2-方情形，但与其他工作不同，在技术上主要采用保密和匿名的 IBE 方案及其用户私钥的盲生成协议，辅以少量易于高效实现的零知识证明协议，由此建立起一个通用的、具有 GUC 安全性的协议方案。本文工作是与其他关系算子的现有工作结合，给出分布式关系数据库系统保密计算完整的、可证明的安全理论方案。

2 基础知识和方法概要

一个关系是具有相同属性的一组元组实例的集合。设 X_1 和 X_2 是 2 个关系，A 是其共同的属性(或属性集合，$A=(A_1,\cdots,A_p)$)，则联结算子 $Join(A{:}X_1,X_2)$ 是这样一种关系代数算子，它将 X_1 和 X_2 映射为一个新的关系 X_3，X_3 的属性包括 X_1 和 X_2 的全部属性，但其中公共属性 A 仅出现一次，且 X_3 的元组由 X_1 和 X_2 的所有这样的元组笛卡儿积构成，这些元组在公共属性 A 上具有相同的值[1,2]。例如，给定 2 个关系 *Incom* 和 *Debt*，*Incom* 具有属性 *customer_id* 和 *income*，*Debt* 具有属性 *customer_ id* 和 *debt*, 则联结算子 *Join*(*customer_id*: *Incom*, *Debt*) 生成的关系 X_3 具有属性 *customer_id*、*income* 和 *debt*; 作为元组的集合，若 *Incom*={(*c*1, 2 500), (*c*2, 3 000), (*c*5, 1 010), (*c*6, 2 000)}，*Debt*={(*c*2, 19 000), (*c*4, 7 000), (*c*5, 88), (*c*7, 100)}(方框标记出 2 个关系在属性 *customer_id* 上相同的值), 于是按定义 X_3 为元组集合{(*c*2, 3 000, 19 000), (*c*5, 1 010, 88)}。在数据库系统的实际应用中，联结算子是最有价值的关系代数算子之一，详细讨论见文献[1,2]。

在分布式数据库系统中，若上例中的 *Incom* 和 *Debt* 分别位于 2 个站点且不允许向任何一方泄露 2 个站点的非公共客户对象的借贷信息，则自然地引出联结算子的保密计算问题。

由于关系算子常常作为复杂数据处理程序的一部分被调用，即关系算子可能运行于任何上下文之中，因此其具有实用价值的安全性质必须在任何(可能是恶意的)运行环境中保持不变，即最好具有 UC/GUC 意义上的安全性[12,13]。最近完善的 GUC 理论[13]具有最现实的安全涵义，该理论是本文进行协议构造和安全证明的基础。

本文基于 IBE 方案及其用户私钥盲生成协议建立关于联结算子的一个非常通用的 2-方保密计算协议，其构造概要描述如下。对 IBE 方案 Π=(Setup, UKG,E,D)，其所谓用户私钥盲生成协议是这样一种协议，其中一方 P_1 持有全局私钥 *msk*，另一方 P_2 向 P_1 提交其用户身份公钥(即身份标识)*a*，该协议使 P_1 为 P_2 计算出正确的用户私钥 *usk*(*a*)= UKG(*msk*, *a*)但却未知 *a* 本身。设站点 P_1 持有关系 X_1，其属性为 *w* 和 *x*；站点 P_2 持有关系 X_2 其属性为 *w* 和 *y*，需求解的问题是在 P_2 上输出 $Join(w{:}\ X_1,\ X_2)$[注1]且 P_2 不能推断出 X_1 的任何未出现在输出结果中的信息(除 X_1 中元组个数外)。设 X_1={(w_1,x_1), (w_2,x_2), (w_3,x_3), (w_4,x_4)}、X_2={(w_2,y_2), (w_4,y_4), (w_5,y_5), (w_6,y_6)}，其中，w_i、x_i 和 y_i 分别表示属性 *w*、*x* 和 *y* 上的值，协议的基本过程是：P_1 生成 IBE 方案的全局公钥/私钥对(*mpk*,*msk*)、向 P_2 传输 *mpk* 和密文 ξ_i=E(*mpk*, w_i,x_i||M_0)(*i*=1,2,3,4)，其中，M_0 是双方公开约定的一个位串，||是位串的连接运算符。P_2 通过该 IBE 方案的用户私钥盲生成协议从 P_1 获得 *usk*(w_2)、*usk*(w_4)、*usk*(w_5)和 *usk*(w_6)，用这些用户私钥解密各 ξ_i，并检验哪些解密输出的后缀为约定的位串 M_0。对本例而言，当 P_2 用 *usk*(w_2)和 *usk*(w_4)进行解密时能从 ξ_2 和 ξ_4 得出具有后缀 M_0 的输出，从而判定相应的前缀 x_2 和 x_4 为 X_1 中相应记录的属性值，即 P_2 能够成功得出 X_1 中的元组子集 $X^0{}_1$={(w_2,x_2), (w_4,x_4)}，并以此为输入之一在本地进行(普通的联结)计算 $Join(w{:}X^0{}_1,X_2)$，得到正确的输出{(w_2,x_2,y_2), (w_4,x_4,y_4)}。直观地说，Π 的匿名性质和保密性质保证 P_2 不能得到 X_1 的那些未出现在结果中的元组信息，而 Π 的用户私钥盲生成协议则保证 P_1 未知关于 X_1 元组的任何信息。严格来说，该协议的正确性

注1 在分布式数据库系统的实际应用中最常用的情形是在多个站点之一生成关系算子的输出，因此本文仅考虑这一情形，但不难推广到多站点输出的情况。

概率不是 1，但 Π 的保密性保证对超对数或多项式长度的字 M_0 的概率与 1 的差是复杂性参数的可忽略函数。如果增加一定的计算复杂度则也存在一个略有不同但完全精确的协议方案，详见第 3 节。

为保证 GUC 安全性，对以上构造需补充一系列的技术性细节，具体论证见第 2 节和第 3 节。在那里可以看到该方案具有常数轮通信复杂度(round- complexity)和线性的消息复杂度 $O(N_1+N_2)$；一方的计算复杂度为 $O(N_1+N_2)$，另一方为 $O(N_1N_2)$，其中 N_1、N_2 为两方关系中的元组数量。注意到 $O(N_1N_2)$也是普通联结计算的渐进复杂度[1,2]，因此除一个常数因子外该方案计算复杂度与普通联结算子的渐进复杂度相似。

3　相关概念与工具

首先约定一组符号：P.P.T. 表示“概率多项式时间”，$x||y$ 表示位串 x 和 y 的联结，$|x|$表示字 x 的位数，$|X|$(X 是集合)表示 X 的基数，$x\leftarrow^{\$}X$ 表示在集合 X 上均匀分布随机采样一个元素 x，k 表示复杂性参数，$\approx^{\text{PPT}}$ 表示 2 个对象计算难辨(computational indistinguishability)，$\approx$ 表示 2 个对象有相同的概率分布(perfect indistinguishability)。

3.1　联结算子的 GUC 理想安全模型

一个协议具有 GUC 安全性是指任何针对该协议的现实攻击都存在一个针对其理想协议的仿真攻击，使任何(恶意)环境都不能有效分辨这 2 种攻击所导致的输出。限于篇幅，假定读者完全熟悉 Canetti 的 UC/GUC 理论的概念和主要结论[12,13]。

以下建立联结算子 2-方保密计算协议的理想模型。P_1*、P_2*表示理想协议的参与方，X_1 和 X_2 表示 P_1*、P_2*分别持有的关系，其属性分别为 $w, u^{(1)}$ 和 $w, u^{(2)}$，其中 w 为公共属性。为论述简洁以下将 $w, u^{(1)}$ 和 $u^{(2)}$均作为单一属性处理，但所有的概念和构造很容易推广到三者都是属性向量的情况。$Join(w{:}X_1,X_2)$表示普通意义的联结算子，其理想的保密计算模型定义为

$$F_{Join}\colon (X_1,X_2)\rightarrow (|w|_2, |X_1|||Join(w{:}X_1,X_2))$$

其中，$|w|_2$ 表示 X_2 中的元组在属性 w 上取不同值的个数。更精确地，设 $N_1=|X_1|, N_2=|X_2|$, S 表示对理想协议的攻击算法，理想协议的程序如下。

