特约稿

基于云计算的可查询加密研究综述

朱艳琴,王琴琴,王婷婷,罗喜召

(苏州大学计算机科学与技术学院,江苏 苏州 215006)

[摘要] 随着云计算时代的到来,吸引着人们将复杂的数据管理外包给云服务器端. 在以更经济、更灵活方式管理数据的同时,个人隐私和数据安全问题一直令人担忧. 加密是一种常用的维护数据隐私的方法,但它不支持有效的数据操作. 可查询加密技术通过加密保证数据的安全性,同时也支持在密文上的某些计算,在一定程度上解决了云计算环境中的隐私保护和数据可用性问题. 本文结合可查询加密的 4 种应用场景,对可查询加密的关键技术和安全问题进行了分析,并对已有的方案做了分类对比,指出了该领域面临的挑战和未来的研究方向.

[关键词] 云计算,可查询加密,密码学

[中图分类号]TP392 [文献标志码]A [文章编号]1001-4616(2014)01-0008-09

A Survey of Searchable Encryption Based on Cloud Computing

Zhu Yanqin, Wang Qinqin, Wang Tingting, Luo Xizhao

(School of Computer Science and Technology, Soochow University, Suzhou 215006, China)

Abstract: With the advent of cloud computing, people are inspired to outsource their data management to the cloud server. While it is a more economical and more flexible way to manage data, privacy security issue has been worrisome. Encryption is a common method for maintaining data privacy, but it does not support efficient data manipulation. Searchable encryption (SE) ensures the security of data by encryption and also supports some computation on the ciphertext. To a certain extent, it solves the privacy protection and data availability in cloud computing environment. With four scenarios, this paper analyzes the main technologies and security problems for searchable encryption. Then classification and comparison on existing schemes are given. At last, the suggestions for future research are put forward.

Key words: cloud computing, searchable encryption, cryptography

作为一种全新的网络计算模式,云计算的出现为学术界、IT 行业乃至全球经济带来了新的机遇和挑战. 在云计算模式下,企业和个人可根据自己的需求购买存储空间和计算能力,而不需花大量资金购买大规模高性能计算机. 相比传统关系数据库,云计算以更经济的方式向用户提供按需服务(存储资源、计算资源和软硬件资源等),大大减少了用户在软硬件维护及升级上的投入成本. 正是由于其特有的优势,云计算得到了 Google、Microsoft、Amazon 等众多信息业巨头的关注,目前已出现了不少实用的在线服务. 在迎来巨大变革及发展前景的同时,云计算的发展也产生了一系列新的问题,其中包括云存储安全问题.

在已经实现的云计算服务中,隐私安全问题尚未完全解决,这也成为阻碍云计算发展和推广的主要因素之一. 云计算的隐私安全问题源于云计算的数据外包和服务租赁的特点. 人们将需要存储或者计算的数据交付给云服务器,获取存储或计算服务的代价是失去了对数据的直接控制能力,这将很可能导致个人隐私数据的泄漏和滥用. 加密是一种常用的隐私保护技术,在云环境的应用中,即在数据外包之前先对其进行加密. 尽管加密能够维护数据安全性,然而目前的大多数加密方案都不支持对密文的运算,如对加密文件进行关键字查询、统计分析、动态操作等,因此大大限制了数据的可用性,严重妨碍了云服务商为用户提

收稿日期:2013-10-10.

基金项目:国家自然科学基金(61373164)、江苏省产学研前瞻性联合研究项目(BY2013030-06)、苏州市应用基础研究计划项目(SYG201238).

通讯联系人:朱艳琴,博士,教授,研究方向:计算机网络与信息安全. E-mai:yqzhu@ suda. edu. cn

供进一步的数据管理和运算服务.

近年来,针对云环境背景下的数据可用性问题,在加密数据上进行运算的可计算加密技术得以提出和发展. 可计算加密技术是一种加密方法,它通过加密保证数据的安全性,且加密后的数据能够支持某些计算. 目前已有的可计算加密技术主要可分为两类:支持查询的加密技术和支持运算的加密技术. 本文主要就可查询加密技术进行分析和总结.

1 概述

与传统关系数据库中的数据查询相比,云环境背景下的数据查询处理具有鲜明的特点,如可扩展性、可用性、在异构环境运行的能力、丰富灵活的用户接口以及高效的数据存储性能等[1].

1.1 可查询加密基本概念

虽然可查询加密方案多种多样,但其基本设定保持一致. 参与方分别是云服务器端(Server)和用户端(User). 用户(如 Alice)将待存储的数据加密之后发送给 Server,Server 存储 Alice 发送过来的数据. 之后,Alice 想要对已存储的密文数据进行某些特定查询,则向 Server 发送一个相关的查询陷门,Server 基于此陷门和 Alice 存储的密文数据进行查询,并将查询结果返回给 Alice,Alice 再解密数据得到明文查询结果. 在上述基础上,衍生出不同复杂程度的设定.

首先,定义一个基本的可查询加密方案的框架,其他的各种变体都是在此基础上拓展延伸的;其次,给出正确性和安全性定义.

