面向托管的数据库即服务系统及其隐私保护技术

陈 萍 张 涛 赵 敏 袁志坚 杨兰娟

(解放军理工大学指挥信息系统学院 南京 210007)

摘 要 数据库即服务(DBaaS)是云计算的一个研究热点,而数据应用托管则是当前DBaaS的一个重要应用领域。针对托管数据隐私保护问题,提出了基于虚拟机和CryptDB系统构建支持多副本的多租户数据托管方法及相应的数据库即服务系统,该系统实现了托管数据的隔离和加密存储并且能基于加密数据执行SQL查询。相关实验表明,和全同态加密系统相比,系统具有较低的性能损耗,较好地解决了隐私保护和实用性问题。

关键词 数据库即服务,隐私保护,CryptDB,虚拟机

中图法分类号 TP315 **文献标识码** A

Database as Service System for Business Database Application Hosting and its Privacy Preservation Mechanism

CHEN Ping ZHANG Tao ZHAO Min YUAN Zhi-jian YANG Lan-juan

(College of Command Information System, PLA University of Science & Technology, Nanjing 210007, China)

Abstract Database as a Service(DBaaS) is becoming a research hotspot of cloud computing, as a main application domain, business database application hosting puts forward the requirements of isolation and privacy preservation on hosting data. To satisfy this requirement, this paper proposed a virtual machine based database hosting method and corresponding DBaaS system based on CryptDB. This system has realized the hosting data encrypted storage and can execute SQL queries based on encrypted data. The experiment shows that compared with fully homomorphic encryption system, the performance of the system has lower loss, and better solve the issue of privacy protection and practical.

Keywords Database as a service, Privacy preservation, CryptDB, Virtual machine

1 引言

云计算作为一种新型的网络计算模式,是当前信息技术领域的热门研究问题。云计算的发展理念符合当前低碳经济与绿色计算的总体趋势,得到了世界各国政府和企业的大力倡导与推动,正带来计算领域、商业领域的巨大变革。随着云计算的发展,其应用范围得到进一步的扩展,工业界开始尝试利用云计算技术提供数据应用托管服务,特别是利用数据库即服务(Database as a Service,简称DBaaS)技术将数据库作为服务提供给客户,降低了数据库应用系统部署及运维的成本[1]。如何基于已有的云计算技术,尤其是虚拟化技术,建立支撑多个企业、组织和部门(租户)的不同行业数据应用并为它们提供隐私保护的数据库即服务系统成为DBaaS推广应用的关键问题。

加密是一种常用的保护用户隐私数据的方法,但是目前的大多数加密方案都不支持对密文的运算,如对加密数据进行检索、统计分析等,严重妨碍了云服务商为用户提供更进一步的数据管理和运算服务,从而削弱了云计算的优势。理想的支持隐私保护的可计算加密方法是全同态加密,不过现阶段全同态加密方案在实用性上还存在问题,所增加的计算时间以万倍计。相对于同态加密方案,CryptDB[2]是麻省理工学院(MIT)计算机科学和人工智能实验室(CSAIL)的一个研究项目,首次解决了实用性问题,它将数据嵌套进多个加密层,允许对加密数据进行简单的操作并且计算时间只增加了15~26%左右。

针对数据库即服务系统中数据的隐私保护问题,基于CryptDB系统,本文设计了一种支持行业数据应用托管的数据库即服务系统,允许用户在公共的IT基础设施上利用虚拟机建立具有良好隐私保护的数据库即服务系统。

本文第2节详细介绍了数据库加密系统CryptDB的基本情况、采用的数据加密策略和加密查询的实现技术等;第3节介绍支持行业数据应用托管及数据隐私保护的数据库即服务系统的总体架构和关键技术;第4节对本文提出的系统进行了实现验证和性能测试;最后对全文内容进行了总结,并提出了下一步的研究方向。

2 CryptDB数据库加密系统

2.1 数据加密策略

CryptDB采用的加密策略有多种,可以分为6类:RND、DET、OPE、HOM、JOIN、Search,这些加密策略使用的加密算法既包括成熟的加密算法比如AES、Paillier等,也包括为特定操作符(join)设计的加密算法,所有加密算法都是对称加密