当从 P_1*收到消息(sid,“input”,P_1*,X_1)时，F_{Join} 记录 X_1 并向 P_2* 和 S 发送消息(*sid*,“input”,N_1)；当从 P_2*收到消息(*sid*,“input”,P_2*,X_2)，F_{Join} 记录 X_2 并向 P_1* 和 S 发送消息(*sid*,“input”,$|w|_2$)。

当从 P_2*收到消息(*sid*,“join”,P_2*)，F_{Join} 向 P_2* 响应消息(*sid*,“join”, $Join(w{:}X_1,X_2)$)。

P_1* 输出$|w|_2$, P_2* 输出 $N_1||Join(w{:}X_1,X_2)$。

将现实的协议记作 ψ，其每个参与方 P_i 在理想模型中对应着参与方 P^*_i，反之亦然。A、S 分别表示对协议 ψ 和 F_{Join} 的攻击算法；Z 表示协议的运行环境，这是一个 P.P.T.算法，用以概括在协议 ψ 运行期间针对协议的一切恶意行为。Z 可与攻击者 A 或 S 相互作用，例如 Z 指示 A 或 S 执行特定的动作、收集 A 或 S 的输出信息等。在 UC 理论中，约定 Z 不访问协议参与方的共享实体(shared functionality)状态，例如对典型的共享字符串(*c.r.s.*)范型，每个共享字符串 σ 仅为协议 ψ 特定实例的参与方所知，但 GUC 理论解除这一约束，具有更现实的意义[13,16]。记 $\text{output}_Z(\psi,A)$表示在环境 Z 和攻击者 A 作用之下，P_1、P_2 的输出所构成的联合随机变量，符号 $\text{output}_Z(F_{\text{INT}},S)$涵义类似。在协议运行期间 Z 也收集这些信息进行处理，其输出分别记作 $Z(\text{output}_Z(\psi,A),u)$ 和 $Z(\text{output}_Z(F_{\text{INT}},S),u)$, u 是辅助输入信息。不失一般性，总可以假设这些输出属于 {0,1}。

定义 1　(GUC 安全[13]) 如果对任何主动的 P.P.T.攻击算法 A 都存在 P.P.T.攻击算法 S，S 入侵的理想参与方恰对应于 A 所入侵的现实参与方，使得对任何 $Z\in$P.P.T.和任何辅助输入信息 u，函数 $|P[Z(\text{output}_Z(\psi,A),u)=1]-P[Z(\text{output}_Z(F_{\text{INT}},S),u)=1]|$ 是复杂性参数 k 的可忽略函数(记做 $\text{output}_Z(\psi, A)\approx^{\text{PPT}}\text{output}_Z(F_{Join},S)$)，则 ψ 定义做与 F_{Join} GUC-相似，简称 ψ *GUC*-安全，记做 $\psi\rightarrow^{\text{GUC}}F_{Join}$。

算法 S(与 A 有关)称为 A 的仿真算法。GUC 相似概念最重要的普遍性质是复合—稳定性定理，概要地说就是：对协议 φ_2、φ_1 和 $\psi(\varphi_1)$，其中 $\psi(\varphi_1)$是以 φ_1 为子协议的复合协议(φ_1-hybrid protocol)，若 $\varphi_2\rightarrow^{\text{GUC}}\varphi_1$，则(在某些自然的技术性条件下)必有 $\psi(\varphi_2/\varphi_1)\rightarrow^{\text{GUC}}\psi(\varphi_1)$，$\psi(\varphi_2/\varphi_1)$表示用 φ_2 对 φ_1 替换而成的新协议[12,13]。

3.2　IBE 方案、匿名性质及其用户私钥盲生成协议

IBE 方案 Π=(Setup, UKG, E, D)是一组 P.P.T.算法，其中，Setup 是全局密钥(亦称主密钥)生成算法，以复杂度参数 k 为输入并输出全局公钥—私钥偶(*mpk*, *msk*)；UKG 是用户私钥生成算法，以全局私钥 *msk*、用户身份标识 a 为输入并输出 a 的私钥

usk(*a*)；E 是加密算法，以全局公钥 *mpk*、用户身份标识 *a* 和消息 *M* 为输入并输出密文 *y*；D 是解密算法，以全局公钥 *mpk*、用户私钥 *usk*(*a*)和密文 *y* 为输入并输出明文 *M*。所有这些算法还满足一致性关系：对 *k*、*a* 和 *M* 恒有

P[(*mpk*,*msk*)←Setup(*k*); *usk*(*a*)←UKG(*msk*,*a*);
y←E(*mpk*, *a*, *M*): D(*mpk*, *usk*(*a*), *y*)=*M*]=1

IBE 方案具有保密性和匿名性，直观地讲，保密性使密文隐藏明文，匿名性则使密文隐藏公钥(对 IBE 方案即身份标识)。

为建立本文的协议，需要构造 IBE 方案的私钥盲生成协议，直观地讲就是一方向另一方(双方均已知全局公钥 *mpk*)提供正确的身份私钥 *usk*(*a*)、但另一方却不向提供方泄露自身的身份标识 *a*。以下建立该协议的理想模型 $F^{\Pi}_{\text{Blind-UKG}}$。

定义 2 (IBE 方案的用户私钥盲生成协议的理想模型) Π 是某个 IBE 方案，P^*_1、P^*_2 是参与方，*sid* 表示会话号、*ssid* 表示子会话号。

P^*_1 生成随机数 ρ、计算出(*mpk*,*msk*)←Setup(ρ)后向 $F^{\Pi}_{\text{Blind-UKG}}$ 提交消息(*sid*,ρ,*mpk*||*msk*)。

$F^{\Pi}_{\text{Blind-UKG}}$ 向 P^*_2 和理想攻击者 S 发送消息(sid, *mpk*)。

每当 P^*_2 向 $F^{\Pi}_{\text{Blind-UKG}}$ 发送请求(*sid*||*ssid*,*a*)(每次 ssid 不同)，$F^{\Pi}_{\text{Blind-UKG}}$ 计算出 *usk*(*a*)←UKG(*msk*,*a*)、向 P^*_2 响应(*sid*||*ssid*,*usk*(*a*))并向 P^*_1 和 S 发送消息(*sid*||*ssid*,*n*)，其中，*n* 初始为 0，每响应一次 *n* 增加 1。

最后 P_1*输出最新的 *n*，P_2*输出所有的 *usk*(*a*)。

私钥盲生成协议是一个有用的工具，例如文献[3]用它构造健忘传输协议(但那里未给出 GUC 安全方案，所用的具体 IBE 方案也非匿名)。

3.3 非可塑零知识证明及其扩展

本节借用文献[14,15]的概念表述，但符号做了适当简化。L 是一个 NP 类语言，R 是与之相伴的 P 类 2-元关系，即 $x\in$L, iff 存在 *w* 使 R(*x*,*w*)=1。对 $x\in$L 记 R(*x*)={ *w*: R(*x* ,*w*)=1}。A、B 是 2 个机器，A(*x*;B)$_{[\sigma]}$表示 A、B 持有公共输入 *x*，共享字串 σ (*c.r.s.*)时 A 与 B 相互作用所导致 A 的输出；$\text{tr}_{\text{A,B}}(x)_{[\sigma]}$ 表示 A、B 持有公共输入 *x* 时 A 与 B 相互作用过程中完整的消息流；当强调 A 具有(对 B 保密的)输入 *y* 时，以上符号也记为 $\text{A}_y(x;\text{B})_{[\sigma]}$和 $\text{tr}_{\text{A}(y),\text{B}}(x)_{[\sigma]}$。设 A=($A_1$, A_2)、B 和 C 都是机器，其中 A_1 可以向 A_2 传递信息/状态，则符号(<B,A_1>,<A_2,C>)表示 A_1 与 B 之间、A_2 与 C 之间相互作用且在此期间 A_1 可向 A_2 传递任何协同信息；若这一过程中 A_2 与 C 之间的消息流为 *tr*、A_2 的最终输出为 *u*、C 的最终输出为 *v*，则(<B, A_1>,<A_2,C>)的输出记作(*u*,*tr*,*v*)。2 个消息流 tr_1 和 tr_2 定义做是匹配，如果 2 个消息流中的每个消息及其顺序都相同，但对应消息的传输方向恰相反。

