基于单断言的安全的密文区间检索

蔡 克 张 敏 冯登国

(中国科学院软件研究所信息安全国家重点实验室 北京 100190)

摘 要 为解决数据外包所带来的敏感数据的数据安全问题,数据所有者更多地选择外包敏感数据的密文.而外包密文的形式,为数据所有者对这些数据的使用带来了不便,如数据所有者无法对密文数据进行区间检索等.目前的密文区间检索方案中,为实现密文的区间检索,服务器需要对区间索引进行多次断言.而断言次数越多,向服务器泄露的信息也越多.文中提出采用单断言实现敏感数据的区间判断,同时使用可逆矩阵对区间索引和区间陷门进行安全保护,不仅减少了整个方案的信息泄露,而且保证了区间索引和区间陷门的安全.文中对方案复杂性进行了分析.该方案在安全性方面的提升并不以效率损失为代价.

关键词 密文检索;区间检索;区间索引;区间陷门

中图法分类号 TP309 **DOI 号**: 10.3724/SP.J.1016.2011.02093

Secure Range Query with Single Assertion on Encrypted Data

CAI Ke ZHANG Min FENG Deng-Guo

(*State Key Laboratory of Information Security, Institute of Software, Chinese Academy of Sciences, Beijing* 100190)

Abstract The outsourcing of sensitive data will bring in some security issues to the data. In order to avoid it, the data owner prefers to encrypt the sensitive data before outsourcing them. Thought the encryption can protect this confidentiality of sensitive data, it is inconvenient for the data owner to use them. For example, the data owner cannot directly make a range query on these encrypted sensitive data any more. To solve it, we usually build interval index for every sensitive data. To realizing whether a sensitive data belongs to a search range, in the previous solutions, the server needs to assert the interval index many times. The more times the server asserts the interval index, the more information that will be revealed to the server. These revealed information risk the confidentiality sensitive data. In this paper, we propose a schema of range query on encrypted data, which reduces the amount of information leakage by only asserting the interval index once. This revealed information such as interval indexes and interval trapdoors, are multiplied with invertible matrix to guarantee the security issues of the sensitive data and the search range. The analyses of complicity can tell the truth, that this schema achieves high level needs of security without losing of efficiency.

Keywords search on encrypted data; range query; interval index; interval trapdoor

1 引 言

外包数据库(Outsourced Database,ODB)和云存储(Cloud Storage)是当前的两个主流数据管理形式.由于数据的外包将数据完全暴露给半可信的服务提供商,使得数据脱离了数据所有者的掌控,数据隐私存在着被泄露的可能.Google、Salesforce 等

收稿日期:2011-08-29;最终修改稿收到日期:2011-09-15.本课题得到核高基重大专项课题(2010ZX01042-001-001-05)、中国科学院知识创新工程领域前沿项目-云计算安全支撑系统(YYYJ-1013)资助.**蔡 克**,男,1986 年生,博士研究生,主要研究方向为数据安全和密文检索.E-mail: caik@is.iscas.ac.cn.**张 敏**,女,1975 年生,博士,副研究员,主要研究方向为数据安全和隐私保护.**冯登国**,男,1965 年生,博士,研究员,主要研究领域为网络与信息系统安全、密码理论与技术.

公司均发生过云存储泄密事件.为保护敏感数据,数据所有者期望在对敏感数据加密后外包存储,使得暴露给服务提供商的是密文形式的数据[1].

外包数据的密文存储,为数据使用带来了诸多不便,如无法对敏感数据密文进行区间检索.通常采用通用的加密算法对敏感数据进行加密保护.我们假设加密算法是安全的.服务器在不了解密钥的情况下,无法直接对密文进行区间检索.如果数据所有者想要获得属于指定区间的数据,其必须向服务器请求全部敏感数据,在本地解密后再进行区间检索.整个过程需要服务器返回大量冗余数据,占用了大量网络带宽和客户端计算资源,这并非是有效的解决方案.

鉴于直接对密文进行区间检索比较困难,目前普遍采用为敏感数据构建区间索引的方法实现区间检索.区间索引基于敏感数据明文构建,在一定程度上包含了敏感数据的相关信息.如果区间索引不安全,那么区间索引的外包,大大增大了服务器破解敏感数据的概率.在对数据安全要求比较高的场景下,不仅需要敏感数据可用,同时还要考虑其安全.应该在保证安全性的前提下提升敏感数据的可用性.在客户端进行区间检索的解决方案[2]不符合数据外包应用场景的要求,不在本文考虑范围之内.

早期对密文区间检索的研究主要集中于保证敏感数据的可用性.最简单的是直接应用保序加密算法[3-4].区间索引继承了敏感数据的大小排列性质,服务器可以直接基于密文进行大小比较,从而获得检索结果.该类方案使得敏感数据的大小信息完全暴露给服务器,安全性方面的牺牲比较大.Hacigümüş 等人提出通过对敏感数据进行分桶[5-8]的方式实现密文区间检索,各区间索引组成了对敏感数据的一次划分.这个方案中,服务器可以直接获得敏感数据的归并特征,即多个敏感数据对应同一个区间索引.归并特征数量少的特点,使得破解难度很低.而且归并特征的破解对获得敏感数据的分布规律非常有利.后来,有学者注意到了密文区间检索方案在安全性方面的缺陷,提出了通过编码[9]的方式来实现密文区间检索.该方案通过对敏感数据进行前缀保序编码获得区间索引,使用前缀匹配获得区间检索结果.安全性有所提升,但是在获知多个区间陷门的情况下,服务器可以获得敏感数据的排列信息[9].

现有方案中,要实现密文区间检索,服务器均需要对区间索引进行多次断言.每次断言都意味着服务器要进行一次判断,如等值匹配判断或大小比较判断等.任何一次断言都会向服务器泄露敏感数据的相关信息,如分桶[1-3]的方案中,检索区间涵盖多个分桶,检索时需要将区间索引与区间陷门的每个值进行一次断言,泄露了敏感数据属于其中某个分桶的信息,而不是只泄露敏感数据是否属于检索区间的信息.所以断言次数越多,检索过程向服务器泄露的信息也就越多,泄露信息的增多,增大了密文区间检索方案的安全隐患.

本文针对上述的问题,第一个提出基于单断言来实现密文的区间检索,即服务器只需要对区间索引进行一次断言,就可以获知敏感数据是否属于该检索区间.这在一定程度上减少了区间检索过程中的信息泄露,使得密文区间检索方案的安全性得到了一定程度的保障.对于必须提供给服务器的信息,比如区间索引和区间陷门,方案对这些信息进行了相应的保护处理.该方案中,敏感数据的排列信息和敏感数据的归并信息不会被泄露,与现有方案相比,它在不影响敏感数据可用性的情况下,获得了更高的安全性.

本文在第 2 节中对方案的安全模型进行假设,同时针对密文区间检索的特征,给出了相关的安全性描述的定义;第 3 节分析方案的基本框架,并描述了区间检索方案中涉及的一些方法;第 4 节对本方案中的区间索引和区间陷门的安全性进行分析;第 5 节对本方案的时间和空间复杂性进行分析,并与现有方案进行比较;第 6 节总结全文并给出对后续工作的展望.

2 相关知识

2.1 应用场景假设

本文考虑的应用场景包含如下两种角色:数据所有者(Data Owner,DO)和服务提供商(Service Provider,SP).如图 1 所示,其中 DO 是敏感数据的拥有者,该角色不具备强大的数据存储和管理能力.为获得针对敏感数据密文的区间检索能力,该角色在将密文数据外包的同时,为敏感数据构建了区间索引并进行上传.SP 是服务提供商,具有强大的存储和计算能力.对于 DO 发起的区间检索请求,SP 承担大量的处理任务,并向 DO 返回检索结果.