定义 1(可查询加密方案) 可查询加密方案 $\sum=(\mathrm{Gen},\mathrm{Enc},\mathrm{Trapdoor},\mathrm{Search},\mathrm{Dec})$ 由以下 5 个算法组成:

(1)$K\leftarrow\mathrm{Gen}(U,d)$:密钥生成算法为用户 U 产生密钥 K,d 为安全参数.

(2)$(\gamma,c)\leftarrow\mathrm{Enc}(K,D)$:加密算法可能为概率算法,以密钥 K 和明文集合 D 作为输入,输出密文 c 和辅助信息 γ.

(3)$\tau_w\leftarrow\mathrm{Trapdoor}(K,w)$:查询陷门算法可能为概率算法,为待查询关键字 w 生成一个对应的陷门 τ_w.

(4)$I_w\leftarrow\mathrm{Search}(\gamma,c,\tau)$:查询算法为确定算法,以密文、辅助信息和查询陷门作为输入,输出包含所查关键字 w 的密文的标识符集合 I_w.

(5)$m\leftarrow\mathrm{Dec}(K,c)$:解密算法为确定算法,对于密文 c,解密结果为 m.

定义 2(正确性) 可查询加密方案 $\sum=(\mathrm{Gen},\mathrm{Enc},\mathrm{Trapdoor},\mathrm{Search},\mathrm{Dec})$ 是正确的:

(1)$\forall m\in D,\exists\,\mathrm{Dec}(\mathrm{Enc}(K,m))=m$;

(2)$\mathrm{Search}(\gamma,C,\tau_w)=D(w)$,其中是 $D(w)$ 是包含所查关键字 w 的文件集合.

定义 3(安全性) 可查询加密方案 $\sum=(\mathrm{Gen},\mathrm{Enc},\mathrm{Trapdoor},\mathrm{Search},\mathrm{Dec})$ 是安全的:

(1)密文不会泄漏任何信息,加解密算法至少是 IND-CCA 安全的;

(2)查询过程不会泄漏任何信息,不能推断出所查内容;

1.2 可查询加密应用场景

1.2.1 个人外包数据库

假设 Alice 经常出差,希望无论在何时何地都能够访问她的数据库. 为此,Alice 可以把自己的个人数据库外包给第三方服务商. 为防止隐私泄漏,需要在外包之前加密数据库. Alice 可以使用可查询加密方案加密数据库,其后当向服务器提出查询要求时,服务器在加密数据库上进行查询,并返回符合查询要求的结果. 在这种应用场景下,可抽象出以下安全性需求:

(1)仅数据拥有者可加解密文件和进行数据查询;

(2)包括服务器在内的其他任何实体都不能获悉查询内容.

这一应用场景首次在文献[2]中被提出,推动了对称可查询加密方案的深入研究.

1.2.2 小组成员授权

这一场景是在个人外包数据库上的扩展. 假设 Alice 在云端存储了一些加密数据,不仅她自己可以对数据进行查询,还允许某些特定人士也可查询她的加密数据. Alice 可以使用可查询加密方案加密数据,并

将查询权利授权给一个特定小组,这样该小组内的所有成员都可以对 Alice 存储的加密数据进行查询与阅读. 在这种应用场景下,可抽象出以下安全性需求:

(1)仅数据拥有者可加密文件;

(2)仅数据拥有者和得到授权的用户可进行数据查询和解密文件,且支持不同程度的授权;

(3)除数据拥有者和得到授权的用户之外,包括服务器在内的其他任何实体都不能获悉查询内容.

1.2.3 邮件路由服务

在所有外包服务中,Email 是应用最广泛的例子之一. 任何拥有 Alice 邮件地址的用户都可以向 Alice 发送邮件,所有用户的 Email 都由服务提供商管理,因而服务提供商通常拥有对其用户群的明文 Email 的访问权,这就难免会导致敏感信息的泄漏. 现假设有服务提供商可以提供安全 Email 服务,允许用户收取加密过的 Email. 在这种场景下,Alice 可使用可查询加密方案加密 Email 或收取其他用户发送的密文 Email,之后可向 Email 服务提供商提出查询请求,提供商在密文 Email 上执行查询操作并返回符合查询要求的结果. 在这种应用场景下,可抽象出以下安全性需求:

(1)任何实体都可以生成加密文件;

(2)仅数据拥有者可进行数据查询和解密文件,此外,可能允许服务提供商扫描其文件以探测病毒或恶意软件,但不允许获悉相关明文信息;

(3)包括服务器在内的其他任何实体都不能获悉查询内容.

这一应用场景首次在文献[3]中被提出,推动了非对称可查询加密方案的深入研究.