设 A 是机器，符号 A 表示这样一种机器，它接受 2 种形式的输入：第 1 种形如(“start”,*i*,*x*,*w*)，使 A 启动一个 A 的新实例并对该实例赋予唯一标识 *i*、公开的输入 *x* 和私有的输入 *w*；第 2 种形如(“message”,*i*,*m*)，使 A 向标识为 *i* 的机器实例 A_i 发送消息 *m* 并返回 A_i 对 *m* 的响应。

定义 3 (零知识证明协议[14,15]) ZPoK_R=(D_{crs},P,V,Sim=(Sim_1,Sim_2))是一组 P.P.T.算法，D_{crs} 是 c.r.s. σ 的生成算法；P 的输入为(σ,*x*,*w*)，(*x*,*w*)$\in$R；V 的输入为(σ,*x*), $x\in$L；Sim_1 和 Sim_2 分别是仿真陷门生成算法和仿真算法，Sim_1(*k*)输出(σ,*s*)、Sim_2 以 $x\in$L 和(σ,*s*)为输入。因为除 D_{crs} 和 Sim_1 外所有其他算法都以 σ 为输入之一，所以除非强调否则不再明显写出 σ。ZPoK_R 定义做 2-元关系 R 或相应的语言 L 的零知识证明协议，如果满足如下。

1) 对任何 $x\in$L、$\sigma\leftarrow D_{crs}$ 有 P[$V(x;P)_{[\sigma]}$=1]=1。

2) 对任何 P.P.T.算法 A、$x\notin$L、$\sigma\leftarrow D_{crs}$ 有 P[$V(x;A)_{[\sigma]}$=1]=0 [注2]。

3) 对任何 P.P.T.算法 A，A 的输出为 0 或 1，ε 表示空串，函数

|P[$\sigma\leftarrow D_{crs}$; *b*←A(ε;P)$_{[\sigma]}$: *b*=1]−P[(σ,*s*)←Sim_1(*k*); *b*←A(ε;Sim_2(*s*))$_{[\sigma]}$: *b*=1]|

总是复杂性参数 *k* 的可忽略函数。

本文需要的是非可塑零知识证明协议 NMZPoK_R=(D_{crs},P,V,Sim=(Sim_1,Sim_2),Ext=(Ext_1,Ext_2))，其中，(D_{crs},P,V,Sim=(Sim_1,Sim_2))构成一个满足以上定义的零知识证明协议；P.P.T.算法 Ext_1 和 Ext_2 统称知识提取算法，前者输出(σ,*s*,τ)(其中，*s*、τ 分别对仿真算法和提取算法起陷门作用)，后者以 $x\in$L 和(σ,τ)为其输入之一，另一输入分量是该协议的消息流，输出记作(*b*, *w*)。NMZPoK_R 定义做非可塑的，是指它进一步满足以下条件。

4) Sim_1 的第一个输出分量 $\sigma\approx\text{Ext}_1$。

注2 严格地说，这时应称该协议为“零知识 argumen”，但本文今后仍使用“零知识证明”这一词汇。

5) 对任何 τ，V 输出 $b\approx \mathrm{Ext}_2$。

6) 存在复杂性参数 k 的可忽略函数 $\eta(k)$，称为差错函数，使任何 P.P.T.算法 $A=(A_1,A_2)$有 $P[(\sigma, s, \tau)\leftarrow \mathrm{Ext}_1(k);(x,tr,(b,w))\leftarrow(<\mathrm{Sim}_2(s),A1>,<A_2,\mathrm{Ext}_2(\tau)>)_{[\sigma]}:b=1\wedge R(x,w)=1\wedge tr$ 与 $\mathrm{Sim}_2(s)$产生的任何消息流不匹配]> $P[(\sigma,s)\leftarrow \mathrm{Sim}_1(k); (x, tr, b)\leftarrow(<\mathrm{Sim}_2(s), A_1>,<A_2,V>)_{[\sigma]}:b=1\wedge tr$ 与 $\mathrm{Sim}_2(s)$产生的任何消息流不匹配]-$\eta(k)$。

回顾记号 tr 是 A_2 与 V(或 Ext_2)之间的消息流，x 是 A_2 的输出，b 是 V 的输出，(b,w)是 Ext_2 的输出。

文献[14, 15]基于所谓 simulation-sound tag-based 承诺方案和 Ω-协议给出了以上非可塑性零知识证明协议的一类高效的通用构造。为构造本文的 GUC 安全方案，需要将以上非可塑零知识证明协议的概念进一步推广到下面的带身份标识扩展的非可塑零知识证明协议(IA-NMZPoK, identity- augmented non-malleable zero-knowldege proof)。

定义 4　(R 的 IA-NMZPoK 协议) R 的 IA-NMZPoK 协议 $\mathrm{IA\text{-}NMZPoK}_R=(D, Setup,UKG,P, V, Sim=(Sim_1,Sim_2), Ext=(Ext_1, Ext_2))$ 是一组 P.P.T.算法，其中 Setup(k)生成全局公钥/私钥对 (mpk,msk)；对协议双方的身份标识 $id\in\{P,V\}$，UKG(msk,id)生成 id 的私钥 $usk(id)$；Sim_1 以 $usk(V)$为输入，Ext_1 以 $usk(P)$为输入。除 Setup 以外其他算法均以(mpk,σ)为输入之一(以下不再明确表出)。除以上特性之外，IA-NMZPoK 协议还满足前述所有性质 1)～性质 6)。

不难看出这里定义的 IA-NMZPoK 协议工作于文献[13]中建立的 ACRS(augmented c.r.s.)范型，其 ACRS 即为该协议的全局公钥 mpk。根据 ACRS 范型的定义，仅有被入侵(corrupt)的验证方 V 能运行仿真算法 Sim (其中 Sim_1 以 $usk(V)$为输入)、被入侵的证明方 P 能运行知识提取算法 Ext (其中 Ext_1 以 $usk(P)$为输入)，这是论证 GUC 安全性的需要。

3.4　非交互式身份基承诺方案

所需要的最后一项工具是最近由文献[13]定义、文献[16]进一步完善的带陷门的非交互式身份基承诺方案(IBTC, non-interactive identity-based trapdoor commitment sheme)。

定义 5　(IBTC 方案[13]) IBTC 方案 IBTC=(D, Setup, UKG, Cmt, Vf, FakeCmt, FakeDmt) 是一组 P.P.T.算法，其中 D(k) 生成身份标识 id，Setup(k) 生成全局钥/私钥对(mpk, msk)，UKG(msk,id) 计算 id 的私钥 $usk(id)$，Cmt(mpk,id,M)计算消息 M 的承诺字/验证字的对(cmt,dmt)，Vf(mpk, id, M, cmt, dmt) 输出 0 或 1，用以验证 cmt 是否为消息 M 在 id.上的承诺字。这些算法满足一致性，即对任何 M 恒有

$P[(mpk,msk)\leftarrow \mathrm{Setup}(k);(cmt,dmt)\leftarrow \mathrm{Cmt}(mpk,id,M): \mathrm{Vf}(mpk, id, M, cmt, dmt)=1]=1$。

算法 FakeCmt(mpk, id, $usk(id)$)输出($\overline{cmt}$,λ)，FakeDmt(mpk, M, $\lambda,\overline{cmt}$) 输出 $\overline{d}$ (不失一般性，总可以设 λ 含 $id||usk(id)$为其分量之一，因此今后的表达式中不再把 id 和 $usk(id)$显式表示为 FakeDmt 的输入)。IBTC 方案是安全的，是指其满足以下所有性质。

1) 隐藏性(hiding)：对任何 id 和 M_0、$M_1(cmt_i, dmt_i)\leftarrow \mathrm{Cmt}(mpk,id,M_i)$, i=0,1,有 $cmt_0\approx^{\text{P.P.T.}} cmt_1$。

2) 绑定性(binding): 对任何 P.P.T.算法 A， 函数 $Adv^{binding}_{IBTC,A}(k)\equiv P[(mpk, msk)\leftarrow \mathrm{Setup}(k); (id^*, cmt^*, M_0^*, d_0^*, M_1^*, d_1^*)\leftarrow A^{\mathrm{UKG}(msk,.)}(mpk)$: A 不向其 oracle-U(msk,.)询问 $id^*\wedge M_0^*\neq M_1^*\wedge \mathrm{Vf}(mpk, id^*, M_0^*, cmt^*, d_0^*)=\mathrm{Vf}(mpk, id^*, M_1^*, cmt^*, d_1^*)=1]$ 是复杂性参数 k 的可忽略函数。

