一种全同态加密的安全内积计算方案

邓 江，许春香，杨浩淼

(电子科技大学计算机科学与工程学院 成都 611731)

【摘要】在云计算环境下密文top-*k*检索的众多方法中，该文聚焦于同态加密方法，该公钥加密方法具有不解密就能对密文进行操作的优点。在密文top-*k*查询中，内积相似性是度量索引向量和查询向量的相似性的最常用的一个指标。该文提出一个安全计算两向量内积相似性的方案，该方案使用基于环上错误学习问题的批处理和打包的同态加密来保护隐私。与其他方法相比，该方案具有通信代价低和计算代价低的优点。

关 键 词 中国剩余定理； 全同态加密； 环上错误学习问题； 单指令多数据流

中图分类号 TP309 **文献标志码** A doi:10.3969/j.issn.1001-0548.2016.05.017

A Secure Computation Scheme of Inner Product Based on Fully Homomorphic Encryption

DENG Jiang, XU Chun-xiang, and YANG Hao-miao

(School of Computer Science and Engineering, University of Electronic Science and Technology of China Chengdu 611731)

Abstract Among many approaches to solve the problem of top-*k* retrieval over encrypted cloud data, we focus on an approach with homomorphic encryption, which is public key encryption supporting some operations on encrypted data. In top-*k* retrieval of encrypted data, the inner product is often used as a metric to compute the similarity between the file feature vector and the query vector. In this paper, we propose an efficient scheme to compute the inner product on encrypted data using the homomorphic encryption based on the learning with errors over ring (RLWE) problem, in which batch and packing techniques are adopted to achieve lower computation and communication cost.

Key words Chinese remainder theorem; fully homomorphic encryption; learning with errors over ring; single instruction multiple data

在云计算中，由于数据如何在云中存储对用户来说是不透明的，因此，用户常常担心其存储在云中数据的安全性。虽然可以通过认证、授权、访问控制等传统数据安全方法来保护存储数据，但这些都是建立在云服务提供商是可信任基础上的，而云服务提供商并不都像用户所希望的那样值得信任。2013年的棱镜门事件，就是这样一个侵犯云安全的典型案例。当云存储提供商不那么可信时，保障云安全的一种方法，是对所有的云数据进行加密，但加密限制了数据的使用，特别是查询和索引数据会变得异常困难。因此，如何在云计算环境下进行密文检索是一个具有挑战性的问题。

密文检索的另一重要问题是多关键字排序。从云中检索返回的大量文档，需要进行相关性排序，使用户快速找到最相关的*k*个结果(top-*k*检索)。另一方面，由于单一的关键字搜索往往会产生过于粗糙的结果，为提高搜索结果的准确性以及提升用户搜索体验，还需支持多个关键字的搜索。现在网络搜索引擎(如：谷歌)的通常做法是，用户提供一组关键字来检索最相关的数据。对于多关键字排序的检索，常常采用坐标匹配[1]的原则，即尽可能多的匹配，来度量相似性。在具体实现中，内积相似性是最常用的一种方法，已被广泛用于在云计算环境下的密文top-*k*检索。

在云计算的密文top-*k*检索中，通过云存储文档的特征向量 $\boldsymbol{V}=(V_1,V_2,\cdots,V_l)$ 和用户查询向量 $\boldsymbol{W}=(W_1,W_2,\cdots,W_l)$ 中的各个坐标尽可能多的匹配，来返回最相似的*k*个文档。如何度量这种相似性？一种最常用的方法是计算特征向量和查询向量的内积 $<\boldsymbol{V},\boldsymbol{W}>=\sum_i V_i\cdot W_i$，内积大者更相似。为了保护

收稿日期：2015－03－11；修回日期：2015－06－20
基金项目：国家自然科学基金面上项目(61472065, 61370203)
作者简介：邓江(1982－)，男，博士，主要从事信息安全方面的研究.