1.2.4 个人医疗记录

基于因特网的个人医疗记录(PHR,Personal Healthcare Record)通过浏览器或某些应用接口为用户存储 PHR,允许信息的访问和修改,并支持信息共享. 以用户 Alice 为例,她的 PHR 数据有很多来源,例如医生的处方、医院的治疗记录以及家用感应器的监视结果等等. Alice 的 PHR 数据可由她本人或其他人发送到她的账户,她可以使用可查询加密方案加密 PHR 数据并存储在云端,在需要时再进行相关查询. 假设另一用户 Bob 也拥有一个 PHR 账户,且 Alice 的医疗记录有部分与 Bob 类似,对 Bob 的医疗有参考价值,若 Bob 拥有对 Alice 医疗记录的访问权,他就可查询 Alice 的医疗记录以获取相关信息. 但数据共享必定会加大信息泄漏和滥用的可能性,因此在引入可查询加密方案的同时进行访问权控制是非常必要的. 在这种应用场景下,可抽象出以下安全性需求:

(1)任何实体都可以生成加密文件;

(2)仅数据拥有者和得到授权的用户可进行数据查询和解密文件,且支持不同程度的授权;

(3)除数据拥有者和得到授权的用户之外,包括服务器在内的其他任何实体都不能获悉查询内容.

这一应用场景推动了非对称可查询加密设定下联合数据库[4-6]的深入研究.

2 可查询加密关键技术

现有的可查询加密方案主要可分3类:(1)工作于数据结构层面的方法,通过加入额外的数据结构信息来进行辅助查询;(2)开发新颖的加密原语,如全同态加密[7],利用其本质特性进行查询;(3)作为一种补充方法,Raykova 等人[8]通过引入一个可信代理为用户的查询请求二次加密,虽然同时隐藏了查询内容和用户身份,但在实际应用环境中可信的第三方难以实现,因此该方法并无太大的实用价值.

2.1 基于数据结构的方案

2.1.1 全域查询

全域查询技术,即循序地扫描每一数据项以测试某些条件. 例如,测试某些数据项是否出现了超过阈值的次数,某些数据项是否不存在. 全域查询以块的形式加密数据项,查询时间与数据项总数线性相关,查询条件比较灵活,可以是动态定义的任何形式. 其缺点是当数据项很多时效率低下.

全域查询应用最广泛的技术之一是维持顺序的加密形式[9-11],即按序加密使得密文与明文序列相对应,这种方式的加密技术安全性较弱. 伪随机置换[12]是一个有效的基于密钥的置换,可被用于打乱明文顺序[13],在一定程度上维护了明文信息的安全性. 此外,健忘的 RAMs[14]技术为全域查询提供了支持,在查询过程中不会泄漏任何信息,实现了可查询加密的最佳安全性,但这种强安全性是以对数级的交互操作为

代价的,难以付诸现实.

全域查询首先需要考虑的实际问题是加密块大小该如何决定,这是因为该加密技术以特定长度的二进制串作为输入,意味着在加密之前需要将数据分割成块状. 文献[2]以单词为分割单元,仅支持精确匹配,为支持模糊匹配可能需要以字母为单元的分组加密,而分组的粒度可能会导致更多的安全问题,如统计分析. 实际上,加密块的大小不仅受到加密算法的限制,也取决于所支持的查询条件. 有文本、图片、影音等多种类型的数据,如何在不降低安全性和效率的前提下支持多种条件的查询是一个实际的问题. 其次需要考虑的问题是计算和存储效率. 对于以单词为单元的方案,假设有 n 个单词,加密就需要 $O(n)$ 次操作,且存储代价将远远比明文大. 影响效率的瓶颈在于查询,每次查询都需要 $O(n)$ 次操作,当数据项总数很大时,效率将非常低.

2.1.2 索引查询

基于索引的查询或者基于关键字的查询技术,是为数据项建立一个由关键字列表构成的索引,查询时用测试索引代替测试数据项内容. 有正排索引和倒排索引两种索引方法. 后者在可查询加密领域应用较广泛. 通常还可将其他数据结构与索引相结合,如文献[15]中构造了基于倒排索引的合并树,从而节省了存储空间. 通过加密索引和关键字,无需逐一浏览数据项内容,因而大大提高了查询效率. 其缺点是查询条件不如全域查询灵活,性能取决于关键字的选择和索引的建立、维护.

伪随机函数[12]是一个有效的基于密钥的函数,通常用于构造查询索引,具有与真随机的不可区分性. 因此在未知密钥的前提下,难以伪造有效的查询陷门,从而保证了陷门的安全性,使得没有相关密钥的用户不能够进行有效查询. 基于密钥的哈希函数是另一种常用于构造索引的方法,哈希函数具有很好的查找效率,但可能会产生冲突,致使不同关键字处理函数的结果映射在同一位置上,可查询加密方案是不允许出现此类现象的,因此必须选择耐碰撞的哈希函数. 目前也有一些方案[16,17]是基于双线性映射[18]的,利用群上双线性映射的特殊性质构造密文和查询索引,使得两者之间有某种联系,在经过一系列计算之后能够找到匹配的结果. 这类方案安全性高,但算法相对复杂,效率也就稍低.