3) 带陷门的两可性(equivocability): 对任何 P.P.T.算法 $A=(A_1,A_2)$，以下实验总使$|P[b^*=b]-1/2|$是 k 的可忽略函数：

$(mpk, msk)\leftarrow \mathrm{Setup}(k)$;
$(St, id^*, M^*)\leftarrow A_1(mpk,msk)$;
$\mathrm{usk}(id^*)\leftarrow \mathrm{UKG}(msk,id^*)$;
$(\overline{cmt},\lambda)\leftarrow \mathrm{FakeCmt}(mpk,id^*,usk(id^*))$;
$d_1\leftarrow \mathrm{FakeDmt}(mpk, M^*, \lambda, \overline{cmt})$; $d_0\leftarrow^{\$}\{0,1\}^{|d_1|}$;
$b\leftarrow^{\$}\{0,1\}$;
$b^*\leftarrow A_2(St, d_b)$;

不难看出 3)蕴涵 $P[\mathrm{Vf}(mpk,id^*, M^*, \overline{cmt}, d_1^*)=1]>1-\gamma(k)$，其中 $\gamma(k)$是 k 的可忽略函数。IBTC 方案的高效构造可参见文献[13,16]。

4　联结算子保密计算协议的通用构造

这一节建立联结算子的保密计算协议 Ψ。设 P_1、P_2 表示 Ψ 的参与方，分别持有关系 X_1 和 X_2，2 个关系分别具有属性 w, x 和 w, y，其中 w 为公共属性(与 2.1 节类似，以下将 w、x 和 y 均作为单一属性处理但所有概念和构造很容易推广到三者都是属性向量的情况)。进一步地设 $X_1=\{(u_1,x_1),\cdots,(u_{N_1}, x_{N_1})\}$、$X_2=\{(v_1,y_1),\cdots,(v_{N_2}, y_{N_2})\}$，其中，$u_i, v_j$ 是属性 w 上的值，x_i 是属性 x 上的值，y_j 是属性 y 上的

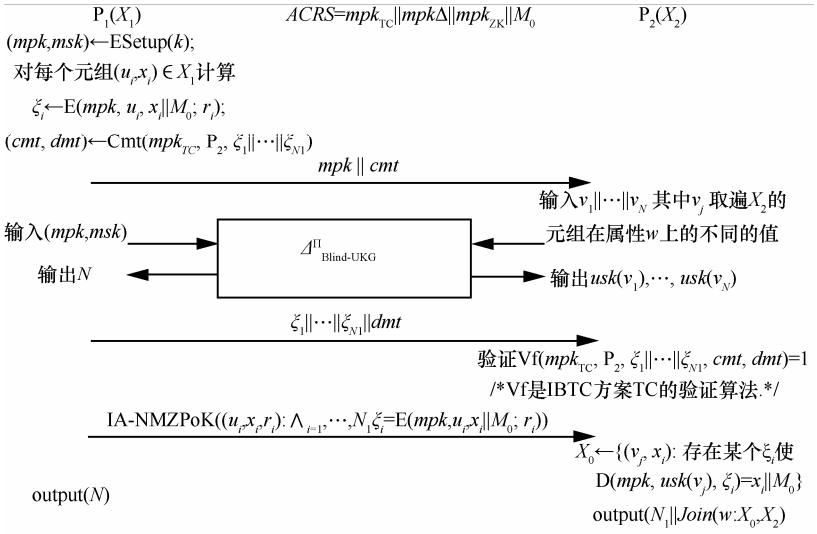

图 1 联结算子的 2-方保密计算协议

值。Π=(ESetup,UKG,E,D)是某个 ANO_CPA 匿名且 IND_CPA 保密的 IBE 方案，$\Delta^{\Pi}_{\text{Blind-UKG}}$ 表示 Π 的用户私钥盲生成协议。TC=(D,TSetup, UKG,Cmt, Vf,FakeCmt, FakeDmt)是某个 IBTC 方案(定义 5)。M_0 是一个公开给定的位串，长度为复杂性参数的多项式。表达式 IA-NMZPoK(w:$R(x,w)$=1)表示 R 的带身份标识扩展的非可塑零知识证明协议，其中，x 公开，w 是 x 的 witness。图 1 给出该过程。

4.1 联结算子的 2-方保密计算协议 Ψ

Ψ 具 ACRS 范型，其 ACRS 为 $mpk_{\text{TC}}||mpk_{\Delta}||mpk_{\text{ZK}}||M_0$，其中 mpk_{TC}、mpk_{Δ}、mpk_{ZK} 分别是 TC、$\Delta^{\Pi}_{\text{Blind-UKG}}$ 和 IA-NMZPoK 的全局公钥。*ACRS* 在协议初始化阶段被获取。在会话阶段 P_1、P_2 的计算程序如下。

1) P_1 计算 IBE 方案 Π 的全局公钥/私钥对 $(mpk,msk)\leftarrow\text{ESetup}(k)$；对每个元组 $(u_i,x_i)\in X_1$ (i=1,…,N_1)计算密文 $\xi_i\leftarrow\text{E}(mpk,u_i,x_i||M_0; r_i)$[注3]，其中，$r_i$ 表示加密过程所用的随机数且这些随机数彼此独立；接下来计算 $(cmt,dmt)\leftarrow\text{Cmt}(mpk_{\text{TC}}, P_2, \xi_1||\cdots||\xi_{N_1})$，向 P_2 发送 $mpk||cmt$。

2) P_1、P_2 运行 Π 的用户私钥盲生成协议 $\Delta^{\Pi}_{\text{Blind-UKG}}$，其中 P_1 (作为用户私钥生成者的角色)向该协议提供输入(mpk,msk)，P_2 (作为用户私钥获取者的角色)向 $\Delta^{\Pi}_{\text{Blind-UKG}}$ 提供输入 $v_1,\cdots,v_N$ (即 $v_1,\cdots,v_{N2}$ 中那些互不相同的值)。当 $\Delta^{\Pi}_{\text{Blind-UKG}}$ 终止时，P_1 取值为 N，P_2 取值为 $usk(v_1),\cdots, usk(v_N)$。

3) P_1 向 P_2 发送 $\xi_1||\cdots||\xi_{N_1}||dmt$。

4) P_2 验证 $\text{Vf}(mpk_{\text{TC}}, P_2, \xi_1||\cdots||\xi_{N_1}, cmt, dmt)$=1 是否成立，不成立则终止会话。

5) P_1、P_2 运行零知识证明协议 IA-NMZPoK $((u_i,x_i,r_i)$: $\wedge_{i=1,\cdots,N_1}\xi_i=\text{E}(mpk, u_i, x_i||M_0; r_i))$，其中 P_1 作为证明方、P_2 作为验证方。若验证成功则 P_2 用第 2)步中得到的各 $usk(v_j)$解密其在第 3)步接收到的各 ξ_i，对那些在解密结果中后缀为 M_0 的输出，即对 $\text{D}(mpk,usk(v_j),\xi_i)=x_i||M_0$ 的情形，P_2 创建一个元组 (v_j, x_i)，所有这样生成的元组构成一个集 X_0(不难看出 $X_0\subset X_1$)，最后 P_2 做普通的联结计算 $Y_0\leftarrow Join(w:X_0,X_2)$。

6) P_1 输出 N，P_2 输出 Y_0。

在证明安全性之前对协议 Ψ 做以下几点讨论。

1) 不难看出除非存在 $u_i\neq v_j$ 使 $\text{D}(mpk, usk(u_i), \xi_i)$和 $\text{D}(mpk,usk(v_j),\xi_i)$有相同的后缀 M_0，否则 P_2 的输出 Y_0 就是 $Join(w:X_1,X_2)$。运用简单的反证法程序容易证明：对任何 IND_CPA 保密的 IBE 方案存在 $u_i\neq v_j$ 使 $\text{D}(mpk, usk(u_i), \xi_i)$和 $\text{D}(mpk, usk(v_j),\xi_i)$有相同多项式长度的后缀的概率必是复杂性参数 k 的可忽略函数(小于 $Adv_{\Pi}^{IND\_CPA}(k)$)，即 P_2 输出 $Join(w:X_1,X_2)$的概率与 1 仅差一个 k 的可忽略函数。另一方面，不难给出一种计算复杂度较高的改进方案(见第 4 节)，使 P_2 恒输出 $Join(w:X_1,X_2)$并具有相同的安全性。从实用性考虑，协议 Ψ 实际上完全适用。