用户的隐私，无论是特征向量$\boldsymbol{V}$和查询向量$\boldsymbol{W}$都应该加密。然而加密限制了数据的使用，云服务器很难采用传统的加密方法来安全计算内积。而全同态加密(fully homomorphic encryption, FHE)在不解密的情况下能够直接基于密文计算[2-7]。于是一些学者使用全同态加密的方法来安全计算两个密文向量内积。

1 相关研究

文献[8]基于整数上的全同态加密来安全计算内积，首先使用全同态加密(FHE)对向量的每一坐标分别加密得到$\boldsymbol{V}'=(V_1',V_2',\cdots,V_l')$和$\boldsymbol{W}'=(W_1',W_2',\cdots,W_l')$，其中用$\boldsymbol{X}'$表示对$X$的加密，然后同态计算$<\boldsymbol{V}',\boldsymbol{W}'>$。然而，这种方法虽然简单，但存在巨大问题。首先，当向量维数l比较大时，需要传送$2l$个密文，通信代价很大；在用户端需要计算$2l$个加密，在云服务器端，为了同态计算内积，服务器需要做l个同态乘和$l-1$个同态加，计算量都较大。此外，基于整数的全同态加密的安全性也是值得怀疑的。文献[9]就给出了基于整数的全同态加密的密码分析。

文献[10]在基于生物认证的认证中，利用基于理想格的全同态加密来安全计算内积。使用打包的技巧来将一个明文向量的所有坐标打包到一个密文中。但是该方案的计算效率并不高，如对两个2 048维的bit向量，为了安全计算内积，其预计计算的时间需要38 s。此外，该方案也不适合计算其他相似性指标，如余弦相似性。

因此，本文提出了一种基于全同态加密的安全计算内积方案，该方案计算量低、通信开销小、针对整数向量而不仅是bit向量，安全性高。

2 本文方案

本文使用基于格上的环上错误学习问题(learning with errors over ring, RLWE)的同态加密方法，安全高效地计算两个高维向量的内积。其主要思想是用户首先将明文向量的所有坐标打包到一个密文里，然后云服务器以单指令多数据流(simple instruction multiple data，SIMD)的方式并行计算两个向量的内积。方案具体包括3个部分：基于RLWE的Somewhat同态加密方案的构造、打包和SIMD并行化技术和安全内积计算。

2.1 基于RLWE的Somewhat同态加密方案的构造

本文基于RLWE来构造Somewhat同态加密方案，该方案基于方案SHE(somewhat homomorphic encryption)[11]上为多项式时间算法的六元组(Setup, KeyGen, Enc, Dec, Add, Mult)，具体描述如下：

参数设置(SHE.Setup)：根据输入的安全参数λ，建立多项式环R和离散高斯分布X。设多项式x^n+1的次数$n=n(\lambda)$，离散高斯分布$X=X(\lambda)$，使得RLWE问题对已知格攻击是2^λ级安全的。令$R=\mathbb{Z}[x]/(x^n+1)$，明文空间为$R_2=\mathbb{Z}_2[x]/(x^n+1)$，密文空间则为$R_q=\mathbb{Z}_q[x]/(x^n+1)$，输出公共参数$\text{params}=(q,n,X)$。

密钥生成(SHE.KeyGen)：在高斯分布X上随机选择一个元素$s\leftarrow X$作为私钥，然后在R_q上随机均匀选取一个元素$a_1\leftarrow R_q$，同时在高斯分布X上随机选取一个差错$e\leftarrow X$，计算$a_0=-a_1s+2e$。令私钥$\text{sk}=s$，公钥$\text{pk}=(a_0,a_1)$。

加密(SHE.Enc)：给定单比特消息$m\in\{0,1\}$，在高斯分布X上分别随机选取元素$u\leftarrow X$和元素$r\leftarrow X$，根据公钥$\text{pk}=(a_0,a_1)$分别计算出密文$c_0=a_0u+m$与$c_1=a_1u+2r$，输出密文$c=(c_0,c_1)$。

解密(SHE.Dec)：根据给定的密文$c=(c_0,c_1)$，利用私钥$\text{sk}=s$，计算出$m'=((c_0+c_1s)\bmod q)\bmod 2$。