基于索引的查询无需考虑全域查询所带来的那几种问题,但也面临着其他的挑战. 首先,关键字空间大小有限且分布不均匀,同时关键字集合可能是公共信息,这些因素将很可能导致安全问题. 例如,当很多个查询对应于相同的查询结果,服务器可以推测该关键字具有很高的分布率,根据公开的词频统计信息就可能猜测出潜在关键字. 因此,为防止信息泄漏,就需要维护查询结果保密性. 其次,索引也可能泄漏信息. 就正排索引来说,每个文件包含不同数量的关键字,则在索引中其所对应的关键字列表也就长度不一,导致了信息重要性高低的泄漏. 针对此类问题有一些缓解的方法,诸如向索引中添加随机值、填充数据项标识符、用固定大小的字典表示关键字存在与否等等. 简而言之,就是将索引从不同长度转变为相同长度. 还有一个很重要的问题是索引更新问题. 对于正排索引,索引更新是比较直观可行的,但对倒排索引来说就比较复杂了. 文献[19,20]提出了使用新密钥添加新索引的方法,结果是旧的陷门不能再用于查询,带来了密钥管理和查询灵活性等问题.

2.2 基于全同态加密的方案

同态加密(Homomorphic Encryption)是当下的一大热点技术. 它是一种加密形式,允许人们对密文进行特定的代数运算,得到仍然是加密的结果,与对明文进行同样的运算再将结果加密一样. 换言之,这项技术使人们可以在加密的数据中进行诸如检索、比较等操作,得出正确的结果,而在整个处理过程中无需对数据进行解密. 其意义在于,真正从根本上解决了将数据及其操作委托给第三方时的保密问题,例如对于各种云计算的应用.

形式化描述如下:记加密操作为 E,明文为 m,加密得 e,即 $e=E(m)$,$m=E'(e)$. 已知针对明文有操作 f,针对 E 可构造 F,使得 $F(e)=E(f(m))$,这样 E 就是一个针对 f 的同态加密算法.

找到这样的 E 并不容易,这一直是密码学领域的一个重要课题,以往人们只找到一些部分实现这种操作的方法. 2009 年 9 月,Craig Gentry[7]从数学上提出了“全同态加密”的可行方法,即可在不解密的条件下对加密数据进行任何可以在明文上进行的运算,使这项技术取得了决定性的突破. 最初的方案依赖矩阵和矢量,每一步都要分别计算每个元,这已经足够复杂;计算完矩阵后还要处理数据本身,使得计算更加复杂. 这使得矩阵和矢量加密方法实用性不强. Smart 与 Vercauteren[21]改写了加密方法,免去了复杂的计算,

使得 Gentry 的理念得以在电脑上实施和测试. 但这一方案也有其局限性,随着计算步骤的增加,连续加密的计算结果质量在下降.

从辩证角度来看,全同态技术也有其弊端,诸如算法复杂、计算效率低、难以实际实现等,但全同态技术发展很快,IT 界对此充满了兴趣. 人们正在研究更完善的实用技术,这对信息技术产业具有重大价值.

3 可查询加密分类

云计算的特殊环境对可查询加密方案提出了迫切需求,目前已涌现出一大批可查询加密的相关方案,各种方案各有特点,各有优势.

3.1 对称性

基于"谁可以为外包数据库生成加密文件"以及"谁可以向第三方服务器提交查询请求"两个问题,可查询加密方案可分为对称可查询加密(Searchable Symmetric Encryption, SSE)和非对称可查询加密(Searchable Asymmetric Encryption, SAE)两类.

3.1.1 对称可查询加密

在对称可查询加密环境设定下,加密之前用户可以任意形式组织数据并附加数据结构(如索引)以允许快速查询,再用私钥(private key)加密数据和附加信息,并发送给服务器存储,因此只有拥有私钥的用户才能够生成有效的查询陷门,对服务器提出合理的查询请求以访问所存信息,没有私钥的用户难以伪造出有效的陷门进行查询.

Song 等人[2]首次提出了实用的 SSE 方案. 在他们的方案中,文件中的每个词按序被加密成两层结构,内层使用验证结构,服务器用陷门剥去外层结构再检查内层. 若陷门和密文是由同一词得来,则内层验证结构将得以保存. Goh[22]率先提出基于安全索引的 SSE 方案,该方案将一份文件中的所有词存入一个布隆过滤器[23]作为其安全索引. Goh 同时提出了 IND-CKA 和 IND2-CKA 的安全模型,并证明了他的方案是 IND-CKA 安全的. Curtmola[20]利用倒排索引构建了查询索引,并介绍了非适应性和适应性两种安全模型,且利用广播加密[24]拓展到多用户查询场景. 在基本 SSE 方案的基础上,多用户的对称可查询加密方案相继被提出,如文献[17]中所有文件是使用不同密钥加密的,通过将使用用户密钥生成的查询标记转变成用文件密钥生成的查询标记,允许用户在其拥有访问权的文件上进行查询,细化了用户的查询权利. 文献[25]中利用了代理加密的概念,让服务器端进行二次加密,使得每个用户都可以成为读者和写者,从而缓解了由密钥共享带来的安全威胁.

在这种设定下,算法较简单,用户的初始工作(如预处理)至少和数据大小一般大,而用户和服务器所需做的后续工作(如访问数据)则非常小. 更重要的是,查询过程中无需过多的交互操作,且所有有关用户访问模式的信息都被隐藏了. 但对称可查询加密方案适用范围较窄,在现实中的应用有较大的限制.