SP 是半可信的,即 SP 会忠实地执行 DO 提交

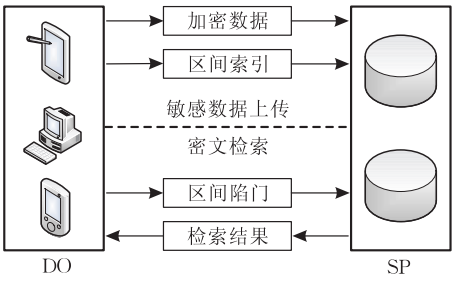

图 1　基本应用场景

的区间检索请求，并返回根据检索请求得到的检索结果. 但是 SP 对敏感数据比较好奇，可能会利用所掌握的一切信息（如敏感数据的区间索引和区间陷门等）来进行分析，期望获得真实的敏感数据.

当前我们主要考虑唯密文模型，即 SP 只了解存放在 SP 的敏感数据的区间索引以及 DO 进行区间检索时提供的区间陷门信息. SP 事先并不了解敏感数据的值域、分布情况、敏感数据与索引的对应关系等信息. 所以，本文需要考虑唯密文模型下的区间索引和区间陷门的安全性.

2.2　安全性定义

我们主要研究唯密文模型下的区间索引和区间陷门的安全. 现有研究中涉及索引和陷门安全的概念主要有：索引安全、陷门安全和关键词安全[10-11]. 索引安全是在密文文档的关键词检索中提出的，考虑的是关键词索引是否会泄露关键词与加密文档之间的对应关系；关键词安全考虑的是检索关键词，要求检索陷门不能泄露检索关键词；而陷门安全要求敌手无法根据已知的陷门生成未知的有效陷门.

上述概念都是针对密文检索领域的关键词检索提出的，比较适合于用于描述关键词检索中的索引和陷门的安全. 由于这些概念不能完全覆盖常见的区间检索中的安全隐患，所以这些概念在区间检索领域的使用具有一定的局限性. 密文区间检索拥有下述特征. 一方面区间检索不再是简单的单值匹配，需要考虑区间陷门是否会泄露检索区间上下界的信息，以及是否会泄露检索区间的归并特征；另一方面，敏感数据之间还附带了具有某种实际意义的大小关系，使得返回的检索结果不再只包含单值敏感数据. 所以我们不仅需要考虑区间索引是否会泄露敏感数据的排列信息，而且需要考虑在检索过程是否会泄露检索结果中各个敏感数据的排列信息.

本文针对密文区间检索的独特安全隐患，提出了相关的安全性描述. 它主要分区间索引和区间陷门两个方面. 其一，区间索引是否会泄露敏感数据的排列信息；其二，区间陷门是否会泄露检索区间的归并特征，是否会泄露检索结果中敏感数据的排列信息.

定义 1（区间索引的排列安全）.　已知一个密文区间检索方案的区间索引集合 $I=\{I_1,I_2,\cdots,d_n\}$ 和函数 $F(d,k)$，敏感数据集合 $D=\{d_1,d_2,\cdots,d_n\}$ 和密钥 k 均未知，其中 $I_i=F(d_i,k)$，$i\in[1,n]$，根据 I 和 F 如果能得到如下结果

（1）$I_j \nrightarrow d_j$，其中 $j\in[1,n]$；

（2）$P(d_i>d_j)=0.5$，其中 $i\in[1,n]$，$j\in[1,n]$，$i\neq j$；

（3）$P(d_i>d_j \mid d_k>d_l)=P(d_i>d_j)$，其中 $i\in[1,n]$，$j\in[1,n]$，$i\neq j$，且 $\{i,j\}\neq\{k,l\}$，

那么我们称该区间索引是排列安全的.

其要求：（1）敌手无法根据区间索引直接获得真实的敏感数据；（2）无法根据区间索引获得任意两个敏感数据之间的大小关系，即敏感数据的大小关系不可区分；（3）即使敌手获得某两个敏感数据之间的大小关系，剩下的敏感数据的之间的大小关系仍然是不可区分的，即无法判断其余任意两个敏感数据之间的大小关系. 所以，如果区间索引是排列不安全的，那么敌手可以获得所有敏感数据的一个排列，排列可能是升序的也可能是降序的.

敏感数据排列信息的泄露，使敌手在了解所有敏感数据分布信息的情况下，可以方便地获得真实的敏感数据.

定义 2（区间陷门的归并特征安全）.　已知一个密文区间检索方案中区间陷门 $T(T_1,T_2,\cdots,T_t)$，如果

（1）$T \nrightarrow H-L$，其中 L 和 H 分别是区间陷门 T 对应的检索区间 $[L,H]$ 的下界和上界；

（2）$L,H,K\rightarrow T_i$，其中 $i\in[1,t]$，K 为密钥，

那么我们称该区间陷门是归并特征安全的.

区间陷门的归并特征安全是指区间陷门只能由区间上下界生成，并且不会泄露检索区间大小. 所以，敌手无法根据区间陷门获得检索区间真实的上下界以及检索区间的大小，并且无法获得检索区间内敏感数据的归并特征.

密文的区间检索不同于关键词检索，区间检索涉及到区间陷门是否泄露敏感数据的归并特征的问题. 由于每个归并特征均表示一组敏感数据，每次检索也是基于归并特征进行的. SP 可以通过多个区间陷门获得归并特征排列信息，这对获得真实的敏感数据非常有利.

定理 1. 设存在 n 个互不相等的数字，只需要 $n-1$ 个特定的区间检索就可以获得这 n 个数字的排列.

证明. 设 n 个数字为 $A=\{a_1,a_2,\cdots,a_n\}$，存在的 $n-1$ 个区间检索 $Q=\{q_1,q_2,\cdots,q_{n-1}\}$ 对应的检索结果为 $R=\{r_1,r_2,\cdots,r_{n-1}\}$. 其中 q_i 的检索结果 $r_i=\{a_i,a_{i+1}\mid 1\leq i\leq n\}$. 那么，根据 r_1,r_2 可以得到 $\{a_1<a_i\mid i=2,3\}$ 或 $\{a_1>a_i\mid i=2,3\}$，根据 r_2,r_3 可以得到 $\{a_2<a_i\mid i=3,4\}$ 或 $\{a_2>a_i\mid i=3,4\}$，…. 而且如果 $a_1<a_2$，那么 $\{a_i<a_j\mid i<j\}$. 否则 $\{a_i>a_j\mid i<j\}$. 根据这些可以得出 $a_1,a_2,\cdots,a_n$ 是以升序或降序排列的信息. 所以，只需要 $n-1$ 个特定的区间检索就可以获得 n 个数字的排列信息. 证毕.

由定理 1 可以看到，假如 n 是指敏感数据个数，且 n 比较大，那么这时需要大量的特定区间检索才能让 SP 获得敏感数据的排列信息. 而当 n 是指归并特征个数时，由于每个归并特征代表了一组敏感数据，敏感数据比较多时，对应的归并特征也不会很多，即 n 比较小. 这时根据少量的区间检索就可以获得归并特征的排列信息. 根据归并特征的排列信息可以获得敏感数据的分布情况. 这对于获得真实的敏感数据非常有利. 所以，一个区间检索方案的区间陷门满足归并特征安全是非常有必要的.

定义 3(区间陷门的检索结果安全). 已知一个区间陷门 T 和基于该陷门进行检索的检索结果集 $R=\{r_1,r_2,\cdots,r_s\}$，而检索结果集对应的敏感数据集为 $D=\{d_1,d_2,\cdots,d_s\}$，如果根据 T 和 R 可以得到

(1) $P(d_i>d_j)=0.5$，其中 $i\in[1,n],j\in[1,n],i\neq j$；

(2) $P(d_k>d_l\mid d_i>d_j)=P(d_k>d_l)$，其中 $i\in[1,n],j\in[1,n],i\neq j$，且 $\{i,j\}\neq\{k,l\}$，

那么我们称区间陷门是检索结果安全的.

区间陷门的检索结果安全是指区间陷门和检索结果集不会泄露敏感数据的大小关系，而且这个判断不受其它敏感数据的大小关系影响.