2) 以上对 Ψ 的构造不依赖于随机 oracle，并且

注3 与该领域中的约定类似，这里隐含的使用某种一一对应的有效明文编码函数 F，为符号简洁，今后只采用表达式 $\text{E}(mpk,u_i,x_i||M_0; r_i)$而非 $\text{E}(mpk,u_i,F(x_i||M_0); r_i)$。当随机数 r 对论证不重要时采用表达式 $\text{E}(mpk,u_i,x_i||M_0)$。

存在非随机 oracle 范型的匿名 IBE 方案[17,18]及其用户私钥盲生成协议的不依赖于随机 oracle 的实现，以及相应的零知识证明协议的不依赖于随机 oracle 的构造，因此 Ψ 具有标准范型。

3) 在 Ψ 的构造中所出现的所有子协议(IBE 的用户私钥盲生成协议和零知识证明协议)都存在 $O(1)$通信复杂度(round-complexity)的实现，因此 Ψ 具有常数轮通信复杂度。不难看出 Ψ 具有线性的消息复杂度 $O(N_1+N_2)$。在计算复杂度方面，P_1 为 $O(N_1+N_2)$次加密计算而 P_2 为 $O(N_1N_2)$次解密计算，N_1、N_2 为两方关系中的元组数量，更细致的分析依赖于该方案的具体实例，但尽管如此，注意到 $O(N_1N_2)$也是普通联结计算的渐进复杂度[1,2]，因此除一个常数因子(取决于具体的工具性密码方案)外，该方案计算复杂度与普通联结算子的渐进复杂度相当，而消息复杂度多一个 $O(N_2)$项。

4) Ψ 显然是一个所谓 $\Delta^{\Pi}_{\text{Blind-UKG}}$-hybrid 协议。Ψ 的安全证明要求 $\Delta^{\Pi}_{\text{Blind-UKG}} \rightarrow^{\text{GUC}} F^{\Pi}_{\text{Blind-UKG}}$ (参见 2.1 节和 2.2 节)，但实际上仅仅 $\Delta^{\Pi}_{\text{Blind-UKG}} \rightarrow^{\text{GUC}} F^{\Pi}_{\text{Blind-UKG}}$ 还不足以保证 Ψ 为 GUC 安全，只能保证 Ψ“半 GUC-安全”(即入侵 P_1 的攻击者能被理想协议的攻击者完全仿真但对入侵 P_2 的攻击者则不能，仅能证明 P_1 的数据对攻击者保密)。

4.2 GUC 安全性的证明

要证明完整的 GUC 安全性，需要 $\Delta^{\Pi}_{\text{Blind-UKG}}$ 具有更强的性质，为此建立以下的概念。

定义 6 (带提取算法的 IBE 用户盲私钥生成协议) 对 IBE 方案 Π=(ESetup,UKG,E,D)及其用户私钥盲生成协议 $\Delta^{\Pi}_{\text{Blind-UKG}}$，设 P_1、P_2 是 $\Delta^{\Pi}_{\text{Blind-UKG}}$ 的参与方，其中 P_2 提供用户身份公钥 a 并从 P_1 得到对应的私钥 $usk(a)$，P_1 持有全局私钥 msk 并为 P_2(盲)计算 $usk(a)$。$\Delta^{\Pi}_{\text{Blind-UKG}}$ 定义做带提取算法，如果存在一组 P.P.T.算法 Setup_Δ, UKG_Δ, $\text{Ext}_\Delta=(\text{Ext}_1,\text{Ext}_2)$ 和一个称为差错函数的可忽略函数 $\delta(k)$具有以下性质。

1) $\text{Setup}_\Delta(k)$ 生成协议的全局公钥/私钥对 (mpk_Δ,msk_Δ)。

2) $\text{UKG}_\Delta(msk_\Delta,id)$仅在 $id=P_2$(用户私钥接收方的身份标识)时生成陷门 $usk_\Delta(P_2)$，否则输出为空。

3) 对 IBE 的任何身份公钥 a、诚实的 P_1 和任何 P.P.T.算法 A，$\text{Ext}_1(usk(P_2))$输出(σ,τ)并(借用 2.3 节的记号)满足 $P[\text{Ext}_2(mpk||\tau; A(a))_{[\sigma]}=a]>P[A_a(mpk; P_1(mpk,msk))_{[\sigma]}=\text{UKG}(msk,a)]-\delta(k)$，其中，$(mpk,msk)$ 是 P_1 持有的 IBE 方案 Π 的全局公钥/私钥对(其中 mpk 公开)。

定理 1 设 IBE 方案 Π=(ESetup,UKG,E,D) IND_CPA 保密且 ANO_CPA 匿名，其用户私钥盲生成协议 $\Delta^{\Pi}_{\text{Blind-UKG}} \rightarrow^{\text{GUC}} F^{\Pi}_{\text{Blind-UKG}}$ 并满足定义 1，其提取算法和差错函数分别记做 $\text{Ext}_\Pi=(\text{Ext}_{\Pi,1},\text{Ext}_{\Pi,2})$ 和 $\delta(k)$； IA-NMZPoK$((u_i,x_i,r_i)$: $\xi_i=\text{E}(mpk, u_i, x_i||M_0; r_i)$, $i=1,\cdots,N_1)$是 IA-NMZPoK 协议；TC=(D,TSetup, UKG,Cmt,Vf,FakeCmt,FakeDmt)是安全的 IBTC 方案，则在静态入侵(static corruptions)模型中有 $\Psi \rightarrow^{\text{GUC}} F_{Join}$。

证明 设 A 表示 Ψ 的攻击者，以下对 A 入侵 P_1 和 P_2 的情形分别讨论如何构造对理想协议 F_{Join} 的仿真攻击 S^A。为叙述清晰，以下用 P_1*、P_2*表示理想协议的参与方，分别与 P_1、P_2 对应。

在协议初始化阶段所有参与方从 $\bar{G}^{Setup,UKG}_{acrs}$ 获取公开的字串 $ACRS$，其分量包括 TC 的全局公钥 mpk_{TC}、$\Delta^{\Pi}_{\text{Blind-UKG}}$ 的全局公钥 mpk_Δ、IA-NMZPoK 的全局公钥 mpk_{ZK} 和位串 M_0。该 $ACRS$ 对应的全局私钥 msk 显然是 $msk_{\text{TC}}||msk_\Delta||msk_{\text{ZK}}$，且 $\bar{G}^{Setup,UKG}_{acrs}$ 的参数算法 UKG(msk,id) 生成 $usk(id)=usk_{\text{TC}}(id)||usk_\Delta(id)||usk_{ZK}(id)$，其中 $id\in\{P_1,P_2\}$，$usk_{\text{TC}}(id)$、$usk_\Delta(id)$ 和 $usk_{\text{ZK}}(id)$ 分别是 TC、$\Delta^{\Pi}_{\text{Blind-UKG}}$ 和 IA-NMZPoK 的与 id 对应的用户私钥。

1) A 入侵 P_1: 首先在 Ψ 调用理想子协议 $F^{\Pi}_{\text{Blind-UKG}}$ 的情况下进行分析，然后应用 GUC 复合—稳定性定理完成最后的证明。设 A(即 P_1)上的关系 $X_1=\{(u_1*,x_1*),\cdots,(u_{N_1}*,x_{N_1}*)\}$，$P_2$ 上的关系 $X_2=\{(v_1*, y_1*),\cdots,(v_{N_2}*, y_{N_2}*)\}$，构造对理想模型 F_{Join} 的攻击算法 S_1，S_1 入侵 P_1*并以黑箱方式调用算法 A。S_1 仿诚实的参与方 P_2 与 A 相互作用，具体程序如下。

当收到来自 F_{Join} 的消息(sid,“input”,N_2)时，S_1 通过提交消息(“retrieve”,sid,P_1)从 $\bar{G}^{Setup,UKG}_{acrs}$ 获取 $usk(P_1)=usk_{\text{TC}}(P_1)||usk_\Delta(P_1)||usk_{\text{ZK}}(P_1)$，计算$(\sigma,s,\tau)\leftarrow$ IA-NMZPoK::$\text{Ext}_1(usk_{ZK}(P_1))$