到稿日期:2013-02-01 返修日期:2013-04-15

陈 萍(1976—),女,硕士,副教授,主要研究方向为信息安全、数据工程,E-mail:chenpin0361@sina.com;张 涛(1974—),男,博士,教授,主要研究方向为信息安全、数据工程。

算法。下面介绍6类加密策略。

(1)RND(Random):是CryptDB提供的最安全的加密策略。RND的加密特征是两个相同的明文加密后密文不同,因而不能有效地支持基于密文的任何计算。RND的实现算法主要采用分组密码算法AES或Blowfish,算法采用CBC模式并且需要一个初始化向量 IV。

(2)DET(Deterministic):加密特征是相同的明文加密后的密文相同。采用这种加密策略会泄露部分信息给DBMS,因此相对来说安全性稍弱。但是这种加密方式允许DBMS服务器基于加密数据执行等值检查计算,包括等值连接、Group By、Count、Distinct等操作。DET的实现采用分组密码算法AES,算法采用CMC模式并且不需要初始化向量 IV。

(3)OPE(Order-preserving encryption):加密特征是明文的大小排序关系依然保留在密文中,亦即:$\forall x \forall y(x<y) \rightarrow OPE_k(x)<OPE_k(y)$,因此这种加密策略支持范围查询,也就是如果查询明文$[c1,c2]$内的数据可以等价地转化为查询密文范围$[OPE_K(c1),OPE_K(c2)]$内的数据,这种加密策略支持DBMS服务器执行Order by、Min、Max、Sort等查询。OPE的实现算法是Boldyreva等[3]在2009年提出的算法。

(4)HOM(Homomorphic encryption,同态加密):该加密策略允许服务器基于加密数据执行计算而最终的结果由CryptDB proxy解密。虽然全同态加密由于性能的原因不可取,但是对于有限的几个操作实现全同态加密是可行的。例如 $HOM_k(x)\times HOM_k(y)=HOM_k(x+y)$,为了计算 $x+y$,CryptDB proxy用UDF代替SUM(x,y)操作(即 $x+y$),该UDF执行 $HOM_k(x)\times HOM_k(y)$,解密后即为 $x+y$ 的值。通过DBMS服务器分别返回几个数的和以及参与运算数据的数目。HOM还能够用来计算平均值,但是对于变量值的自增长比如set id=id+1,目前还不支持。HOM的实现算法是Paillier[4]算法。

(5)JOIN(JOIN和OPE-JOIN):在两列之间执行等值连接需要一个单独的加密算法,原因是如果采用DET加密策略,为了防止获知列之间的关系,不同的列加密密钥是不同的,因此不能直接基于DET加密策略执行连接。JOIN加密策略除了适用于连接操作外,对于所有DET所支持的查询操作也都支持,而OPE-JOIN支持不等值连接。

(6)Search:这种机密模式是用来支持执行SQL中的Like查询的。Search只支持关键字搜索,不支持正则表达式匹配。加密实现算法是Song等[5]提出的算法。

2.2 可动态调整的加密策略

如果应用程序对某列数据的查询不涉及比较和排序操作,考虑到隐私保护的需要,该列应该用安全性最强的RND加密策略;而当查询涉及到等值比较但不要求排序时,采用DET加密就足够了。但是查询并不总是能提前知道,所以需要一个自适应调整方案,CryptDB可以基于查询动态调整DBMS中数据的加密策略,将加密策略调整到能够运行所请求的查询的最安全的加密策略。

