同态加密原理及应用

叶明聪 1901210694
刘华兴 1901210446

1 同态加密背景

随着互联网的发展和云计算概念的诞生，以及人们在密文搜索、电子投票、移动代码和多方计算等方面的需求日益增加，同态加密(Homomorphic Encryption)变得更加重要。同态加密是一类具有特殊自然属性的加密方法，此概念是 Rivest 等人在 20 世纪 70 年代首先提出的，与一般加密算法相比，同态加密除了能实现基本的加密操作之外，还能实现密文间的多种计算功能，即先计算后解密可等价于先解密后计算。

这个特性对于保护信息的安全具有重要意义，利用同态加密技术可以先对多个密文进行计算之后再解密，不必对每一个密文解密而花费高昂的计算代价；利用同态加密技术可以实现无密钥方对密文的计算，密文计算无须经过密钥方，既可以减少通信代价，又可以转移计算任务，由此可平衡各方的计算代价；利用同态加密技术可以实现让解密方只能获知最后的结果，而无法获得每一个密文的消息，可以提高信息的安全性。正是由于同态加密技术在计算复杂性、通信复杂性与安全性上的优势，越来越多的研究力量投入到其理论和应用的探索中。

近年来，云计算受到广泛关注，而它在实现中遇到的问题之一即是如何保证数据的私密性，同态加密可以在一定程度上解决这个技术难题。

2 同态加密的原理

2.1 同态加密的定义

Craig Gentry 给出的直观定义：

A way to delegate processing of your data, without giving away access to it.

定义 1 同态性：

对于加密算法ε和明文域 P_ε上的运算。,若$\forall p_1, p_2, \dots, p_n \in P_\varepsilon$都满足式(1):

$$\mathrm{Dec}_\varepsilon(\mathrm{key},\ \mathrm{Cal}_\varepsilon((c_1, \dots, c_n),。)) = (P_1, \dots, P_n)。 \quad (1)$$

其中， Cal_ε为密文运算算法， key 为密钥;

定义 2 部分同态加密算法：

对于加密算法ε和明文域 P_ε上的运算(+,*),若$\forall p_1, p_2, \ldots, p_n \in P_\varepsilon$仅满足加法或者乘法运算在式(1)中成立,则称加密算法ε为满足部分同态的加密算法

一般的加密方案关注的都是数据存储安全。没有密钥的用户，不可能从加密结果中得到有关原始数据的任何信息。我们注意到，这个过程中用户是不能对加密结果做任何操作的，只能进行存储、传输。对加密结果做任何操作，都将会导致错误的解密，甚至解密失败。

同态加密方案最有趣的地方在于，其关注的是数据处理安全。同态加密提供了一种对加密数据进行处理的功能。也就是说，其他人可以对加密数据进行处理，但是处理过程不会泄露任何原始内容。同时，拥有密钥的用户对处理过的数据进行解密后，得到的正好是处理后的结果。

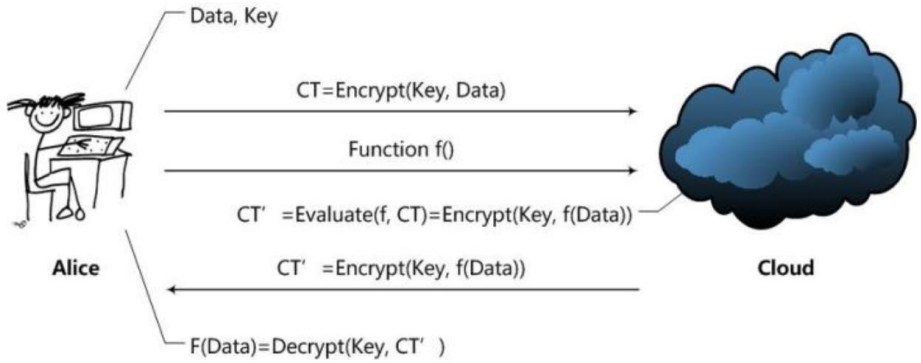

图 2-1 云场景下同态加密过程

以云计算应用场景为例[1],如图 1 所示。Alice 通过 Cloud，以 Homomorphic Encryption（以下简称 HE）处理数据的整个处理过程大致是这样的：

(1)Alice 对数据进行加密。并把加密后的数据发送给 Cloud；

(2)Alice 向 Cloud 提交数据的处理方法，这里用函数 f 来表示；

(3)Cloud 在函数 f 下对数据进行处理，并且将处理后的结果发送给 Alice；

(4)Alice 对数据进行解密，得到结果。

2.2 密文计算

密文计算(cipher-text computation)是指在密文域上所进行的计算以及具有访问权限的用户对密文域上的计算结果可确认并可解密获得对应的明文.为了确保用户隐私数据安全，需要将隐私数据进行加密处理后上传到云端存储,参与计算的密文数据主要包括两部分，分别由用户直接提供以及通过密文检索得到的数据(服务商受托方对用户交付的数据作外包计算)[2]。同态加密为云计算环境中存储与外包计算等服务的隐私安全问题提供了良好的解决方案，理论上利用全同态加密算法能够从根本上解决在第三方不可信或半可信平台上进行数据存储和数据