密文的区间检索不同于关键词检索，区间检索中，各检索结果之间具有大小关系. 如果区间检索的结果之间可以简单地判断出大小关系，那么只要获得多个特定区间陷门，敌手就可以获得所有敏感数据的顺序信息. 所以，区间陷门的检索结果安全对于区间检索的安全性非常重要.

定义 4(区间检索的唯密文安全). 如果一个密文区间检索方案在唯密文场景下，其区间陷门是排列安全的，区间陷门是归并特征安全和检索结果安全的，那么称该方案是唯密文安全的.

由上述可以看出，无论是区间陷门的排列安全，还是区间陷门的归并特征安全和检索结果安全，均可以归约为尽可能地保护敏感数据的大小关系. 由于密文区间检索与关键词检索相比，其特征在于敏感数据之间是具有某种实际意义的大小关系的. 所以区间检索的唯密文安全的描述重心就是从区间索引和区间陷门方面考虑，尽可能保护敏感数据的大小关系信息不被敌手获得.

2.3 单断言实现区间判断

当前的密文区间检索方案都是基于多次断言实现的. 桶式索引的方案[5]，检索时需要对检索区间中各个预定义的桶标识进行断言；编码索引方案[9]，检索时需要对检索区间包含的所有最小前缀编码进行断言；而保序加密[1-2]，需要对索引分别进行区间上下界两次断言判断. 所以，当前并不存在基于单断言实现密文的区间检索方案.

本节将介绍一种可以将区间匹配转换成单断言判断的方法. 该断言将一个被断言值和一个断言区间作为输入，服务器进行相关计算和单次比较判断获得断言结果. 其中被断言值即为敏感数据的区间索引，断言区间即为检索区间对应的区间陷门，断言结果就是敏感数据是否属于检索区间. 该方法是基于单位圆的不同半径的位置关系转化而来的.

定理 2. 由图 2，已知在一个原点为 O 的坐标系中存在一个单位圆，上半圆周上有 A、B、C 3 个不同的点，OA 和 OB，OB 和 OC，OA 和 OC 之间的夹角分别为 $\theta_1,\theta_2,\theta_3$，其中 $0<\theta_i<\pi,i=1,2,3$. 那么当且仅当 $\cos\theta_3<\cos\theta_1\cos\theta_2$ 时，OB 位于 OA 和 OC 之间(仅考虑上半圆).

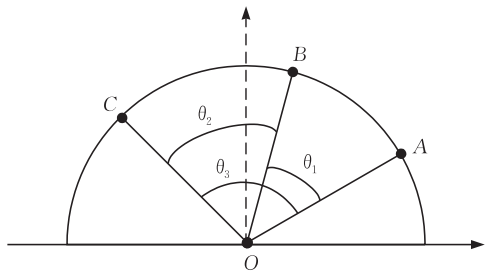

图 2 单位圆不同半径位置关系

证明. 由于 $\theta_1\in(0,\pi)$ 且 $\theta_2\in(0,\pi)$，那么 $\sin\theta_1>0,\sin\theta_2>0$，即 $\sin\theta_1\sin\theta_2>0$.

由于 A、B、C 是 3 个不同的点，OA、OB、OC 并不会重叠，即 $\cos\theta_3\neq\cos\theta_1\cos\theta_2$.

不妨设 OB 位于 OA、OC 之间，那么 $\theta_3=\theta_1+$

θ_2，$\cos\theta_3=\cos(\theta_1+\theta_2)=\cos\theta_1\cos\theta_2-\sin\theta_1\sin\theta_2$，由 $\sin\theta_1\sin\theta_2>0$，得 $\cos\theta_3<\cos\theta_1\cos\theta_2$.

当 OB 位于 OA、OC 之外时，$\theta_3=|\theta_1-\theta_2|$，$\cos\theta_3=\cos(\theta_1-\theta_2)=\cos\theta_1\cos\theta_2+\sin\theta_1\sin\theta_2$，由 $\sin\theta_1\sin\theta_2>0$，得 $\cos\theta_3>\cos\theta_1\cos\theta_2$.

所以，当且仅当 $\cos\theta_3<\cos\theta_1\cos\theta_2$ 时，OB 位于 OA 和 OC 之间.　　证毕.

由定理 2，断言 $\cos\theta_3<\cos\theta_1\cos\theta_2$ 可以用来表示半径线段 OB 位于线段 OA、OC 之间，OB 的位置看作是要被检索的值，而 OA 和 OC 的位置看作为检索区间，断言 $\cos\theta_3<\cos\theta_1\cos\theta_2$ 可以表示一个区间判断. 为更贴切地表达被检索值和检索区间的关系，我们将 3 个线段之间的位置关系用线段与横轴正半轴的夹角来表示.

推论 1.　已知在一个原点为 O 的坐标系中存在一个圆，上半圆周上有 V、V_L、V_H 3 个不同的点，其 OV、OV_L、OV_H 与横坐标的夹角分别是 θ、θ_L、θ_H，其中 $0<\theta,\theta_L,\theta_H<\pi$. 那么当且仅当 $\cos(\theta_H-\theta_L)<\cos(\theta-\theta_L)\cos(\theta-\theta_H)$ 时，$\theta_L<\theta<\theta_H$.

证明.　由于 OV 与 OV_L、OV 与 OV_H、OV_L 与 OV_H 的夹角分别是 $\theta-\theta_L$，$\theta-\theta_H$，$\theta_H-\theta_L$. 由定理 2 得，当且仅当 $\cos(\theta_H-\theta_L)<\cos(\theta-\theta_L)\cos(\theta-\theta_H)$ 时，OB 位于 OA 和 OC 之间，即 $\theta_L<\theta<\theta_H$.　证毕.

3　方　案

3.1　基本框架

一个唯密文安全的密文区间检索方案，不仅要求构造的区间索引是排列安全的，而且构造的陷门是归并特征安全和检索结果安全的. 本文将给出一种将区间判断转换成单断言的方法，并基于该方法构建一个安全的密文区间检索方案.

密文的区间检索的方案构建与通用的密文检索方案[12]类似，涉及的两个流程分别是构建区间索引和区间检索. 对这两个流程进行分解，可以获得如下几个过程：

1. Setup. 初始化并生成一个密钥 K.

2. BuildIndex(v,K). DO 根据敏感数据 v 生成与其对应的索引，并使用密钥 K 进行加密保护，得到区间索引 I.

3. Trapdoor(L,H,K). DO 根据检索区间上下界 L 和 H，生成与之对应的陷门，并使用密钥 K 进行加密保护，得到区间陷门 T.

4. Search(I,T). SP 根据区间索引 I 和区间陷门 T 进行单断言判断，获知区间索引对应的敏感数据是否满足检索区间.

构建具体方案时，首先 DO 为外包的敏感数据构建索引，涉及 Setup 和 BuildIndex 两个过程. 在初始化加密密钥后，对每个敏感数据进行区间索引的构建，并将这些区间索引上传到 SP. 然后，DO 发起基于 $[L,H]$ 的区间检索，其涉及 Trapdoor 和 Search 两个过程，即 DO 在基于 L 和 H 构建了区间陷门后，将区间陷门提交给 SP 进行检索. SP 对区间索引和区间陷门进行判断，获得被检索的敏感数据是否属于检索区间的信息.

3.2　区间判断到单断言的映射

上一节介绍了一种可以获得区间判断效果的断言，我们虽然可以通过单个断言 $\cos(\theta_H-\theta_L)<\cos(\theta-\theta_L)\cos(\theta-\theta_H)$ 来表示一个区间判断 $\theta_L<\theta<\theta_H$，但是由推论 1 知该区间判断的数值存在着大小的限制. 由定理 2 可以知道，被判断的值只能属于 $(0,\pi)$. 对于大部分的应用场景，$(0,\pi)$ 无法表示敏感数据的值域，所以需要一种将值域 D 的数值映射到 0 到 π 之间的方法，我们设计了如下的值的圆周映射方法.