求和(SHE.Add)：输入两个密文c_1,c_2，输出这两个密文的和c_1+c_2。

乘积(SHE.Mult)：输入两个密文c_1,c_2，输出这两个密文的乘积$c_1\times c_2$。

2.2 打包和SIMD技术

文献[12]在基于理想格的全同态加密方案引入了打包和单指令多数据流(simple instruction multiple data, SIMD)方法，本文将该方法用于上述基于RLWE的SHE方案中，具体过程描述如下。

分解明文空间$R_2=\mathbb{Z}_2[x]/(x^n+1)$：虽然多项式$x^n+1$在$\mathbb{Z}$上是不可约的，但可选择合适的$n$值，使$x^n+1$在$\mathbb{Z}_2$上是可约的。由此，将其分解为$x^n+1=f_1(x)f_2(x)\cdots f_l(x)$，其中$f_i(x)$，$i=1,2,\cdots,l$的次数为$d=n/l$。于是明文空间被分解为$l$个明文槽：$\mathbb{Z}_2[x]/(x^n+1)=\mathbb{Z}_2[x]/(f_1(x))\times\mathbb{Z}_2[x]/(f_2(x))\times\cdots\times\mathbb{Z}_2[x]/(f_l(x))$。即在原来的明文空间中，明文$m(x)$是$R_2$上的多项式，有$n$个比特，通过中国剩余定理(Chinese remainder theorem, CRT)，$m(x)$可被分解为：$m_1(x),m_2(x),\cdots,m_l(x)$，分别对应到$l$个明文槽$\mathbb{Z}_2[x]/(f_1(x)),\mathbb{Z}_2[x]/(f_2(x)),\cdots,\mathbb{Z}_2[x]/(f_l(x))$，对每个$m_i(x)$，$i=1,2,\cdots,l$，均有$d$个比特。令明文$m(x)$对应的密文为$c(x)$，通过这种方式，可将$l$个明文

$m_1(x),m_2(x),\cdots,m_l(x)$ 打包到一个密文 $c(x)$。

在对打包后的密文进行操作时，加同态和乘同态能在这些明文槽上并行地执行，这样极大地减轻了计算量。具体过程为：令一个打包后的密文为 $c(x)$，对应在 l 个明文槽的明文分别为 $m_1(x),m_2(x),\cdots,m_l(x)$，令另一个打包后的密文为 $c'(x)$，对应在 l 个明文槽的明文则分别为 $m'_1(x),m'_2(x),\cdots,m'_l(x)$，加同态SHE. Add (c,c') 对应在各个明文槽的明文就分别为：$m_1(x)+m'_1(x),\ m_2(x)+m'_2(x),\cdots,m_l(x)+m'_l(x)$，可以得到乘同态SHE. Mult (c,c') 对应在各个明文槽的明文为：$m_1(x)\times m'_1(x) \bmod f_1(x),\ m_2(x)\times m'_2(x) \bmod f_2(x),\cdots,\ m_l(x)\times m'_l(x) \bmod f_l(x)$。因此同态操作可以看成是在明文槽上使用SIMD并行执行的。

自同构映射能使对应在各个明文槽的明文发生置换：当明文为 $m(x)$，其对应在各个明文槽的明文就分别为 $m_1(x),m_2(x),\ m_3(x),\cdots,m_l(x)$，Frobenius自同构映射 $m(x)$，使 $m(x)\to m(x^{2^i})$ 明文槽循环移动 i 个明文槽。如，当 $i=1$ 时，$m(x)$ 的明文槽就循环移动1个位置，其对应于各个明文槽的明文就发生变化，变更为 $m_2(x),m_3(x),\cdots,m_l(x),m_1(x)$。因此，就可以通过操纵相应的密文 $c(x)$，并利用密钥更新的技巧(密钥 $s(x^k)$ 从更新为 $s(x)$)，使各个明文槽中的明文发生置换。

2.3 安全内积计算

如前所述，基本方案只能对向量中的每一个坐标分别进行加密。本文将向量打包成一个单一的密文，再以SIMD并行的方式计算两个向量的内积。计算内积的过程如下：

1) 系统选择合适的参数 n 和 l，使得明文空间 $R_2=\mathbb{Z}_2[x]/(x^n+1)$ 分解为 l 个明文槽 $\mathbb{Z}_2[x]/(f_1(x)),\ \mathbb{Z}_2[x]/(f_2(x)),\cdots,\mathbb{Z}_2[x]/(f_l(x))$，其中 $f_i(x)$，$i=1,2,\cdots,l$ 的次数为 $d=n/l$；