3.1.2 非对称可查询加密

SAE 通常也用 PEKS 表示(Public Key Encryption with Keyword Search). 非对称可查询加密是在用公钥(public key)加密的数据上进行查询,即用户用于加密的密钥和用于解密的密钥是不同的. 当公钥公布之后,任何人都可以使用公钥生成密文,而只有拥有私钥的人才可以生成查询陷门在密文上进行查询和解密查询结果. 一个典型的应用是 1.2.3 节中的邮件路由服务,1.2.4 节中的应用场景也是 SAE 的另一种变体.

Boneh[3]等人于 2004 年首次介绍了非对称可查询加密方案. 但该方案在很多方面都有所限制,如仅考虑单个关键字查询、只支持精确匹配、会泄漏用户的访问模式等. 他在 2007 年提出的方案[26]中隐藏了用户的访问模式,要求消息发送方和服务器参与交互协议,因而需要更大的查询消耗. 之后,专家们为解决陷门的安全传输问题提出了 dPEKS(PEKS with designated tester)方案[27],为应对离线消息恢复攻击提出了 rPEKS(PEKS with registered keywords)方案[28]等等. 以上都是基于索引的方案,不少非对称环境下的全域查询方案也相继被提出. 例如,Bellare[29]等人提出了确定性全域查询加密的概念,允许任意实体生成查询,但用这种方法加密的密文是确定性的,安全强度相对弱了一些. Ibraimi[30]提出的 PKEDS(Public Key Encryption with Delegated Search)允许服务器直接在密文上搜索,但安全模型仅考虑了单向性,安全性比文献[29]更弱.

在这种设定下,算法更加复杂,一般无论是加解密还是查询操作消耗的代价都相对较大,所需的用

户—服务器间的交互操作可能更多. PEKS 及其变体也大多存在访问模式泄漏的弊端;但非对称可查询加密适用的范围明显更广,变体更多,在现实生活中适用性也更强,具有更大的推广前景.

3.2 功能性

就“可查询加密方案支持什么样的查询”这一问题,可按照不同功能大致将已存在的可查询加密方案分成以下几类.

3.2.1 支持精确/模糊查询

精确查询意味着只返回完全匹配所查询关键字的文件,不支持微小错误和格式不一致. 目前的很多可查询加密方案都只支持精确查询. 但是,输入错误和格式问题是很常见的,如将“Angel”误写为“Angle”,如“multiuser”和“multi-user”,这时支持精确查询的方案就不能返回正确的结果,这显然不是用户乐意看到的. 支持模糊查询的方案就可以解决这类问题,在出现微小错误和格式不一致的情况下也能返回满足用户查询内容的结果.

Li[31]首次提出了支持模糊查询的方案,在该方案中使用通配符和编辑距离进行匹配,服务器能够返回与所查关键字相似的结果. 代价是需要更大的存储空间和计算时间,但这对于云计算庞大的存储能力和计算能力来说就不值一提了. 也有一些其他方案做了相关探讨,如文献[32]. 但对此的研究还处于初步阶段,有待进一步完善.

3.2.2 支持单字/多字查询

基于关键字的查询可分为单一关键字查询和多关键字查询两种,单字查询是多字查询的一种特例,即支持多字查询的方案必定也支持单字查询. 尽管目前大部分方案都只支持单字查询,但也已涌现出一些关于多字查询的技术. 一种直观的方法是直接扩展单字查询方案,即为每个关键字生成一个索引,同时匹配所有索引的文件以满足多字查询需求. 很明显这种方法效率低下.

文献[32]提出了两个支持多字查询的方案. 在第一个方案中,每次生成查询都需要关于同步查询关键字数量的线性次求幂操作,而第二个方案的通信代价则为关键字域大小的数量级. 方案[16]在双线性映射的基础上使用异或操作,使得在支持多字查询的同时将时间消耗降低到常量级别,几乎达到与单字查询同样的效率,但用户需要事先了解包含所有有效关键字的列表及其所处位置,这在某些场合是不适用的. 方案[32]根据关键字的哈希值,将它们分布到固定数量的密钥桶中,从而达到同一陷门可用于多个查询的效果,大大提高了查询效率. 但该方案有一定的错误率,可能出现假阳性事件.

多关键字查询的研究与发展正是来自于用户需求的驱使,因为在很多情况下,用户需要查找出同时包含几个关键字的文件,而不是被约束在单一关键字下. 多关键字查询更符合用户的查询习惯和需求,具有更大的应用前景.

3.2.3 支持静态/动态查询

所谓的静态查询,即只支持在固定密文上的查询,只允许一次存储,当需要加入新的密文或删除旧的密文时就不再适用查询功能了,必须在重新生成辅助信息之后才能返回正确的查询结果;反之,动态查询就实现了在增加或删除密文之后仍可进行查询的效果.