为避免混淆这里用 $\Gamma::f$ 表示协议或方案 Γ 的算法 f，下同)，随机生成 N_2 个元组$(v_1, y_1),\cdots, (v_{N_2}, y_{N_2})$ 然后启动 A。

S_1 截获 A 发出的消息 $mpk||cmt$，以诚实参与方的角色通过理想子协议 $F^{\Pi}_{\text{Blind-UKG}}$ 与 A 相互作用，得到 $usk(v_1),\cdots,usk(v_{N2})$。

S_1 截获 A 发出的消息 $\xi_1||\cdots||\xi_{N_1}||dmt$，验证

Vf(mpk_{TC}, P_2, $\xi_1||\cdots||\xi_{N_1}$, *cmt*,*dmt*)是否为 1，作为诚实的验证方参与零知识证明协议 IA-NMZPoK((*u**$_i$, *x**$_i$, r_i): ξ_i=E(*mpk*, u_i*, *x**$_i||M_0$; r_i), *i*=1,⋯,N_1)，并在此期间调用知识提取算法 IA-NMZPoK::Ext$_2$(τ)(τ 是前面生成的陷门)提取出(*u**$_i$,*x*$_i$*,r_i), *i*=1,⋯,N_1 (以下仅用到 *u**$_i$ 和 *x*$_i$*)。

S_1 向 F_{Join} 发送消息(*sid*,“input”,{(*u**$_1$,*x*$_1$*),⋯,(u_{N_1}*,*x**$_{N_1}$)}，然后输出 A 的输出。

用 tr(*A*,S_1)表示 S_1 和 A 之间的消息流，tr$^{\psi}$(*A*, P_2(X_2))表示 A 与 P_2(X_2)(与理想参与方 P_2*有相同输入集合 X_2 的现实参与方)相互作用所生成消息流，从 A 的观点看，tr(A,S_1)和 tr$^{\psi}$(A, P_2(X_2))之间仅有的差别在于，前者对 $F^{\Pi}_{\text{Blind-UKG}}$ 提交的输入是{v_1,⋯, v_{N2}}，而后者提交的是{v_1*,⋯,v_{N2}*}，但由 $F^{\Pi}_{\text{Blind-UKG}}$ 的定义除元组数量 N_2 之外，A 对对方输入的数据集合完全不可见，因此 tr(A,S_1)≈tr$^{\psi}$(A,P_2(X_2))，特别地，A 因与 S_1 相互作用所导致的输出(在现实的协议 Ψ 中)A 因与 P_2(X_2)相互作用所导致的输出具有相同的概率分布。设 IA-NMZPoK 协议的差错函数为 η(参见定义 3 性质 6)，IBTC 方案 TC 的抗绑定优势函数为 $Adv_{TC}^{binding}$(参见定义 5 性质 2))，两者都是 *k* 的可忽略函数，由标准的反证法程序不难得出 S_1 正确提取出 A 的全部数据项(*u**$_1$, *x*$_1$*),⋯, (*u**$_{N_1}$,*x*$_{N_1}$*)的概率 ≥P[P_2(*mpk*$||\xi_1||\cdots||\xi_{N1}$;A)=1]−$N_1(\eta + Adv_{TC}^{binding})$ ≥ P[P_2 输出 *Join*(*w*:X_1,X_2)]−$N_1(\eta+ Adv_{TC}^{binding})$，从而在理想模型中，当有 S_1 攻击的情况下，P*$_2$(X_2)输出 *Join*(*w*: X_1,X_2)的概率与在协议 Ψ 中，当有 A 攻击的情况下 P_2(X_2)输出 *Join*(*w*:X_1, X_2)的概率之间至多相差 $N_1(\eta+ Adv_{TC}^{binding})$，这是复杂度参数 *k* 的一个可忽略函数。再注意到 2 种情况下 S_1 和 A 的输出相同，综合前述对任何 P.P.T.环境 Z 便有 output$_Z$(ψ,A)≈^{PPT} output$_Z$(F_{Join}, S_1)，即 Ψ→$^{GUC}F_{Join}$。

最后，若 Ψ 调用的是现实的协议 $\Delta^{\Pi}_{\text{Blind-UKG}}$，则由 $\Delta^{\Pi}_{\text{Blind-UKG}}$→$^{GUC}F^{\Pi}_{\text{Blind-UKG}}$ 及 GUC 复合—稳定性定理仍有 output$_Z$(ψ,A)≈^{PPT} output$_Z$(F_{Join},S_1)成立。不难估计出 S_1 的计算时间 $T_{S1}=T_A+O(N_2+N_1T_e)$，其中，T_A 和 T_e 分别是现实攻击者 A 和知识提取算法的计算时间。

2) A 入侵 P_2: 设 A(即 P_2)上的关系 X_2={(*v**$_1$, *y**$_1$),⋯,(*v**$_{N_2}$, *y**$_{N_2}$)}，P_1 上的关系 X_1= {(*u**$_1$, *x**$_1$),⋯, ((*u**$_{N_1}$, *x**$_{N_1}$)，构造对理想模型 F_{Join} 的攻击算法 S_2，S_2 入侵 P_2*、通过提交消息(“retrieve”,*sid*,P_2)从 $\bar{G}_{acrs}^{Setup,UKG}$ 获取 *usk*(P_2)=*usk*$_{TC}$(P_2)||*usk*$_{\Delta}$(P_2)|| *usk*$_{ZK}$(P_2)、计算(σ,*s*)←IA-NMZPoK::Sim$_1$(*usk*$_{ZK}$(P_2))并以黑箱方式调用算法 A。S_2 仿诚实的参与方 P_1 与 A 相互作用，具体程序如下。

当从 F_{Join} 收到消息(*sid*,“input”, N_1)，S_2 随机生成元组(u_1, x_1),⋯,(u_{N_1}, x_{N_1})、计算(*mpk*, *msk*)←ESetup(*k*)、对每个(u_i, x_i)计算ξ_i←E(*mpk*,u_i, $x_i||M_0$; r_i)，其中，r_i 是独立的随机数、计算(cmt^0, λ)←FakeCmt(*mpk*$_{TC}$, P_2, *usk*$_{TC}$(P_2))，启动 A 并向 A 发送消息 *mpk*||cmt^0。

S_2 以用户私钥生成者的角色与 A 运行协议 $\Delta^{\Pi}_{\text{Blind-UKG}}$ 并调用提取算法 $\Delta^{\Pi}_{\text{Blind-UKG}}$::Ext$_{\Delta}$ (*usk*$_{\Delta}$(P_2))提取出 A 的 *v**$_1$,⋯,*v**$_N$ (*N* 是互不相同的 *v**$_i$ 的个数)，随机生成 y_1,⋯, y_N，向 F_{Join} 提交消息(*sid*, “input”, P_2*, {(*v**$_1$, y_1),⋯,(*v**$_N$, y_N)})。

S_2 向 F_{Join} 发送消息(*sid*,“*Join*”,P_2*)并得到响应{(u^*_{j1}, x^*_{j1}, y_{j1}),⋯, (u^*_{jt}, x^*_{jt}, y_{jt})}(即 X_1 和 {(*v**$_1$, y_1),⋯,(*v**$_N$, y_N)}的联结，特别地，该响应表明有 *t* 个 v_j 使 $u^*_{j1}=v_{j1}$,⋯, $u^*_{jt}=v^*_{jt}$。为符号简洁记该响应为{(*u**$_1$, *x**$_1$, y_1),⋯, (*u**$_t$, *x**$_t$, y_t)}。

S_2 计算 ξ^*_i←E(*mpk*, *u**$_i$, *x**$_i||M_0$; r_i*), *i*=1,⋯, *t*，其中，r_i*为独立的随机数；从前面计算出的 ξ_i 中任选 *t* 个替换成这里的 ξ_i*，其他 N_1-t 个 ξ_i 不变，由此形成的新的密文序列记做 $\xi'_1||\cdots||\xi'_{N_1}$；计算 dmt^0← FakeDmt(*mpk*$_{TC}$, $\xi'_1||\cdots||\xi'_{N_1}$, λ, cmt^0)，向 A 发送 $\xi'_1||\cdots||\xi'_{N_1}||dmt^0$；调用零知识证明仿真算法 IA-NMZPoK::Sim$_2$($\xi'_1||\cdots||\xi'_{N_1}$, *s*)，并以此与 A 运行 IA-NMZPoK 协议(由 S_2 入侵 P_2*、ACRS 范型的定义及定义 4，S_2 这时持有 *usk*$_{ZK}$(P_2)，从而恰可以运行零知识证明的仿真算法)，其中，ξ'_i=E(*mpk*, u^0_i, $x^0_i||M_0$; r'_i)对 *t* 个 *i* 有 $u^0_i=u^*_i=v^*_i$ 和 $x^0_i=x^*_i$，对其他 N_1-t 个 *i* 有 $u^0_i=u_i$ 和 $x^0_i=x_i$。

