数据库应用多方安全计算的理论探究

作者、摘要、关键词、英文题名、英文摘要、英文关键词、正文、参考文献、附录和致谢

摘要：当前数据信息价值通过挖掘和分析而不断地被人们所重视，因此作为数据信息的载体，对数据库资源进行整合并对多数据库进行联合计算将会是未来信息发展的重点。本文通过数据扰乱、OT不经意传输和解决当数据库群体面临资源共享需求时进行共同计算并同时保护己方隐私问题，

介绍：

多方安全计算概念的提出源于1982年YAO A.C在23届IEEE上做出的关于《Protocols for secure computations》的报告[1]。随后在1987年由O.Goldreich、S.M icali和A.Wigderson三位学者提出了密码学安全的可以计算任意函数的多方安全计算协议。自此之后，多方安全计算协议又得到了诸多有益的改进，在现实生活中的应用也愈加成熟。在实现基本加法、乘法运算的基础上，在几何计算、隐私保护、统计分析方面都有着广泛的应用，并且形成了电子投票、群包含验证等应用协议。

下文将从多方安全计算的理论基础和基本算法入手，介绍经典的双方求和安全计算协议和双方比较安全计算协议，并推广为多方安全计算协议，其中将详细阐明所涉及的一些重要方法，并且我们将多方求和协议应用在数据库信息查询中，实现多数据库之间的共享计算。

背景：

信息技术的快速发展使得信息数据库作为21世纪新的重要资源而得到了人们的重视，每个数据库中含有的信息就是数据库的核心所在，也是数据拥有者的资本。但单个数据库的存储量终究受限于存储条件、成本、信息源。因此信息资源共享成为新技术的必要手段。但是各个数据库的拥有者都要求本身私有信息的不可泄露或知识版权等问题，在资源共享过程中完成公共查询要求并不泄露自身的保密数据，这正是安全多方计算绝佳的应用场景。

数据库资源共享的需求，使得实现数据库间的安全多方计算的要求应运而生。

OT不经意传输

零知识证明[]

零知识证明是指知识的掌握者通过交互或不交互的方式使验证者相信自己拥有知识，并不能使验证者获取自己的知识信息。

以下为QR零知识证明系统的运作过程：

步骤1：Alice随机选择整数r，并将发送给Bob；

步骤2：Bob在{0、1}中随机选择数字i，发送给Alice；

步骤3：Alice计算,发送给Bob；

步骤4：Bob验证成立；

以上步骤重复m次，验证通过即Bob相信Alice中x为n的QR。

一份零知识证明协议具有完全性、合理性和零知识性，完全性是指如果证明结果命题为真，则验证方概率为1的接受证明成立；合理性是指如果证明结果命题为假，则验证方承认证明结果为真的概率最多为；零知识性是指，证明过程没有泄露包括知识在内的额外知识。零知识性的证明过程即：

通过构造模拟机M，其不知道是否x为n的QR，但是通过随机产生的方式参与到证明体系中来，模拟Alice完成验证过程，在m次步骤中产生的数据和真实协议中产生的数据统计不可区分，则认为零知识性成立。

据以上知识，2001年Du W L和Atallah M J在（Computer Security Foundations Symposium）报告中提出了以下双方安全线性方程组计算协议，满足Alice和Bob在不暴露自身隐私数据的情况下，计算线性方程组（M1+M2）x=V1+V2。

核心构建：

在隐私数据拥有者Di彼此之间可分布式交换流量的（i>2）环境下，实现数据交换并计算所拥有隐私数据矩阵mi之和的多方安全求和计算协议：

步骤1：各节点Di拥有隐私数据矩阵mi，矩阵mi规模m\*n，，协商参数p，使得穷搜2^p个矩阵之和是不可能的；

步骤1.1：各节点生成矩阵子列满足；

步骤1.2：各节点生成随机矩阵列，满足 ；

步骤2：首节点生成扩展矩阵组，其中；而为同规格随机矩阵，标记矩阵记录扩展矩阵分组中所在位置；

步骤2.1：首节点将扩展矩阵发送给下一节点，节点对扩展矩阵、矩阵子列和随机矩阵列进行矩阵加法运算；

步骤2.2：节点将从节点收到的扩展矩阵和本节点矩阵子列、随机矩阵按步骤2.1进行矩阵加法运算；

步骤3：首节点收到遍历参与节点后的扩展矩阵列，通过自身矩阵还原原矩阵；

步骤3.1：首节点依次接收节点的随机矩阵，并计算；

步骤4：公布求和结果。

安全度分析：

分析：

对数据库中信息的数据特征进行保护的技术有很多，就目前已经做的工作来说，针对不同方面大致从以下几个角度出发：

1. 数据扰乱技术[1]。数据库中的数据多为真实有效的数据，可以直接反映出数据对象的全部特征，但在范围性查询过程中，往往只需要提取数据整体的某些数据特征，因此通过向数据添加噪声来增加扰乱，使得扰乱后的数据和原始数据的某些基本性质保持不变，而其他信息特征的失真使得攻击者无法重构原始数据。

但数据扰动技术的缺点在于使用者必须提前知晓应用的全部特征结构，同时需要对某次查询的目标特征十分清楚，保证产生的噪音不会影响某些不易察觉的数字特征，并且针对每个应用的具体特征的差异，每次不同的应用都需要重建。显然不利于多样环境的推广。

2. 数据加密技术[2]：一些加密算法的优良特性使得对加密过的数据的计算不会影响原始数据的计算结果，如RSA加密算法、ElGamal同态加密方案等算法的同态加密特性可以满足对数据的乘积运算同态。

但数据加密技术的缺点在于运算过程时间长、消耗大，对大数据库之间的交互计算来说，是不现实的。在通常的应用中，哪怕是简单的加密通信，也常使用对称加密算法进行通讯，而只使用非对称加密算法进行密钥协商。

相较而言，通过多方安全计算算法所实现的数据库数据传输过程。对比数据扰乱技术，适用性强，可针对不同环境下的数据实体进行操作，计算的结果与对数据原型直接计算的结果一致，区别只在于计算过程不直接暴露数字实体，因此计算结果不会出现偏差，并且计算过程中不需要针对某一数据特征对数据进行修改，因而不会影响对结果的再分析；对比密码加密技术，不需要对选取加密算法的特征有特别的要求，加密过程不涉及或少涉及对数据的数学性操作，多为加减简单运算，考虑到数据库操作多为大数据，计算总量巨大，对提供多用户查询的数据库来说可以有效削减加密方式所带来的负载过重的问题。

应用:

参考文献：

[1] YAO A C. Protocols for secure computations. In: Proceedings of 23th Annual IEEE Symposium on Foundations of Computer Science. Chicago, USA, 1982.160-164.

[1] Du W L, Atallah M J., Privacy-preserving cooperative scientific computations,In CSFW 2001，Nova Scotia, Canada 2001,pp.273-282.

[]Quisquater, Jean-Jacques; Guillou, Louis C.; Berson, Thomas A. (1990). "How to Explain Zero-Knowledge Protocols to Your Children". Advances in Cryptology - CRYPTO '89: Proceedings 435: 628–631.

[1] 周水庚, 李丰, 陶宇飞, 肖小奎. 面向数据库应用的隐私保护研究综述[J]. 计算机学报, 2009, 32(5): 847-861

[] 安全多方计算及其应用研究[D]. 北京邮电大学:孙茂华, 2013. 46-51

[] 1.冯国登. Cryptography: Theory and Practice, Third Edition][M]. 加拿大:Douglas R. Stinson, 2008. 157-158