Seny 等人[32]首次提出了动态可查询加密方案(Dynamic SSE),实现了对密文的更新(增加或删除文件)以及在此基础上的再次查询,该方案不仅达到了 CKA2 安全性,还具有较高的执行效率. 之后,他们又基于红黑树构造出了并行的 DSSE 方案[33],执行时间与服务器数量成反相关,在大大提高效率的前提下并未削弱安全强度. 但基于红黑树的加密方案在空间复杂度上有所提高,一种可行的降低数据大小的方法是使用二层树的数据结构.

目前还只有少量支持动态可查询加密技术的方案,这是一个非常新的研究领域. 但毋庸置疑,用户的需求变化多端,不仅要求文件的存储能力,而且对文件的增删查改也有很强烈的愿望. 目前的动态查询技术主要是针对整个文件的增加和删除,还没有上升到对文件内容的修改层面.

3.2.4 支持可验证性查询

在享受商业云计算服务带来便利的同时,我们注意到云服务器端可能是自私的,它会为了谋求节省计算能力和下载带宽等利益而不完全执行用户的要求,可称之为“半诚实而好奇(semi-honest-but-curious)”的服务器. 此类服务器有以下几个属性:不会修改或删除所存内容;尝试从存储数据和查询操作中提取敏

感信息;可能只诚实地执行一部分查询操作,返回一部分查询结果. 为了与之抗衡,提出了可验证可查询加密(Verifiable SSE)的概念,即服务器需要向用户提供查询结果正确性和完整性的证据,用户可以证明其是否可信.

Chai[34]首次构造了VSSE方案,实现了基本的数据安全性和查询可验证性. 该方案使用特里树结构并维护一张字母表,树中每个节点都对其子节点所对应的字母序号序列进行哈希运算,自上而下记录了字母出现的次序,从而使得用户可以验证查询结果是否正确和完整. 陷门生成和验证可以在常量时间内完成,但预处理时间随数据集合大小呈线性增长,且由于是工作在数据结构层面,因而有额外的存储代价. 文献[35]通过使用不同密钥生成不同关键字的HMAC,实现了在不解密密文的前提下就可确认是否包含所查关键字,同时将每个关键字与一个计数器相关联以确保查询结果的完整性. 该文提出了基于属性加密的VSSE方案,分别针对KP-ABKS和CP-AKBS构造了两种强制执行访问控制策略,使得查询者无需与数据拥有者进行交互即可查询. 该方案算法较复杂,时间和空间复杂度都不甚理想. 方案[13]拓展到结构化层面,实现了对文本和图片的查询验证,算法也相对较简单.

VSSE技术的研究还处于起步阶段,但具有很大的潜在价值,值得更多的关注和探讨. 云服务的不完全可信和半诚实性制约了云计算的发展,若不能较好地解决可信问题,云计算将面临严峻的挑战. 将可验证性与可查询加密相结合,可以构建出更多VSSE方案,以同时解决云环境下的查询和验证问题.

4 分析与展望

4.1 典型方案分析

作为云计算服务发展进程中重要的一环,可查询加密技术引起了工业界和学术界的广泛关注. 本文依据不同的应用场景和安全需求,选取了多个典型方案,对其是否采用对称可查询加密以及支持什么样的查询进行归纳和分析,并对比了不同方案的安全性和时间复杂度,如表1所示.

表1 典型方案对比

Table 1 Comparisons of typical schemes

方案	查询技术	对称性	功能性				安全性	查询时间
			模糊	多字	动态	可验证		
SWP[2]	全域查询	对称	×	×	×	×	CPA	$O(n)$
GO[14]	全域查询	对称	×	×	×	×	Optimal	$O(\log^3 n)$
PKEET[4]	全域查询	非对称	×	×	×	×	OW-CCA	$O(\log n)$
BBO[29]	全域查询	非对称	×	×	×	×	PRIV-CCA	$O(\log n)$
SSE-1[20]	索引查询	对称	×	×	×	×	CKA1	$O(1)$
DSSE[32]	索引查询	对称	×	×	√	×	CKA2	$O(w)$
VSSE[34]	索引查询	对称	×	×	×	√	CKA2	$O(1)$
LWW[31]	索引查询	对称	√	×	×	×	CKA1	$O(l^d)$
BCO[3]	索引查询	非对称	×	×	×	×	CKA2	$O(m)$
YLW[16]	索引查询	非对称	√	√	×	×	CKA1	$O(m)$

注:CPA:Chosen Plaintext Attack
OW-CCA:One Way-Chosen Ciphertext Attack
PRIV-CCA:Privacy-Chosen Ciphertext Attack
CKA1:Chosen Keyword Attack
CKA2:Adaptive Chosen Keyword Attack
n:明文空间字词总数;
w:包含所查关键字的文件总数;
l:关键字总数;
d:编辑距离;
m:数据记录总数.

4.2 可查询加密展望

对于可查询加密技术的研究工作仍处于起步阶段,存在着大量有价值的研究问题.