操作时的隐私保护问题，用户将计算请求 F 和⟨⟩密文$\langle c_1,\dots,c_t\rangle=(\mathrm{Enc}(m_1),\dots,\mathrm{Enc}(m_t))$发送给云端,在云端对密文直接进行任意运算，所得到的密文结果与明文运算后结果一致，即 $\mathrm{Y}(\mathrm{K},\mathrm{F},(\mathrm{Enc}(m_1),\dots,\mathrm{Enc}(m_t)))=\mathrm{E}\big(\mathrm{K},\mathrm{F}(m_1,\dots,m_t)\big)$，同态的概念源于近世代数中群与环的同态，设⟨H1, ∘⟩、⟨H2, *⟩为两个代数结构，f：H1 → H2为 H1 到 H2 的一个映射，∀a,b∈H1 都有 $\mathrm{f}(\mathrm{a}\circ\mathrm{b})=\mathrm{f}(\mathrm{a})*\mathrm{f}(\mathrm{b})$，则称 f：H1 → H2是一个同态映射。一个同态加密算法ε包括 4 个部分，分别是密钥生成算法Gen_ε、加密算法Enc_ε、解密算法Dec_ε、密文运算算法Cal_ε [3]。

(1) 密钥生成算法Gen_ε:U→key 表示用户通过输入参数 U 生成密钥 key;

(2) 加密、解密算法: Enc_ε：$(\mathrm{key},\mathrm{P}_\varepsilon)$ Cε ,Decε: $(\mathrm{key},\mathrm{C}_\varepsilon)\rightarrow\mathrm{P}_\varepsilon$，$P_\varepsilon$为明文空间， C_ε为密文空间;

(3) 计算算法: Cal_ε: $(P_\varepsilon,\ F_\varepsilon\ )\rightarrow(C_\varepsilon,\ F_\varepsilon)$, $\circ F_\varepsilon$,$(p_1,p_2,\dots,p_n)\in\mathrm{P}_\varepsilon$, F_ε是P_ε上的运算集合，对于$\circ\in F_\varepsilon$ $(p_1,p_2,\dots,p_n)\in P_\varepsilon$，$Cal_\varepsilon$将$P_\varepsilon$上进行的运算∘转化为$C_\varepsilon$上运算再进行计算，结果是等价的。

2.3 Paillier 加密方案

密钥生成过程$\mathrm{Gen}_{\mathrm{Paillier}}$ 如下:

随机选取 p、q 两个大素数以及$\mathrm{g}\in\mathrm{Z}_{n^2}$ ，令 n＝pq，λ =(p-1)(q-1)，设函数 $l(u)=\frac{u-1}{n}$， 且 g、n 满足:

$$\gcd(l(g^{\lambda}\bmod n^2),n)=1 \qquad (2)$$

这里，公钥为 pk = (n, g),私钥 sk=λ.

加密算法$\mathrm{Enc}_{\mathrm{Paillier}}$:随机选取整数$\mathrm{r}\in z_{n^2}^*$ ， 对于明文$m\in\mathrm{Z}_n$， 加密后的密文 c 为：

$$c=\ g^m r^n\ mod\ n^2 \qquad (3)$$

式(3)中，$\mathrm{c}\in z_{n^2}^*, z_{n^2}^*$为小于 n^2且与 n^2互素的正整数集合.

解密算法Dec_{Paillier}：对于密文 c，其对应的明文 m 为

$$m=\frac{l(c^{\lambda}\bmod n^2)}{l(g^{\lambda}\bmod n^2)}\ \bmod \mathrm{n} \qquad (4)$$

同态属性分析：

由于$\mathrm{Enc}(\mathrm{m}_1)\mathrm{Enc}(\mathrm{m}_2)=(g^{\mathrm{m}_1}r_1^n)(g^{\mathrm{m}_2}r_2^n)=g^{\mathrm{m}_1+\mathrm{m}_2}(\mathrm{r}_1\mathrm{r}_2)^n=\mathrm{Enc}(\mathrm{m}_1+\ \mathrm{m}_2)\bmod n^2$，即明文加法运算对应密文乘法运算，所以方案具备加法同态性:

$$Dec_{paillier}(\mathrm{key},Cal_{paillier}((c_1,\dots,c_n),\mathrm{x}))=(m_1,\dots,m_n)+ \qquad (5)$$