2) 用户 A 拥有向量 $\boldsymbol{V}$，对于 l 维向量 $\boldsymbol{V}=(V_1,V_2,\cdots,V_l)\in\mathbb{Z}^l$，将 V_i 分别编码为 $m_i(x)$，其中，$i=1,2,\cdots,l$；

3) 通过中国剩余定理将 $m_i(x)$ 打包为 $m(x)$，并调用加密算法得到 $V(x)\leftarrow \mathrm{SHE.Enc}(pk,m(x))$；

4) 类似地，用户 B 通过同样步骤将 l 维向量 $\boldsymbol{W}=(W_1,W_2,\cdots,W_l)\in\mathbb{Z}^l$ 打包到 $W(x)$，本文以用户 A 将向量 $\boldsymbol{V}$ 打包为 $\boldsymbol{V}(x)$ 为例，具体过程如图1所示；

5) 云服务器调用SHE.Mult $(V(x),W(x))$，如表1所示，这一步相当于以SIMD的方式并行计算 $V_1(x)\times W_1(x) \bmod f_1(x)$，$V_2(x)\times W_2(x) \bmod f_2(x)$，$\cdots$，$V_l(x)\times W_l(x) \bmod f_l(x)$；

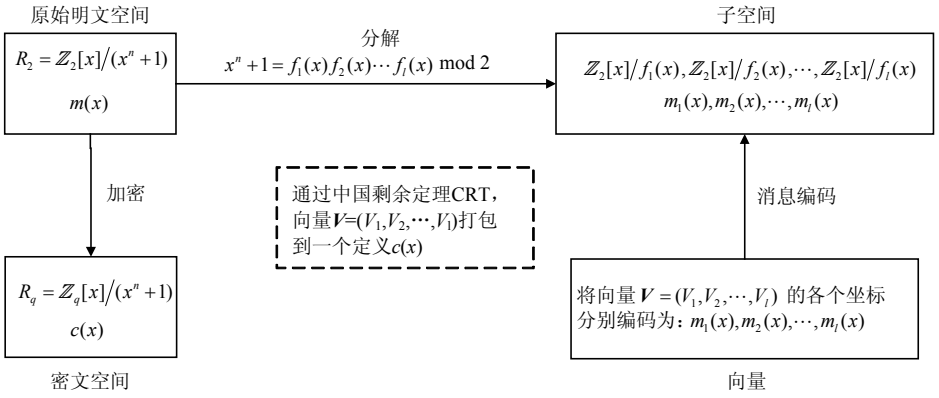

图1 用户将明文向量的所有坐标打包到一个密文

表1 基于SIMD并行计算 $\{V_i\cdot W_i,1\leqslant i\leqslant l\}$

打包后的密文	对应于各自空间的明文
$V'(x)$	$(V_1(x),V_2(x),\cdots,V_l(x))$
$W'(x)$	$(W_1(x),W_2(x),\cdots,W_l(x))$
乘同态 $V'(x)W'(x)$	$(V_1(x)W_1(x) \bmod f_1(x),(V_2(x)W_2(x) \bmod f_2(x),\cdots,V_l(x)W_l(x) \bmod f_l(x))$

6) 对上一步的结果调用 $\log l$ 个Frobenius自同构映射以及 $\log l$ 个同态加法，同态评估得到 $c(x)$，其对应的明文为 $m(x)$；

7) 通过调用算法SHE.Dec $(\mathrm{sk},c(x))$，计算出 $m(x)$ 为：$m(x)=\sum_{i=1}^{l}V_i(x)\cdot W_i(x)$；

8) 令 $x=2$，代入 $m(x)$，得到 $<\boldsymbol{V},\boldsymbol{W}>$。

3 方案比较

将本文方案和已有的两个基于全同态加密的安全内积计算方案进行比较，结果如表2、表3所示。

表 2 三种基于全同态加密的方案比较

方案	所使用的FHE	是否针对整数向量	是否使用打包	是否使用SIMD
文献[8]	基于整数	是	否	否
文献[10]	基于理想格	否	是	否
本文方案	基于RLWE	是	是	是