在CryptDB系统中,每一个数据嵌套进多个加密圈(加密圈由多个加密策略层组成,越往外层加密策略的安全强度越高,也形象地称为"洋葱加密圈"),每一种加密圈支持一类操作,通常有4种加密圈,如图1所示。其中Eq加密圈支持等值比较操作;包括等值选择和等值连接操作;Ord加密圈支持排序操作,Search加密圈支持字符串型数据的关键字匹配操作,该加密层对整数列没有意义;Add加密圈支持数值型数据的求和计算,对于字符串列是没有意义的。图2中Employees表包括两列,其中ID列定义为整数型,而Name列定义为字符串型。因此对于ID列的值,需要分别记录其对应的Eq、Ord和Add加密层的值,以及RND加密策略所需要的初始化向量 IV,在图中对应的列名分别为C1-Eq,C1-Ord,C1-Add,C1-IV。而Name列的值需要分别记录其对应的Eq、Ord和Search加密层的值,以及RND加密策略所需要的初始化向量 IV,在图中对应的列名分别为C2-Eq,C2-Ord,C2-Search,C2-IV,这样就完成了原始的Employees表的加密。这种加密方式不仅对数据进行了隐藏,而且加密后列的数量、列名、表名都发生了改变,进一步保护了隐私数据。

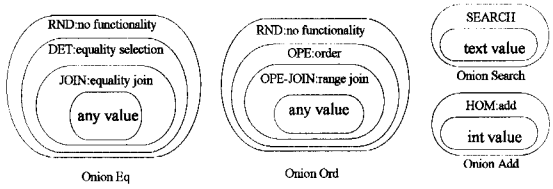

图1 洋葱加密圈

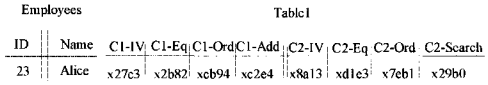

图2 加密示例

CryptDB对于表中同一列的值在相同的加密层上使用相同的密钥来加密,而不同的表、列、加密圈、加密层则采用不同的加密密钥。所有这些加密秘钥来源于主键MK。例如,对于表 t、列 c、洋葱 o 和加密层 l,使用的密钥:

$$K_{t,c,o,l}=PRP_{MK}(\text{table } t,\text{column } c,\text{onion } o,\text{layer } l)$$

其中,PRP 是一个伪随机序列,主密钥保存在JDBC或ODBC客户端,在对托管数据加密时将其传递给CryptDB,CryptDB根据上述公式获得加解密的密钥。

每个加密圈最外层采用最安全的加密方案,在初始情况下,所有的数据一直加密到最外层,比如Eq加密圈加密过程为数据依次利用JOIN加密策略、DET和RND加密策略得到加密后的结果。当接收到SQL查询时,CryptDB根据查询类型决定是否需要剥除当前加密层,例如如果一个查询需要在列 c 上执行谓词 P,CryptDB首先需要确定利用哪个加密圈的哪个加密层来执行 P,如果 c 当前不在合适的加密层,则剥掉一些加密层直到允许 P。由于采用的加密策略都是对称加密算法,因此剥除加密层实际上就是对数据进行解密。CryptDB通过在DBMS服务器上执行UDF实现加密层解密,例如,在图2中,为了将第2列从Ord加密圈的RND层剥除到OPE层,CryptDB proxy提交以下查询到服务器,查询中的DECRYPT_RND为UDF,K 是密钥生成公式计算得到的密钥。

```
UPDATE Table1
SET C2-Ord=DECRYPT_RND(K,C2-Ord,C2-IV)
```

3 支持行业数据应用托管及数据隐私保护的数据库即服务系统

3.1 系统设计

行业数据应用托管不同于面向公众的数据托管(如Amzon的RDS等互联网公司提供的数据托管服务),在数据服务

方面存在数据隔离要求、性能隔离要求、可靠性保障以及数据隐私保护要求等特定要求。

针对数据隔离、性能隔离、可靠性保障要求，文献[6]中提出了基于虚拟机的数据应用托管方法及相应的数据库即服务系统。为了还能支持托管数据的隐私保护，本文在文献[6]中提出的系统基础上设计了图3所示的数据库即服务系统。该方法的基本思想是通过为租户提供基于独立虚拟机的数据托管环境，满足不同租户数据库之间的数据隔离和性能隔离，并且由于不同租户数据库可存在于同一服务器，从整体上减少了系统资源(CPU、内存和IO)的使用量；同时为了提高托管数据的可靠性保障，每个租户数据库至少建立两个数据库副本，租户的数据库副本部署在虚拟机上；为了满足托管数据的隐私保护问题，系统在租户端通过CryptDB proxy对数据进行加密，而托管服务端可基于加密数据执行SQL查询，实现了服务端在不解密的情况下获得查询结果，查询结果经过CryptDB proxy解密后交给应用程序。支持数据隐私保护的数据库即服务系统包括以下3个层次的内容：