4.2.1 索引管理技术

目前在可查询加密技术中,针对基于索引的查询技术已取得了较好的研究进展,相对于全域查询技术而言有更大的发展空间,但在以下两个方面还有待于进一步研究:首先,已存的索引方案大多以关键字列表为基础,将关键字与文件标识符绑定在一起,比较适合于相对稳定的数据. 但对于数据频繁更新的状况,索引更新维护的代价比较高. 因此,在云计算环境下,如何设计支持频繁更新的索引方案是一个富有挑战性的工作. 其次,现有的索引技术大多只支持布尔查询、范围查询等简单的查询操作,无法对一些复杂的查询提供良好的支持,如模糊查询等. 然而,在一些特定的情境下往往需要支持一些相对复杂的查询,因此,

针对不同的应用需求设计支持复杂查询的方案具有重要意义.

4.2.2 多功能查询技术

无论何种产品,用户希望的总是功能越多越好,功能越多,吸引力就越大.对于查询功能的要求也是如此:要求即使在有微小输入错误的前提下依然能够查找到真正想要查找的内容,这是模糊查询;不满足于仅仅只支持单个字的查询,这是要求能够查找出同时包含多个字的文件,即多关键字查询;在已存储密文上加入新的数据或者删除旧的内容之后,不需要做太多的变动或修改就能保持查询的正确性,这就要求满足查询的动态变化;对于服务器查询的返回结果存有疑问,不能确信是否正确和完整,这就要求服务器提供可验证功能,等等.因此,如何将这些功能都融合到一起,设计出能够符合大多数人查询习惯的可查询加密方案,这也将成为今后深入研究的一大方向.

4.2.3 全同态加密技术

全同态加密是一种新兴的计算形式,具有直接操作密文而不需要解密的优越性质,它的密文运算性质使得它在云计算、密文查询、电子投票和多方计算等领域都有着重要的应用.在可查询加密领域,可以以全同态加密技术为基础,在此之上构造多样化的查询方案,目前也有了初步的进展.然而,已经提出的全同态加密方案多是依据 Gentry 的全同态加密思路构造而成,方案复杂、执行效率低,与实际应用还有天壤之别.如何改进全同态加密方案的执行效率与安全性,已经成为当前全同态加密技术研究的重点与难点.因此,全同态加密技术的发展也会影响可查询加密技术的前景.

4.2.4 查询优化

查询处理方法和优化策略对云数据查询管理来说是一个至关重要的问题.目前的可查询加密技术对云计算环境下数据存储的特点考虑得较少,没有很好地利用其优越性.云环境中数据量大且分布存储,为了达到较高的可用性和容错性,多处冗余备份是必要的,利用数据冗余备份的特点对查询进行并行化处理和传输非常具有现实意义.进行查询优化时,增加并行度可以充分利用云端的计算资源,提高查询性能,但是一味的增加并行度也是过犹不及的,很可能会因数据传输数量的不同而造成网络拥塞或计算节点空闲.平衡查询并行度和数据传输代价是不容忽视的.此外,对于不同的查询要求,选择相应不同的查询方法,也是节省计算资源的一个途径.

5 结论

随着信息产业的不断发展,云计算在取得巨大成功的同时,数据安全性和可用性问题日益成为阻碍其飞速成长的拦路虎.可查询加密技术在通过加密保证数据安全的同时,也支持在密文上的相关查询,在一定程度上对云计算背景下的数据安全性和可用性起到了保障作用.本文对近几年来国内外在云数据查询领域的主要研究成果进行了总结,综述了该技术的应用场景,并对相关技术进行了分析对比,最后指出该领域仍然存在的问题和发展前景.总的来说,云计算背景下的可查询加密技术还处在起步阶段,大量有价值的关键问题仍有待于深入研究.该领域为研究者提供了广阔的探索空间.

[参考文献]

[1] 史英杰,孟小峰.云数据管理系统中查询技术研究综述[J].计算机学报,2013,36(2):209-225.

[2] Song D X,Wagner D,Perrig A. Practical techniques for searches on encrypted data[C]//2000 IEEE Symposium on Security and Privacy. Berkeley,CA:IEEE,2000:44-55.

[3] Boneh D,Di Crescenzo G,Ostrovsky R,et al. Public key encryption with keyword search[C]//Advances in Cryptology-Eurocrypt 2004. Berlin Heidelberg:Springer,2004:506-522.

[4] Yang G,Tan C H,Huang Q,et al. Probabilistic public key encryption with equality test[C]//Topics in Cryptology-CT-RSA 2010. Berlin Heidelberg:Springer,2010:119-131.

[5] Tang Q. Towards public key encryption scheme supporting equality test with fine-grained authorization[C]//Information Security and Privacy. Berlin Heidelberg:Springer,2011:389-406.

[6] Tang Q. Public key encryption schemes supporting equality test with authorisation of different granularity[J]. International Journal of Applied Cryptography,2012,2(4):304-321.

[7] Gentry C. A Fully Homomorphic Encryption Scheme[M]. South Carolina:BiblioBazaar,2009.