3 同态加密的应用

同态加密技术在分布式计算环境下的密文数据计算方面具有比较广泛的应用领域，比如云计算、多方保密计算、匿名投票等。

同态加密的一般性框架如下所示：

加密数据处理方法简述如下：假设存在同态加密函数$Enc_k(x)$，首先用户用自己的私钥 k 对需要处理的数据进行同态加密$c = Enc_k(m)$，然后将加密数据 c 上传到云端服务器。服务器能够对加密数据 c 直接进行处理，得到$c' = f(c) = f(Enc_k(m))$，然后将处理后的密文c'返回给用户。用户收到$c' = f(c) = f(Enc_k(m)) = Enc_k\big(f(m)\big)$后，利用自己的私钥 k 对其进行同态解密，得到已经处理好的明文数据$f(m)$。

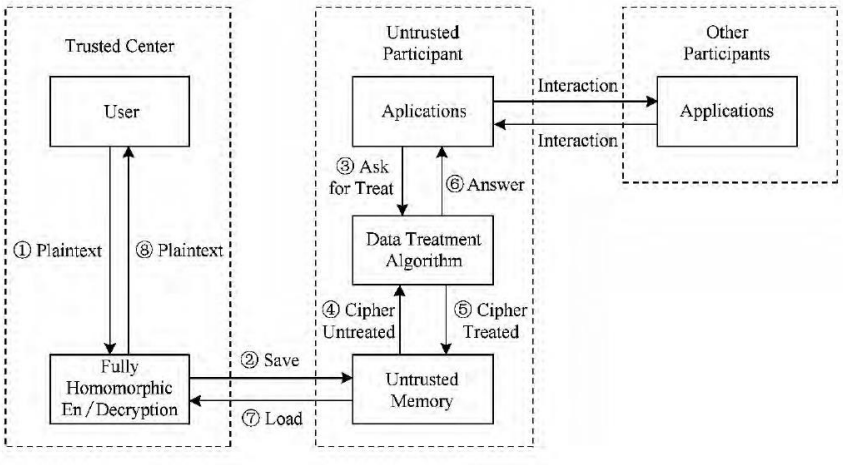

图 3-1 同态加密的一般性框架

3.1 同态加密与机器学习

机器学习模型一般都很大，而用户的设备实际上可能没有足够的存储空间或算力来运行模型。机器学习模型一般都会频繁地更新，可能不会想在网络上频繁传输这么大的模型。开发机器学习模型需要大量时间和计算资源，你可能会想通过向使用该模型的用户收费来收回成本。

接下来，常用的解决方案是将模型作为应用程序接口（API）在云上公开。在过去几年间，这些“机器学习即服务”产品如雨后春笋般涌现，每个主要的云平台都会为企业级开发者提供这样的服务[4]。

但这类产品的潜在用户所面对的困境也是显而易见的——处理用户数据的远程服务器可能并不可信。这样就会存在明确的伦理和法律的分歧，从而限制这种解决方案的有效范围。在受监管的产业（尤其是医疗业和金融业）中，一般是不允许将病患或金融数据发送给第三方进行处理的。

同态加密可以在无需进行解密的情况下，直接计算加密数据。在训练图像数据中，用户可以将加密数据（例如图像）传递给云 API，以此运行机器学习模型，并返回加密的答案。整个过程中都没有解密用户数据，尤其是云服务商既不能访问原始图像，也不能解码计算得到的预测值。

3.2 安全云计算与委托计算

同态技术在该方面的应用可以使得我们在云环境下，充分利用云服务器的计算能力，实现对明文信息的运算，而不会有损私有数据的私密性[1]。例如医疗机构通常拥有比较弱的数据处理能力，而需要第三方来实现数据处理分析以达到更好的医疗效果或者科研水平，这样他们就需要委托有较强数据处理能力的第三方实现数据处理（云计算中心），但是医院负有保护患者隐私的义务，不能直接将数据交给第三方。在同态加密技术的支持下，医疗机构就可以将加密后的数据发送至第三方，待第三方处理完成后便可返回给医疗结构。整个数据处理过程、数据内容对第三方是完全透明的。

3.3 文件存储与密文检索

用户可以将自己的数据加密后存储在一个不信任的远程服务器上，日后可以向远程服务器查询自己所需要的信息，存储与查询都使用密文数据，服务器将检索到的密文数据发回。用户可以解密得到自己需要的信息，而远程服务器却对存储和检索的信息一无所知。此种方法同样适用于搜索引擎的数据检索。

3.4 安全多方计算协议设计的工具

所谓安全多方计算就是分别持有私有数据 x1,x2,…,xn 的 n 个人，在分布式环境中协同计算函数 f (x1,x2,…,xn) 而不泄露各方的私有数据。以同态技术加密的密文数据计算不仅可以满足安全多方计算协议设计中保护各方隐私的需要，还能避开不经意传输协议而大大提升协议效率。