值的圆周映射(F). 假设映射关系原像的值域为 D，二维坐标系中存在一个半径为 D 的上半圆，在横轴找到映射值 v，取其与 O 相对称的 $-v$，并将其投影到上半圆周获得点 A，那么 OA 与 x 正半轴的夹角 θ 就为 v 的映射值.

如图 3，$\cos\theta=\frac{-v}{D}$，那么 $F(v)=\theta=\arccos\frac{-v}{D}$，其中 v 是要进行映射的值，D 是 v 所在值域的大小.

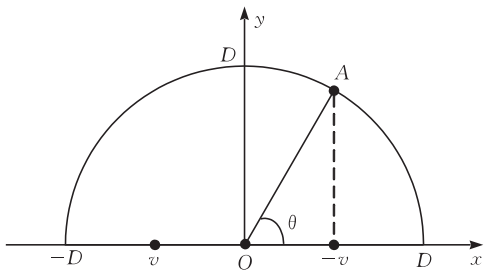

图 3　值的圆周映射

将值域为 D 的数值映射到 0 到 π 之间的方法有很多. 我们这里要求的方法是在映射后，断言 $\cos(\theta_H-\theta_L)<\cos(\theta-\theta_L)\cos(\theta-\theta_H)$ 仍然可以表示区间判断 $\theta_L<\theta<\theta_H$，其中 $0<\theta,\theta_L,\theta_H<\pi$，即该映射必须维持断言对于区间关系的判断. 所以，我们定义了区间关系维持映射来表述这样的映射.

定义 5(区间关系维持映射).　一个映射 F，对于任意区间 $[L,H]$ 和值 v，至少存在一个断言 $AS_{F(L)-F(H)}(F(v))$ 与该区间相对应，使得只有当 $L<v<H$ 时，断言 $AS_{F(L)-F(H)}(F(v))$ 才为真，那么

该映射 F 被称为是区间关系维持映射.

区间关系维持映射表述了,数据进行映射之后,仍然存在一个可以表示区间关系的判断的断言.

定理 3. 值的圆周映射是一个区间关系维持映射.

证明. 如果存在 $v_1>v_2$,由于值的圆周映射 $\theta=\arccos\frac{-v}{D}$ 在 $v\in[-D,D]$ 时是单调递增的,那么 v_1,v_2 经过映射后,对应的映射值 θ_1 和 θ_2 满足 $\theta_1>\theta_2$,其中 $\theta_1\in[0,\pi]$,且 $\theta_2\in[0,\pi]$.

假设存在 3 个值 v,v_L,v_H,其中 $v_L<v<v_H$,那么对应的映射值分别为 θ,θ_L,θ_H,且满足关系 $\theta_L<\theta<\theta_H$.

由推论 1 可知,当且仅当 $\cos(\theta_H-\theta_L)<\cos(\theta-\theta_L)\cos(\theta-\theta_H)$ 时,满足 $\theta_L<\theta<\theta_H$,即 $v_L<v<v_H$. 所以,区间判断可以由断言 AS:$\cos(\theta_H-\theta_L)<\cos(\theta-\theta_L)\cos(\theta-\theta_H)$ 来表示.

那么对于值的圆周映射而言,任意区间 $[v_L,v_H]$ 和值 v,对应的映射值分别为 $[\theta_L,\theta_H]$ 和 θ. 必然存在与区间 $[\theta_L,\theta_H]$ 对应的断言 AS:$\cos(\theta_H-\theta_L)<\cos(\theta-\theta_L)\cos(\theta-\theta_H)$,使得只有当 $v_L<v<v_L$ 时,断言 AS 才为真. 所以,值的圆周映射是一个区间关系维持映射,其对应的断言为 $\cos(\theta_H-\theta_L)<\cos(\theta-\theta_L)\cos(\theta-\theta_H)$. 证毕.

3.3 映射值的加密保护

单断言 $\cos(\theta_H-\theta_L)<\cos(\theta-\theta_L)\cos(\theta-\theta_H)$ 可以实现区间判断. 如果直接应用于区间检索方案中,虽然可以达到区间检索的目的,但无法满足区间索引安全的要求. SP 进行断言判断时,需要 DO 向其提供敏感数据的映射值,SP 可以利用映射值,方便地获得真实的敏感数据. 所以,需要对上述断言进行相应的保护,使其满足安全性的要求.

由于在断言 AS 中,θ_H 和 θ_L 除了大小不同外,完全等价,下面以 $\cos(\theta-\theta_L)$ 作为例子进行分析,$\cos(\theta-\theta_L)$ 的变换与之类似.

通过如下公式进行 θ 和 θ_L 的分离,并通过可逆矩阵 $\boldsymbol{M}$ 对分解的矩阵进行保护[12].

$$\begin{aligned}\cos(\theta-\theta_L)&=\cos\theta\cos\theta_L+\sin\theta\sin\theta_L\\&=[\cos\theta\quad\sin\theta]\begin{bmatrix}\cos\theta_L\\\sin\theta_L\end{bmatrix}\\&=[\cos\theta\quad\sin\theta]\boldsymbol{M}\boldsymbol{M}^{-1}\begin{bmatrix}\cos\theta_L\\\sin\theta_L\end{bmatrix}.\end{aligned}$$

同理得

$$\cos(\theta-\theta_H)=[\cos\theta\quad\sin\theta]\boldsymbol{M}\boldsymbol{M}^{-1}\begin{bmatrix}\cos\theta_H\\\sin\theta_H\end{bmatrix}.$$

上述将 $\cos(\theta-\theta_L)$ 分解成 $[\cos\theta\quad\sin\theta]\boldsymbol{M}$ 和 $\boldsymbol{M}^{-1}\begin{bmatrix}\cos\theta_L\\\sin\theta_L\end{bmatrix}$,不仅实现了角度 θ 和 θ_L 的分离,而且实现了对两者的加密保护. 矩阵加密的形式可以简化为 $\boldsymbol{C}=\boldsymbol{P}\times\boldsymbol{M}$,其中 $\boldsymbol{C}$ 为密文,$\boldsymbol{P}=[\cos\theta\quad\sin\theta]$,$\boldsymbol{M}$ 为 2×2 的可逆矩阵,在本文假设的唯密文场景下,如果敌手不知道 $\boldsymbol{M}$,那么其无法恢复出 $\boldsymbol{P}$ 的值[12]. 所以在本文的场景下,我们可以认为矩阵加密是安全的. 上述对 θ 和 θ_L 的分离,使得断言 AS 可以适用于密文区间检索的应用场景,可以将 $[\cos\theta\quad\sin\theta]\boldsymbol{M}$ 作为区间索引,而矩阵加密的应用,使得在唯密文的场景下,$[\cos\theta\quad\sin\theta]\boldsymbol{M}$ 的泄露不会影响到真实的敏感数据的安全.

3.4 安全的密文区间检索方案

在本节中,我们结合上述的区间关系维持映射和加密保护来构建一个基于单断言的安全的密文区间检索方案(SRQSAE 方案). 该方案在唯密文的场景下,区间索引本身不会直接泄露真实的敏感数据,不会泄露敏感数据的大小关系,而且检索区间的陷门可以维护敏感数据的大小关系. 其主要包含了如下几个步骤:

1. Setup. 生成一个 2×2 的可逆矩阵 $\boldsymbol{M}$,计算获得 $\boldsymbol{M}^{-1}$,得到相关密钥 $K=\{\boldsymbol{M},\boldsymbol{M}^{-1}\}$.

2. BuildIndex(v,K). 对 v 进行值的圆周映射,获得映射值 θ 及 $[\cos\theta\quad\sin\theta]$,然后进行矩阵的乘法运算,获得 v 对应的区间索引 $\boldsymbol{I}=[\cos\theta\quad\sin\theta]\boldsymbol{M}$.