[8] Raykova M,Vo B,Bellovin S M,et al. Secure anonymous database search[C]//Proceedings of the 2009 ACM Workshop on

Cloud Computing Security. New York:ACM,2009:115-126.
[9] Agrawal R,Kiernan J,Srikant R,et al. Order preserving encryption for numeric data[C]//Proceedings of the 2004 ACM SIGMOD International Conference on Management of Data. New York:ACM,2004:563-574.
[10] Boldyreva A,Chenette N,Lee Y,et al. Order-preserving symmetric encryption[C]//Advances in Cryptology-EUROCRYPT 2009. Berlin Heidelberg:Springer,2009:224-241.
[11] Tang Q. Privacy preserving mapping schemes supporting comparison[C]//Proceedings of the 2010 ACM Workshop on Cloud Computing Security Workshop. New York:ACM,2010:53-58.
[12] Katz J,Lindell Y. Introduction to Modern Cryptography[M]. Florida:CRC Press,2008.
[13] Mohamad M S,Poh G S. Verifiable structured encryption[C]//Information Security and Cryptology. Berlin Heidelberg:Springer,2013:137-156.
[14] Goldreich O,Ostrovsky R. Software protection and simulation on oblivious RAMs[J]. Journal of the ACM(JACM),1996,43(3):431-473.
[15] Lu H,Gu D,Jin C,et al. Reducing extra storage in searchable symmetric encryption scheme[C]//2012 IEEE 4th International Conference on Cloud Computing Technology and Science(CloudCom). Taipei:IEEE Computer Society,2012:255-262.
[16] Yang Y,Lu H,Weng J. Multi-user private keyword search for cloud computing[C]//2011 IEEE Third International Conference on Cloud Computing Technology and Science(CloudCom). Athens:IEEE Computer Society,2011:264-271.
[17] Popa R A,Zeldovich N. Multi-key searchable encryption[J/OL]. Cryptology ePrint Archive,Report 2013/508. [2013-10-10]http://eprint. iacr. org.
[18] Boneh D,Franklin M. Identity-based encryption from the Weil pairing[C]//Advances in Cryptology—CRYPTO 2001. Berlin Heidelberg:Springer,2001:213-229.
[19] Chang Y C,Mitzenmacher M. Privacy preserving keyword searches on remote encrypted data[C]//Applied Cryptography and Network Security. Berlin Heidelberg:Springer,2005:442-455.
[20] Curtmola R,Garay J,Kamara S,et al. Searchable symmetric encryption:improved definitions and efficient constructions[C]//Proceedings of the 13th ACM Conference on Computer and Communications Security. New York:ACM,2006:79-88.
[21] Smart N P,Vercauteren F. Fully homomorphic encryption with relatively small key and ciphertextsizes[C]//Public Key Cryptography - PKC 2010. Berlin Heidelberg:Springer,2010:420-443.
[22] Goh E J. Secure indexes[J/OL]. Cryptology ePrint Archive,Report 2003/216. [2013-10-10]http://eprint. iacr. org.
[23] Bloom B H. Space/time trade-offs in hash coding with allowable errors[J]. Communications of the ACM,1970,13(7):422-426.
[24] Berkovits S. How to broadcast a secret[C]//Advances in Cryptology—EUROCRYPT'91. Berlin Heidelberg:Springer,1991:535-541.
[25] Dong C,Russello G,Dulay N. Shared and searchable encrypted data for untrusted servers[C]//Data and Applications Security XXII. Berlin Heidelberg:Springer,2008:127-143.
[26] Boneh D,Kushilevitz E,Ostrovsky R,et al. Public key encryption that allows PIR queries[C]//Advances in Cryptology-CRYPTO 2007. Berlin Heidelberg:Springer,2007:50-67.
[27] Rhee H S,Park J H,Susilo W,et al. Trapdoor security in a searchable public-key encryption scheme with a designated tester[J]. Journal of Systems and Software,2010,83(5):763-771.
[28] Tang Q,Chen L. Public-key encryption with registered keyword search[C]//Public Key Infrastructures,Services and Applications. Berlin Heidelberg:Springer,2010:163-178.
[29] Bellare M,Boldyreva A,O'Neill A. Deterministic and efficiently searchable encryption[C]//Advances in Cryptology-CRYPTO 2007. Berlin Heidelberg:Springer,2007:535-552.
[30] Ibraimi L,Nikova S,Hartel P,et al. Public-key encryption with delegated search[C]//Applied Cryptography and Network Security. Berlin Heidelberg:Springer,2011:532-549.
[31] Li J,Wang Q,Wang C,et al. Fuzzy keyword search over encrypted data in cloud computing[C]//2010 Proceedings IEEE INFOCOM. San Diego:IEEE,2010:1-5.
[32] Kamara S,Papamanthou C,Roeder T. Dynamic searchable symmetric encryption[C]//Proceedings of the 2012 ACM Conference on Computer and Communications Security. Raleigh:ACM,2012:965-976.
[33] Kamara S,Papamanthou C. Parallel and dynamic searchable symmetric encryption[C]//Financial Cryptography and Data Security. Berlin Heidelberg:Springer,2013:258-274.
[34] Chai Q,Gong G. Verifiable symmetric searchable encryption for semi-honest-but-curious cloud servers[C]//2012 IEEE International Conference on Communications(ICC). Ottawa:IEEE Computer Society,2012:917-922.

[责任编辑:严海琳]