最后 S_2 输出 A 的输出。

设 tr(S_2, *A*)表示 A 和 S_2 之间的消息流，tr$^{\Psi}$(P_1(X_1), A)表示 A 和 P_1(X_1)(与理想协议的参与方 P_1*有相同输入集合 X_1={(*u**$_1$,*x*$_1$*),⋯,(u_{N_1}*,*x**$_{N_1}$)}的现实参与方)之间的消息流。从 A 的观点看，tr(S_2, *A*)和 tr$^{\Psi}$(P_1(X_1), A)之间的差别在于：1) 2 个消息流中的 *cmt* 分量分别是由 FakeCmt 生成的 cmt^0 和 Cmt(*mpk*$_{TC}$, P_2, E(*mpk*, u_1*, x_1*$||M_0$; r_1)||⋯||E(*mpk*, *u**$_{N1}$, *x**$_{N1}||M_0$; r_{N1}))生成的 *cmt*；2) *dmt* 分量分别是由 FakeDmt 生成的 dmt^0 和 Cmt(*mpk*$_{TC}$, P_2, E(*mpk*, u_1*, x_1*$||M_0$; r_1)||⋯ ||E(*mpk*, *u**$_{N_1}$, *x**$_{N_1}||M_0$; r_{N_1}))生成的 *dmt*；3) 2 个消息流的密文分量$\xi_1||\cdots||\xi_{N1}$ 中有 *t* 个 ξ_i 彼此具有相同的

公钥 u^*_i 和明文 $x^*_i||M_0$，其余 N_1-t 个密文则具有不同的公钥和明文；4) 在 2 个消息流中的 IA-NMZPoK 零知识证明协议的消息分量中有 t 个 witness 含相同的 u^0_i 和 x^0_i。

由 IBTC 方案 TC 的带陷门的两性质立得 $\mathrm{tr}(S_2,A)$和 $\mathrm{tr}^{\Psi}(P_1(X_1),A)$中的($cmt$,$dmt$)之间 P.P.T.-难辨。由 IBE 方案的 ANO_CPA 匿名性质和 IND_CPA 保密性质得 $\mathrm{tr}(S_2,A)$和 $\mathrm{tr}^{\Psi}(P_1(X_1),A)$中的 $\xi_1||\cdots||\xi_{N1}$ 之间 P.P.T.-难辨(否则存在 P.P.T.算法以概率 $\rho\geqslant 1/poly(k)$分辨两者，由标准的反证法程序容易导出 $Adv_{\Pi}^{ANO\_CPA}(k)\geqslant\rho/N_1$ 或 $Adv_{\Pi}^{IND\_CPA}(k)\geqslant\rho/N_1$ 至少有一个成立，从而与匿名性或保密性相矛盾)。现在将 $\mathrm{tr}(S_2,A)$和 $\mathrm{tr}^{\Psi}(P_1(X_1),A)$中的消息分量 $\xi_1||\cdots||\xi_{N1}$ 分别记为 $\xi_1^{(1)}||\cdots||\xi_{N_1}^{(1)}$和 $\xi_1^{(2)}||\cdots||\xi_{N_1}^{(2)}$，将两者之中属于零知识证明协议 IA-NMZPoK 的消息流分别记做 $\text{IA-NMZPoK}^{(1)}(=\mathrm{tr}_{S2,A}(mpk||M_0||\ \xi_1^{(1)}||\cdots||\xi_{N_1}^{(1)}))$和 $\text{IA-NMZPoK}^{(2)}(=\mathrm{tr}_{P1,A}(mpk||M_0\ ||\xi_1^{(2)}||\cdots||\xi_{N_1}^{(2)})))$，由以上分析已有 $\xi_1^{(1)}||\cdots||\xi_{N_1}^{(1)}\approx^{\mathrm{PPT}}\xi_1^{(2)}||\cdots||\xi_{N_1}^{(2)}$；进一步由 IA-NMZPoK 的零知识性质有

$$\text{IA-NMZPoK}^{(2)}\approx^{\mathrm{PPT}}\text{IA-NMZPoK}::\mathrm{Sim}_2(\xi_1^{(2)}||\ \cdots||\xi_{N_1}^{(2)},s)$$

由 S_2 的构造则有

$$\text{IA-NMZPoK}^{(1)}=\text{IA-NMZPoK}::\mathrm{Sim}_2(\xi_1^{(1)}||\cdots||\xi_{N_1}^{(1)},s),$$

于是 $\text{IA-NMZPoK}^{(1)}\approx^{\mathrm{PPT}}\text{IA-NMZPoK}^{(2)}$。

由此，在 $\mathrm{tr}(S_2,A)$和 $\mathrm{tr}^{\Psi}(P_1(X_1),A)$中 A 所接收到的消息分量彼此之间 P.P.T.-难辨。

设 δ 表示 $\Delta^{\Pi}_{\text{Blind-UKG}}$ 提取算法的差错函数(它是 k 的可忽略函数)，显然 S_2 正确提取出 A 的一个数据项 v^*_i 的概率不小于 $\mathrm{P}[A(mpk;P_1(mpk,msk))=\mathrm{UKG}(msk,v^*_i)]-\delta$，因此 S_2 正确提取出 A 的数据项 $v^*_1,\cdots,v^*_N$ 的概率不小于 $\mathrm{P}[A(mpk;P_1(mpk,\ msk))=\mathrm{UKG}(msk,\ v^*_i):\ i=1,\cdots,N]-N_2\delta\geqslant\mathrm{P}[P_2$ 输出 $\mathrm{Join}(w:X_1,X_2)]-N\delta$，从而对环境 Z 而言 S_2 攻击理想协议情况下的输出与 A 攻击协议 Ψ 的情况下输出之间 P.P.T.难辨，概率误差至多为 $N_1(k)(\ Adv_{\Pi}^{ANO\_CPA}(k)+Adv_{\Pi}^{IND\_CPA}(k)\ )+N_2\delta\ (N\leqslant N_2)$，结果也是复杂度参数 k 的可忽略函数。再注意到 2 种情况下另一方 $P_1^*(X_1)$和 $P_1(X_1)$总是输出 N_2，便得结论 $\mathrm{output}_Z(\psi,\ A)\approx^{\mathrm{PPT}}\mathrm{output}_Z\ (F_{Join},S_2)$，且不难估计出 S_2 的计算时间 $T_{S2}=T_A+O(N_1+N_2T_{ext})$，其中 T_A 和 T_{ext} 分别是攻击者 A 和提取算法 Ext_{Δ} 的计算时间。

综上所述，便有 $\Psi\rightarrow^{\mathrm{GUC}}F_{Join}$，即 Ψ 与理想协议 F_{Join} GUC 相似。证毕。

5　改进的协议 Ψ^+

Ψ^+仅在以下方面与 Ψ 不同：除构造 Ψ 所基于的那些密码工具外，Ψ^+还利用一个基于身份的 UF_CMA 抗伪造的数字签名方案 Sig=(KG, Sign, Vf)[19]，Sig 的签字公钥即为参与方的身份标识 P_i，对应的签字私钥记做 $sk(P_i)$(实际上仅用到 i=1)。在计算程序上，Ψ^+在以下两步与 Ψ 不同。

1) P_1 计算 IBE 方案 Π 的全局公钥/私钥对(mpk, msk)←ESetup(k)；对每个元组$(u_i,x_i)\in X_1$ (i=1,⋯,N_1)计算 $\xi_i^{(1)}\leftarrow\mathrm{E}(mpk,u_i,M_0)$、$\xi_i^{(2)}\leftarrow\mathrm{E}(mpk,\ u_i,x_i)$、$\sigma_i\leftarrow\mathrm{Sign}(sk(P_1),\xi_i^{(1)}||\xi_i^{(2)})$、$\xi_i\leftarrow\xi_i^{(1)}||\xi_i^{(2)}||\sigma_i$；接下来计算($cmt$, dmt)←$\mathrm{Cmt}(mpk_{TC},\ P_2,\ \xi_1||\cdots||\xi_{N_1})$并向 P_2 发送 $mpk||cmt$。