3. Trapdoor(L,H,K). 将检索区间的上下界分别进行值的圆周映射,获得映射值 θ_L 和 θ_H,再进行矩阵乘法运算,获得陷门 $\boldsymbol{T}_L=\boldsymbol{M}^{-1}\begin{bmatrix}\cos\theta_L\\\sin\theta_L\end{bmatrix}$ 和 $\boldsymbol{T}_H=\boldsymbol{M}^{-1}\begin{bmatrix}\cos\theta_H\\\sin\theta_H\end{bmatrix}$,同时计算陷门 $T_{\text{range}}=\cos(\theta_H-\theta_L)$,最后,区间陷门 $T=\{\boldsymbol{T}_1,\boldsymbol{T}_2,T_{\text{range}}\}$. 其中,陷门 $\boldsymbol{T}_1$、$\boldsymbol{T}_2$ 并不是固定地分别对应着 $\boldsymbol{T}_L$、$\boldsymbol{T}_H$,其对应关系是随机的,当 $\boldsymbol{T}_1$ 对应 $\boldsymbol{T}_L$ 时,$\boldsymbol{T}_2$ 就对应着 $\boldsymbol{T}_H$,否则相反.

4. Search($\boldsymbol{I},T$). 进行单断言 $T_{\text{range}}<(\boldsymbol{I}\cdot\boldsymbol{T}_1)\cdot(\boldsymbol{I}\cdot\boldsymbol{T}_2)$ 判断,如果是真,则返回 yes,否则返回 no.

该方案的主要步骤与一般的密文检索方案[12-14]类似,DO 首先需要初始化一个可逆的矩阵 $\boldsymbol{M}$ 作为区间索引和区间陷门的加密密钥,在外包敏感数据时,通过值的圆周映射和矩阵加密来生成区间索引 $\boldsymbol{I}$. 当 DO 进行检索时,对检索区间的上下界同样进行值的圆周映射和矩阵加密来生成相应的区间陷门 T,并提交给 SP,SP 将每个区间索引 $\boldsymbol{I}$ 都与

区间陷门 T 进行单断言判断，返回判断结果.

4　安全性分析

密文区间检索方案构建的目的是在保证敏感数据机密性的前提下，实现区间检索的目的. 其方案的构建需要考虑在假设的场景下区间索引、区间陷门是否满足安全性的需求. 当前我们假设的场景是唯密文场景，SP 不了解敏感数据的任何信息，但可以获得 DO 的多个检索区间陷门，以期望分析出敏感数据的分布特征或大小排列信息等. 本节我们将对区间索引和区间陷门的安全分别进行分析，验证该方案满足我们假定的安全性的需要.

4.1　区间索引安全

由于本文只考虑唯密文的场景，敌手对敏感数据的分布情况、敏感数据与区间索引的对应关系都不了解. 所以进行区间索引安全考虑时，我们并不需要考虑统计分析攻击、明密文攻击等，只需要考虑区间索引是否会泄露真实的敏感数据，区间索引是否会泄露敏感数据的大小关系信息. 后者的安全性要求比前者要求更高，因为单纯从区间索引的角度，区间索引如果泄露了真实敏感数据，那么敏感数据的大小关系必然泄露，而如果区间索引泄露了敏感数据大小关系，那么真实的敏感数据并不一定会泄露. 所以，对区间索引来说，保护敏感数据大小排列信息的安全性要求更高.

由于敏感数据在经过值的圆周映射后，通过矩阵相乘的方式来生成区间索引，区间索引的安全性依赖于矩阵相乘的安全性. 在上文可以看到，在唯密文的场景下，只要不泄露矩阵相乘中的可逆矩阵，那么敏感数据的机密性可以得到保证，敌手无法根据区间索引获得具体的敏感数据. 所以方案 SRQSAE 中，区间索引不会泄露具体的敏感数据.

定理 4.　已知公式 $\boldsymbol{C}=\boldsymbol{P}\times\boldsymbol{M}$，其中 $\boldsymbol{P}$ 是 1×2 的矩阵 $[\cos\theta \quad \sin\theta]$，$\theta\in(0,\pi)$，$\boldsymbol{M}$ 是 2×2 的可逆的未知矩阵，给定任意两个不同 $\boldsymbol{C}_1$ 和 $\boldsymbol{C}_2$，假设存在与之对应的 θ_1 和 θ_2，那么 θ_1 和 θ_2 的大小关系是不可区分的.

证明.　在可逆矩阵 $\boldsymbol{M}$ 未知的情况下，给定 $\boldsymbol{C}$，必然存在多个可能的 $\boldsymbol{P}$ 满足公式 $\boldsymbol{C}=\boldsymbol{P}\times\boldsymbol{M}$，由于 $\boldsymbol{P}=[\cos\theta \quad \sin\theta]$，所以存在多个可能的 θ. 而对于 $\boldsymbol{C}_1$ 和 $\boldsymbol{C}_2$，在 $\boldsymbol{M}$ 未知的情况下，也可能存在多对 θ_1 和 θ_2 满足上述公式.

对于任意一对满足 $\boldsymbol{C}_1$ 和 $\boldsymbol{C}_2$ 的 θ_1 和 θ_2，可以记为元组 $A=\{\theta_1,\theta_2,\boldsymbol{M}\}$，其中 $[\cos\theta_1 \quad \sin\theta_1]\boldsymbol{M}=\boldsymbol{C}_1$ 且 $[\cos\theta_2 \quad \sin\theta_2]\boldsymbol{M}=\boldsymbol{C}_2$.

如果对于任意满足 $\boldsymbol{C}_1$ 和 $\boldsymbol{C}_2$ 要求的 $A'=\{\theta_1',\theta_2',\boldsymbol{M}\}$，其中 $\theta_1'>\theta_2'$，均存在另一个同意满足 $\boldsymbol{C}_1$ 和 $\boldsymbol{C}_2$ 的元组 $A''=\{\theta_1'',\theta_2'',\boldsymbol{M}\}$，其中 θ_1'' 和 θ_2'' 满足 $\theta_1''<\theta_2''$，那么在 θ_1 和 θ_2 值未知的情况下，即使其满足 $\boldsymbol{C}_1$ 和 $\boldsymbol{C}_2$，θ_1 和 θ_2 的大小关系是不可区分的，因为无法判断 $\theta_1=\theta_1'$，$\theta_2=\theta_2'$ 还是 $\theta_1=\theta_1''$，$\theta_2=\theta_2''$. 下面来证明对于任意的元组 A' 必然存在元组 A'' 与之对应.

不妨设存在满足要求的元组 $A'=\{\theta_1',\theta_2',\boldsymbol{M}'\}$，其对应的可逆矩阵为 $\boldsymbol{M}'=\begin{bmatrix}a & b\\ c & d\end{bmatrix}$. 那么 $\boldsymbol{P}_1'=[\cos\theta_1' \quad \sin\theta_1']$，根据公式可以得到

$$\boldsymbol{C}_1'=\boldsymbol{P}_1'\times\boldsymbol{M}'$$
$$=[a\times\cos\theta_1'+c\times\sin\theta_1' \quad b\times\cos\theta_1'+d\times\sin\theta_1'],$$

同理可得

$$\boldsymbol{P}_2'=[\cos\theta_2' \quad \sin\theta_2'],$$
$$\boldsymbol{C}_2'=[a\times\cos\theta_2'+c\times\sin\theta_2' \quad b\times\cos\theta_2'+d\times\sin\theta_2'].$$

那么对应的，存在元组 $A''=\{\theta_1'',\theta_2'',\boldsymbol{M}''\}$，其中 $\theta_1''=\pi-\theta_1'$，$\theta_2''=\pi-\theta_2'$，矩阵 $\boldsymbol{M}''=\begin{bmatrix}-a & -b\\ c & d\end{bmatrix}$，则 $\boldsymbol{P}_1''=[\cos\theta_1'' \quad \sin\theta_1'']=[\cos(\pi-\theta_1') \quad \sin(\pi-\theta_1')]=[-\cos\theta_1' \quad \sin\theta_1']$，由公式 $\boldsymbol{C}_1''=\boldsymbol{P}_1''\times\boldsymbol{M}''$ 可得

$$\boldsymbol{C}_1''=[-\cos\theta_1' \quad \sin\theta_1']\begin{bmatrix}-a & -b\\ c & d\end{bmatrix}$$
$$=[a\times\cos\theta_1'+c\times\sin\theta_1' \quad b\times\cos\theta_1'+d\times\sin\theta_1']=\boldsymbol{C}_1,$$

那么同理可以得到 $\boldsymbol{P}_2''=[-\cos\theta_2' \quad \sin\theta_2']$，$\boldsymbol{C}_2''=\boldsymbol{C}_2$. 即元组 $A''=\{\theta_1'',\theta_2'',\boldsymbol{M}''\}$ 同样满足 $\boldsymbol{C}_1$ 和 $\boldsymbol{C}_2$ 的要求.