(1)数据托管应用层。该层次由归属不同租户的众多数据库应用系统构成，由于数据托管管理层提供了标准的数据库服务，应用系统可以基于标准的数据库接口(JDBC、ODBC等)来完成对数据存储和数据访问等请求，数据存储和访问请求通常是以标准SQL的形式提交，经过CryptDB系统的加密处理，一方面实现了对存储数据的隐私保护，另一方面还可基于加密数据执行SQL查询。

(2)数据托管管理层。该层次对托管的租户数据库进行统一的管理、监控，并根据数据请求负载对承载数据库的虚拟机进行动态资源调度。

(3)数据托管基础设施层。该层次建立在一组共享的物理服务器基础上，通过虚拟机的形式提供对租户数据库的支撑，包括租户数据库虚拟机的创建及撤销、计算及存储等物理资源的分配与监控。虚拟机数据库中存储的数据均是经过CryptDB加密后的数据，为了支持基于加密数据的SQL查询，该层虚拟机数据库系统配置了CryptDB UDF(User-defined function，用户定义函数)，通过执行UDF实现对加密数据的计算和解密等操作。

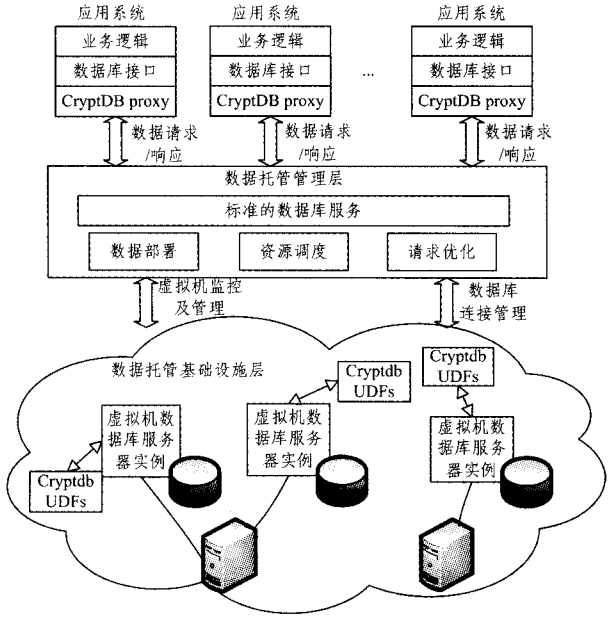

图3 支持隐私保护的数据库即服务系统层次图

上述数据库即服务系统采用CryptDB系统实现数据的隐私保护，包括一个CryptDB proxy和一组UDF。在图3中应用系统通过数据库接口如JDBC、ODBC等提交SQL查询，所有SQL查询首先由CryptDB proxy处理。CryptDB proxy加密查询中涉及到的常量、必要时改变查询操作符、重写查询，以在加密数据上执行查询，同样对于DBMS返回的结果也是首先由CryptDB Proxy解密，解密后传递给应用程序。在这种体系结构下，DBMS中存储的是加密后的数据，所以即使是托管系统的DBMS管理员也无法获得敏感数据，从而保证了数据的隐私性。

3.2 基于加密数据的查询示例

在支持隐私保护的数据库即服务系统中，数据的加解密过程对于应用系统来说是透明的。下面以图2中的加密示例为例介绍系统中SQL语句查询执行的过程。初始情况下，图3中虚拟机DBMS中的数据处于各加密圈的最外层，当应用系统提交如下查询：

```
SELECT ID
FROM Employees
  WHERE Name='Alice';
```