2) P_1、P_2 运行零知识证明协议 IA-NMZPoK $((u_i,x_i,r_i):\xi_i=\xi_i^{(1)}||\xi_i^{(2)}||\sigma_i,\ i=1,\cdots,N_1)$，其中，$\xi_i^{(1)}$、$\xi_i^{(2)}$、$\sigma_i$ 如第 1)步中所述，$r_i=r_i^{(1)}||r_i^{(2)}$是计算 $\xi_i^{(1)}$、$\xi_i^{(2)}$时用到的随机数，P_1 作为证明方、P_2 作为验证方，包括验证各 σ_i 确是 P_1 对 $\xi_i^{(1)}||\xi_i^{(2)}$的数字签名。若验证成功则 P_2 用前面第 2)步中得到的各 $usk(v_j)$解密 $\xi_i^{(1)}$，若得到 M_0 则进一步用 $usk(v_j)$解密 $\xi_i^{(2)}$，并将所得到的 x_i 连同 v_j 创建一个元组(v_j, x_i)，所有这样生成的元组构成一个集合 $X_0(X_0\subset X_1)$，最后 P_2 做普通的联结计算 $Y_0\leftarrow Join(w:X_0,X_2)$。

Ψ^+的其他步骤及初始化阶段与 Ψ 相同。

不难看出 Ψ^+的输出恒为所要求的 $Join(w:X_1,X_2)$，并且 Ψ^+的计算复杂性及消息复杂性的渐进形式与 Ψ 相同但具有更大的常数因子，除此而外其他特点与 Ψ 完全相同。运用完全类似于定理 1 的证明方法和数字签名方案的 UF_CMA 抗伪造性质，容易证明定理 1 对 Ψ^+仍成立。

6　结束语

本文建立了分布式数据库系统中广泛应用的关系联结算子的 2-方保密计算协议，基于保密及匿名的 IBE 方案给出一种通用的有效构造并证明该构造具有 GUC 安全性。本文工作与关于其他关系算子的现有工作结合起来[20,21]，将给出分布式关系数据库系统保密计算的完整、可证明的理论解决方案。

另一方面，本文的工作还有以下诸方面值得继续完善：1) 如何有效推广到任意多方的保密计算情形(这也是目前大部分工作需要进一步解决的问题)；2) 如果能将通用构造中的 IA-NMZPoK 协议

替换为非交互形式，则有可能进一步提高计算及通信效率，在这方面文献[22,23]给出了有意义的参照，但目前尚不足以直接加以应用；3) 如果能将 IBE 方案目前的匿名性质实质性地扩展到针对 UKG 的匿名性并具体构造出满足该性质的保密的 IBE 方案，则本文的通用构造将不再需要用户私钥盲生成协议，从而大大减小通信复杂度和计算复杂度，在这方面文献[23]给出了有启发意义的进展但仍然不足以直接应用(注意在那里较强的安全结论依赖于较复杂的难解性假设和随机 oracle)；4) 应用本文方法及最近建立的匿名 ABE 方案构造带约束条件的联结算子保密计算协议。

参考文献：

[1] GARCIA-MOLINA H, ULLMAN J D, WIDOM J. Database System Implementation[M]. Prentice Hall, 2000.

[2] ÖZSU M T, VALDURIEZ P. Principles of Distributed Database Systems[M]. Springer, 2011.

[3] GREEN M, HOHENBERGER S. Blind identity-based encryption and simulatable oblivious transfer[A]. Asiacrypt'07[C]. 2007. 265-282.

[4] SHI E, BETHENCOURT J, CHAN T H H, *et al*. Multi-dimensional range query over encrypted data[A]. IEEE Symposium on Security and Privacy[C]. 2007. 350-364.

[5] BONEH D, WATERS B. Conjunctive, subset, and range queries on encrypted data[A]. Theory of Cryptography[C]. Springer Berlin Heidelberg, 2007.535-554.

[6] ABDALLA M, BELLARE M, CATALANO D, *et al*. Searchable encryption revisited: Consistency properties, relation to anonymous IBE, and extensions[J]. Journal of Cryptology, 2008, 21(3): 350-391.

[7] BLAKE I F, KOLESNIKOV V. Conditional encrypted mapping and comparing encrypted numbers[A]. Proceedings of the 10th international conference on Financial Cryptography and Data Security[C]. Springer-Verlag, 2006.206-220.

[8] HAZAY C, LINDELL Y. Efficient protocols for set intersection and pattern matching with security against malicious and covert adver-saries[A]. Theory of Cryptography[C]. Springer Berlin Heidelberg, 2008.155-175.

[9] KISSNER L, SONG D. Privacy-preserving set operations[A]. Cryptology–CRYPTO 2005[C]. Springer Berlin Heidelberg, 2005.241-257.

[10] LINDELL Y, PINKAS B. Privacy preserving data mining[J]. Journal of Cryptology, 2002, 15(3): 177-206.

[11] GOLDREICH O. Foundations of Cryptography: Volume 2, Basic Applications[M]. Cambridge University Press, 2009.

[12] CANETTI R. Universally composable security: a new paradigm for cryptographic protocols[A]. IEEE Symposium on Foundations of Computer Science IEEE[C]. 2001.136-145.

[13] CANETTI R, DODIS Y, PASS R, *et al*. Universally composable security with global setup[A]. TCC'07[C]. 2007.61-85.

[14] GARAY J A, MACKENZIE P, Yang K. Strengthening zero-knowledge protocols using signatures[A]. Advances in Cryptology—Eurocrypt 2003[C]. Springer Berlin Heidelberg, 2003.177-194.

[15] MACKENZIE P, YANG K. On simulation-sound trapdoor commitments[A]. Cryptology-EUROCRYPT 2004[C]. Springer Berlin Heidelberg, 2004.382-400.

[16] DODIS Y, SHOUP V, WALFISH S. Efficient constructions of composable commitments and zero-knowledge proofs[A]. Proc Crypt'08[C]. 2008.515-535.

[17] GENTRY C. Practical identity-based encryption without random oracles[A]. Eurocrypt'06[C]. 2006.445-464.

[18] BOYEN X, WATERS B. Anonymous hierarchical identity-based encryption (without random oracles)[A]. Crypto'06[C].2006. 290-307.

[19] BELLARE M, NAMPREMPRE C, NEVEN G. Security proofs for identity-based identification and signature schemes[J]. Journal of Cryptology, 2009, 22(1): 1-61.

[20] 李顺东, 王道顺. 基于同态加密的高效多方保密计算[J]. 电子学报, 2013, 41(4): 798-803.
LI S D, WANG D H. Efficient secure multiparty computation based on homomorphic[J]. Encryption Acta Electronica Sinica 2013, 41(4): 798-803.

[21] 夏峰, 杨波, 张明武等. 基于 LWE 的集合相交和相等的两方保密计算[J]. 电子与信息学报, 2012, 34(2): 462-467.
XIA F, YANG B, ZHANG M W, *et al*. Secure two-party computation for set intersection and set equality problems based on LWE[J]. Journal of Electronics & Information Technology,2012, 34(2): 462-467.

[22] GROTH J, SAHAI A. Efficient non-interactive proof systems for bilinear groups[A]. Proc Eurocrypt[C]. 2008.415-432.

[23] 田有亮, 彭长根, 马建峰等. 通用可组合公平安全多方计算协议[J]. 通信学报, 2014, 35(2): 54-62.
TIAN Y L, PENG C G, MA J F, *et al*. Universally composable secure multiparty computation protocol with fairness[J]. Journal on Communications, 2014, 35(2): 54-62.

[24] IZABACHENE M, POINTCHEVAL D. New anonymity notions for identity-based encryption[A]. Security and Cryptography for Networks[C]. Springer Berlin Heidelberg, 2008.375-391.

作者简介：

田园（1966-），男，湖北武汉人，大连理工大学副教授，主要研究方向为多方保密计算、多用户信息论。

孙荣辛（1991-），男，吉林公主岭人，大连理工大学硕士生，主要研究方向为格密码学、多方保密协议。

蔡悟洋（1991-），男，陕西安康人，大连理工大学硕士生，主要研究方向为云计算安全技术、MIMO 技术。