由上可以看出，对于任意的满足要求的元组 A'，必然存在对应的另一个满足条件的元组 A''. 如果 $\theta_1'>\theta_2'$，那么必然可以得到 $\pi-\theta_1'<\pi-\theta_2'$，而 $\pi-\theta_1'$ 和 $\pi-\theta_2'$，即为 A'' 中对应的角度值 θ_1'' 和 θ_2''.

对于任意满足要求的 $\theta_1'>\theta_2'$，必然存在另一对满足要求的 $\theta_1''<\theta_2''$. 所以，θ_1 和 θ_2 的大小是不可区分的.　　证毕.

由定理 4 的证明过程可知，对于任意两个区间索引，其对应的敏感数据的大小关系是不可区分的. 上述过程中，我们为一对已知敏感数据找出另一对具有相同区间索引但大小相反的敏感数据. 所以，在 M 安全的情况下，该证明过程不受其它敏感数据大小关系的影响，即敏感数据的大小不可区分性不受其它敏感数据大小关系的影响. 而且由于该区间索引是通过矩阵相乘来生成的，在唯密文的场景下，不会泄露真实的敏感数据. 所以方案 SRQSAE 中的区

间索引达到了排列安全的要求.

4.2 区间陷门安全

对于区间陷门的安全性,不仅仅需要考虑区间陷门是否会泄露区间上下界,而且需要考虑区间陷门是否泄露归并特征的信息以及检索结果之间的大小关系信息.本文提出的方案中区间陷门是满足这些要求的,即满足定义 2 和 3 所说的区间陷门的归并特征安全和检索结果安全的要求.下面将从这两个方面进行分析.

定理 5. 已知存在两个值v_1,v_2,映射关系$F(v)=\arccos\frac{-v}{D}$,那么根据三角计算$E=\cos(F(v_1)-F(v_2))$的计算值,敌手无法判断$|v_1-v_2|$值的大小.

证明. 不妨设 v_1,v_2 的三角计算结果 E 为 $value$,即 $value=\cos(F(v_1)-F(v_2))$.

由 $value=\cos(F(v_1)-F(v_2))$ 可以得到,$value=\cos\left(\left(\frac{\pi}{2}-F(v_2)\right)-\left(\frac{\pi}{2}-F(v_1)\right)\right)$,那么不妨设存在另外两个值 v_1',v_2',使得 $\arccos\frac{-v_1'}{D}=\frac{\pi}{2}-F(v_1)$,$\arccos\frac{-v_2'}{D}=\frac{\pi}{2}-F(v_2)$,那么它们对应的三角计算 E 的结果

$$\begin{aligned}\cos(F(v_1')-F(v_2'))&=\cos\left(\arccos\frac{-v_1'}{D}-\arccos\frac{-v_2'}{D}\right)\\&=\cos\left(\frac{\pi}{2}-F(v_1)-\left(\frac{\pi}{2}-F(v_2)\right)\right)\\&=\cos(F(v_1)-F(v_2))=value.\end{aligned}$$

所以,至少存在两对满足三角计算结果为 E 的 v_1,v_2,如果$|v_1-v_2|\neq|v_1'-v_2'|$,那么我们可以认为敌手无法根据 $\cos(F(v_1)-F(v_2))$ 判断 $|v_1-v_2|$ 的值.

由$D\sin(F(v))=D\sqrt{1-\cos^2F(v)}=\sqrt{D^2-v^2}$得

$$|v_1'-v_2'|=\left|D\cos\left(\frac{\pi}{2}-F(v_2)\right)-D\cos\left(\frac{\pi}{2}-F(v_1)\right)\right|=$$

$$|D\sin(F(v_2))-D\sin(F(v_1))|=|\sqrt{D^2-v_2^2}-\sqrt{D^2-v_1^2}|.$$

那么,当 v_1,v_2 取某个固定值,值域 D 的大小发生变化时,$|v_1'-v_2'|$ 的值也随着变化,所以$|v_1-v_2|\neq|v_1'-v_2'|$.

因此,敌手无法根据 $\cos(F(v_1)-F(v_2))$ 判断 $|v_1-v_2|$ 的大小. 证毕.

由上述定理 5,如果 v_1,v_2 为某个检索区间的上下界时,那么向敌手泄露三角计算 E:$\cos(F(v_1)-F(v_2))$,并不影响检索区间的安全性,因为敌手无法判断 $|v_1-v_2|$ 的大小,即敌手无法根据三角计算 E 来获得检索区间的大小.

定理 6. 方案 SRQSAE 中的区间陷门是归并特征安全的.

在方案 SRQSAE 中,检索区间的陷门为 $T=\left\{\boldsymbol{M}^{-1}\begin{bmatrix}\cos\theta_L\\\sin\theta_L\end{bmatrix},\boldsymbol{M}^{-1}\begin{bmatrix}\cos\theta_H\\\sin\theta_H\end{bmatrix},\cos(\theta_H-\theta_L)\right\}$或$T=\left\{\boldsymbol{M}^{-1}\begin{bmatrix}\cos\theta_H\\\sin\theta_H\end{bmatrix},\boldsymbol{M}^{-1}\begin{bmatrix}\cos\theta_L\\\sin\theta_L\end{bmatrix},\cos(\theta_H-\theta_L)\right\}$. 根据矩阵相乘的分析,我们可以知道,在唯密文的情况下,根据陷门 SP 无法分析出真实的检索区间上下界.且根据定理 5,SP 无法根据陷门计算出检索区间的大小.所以,SP 无法根据检索区间的真实上下界或区间大小去分析该区间内的归并特征.同时该区间陷门是由检索区间的上下界而非归并特征计算生成,所以,该方案中的区间陷门不会泄露检索区间任何归并特征,其区间陷门是归并特征安全的.

定理 7. 假设由 SP 进行断言 SA':$value<value1\times value2$ 的判断,DO 选择 θ_H,θ_L,θ,计算出 $value,value1,value2$,并提交给 SP,其中 $value=\cos(\theta_H-\theta_L)$,$value1$ 和 $value2$ 分别对应 $\cos(\theta-\theta_L)$和 $\cos(\theta-\theta_H)$中的一个.那么在任意两次断言 SA'为真的情况下,如果 DO 选择的是相同的 θ_H 和 θ_L,不同的 θ,那么 SP 无法判断 DO 在这两次断言中所选择的 θ 的大小关系.

证明. 由于判断相同的 θ_H 和 θ_L 情况下,两次断言为真时,两个 θ 的大小不可区分,所以下面的分析都是针对固定的 θ_H 和 θ_L 进行的,由于固定的 θ_H 和 θ_L 对应相同的 $value$,下文分析时不再考虑 $value$.

我们使用断言元组 B 来表示某次断言判断过程中,SP 获得的数据,$B=\{value_1,value_2\}$.由于 DO 在提交数据给 SP 时,$value_1$ 和 $value_2$ 的顺序是随机的,所以对于 $B_1=\{value_1,value_2\}$ 和 $B_2=\{value_2,value_1\}$,SP 认为其是无差别的,即 B_1 等价于 B_2.断言元组 B 中的数据前后顺序的变化不影响 B.由于在一次断言中,SP 进行断言数据都包含于断言元组 B 中.所以,如果两次断言中的断言元组 B 相同,SP 无法判断 DO 选择的 θ 是否不同.