3.5 电子投票

基于同态加密技术设计的电子选举方案，统计方可以在不知道投票者投票内容的前提下，对投票结果进行统计，既保证了投票者的隐私安全，有能够保证投票结果的公证[5]。

电子投票在计票的快捷准确、人力和开支的节省、投票的便利性等方面有着传统投票方式无法企及的优越性. 而设计安全的电子选举系统是全同态加密的一个典型应用。下面描述了一个简单的电子选举方案：

1）若有同态函数 Enck(x1+x2)＝Enck(x1)*Enck(x2)，选民将自己的选票进行

加密 Ci＝Enck（Mi），其中Mi∈{0,1}；

2)投票中心收集同态加密后的选民选票Ｃi；投票中心基于全同态加密方案的同态性质对加密后的选票Ｃi进行计票 C=C1*C2*…*Cn，得到经过同态加密后的选举结果Ｃ＝Enck(M1+M2+…+Mn)；3)只有拥有解密密钥的某个可信机构才能够对加密后的选举结果进行解密，公布选举结果．在上述过程中，选票收集与计票完全对加密后的选票数据进行操作，不需要使用任何解密密钥．因此，任何一个主体或机构都可以完成计票员的职责，无论其是否可信．

3.6 数字水印

数字水印技术是指用信号处理的方法在数字化的多媒体数据中嵌入隐蔽的标记，这种标记通常是不可见的，只有通过专用的检测器或阅读器才能提取．如何应对复杂网络环境下数据隐藏与数字水印系统的安全挑战，是目前需要迫切解决的问题[6]。

针对数字水印的一种主要的安全性攻击手段是非授权检测攻击，即攻击者在未经授权的情况下对含有水印的载体进行检测，以确定水印是否存在，进而猜测或破译水印的含义，甚至去除载体中的水印并嵌入一个伪造的水印。文献提出的基于全同态加密的数字水印方案可以有效地抵抗这种攻击。该方案首先利用全同态加密体制对水印信号与原始载体进行加密，然后将加密后的水印嵌入到原始载体中。在用户检测水印之前，必须首先对含有水印的载体进行同态解密，从而保证解密后的水印信号与含水印的载体之间没有明显的相关性．在解密含水印的载体之后，可以通过计算解密后的载体与水印信号之间的相关度，判断水印的存在性进而提取水印。图 3 -2 描述了传统的数字水印方案与基于全同态加密的数字水印方案的区别：

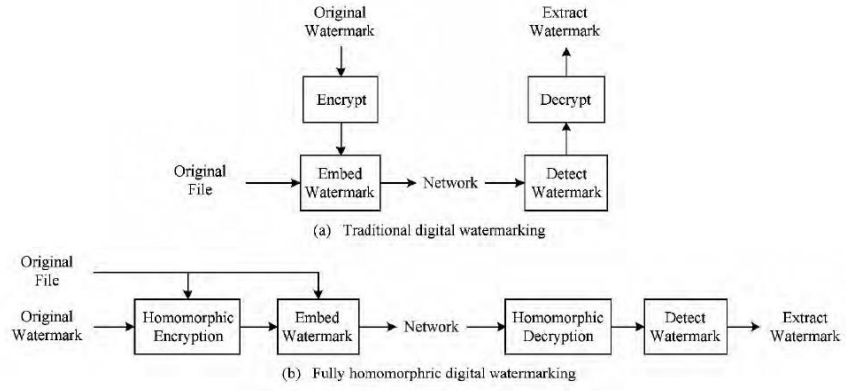

图 3-2 传统数字水印与基于全同态加密的数字水印的区别

参考文献

[1]李宗育,桂小林,顾迎捷,李雪松,戴慧珺,张学军.同态加密技术及其在云计算隐私保护中的应用[J].软件学报,2018,29(07):1830-1851.

[2]刘明洁,王安.全同态加密研究动态及其应用概述[J].计算机研究与发展,2014,51(12):2593-2603.

[3]陈智罡,王箭,宋新霞.全同态加密研究[J].计算机应用研究,2014,31(06):1624-1631.

[4]Chen H, Gilad-Bachrach R, Han K, et al. Logistic regression over encrypted data from fully homomorphic encryption[J]. BMC medical genomics, 2018, 11(4): 81.

[5]Martins P, Sousa L, Mariano A. A survey on fully homomorphic encryption: An engineering perspective[J]. ACM Computing Surveys (CSUR), 2018, 50(6): 83.

[6]Armknecht F, Boyd C, Carr C, et al. A Guide to Fully Homomorphic Encryption[J]. IACR Cryptology ePrint Archive, 2015, 2015: 1192.