查询首先由CryptDB proxy进行处理，由于查询涉及到等值比较，需要将查询条件中的属性Name对应的Eq加密圈的加密层从RND下降为DET层。因此，CryptDB proxy首先向虚拟机DBMS提交如下修改语句：

```
UPDATE Table1
  SET C2-Eq=
    DECRYPT_RND(K_T1,C2,Eq,RND, C2-Eq, C2-IV);
```

接下来提交SQL查询：

```
SELECT C1-Eq, C1-IV
  FROM Table1
    WHERE C2-Eq=x7..d;
```

这里C1对应列ID，x7..d是“Alice”的密文，注意，查询结果需要同时获得*IV*向量以最终能解密。这样，当CryptDB proxy接收到查询结果后，根据保存的各加密圈当前的状态信息，利用合适的解密算法和解密密钥就能获得查询结果的明文信息。

4 实验

在本文设计的数据库即服务系统中，CryptDB给应用系统客户端和DBMS服务端都带来了性能影响，应用系统端性能损耗的原因主要在于查询重写、托管数据加解密等。服务器端主要是由于加密后数据量增大而导致计算和查询时间的损耗。通过一组实验对比测试CryptDB对系统带来的影响，实验结果如图4所示，其中的基准数据是没有隐私保护的数据库及服务系统查询执行情况。实验结果表明，使用CryptDB系统的数据库即服务系统的性能损耗约为20%，相对于同态加密系统来说，这个数据在我们能够容忍的范围内。

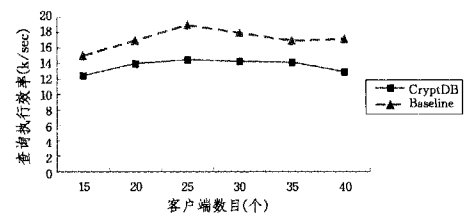

图4 性能开销

(下转第146页)

4.4 实验及结果分析

为了验证漏洞挖掘分析系统 VulAs 的有效性，对含有缓冲区溢出的漏洞程序 strcpy_overflow. exe 进行漏洞的实际挖掘分析。表 1 显示了相应的调用地址、严重程度以及相应的描述。

表 1 程序分析结果的主要内容

Address	Severity	Description
401054	2	UNKNOWN_SOURCE_SIZE: investigated manually
4017f5	8	target buffer is smaller than the source buffer
40182d	2	UNKNOWN_SOURCE_SIZE: investigated manually
4018c8	8	target buffer is smaller than the source buffer
4049c4	2	UNKNOWN_SOURCE_SIZE: investigated manually

由表 1 可见，在 IDA 中查看相应地址发现确实都存在 strcpy()函数的调用，其中在地址 004017F5 调用 strcpy()函数之前，直接将目标 DstBuf 地址传送到 eax 寄存器，但是并没有检查传送内容的长度就直接调用 strcpy()函数，这就为缓冲区溢出造就可能。通过分析说明，漏洞挖掘分析系统对于缓冲区溢出漏洞的检测具有较高的准确性与有效性。

结束语 缓冲区溢出漏洞挖掘分析、利用的研究已经成为当今国内外安全研究的热点，如何解决由缓冲区溢出漏洞所引起的安全问题也成为了安全研究人员的必修课题。本文对缓冲区溢出原理以及漏洞挖掘分析与利用技术进行分析总结，提出了一种基于动静态相结合的漏洞挖掘分析方法，并设计实现了漏洞挖掘分析系统 Vulas，最后通过实验验证了系统的有效性与准确性。

本文的半自动化漏洞挖掘分析系统虽然可以在 Windows 平台下对漏洞挖掘起到一定的作用，但是对于其他平台比如 Linux，还没有进行验证，需要进行下一步的验证。同时对于漏洞挖掘分析系统 Vulas 中核心的漏洞模型库的设计还有待于完善、完整，而对更加合理高效的漏洞特征描述语言的探索则需要继续改进与研究。

参 考 文 献

[1] Aleph One. Smashing The Stack For Fun And Profit [J]. Phrack, 1996, 7(49)