我们使用断言判断元组 A 来表示某次断言为真的判断中涉及的相关数据,即 $A=\{\theta,B\}$,其中 θ 为 DO 选择的数值,其对 SP 保密,B 为断言元组.那么 $A_1=\{\theta_1,B\}$ 和 $A_2=\{\theta_2,B\}$ 是两次不同的断言,因为 DO 选择了不同的 θ,但生成了等价的断言元组 B.对于 SP 而言,A_1 和 A_2 两个断言没有任何区别,是不可区分的,我们可以称 A_2 为 A_1 的 SP 无差别的

不同断言判断元组. 即虽然本质上这两次断言是针对不同的 θ 的进行的，但是由于 DO 提交给 SP 进行断言的数据集合完全相同，SP 无法判断其是否是针对不同的 θ 进行的断言.

对于 DO 选择的数值 θ，由于 θ 和 $\theta_H+\theta_L-\theta$ 生成相同的断言元组 B，即 $\cos((\theta_H+\theta_L-\theta)-\theta_L)=\cos(\theta_H-\theta)=\cos(\theta-\theta_H)$，$\cos((\theta_H+\theta_L-\theta)-\theta_H)=\cos(\theta_L-\theta)=\cos(\theta-\theta_L)$. 即当 DO 分别选择 θ 和 $\theta_H+\theta_L-\theta$ 进行断言时，SP 无法区分. 所以，对任意 θ，都存在一个 $\theta_H+\theta_L-\theta$ 使得针对这两个值进行的断言，SP 无法区分. 对应的断言判断元组分别为 $A=\{\theta,B\}$ 和 $A=\{\theta_H+\theta_L-\theta,B\}$，这两个断言判断元组互为无差别的断言判断元组.

不妨设存在两次针对不同值的 θ' 和 θ'' 的断言，其对应的断言判断元组分别是 $A'=\{\theta',B'\}$ 和 $A''=\{\theta'',B''\}$，其中 $B'=\{value'_1,value'_2\}$，$B''=\{value''_1,value''_2\}$. 那么对于 A' 和 A'' 均存在使得 SP 无从区分的无差别的断言判断元组 A'_a 和 A''_a，其中 $A'_a=\{\theta'_a,B'\}=\{\theta_H+\theta_L-\theta',B'\}$ 和 $A''_a=\{\theta''_a,B''\}=\{\theta_H+\theta_L-\theta'',B''\}$，那么当 $\theta'>\theta''$ 或者 $\theta'<\theta''$ 时，$(\theta_H+\theta_L-\theta')<(\theta_H+\theta_L-\theta'')$ 或者 $(\theta_H+\theta_L-\theta')>(\theta_H+\theta_L-\theta'')$，即 $\theta'_a<\theta''_a$ 或 $\theta'_a>\theta''_a$. 所以，对于任意的一对断言元组 B' 和 B''，如果其分别是针对 $\{\theta',\theta''|\theta'>\theta''\}$ 进行断言，那么必然存在针对 $\{\theta_H+\theta_L-\theta',\theta_H+\theta_L-\theta''|\theta'>\theta''\}$ 的断言，其对应断言元组也是 B' 和 B''. 所以，对于任意一对断言元组，SP 无法判断 DO 选择的 θ 的大小关系.

所以当 DO 选择相同的 θ_H 和 θ_L 时，在任意两次断言为真的情况下，SP 无法判断 DO 在这两次判断中选择的 θ 的大小关系.　　证毕.

定理 8.　方案 SRQSAE 中的区间陷门是检索结果安全的.

在方案 SRQSAE 中，判断一个敏感数据是否属于检索区间时，需要向 SP 提供检索区间陷门 $T=\left\{\boldsymbol{M}^{-1}\begin{bmatrix}\cos\theta_L\\ \sin\theta_L\end{bmatrix},\boldsymbol{M}^{-1}\begin{bmatrix}\cos\theta_H\\ \sin\theta_H\end{bmatrix},\cos(\theta_H-\theta_L)\right\}$ 和敏感数据的索引 $\boldsymbol{I}=[\cos\theta\quad\sin\theta]\boldsymbol{M}$，其中 T 的前两个数据可以交换位置.

由于区间陷门和区间索引本身不会泄露检索区间和敏感数据的真实信息，SP 获得区间陷门和区间索引的信息不会影响到检索区间和敏感数据的安全性. 同时区间陷门和区间索引都是通过矩阵相乘进行保护的. 由定理 5 可以看出，在唯密文的情况下，区间索引和区间陷门不会泄露真实的敏感数据，而且也可以保护敏感数据之间的大小关系.

检索时，需要将检索陷门中的 $\boldsymbol{T}_1$ 和 $\boldsymbol{T}_2$ 分别与 $\boldsymbol{I}$ 进行相乘，获得 $\cos(\theta-\theta_L)$ 和 $\cos(\theta-\theta_H)$ 的值，最后与 $\cos(\theta_H-\theta_L)$ 进行比较. 由于区间陷门和索引本身不会泄露真实的敏感数据以及敏感数据的大小关系，所以 SP 根据区间陷门和区间索引无法获得相应的有效信息. 检索时，还将向 SP 泄露 $\cos(\theta-\theta_L)$、$\cos(\theta-\theta_H)$ 和 $\cos(\theta_H-\theta_L)$ 这 3 个值. 由定理 7，当只泄露 $\cos(\theta-\theta_L)$、$\cos(\theta-\theta_H)$ 和 $\cos(\theta_H-\theta_L)$ 时，SP 完全无法区分两个被检索值之间的大小顺序信息. 由于定理 7 的证明过程也是通过寻找两对大小关系相反，且均满足条件的 θ 来证明的，所以它们的大小关系判断不会受到其它检索结果之间的大小关系的影响.

所以方案 SRQSAE 中的区间陷门是检索结果安全的.

综上，方案 SRQSAE 中的区间索引的安全性可以达到排列安全的要求. 根据定理 6 和定理 8，其区间陷门不仅可以满足区间陷门的归并特征安全的需要，而且对于每次检索，均能达到区间陷门的检索结果安全. 所以，方案 SRQSAE 的安全性满足了区间检索的唯密文安全的要求.

5　复杂性分析

本方案是针对外包数据库(ODB)和云存储(Cloud Storage)考虑的. 由于数据外包主要是在 DO 不具备强大的存储能力或 DO 具有可移动的特点的情况下应用的，所以方案的复杂性考虑主要集中于客户端的存储空间复杂性和服务器计算时间复杂性. 不妨设总的记录数为 n.

进行方案部署时，如果客户端需要存储较多的元数据，那么该方案的布署会降低 DO 的可移动性，降低整个方案的可用性. 本方案中，需要在客户端存储可逆矩阵 $\boldsymbol{M}$ 以及外包数据的值域 D，而逆矩阵 $\boldsymbol{M}^{-1}$ 可以通过 $\boldsymbol{M}$ 计算获得，所以只需要获得 $\boldsymbol{M}$ 和 D，就可以计算获得密钥 $K=\{\boldsymbol{M},\boldsymbol{M}^{-1}\}$ 和 D. 无论外包数据的分布规律如何，该方案在客户端的存储空间复杂性为 $O(4r+r)=O(1)$，r 为 $\boldsymbol{M}$ 和 D 中每个数据需要的存储空间. 而对于每个桶内均有 d 个数据的分桶方案而言，DO 需要了解每个分桶的范围以及桶标识，其客户端的空间复杂性为 $O(tn/d)=O(n)$，t 为客户端保存桶的信息所需要的空间. 编码方案中，客户端需要保存所有数据的对应的长度为

L 的编码以及前缀保序加密的密钥，不妨设 B+树每层均有一个 flip 点，那么密钥的存储空间为 $s\log_2 n$，s 为密钥的长度，客户端空间复杂度为 $O(Ln+s\log_2 n)=O(n)$.