表3 两个方案的性能比较

方案	通信量	计算量
文献[2]	2l个密文	2l个加密，l个同态乘，$l-1$ 个同态加
本文方案	2个密文	2个加密，1个同态乘，log$l$个同态加与自同构映射

从表中可以看出，本文方案使用基于RLWE的FHE计算两个向量的内积，不仅针对bit向量，还可以针对一般的整数向量，这一点与文献[8]的方案相同，而文献[10]的方案只能针对bit向量，因此，本文方案和文献[8]的方案在实际应用中更加广泛。而与文献[8]的方案相比，本文方案在计算中使用了打包技巧，将一个向量的所有坐标打包到一个密文中，同时使用了SIMD技巧以进行并行化处理，使得本文方案有着较高的计算效率。就通信成本而言，两个方案中的明文空间均被划分为l个明文槽，因此本文方案有着较低的通信成本和计算代价。

4 结 束 语

本文提出一个适合于云计算中top-k检索的安全计算两向量内积相似性的方案，该方案使用基于RLWE的批处理同态加密来保护隐私。与其他方案相比，该方案具有通信开销小和计算代价低的优点。此外，在社交网络中的群体挖掘，以及基于生物特征的认证等场景，也需要用到相似性度量。因此下一步的工作将研究如何将本文提出的方案应用到其他场景。

参 考 文 献

[1] WITTEN H, MOFFAT A，BELL T C. Managing gigabytes: Compressing and indexing documents and images[M]. San Francisco: Morgan Kaufmann Publishing, 1999.

[2] GENTRY C. Fully homomorphic encryption using ideal lattices[C]//Proceedings of the 41st Annual ACM Symposium on Symposium on Theory of Computing-STOC'09. [S.l.]: ACM Press, 2009: 169-178.

[3] SMART N P, VERCAUTEREN F. Fully homomorphic encryption with relatively small key and ciphertext sizes[C]//Public Key Cryptography-PKC 2010. Berlin: Springer, 2010: 420-443.

[4] STEHLE D, STEINFELD R. Faster fully homomorphic encryption[C]//Advances in Cryptology-ASIACRYPT 2010. Berlin: Springer, 2010: 377-394.

[5] GU Chun-sheng. Cryptanalysis of the Smart-Vercauteren and Gentry-Halevi's fully homomorphic encryption[J]. International Journal of Security & Its Applications, 2012, 6(2): 176-184.

[6] CORON J S, MANDAL A, NACCACHE D, et al. Fully homomorphic encryption over the integers with shorter public keys[C]//Advances in Cryptology-CRYPTO 2011. Berlin: Springer, 2011: 487-504.

[7] GENTRY C, HALEVI S. Implementing Gentry's fully-homomorphic encryption scheme[C]//Advances in Cryptology-EUROCRYPT 2011. Berlin: Springer, 2011: 129-148.

[8] YU J, LU P, ZHU Y, et al. Towards secure multi-keyword top-k retrieval over encrypted cloud data[J]. IEEE Transactions on Dependable and Secure Computing, 2013, 10(4): 239-250.

[9] DING J, TAO C. A new algorithm for solving the general approximate common divisors problem and cryptanalysis of the fhe based on the gacd problem[DB/OL]. [2014-04-07]. http://eprint.iacr.org/2014/042.

[10] YASUDA M, SHIMOYAMA T, KOGURE J, et al. Packed homomorphic encryption based on ideal lattices and its application to biometrics[J]. Security Engineering and Intelligence Informatics, 2013(8128): 55-74.

[11] NAEHRIG M, LAUTER K, VAIKUNTANATHAN V. Can homomorphic encryption be practical?[C]//Proceedings of the 3rd ACM Workshop on Cloud Computing Security Workshop. Rome: ACM Press, 2011: 113-124.

[12] SMART N P, VERCAUTEREN F. Fully homomorphic SIMD operations[J]. Designs, Codes and Cryptography, 2014, 71(1): 57-81.

编 辑 蒋 晓