[2] 邓爽. 缓冲区溢出攻击分析及防范策略研究[D]. 济南：山东大学，2009

[3] 李毅超，刘丹，韩宏，等. 缓冲区溢出漏洞研究与进展[J]. 计算机科学，2008，35(1)：87-89，125

[4] 林志强，夏耐，茅兵，等. 缓冲区溢出研究综述[J]. 计算机科学，2004，31(9)：110-113，160

[5] 王业君，倪惜珍，文伟平，等. 缓冲区溢出攻击原理与防范的研究[J]. 计算机应用研究，2005，22(10)：101-104

[6] 2011 年我国互联网网络安全态势综述[EB/OL]. http://www.cert.org.cn/UserFiles/File/20120319011annualreport.pdf

[7] 彭青白. 缓冲区溢出漏洞的挖掘与利用方法研究[D]. 武汉：华中科技大学，2009

[8] Voas J M, McGraw G. Software Fault Injection: Inoculating Programs Against Errors[M]. John Wiley and Sons, New York, 1998

[9] DaveAitel. TheAdvantages of Block-Based ProtocolAnalysis for Security Testing[R]. Immunity, Inc., 2003

[10] AutoDafe [EB/OL]. http://autodafe.sourceforge.net, http://autodafe.sourceforge.net/docs/autodafe.pdf

[11] Oulu University Secure Programming Group. PROTOS Test-Suite: c06-snmpv1[R]. University of Oulu, Electrical and Information Engineering, 2002

[12] BeyondSecurity. beStrom[EB/OL]. http://www.beyondsecurity.com/bestorm_whitepaper.html

[13] 刘奇旭，张玉清. 基于 Fuzzing 的 TFTP 漏洞挖掘技术[J]. 计算机工程，2007，33(20)：142-147

[14] 李伟明，张爱芳，刘建财，等. 网络协议的自动化模糊测试漏洞挖掘方法[J]. 计算机学报，2011，34(2)：242-255

[15] 杨丁宁，肖晖，张玉清. 基于 Fuzzing 的 ActiveX 控件漏洞挖掘技术研究[J]. 2012，49(7)：1525-1532

[16] Kkqq, bugscam Analysis[J]. 绿盟安全月刊，2004(46)

(上接第 142 页)

结束语 基于虚拟化技术，提出并设计了一种支持行业数据应用托管及数据隐私保护的数据库即服务系统，其允许用户在公共的 IT 基础设施之上利用虚拟机建立具有数据、性能隔离、可靠性保障的独立数据库及相关数据的应用。在数据隐私保护方面，采用 CryptDB 系统对数据进行加密，同时利用多种加密策略以及可动态调整的加密策略技术，解决了基于加密数据执行 SQL 查询的问题。下一步工作将研究在保护数据隐私的同时如何进行数据的优化处理以进一步提高数据处理效率，减少系统开销，提高用户体验。

参 考 文 献

[1] Ashraf A. Deploying database appliances in the cloud[J]. IEEE Data Engineering Bullentin, 2009, 32(1): 13-20

[2] Popa R A, Redfield C M S, Zeldovich N, et al. CryptDB: Protecting Confidentiality with Encrypted Query Processing[C]// Proceedings of the 23rd ACM Symposium on Operating Systems Principles(SOSP 2011). Cascais, Portugal, October 2011

[3] Boldyreva A, Chenette N, Lee Y, et al. Order preserving symmetric encryption[C]// Proceedings of the 28th Annual International Conference on the Theory and Applications of Cryptographic Techniques(EUROCRYPT). Cologne, Germany, April 2009

[4] Paillier P. Public-key cryptosystems based on composite degree residuosity classes[C]// Proceedings of the 18th Annual International Conference on the Theory and Applications of Cryptographic Techniques(EUROCRYPT). Prague, Czech Republic, May 1999

[5] Song D X, Wagner D, Perrig A. Practical techniques for searches on encrypted data[C]// Proceedings of the 21st IEEE Symposium on Security and Privacy. Oakland, CA, May 2000

[6] 王卓昊，王希诚. 面向托管的数据库即服务系统及资源优化技术[J]. 计算机工程与应用，2011，47(27)：19-23