服务器的计算时间主要是在进行密文区间检索时服务器检索花费的时间，其体现了密文区间检索方案的检索效率. 本方案在对一个区间索引进行检索时，无论检索区间的上下界为多少，其只涉及 1×2 矩阵和 2×1 矩阵相乘两次，数字相乘一次和比较运算一次，可以归纳为五次乘法、两次加法和一次比较运算. 其服务器的计算复杂性固定为 $O(5an+2bn+cn)=O(n)$，其中 a、b、c 分别是乘法、加密、比较运算的时间代价. 而桶内数据个数恒定的方案中，如果不对区间索引进行运行时优化处理，最优情况是检索区间只覆盖到一个分桶，其计算复杂性 $O(cn)=O(n)$；最坏的情况是检索区间覆盖所有的分桶，即 n/d 个分桶，其计算复杂性 $O((n/d)cn)=O(n^2)$. 编码方案中，最优的情况是检索区间只能分解出一个检索子区间，即 $O(cn)=O(n)$，而检索区间能分解出的检索子区间的上限为 $2(n-1)$，那么对于每个区间索引，均需要进行 $2(n-1)$ 次匹配，服务器的计算时间复杂性 $O(2(n-1)cn)=O(n^2)$.

表 1 复杂性分析

	客户端空间复杂性	服务器时间复杂性	
		最优情况	最坏情况
本方案	$O(1)$	$O(n)$	$O(n)$
分桶方案*	$O(n)$	$O(n)$	$O(n^2)$
编码方案	$O(n)$	$O(n)$	$O(n^2)$

注：分析上述区间索引时，均不考虑具体实现时对区间索引的优化处理.
*为了简化分析，选择的桶的划分方式为每个桶包含固定数目的数据.

综述所述，由表 1 可以看到，本方案所需要客户端的元数据存储空间不随数据量的变化而变化. 在检索时，服务器的计算复杂性也不随检索区间的变化而变化. 所以，本方案中客户端的空间复杂性和服务器的时间复杂性维持在一个定值上. 与现有的两个方案相比，不仅需要较少的客户端元数据存储空间，且在最坏情况下，服务器的时间复杂性也比较有优势.

6 结 论

在本文中，我们提出了一种从区间判断到单断言的转换方法，并将该方法应用到密文区间检索方案中，获得了基于单断言的密文区间检索方案. 该方案使得区间陷门的构建完全基于检索区间的上下界，避免了区间陷门泄露归并特征的安全隐患，并且减少了客户端元数据的存储空间. 同时针对密文区间检索的特点，提出了相关概念来定义区间索引和区间陷门的安全，并证明本文方案满足这些安全要求. 本文的密文区间检索方案获得了比现有方案更高的安全性. 不仅保护了敏感数据不被泄露，保护了敏感数据的大小关系等信息，使得该方案可以满足高安全的密文区间检索的需要. 同时该方案不差于现有密文区间检索的检索效率，使得该方案具有极大的应用前景.

目前，本文提出的密文区间检索方案限于唯密文的场景，对于服务器了解敏感数据相关背景或服务器了解部分索引与敏感数据的对应关系等场景均未涉及，这部分研究将在后续工作中进行.

参 考 文 献

[1] Song D X, Wagner D, Perrig A. Practical techniques for searches on encrypted data//Proceedings of the 2000 IEEE Symposium on Security and Privacy. Oakland, USA, 2000: 44-55

[2] Damiani E, Vimercati S D C, Jajodia S, Paraboschi S, Samarati P. Balancing confidentiality and efficiency in untrusted relational DBMSs//Proceedings of the 10th ACM Conference on Computer and Communications Security(CCS). Washington, USA, 2003: 93-102

[3] Agrawal R, Kiernan J, Srikant R, Xu Yi-Rong. Order preserving encryption for numeric data//Proceedings of the 2004 ACM SIGMOD International Conference on Management of Data. Paris, France, 2004: 563-574

[4] Boldyreva A, Chenette N, Lee Y, O'Neill A. Order-preserving symmetric encryption//Proceedings of the 28th Annual International Conference on Theory and Applications of Cryptographic Techniques (EUROCRYPT). Cologne, German, 2009: 224-241

[5] HacigümüşH, Iyer B, Li Chen, Mehrotra S. Executing SQL over encrypted data in the database-service-provider model//Proceedings of the 2002 ACM SIGMOD International Conference on Management of Data. Wisconsin, USA, 2002: 216-227

[6] Hore B, Mehrotra S, Tsudik G. A privacy-preserving index for range queries//Proceedings of the 30th International Conference on Very Large Data Bases(VLDB). Toronto, Canada, 2004: 720-731

[7] Wang Jie-Ping, Du Xiao-Yong. LOB: Bucket based index for range queries//Proceedings of the 2008 the 9th International Conference on Web-Age Information Management (WAIM). Zhangjiajie, China, 2008: 86-92

[8] Zhang Yong, Li Wei-Xin, Niu Xia-Mu. A method of bucket index over encrypted character data in database//Proceedings

of the 2007 3rd International Conference on Intelligent Information Hiding and Multimedia Signal Processing. Splendor Kaohsiung, China, 2007: 186-189

[9] Li Jun, Omiecinski E R. Efficiency and security trade-off in supporting range queries on encrypted databases//Proceedings of the 19th Annual IFIP WG 11.3 Working Conference on Data and Applications Security (DBSec). Storrs, USA, 2005: 69-83

[10] Cao Ning, Wang Cong, Li Ming, Ren Kui, Lou Wen-Jing. Privacy-preserving multi-keyword ranked search over encrypted cloud data//Proceedings of the INFOCOM. Shanghai, China, 2011: 829-837

[11] Curtmola R, Garay J, Kamara S, Ostrovsky R. Searchable symmetric encryption: Improved definitions and efficient constructions//Proceedings of the 13th ACM Conference on Computer and Communications Security (CCS). Alexandria, USA, 2006: 79-88

[12] Wong W K, Cheung D W, Kao B, Mamoulis N. Secure *k*NN computation on encrypted databases//Proceedings of the 35th SIGMOD International Conference on Management of Data. Rhode Island, USA, 2009:139-152

CAI Ke, born in 1986, Ph. D. candidate. His research interests include data security and searching on encrypted data.

ZHANG Min, born in 1975, Ph. D., associate professor. Her research interests include data security and privacy protection.

FENG Deng-Guo, born in 1965, professor, Ph. D. supervisor. His research interests include network and information security, cryptography theory and technique.

Background

This work began as a part of the research project "Cloud Storage Security Supporting Platform", which is supported by CAS Innovation Program, and improved in a project supported by the National Science and Technology Major Special Projects of China.

The outsourcing of sensitive data reveals all the data to an untrustworthy server, which risks the confidentiality of the sensitive data. To protect these data, the simplest approach is to encrypt them. The disorder of the encrypted form of sensitive data brings in the hard-to-search problem. This paper focus on how to make a range query on the encrypted data while not bring in much privacy issues.

Several researchers proposed approaches to solve this hard-to-search problem, and these approaches mainly solve the problem through building interval index. Hacigümüş et al. divided the domain of sensitive data by the so called buckets, and the interval index is the identification of the bucket which covers it. Though many researchers did a lot of researches on how to divide the sensitive data into buckets, these approaches still reveal the characteristics of the sensitive data more or less. Li J et al. encoded the sensitive data with an order-preserving method, and generated the interval index by a prefix-order encryption of the code. As the author proved, with some different interval trapdoors, the untrustworthy server may achieve the rank of the prefix which reveals the rank of the sensitive data. Agrawal R et al. proposed an order-preserving encryption algorithm on numberical data. Such algorithms can be used to create the interval index, but they reveal the order of sensitive data directly.

After reviewing current approaches, we can find that range query is mainly made through two methods: one is comparing the interval index with the boundary of the range; the other is listing the characteristics of the range and comparing every one with the interval index. These methods reveal too much information. We find a method which can tell whether a data belongs to a range by judging only once. This single assertion reduces the amount of information to be collected to support the range query. Under this framework, we protected the interval indexes and the interval trapdoors by an invertible matrix. So the total solution not only can protect the rank and characteristic of the sensitive data, but also have a stable efficiency which is not lower than the current approaches.