

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2020 年 11 月 5 日 (05.11.2020)

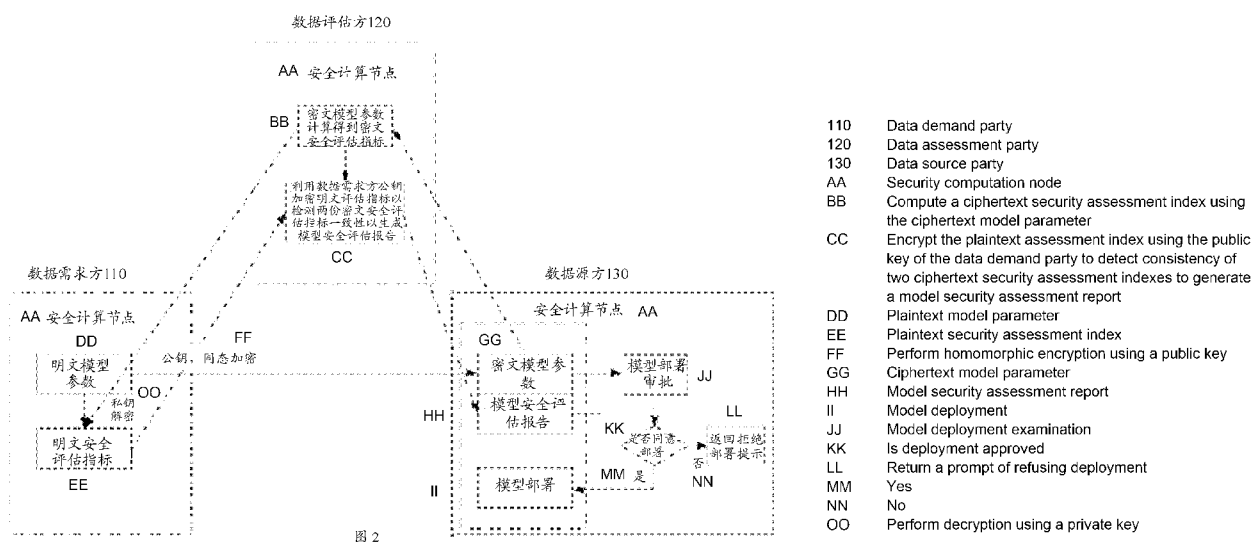


(10) 国际公布号
WO 2020/220754 A1

- (51) 国际专利分类号:
H04L 12/24 (2006.01) **H04L 9/08** (2006.01)
H04L 9/00 (2006.01)
- (21) 国际申请号: PCT/CN2020/071605
- (22) 国际申请日: 2020 年 1 月 12 日 (12.01.2020)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201910365011.2 2019年4月30日 (30.04.2019) CN
- (71) 申请人: 创新先进技术有限公司 (ADVANCED NEW TECHNOLOGIES CO., LTD.) [—/CN]; 开曼群岛大开曼岛乔治镇医院路27号开曼企业中心, Grand Cayman KY1-9008 (KY)。
- (72) 发明人: 阮方圆(RUAN, Fangyuan); 中国浙江省杭州市余杭区文一西路969号3号楼5楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。
- (74) 代理人: 北京博思佳知识产权代理有限公司 (BEIJING BESTIPR INTELLECTUAL PROPERTY LAW CORPORATION); 中国北京市海淀区上地三街9号嘉华大厦B座409, Beijing 100085 (CN)。
- (81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,

(54) Title: SECURITY ASSESSMENT METHOD AND DEVICE FOR ENCRYPTION MODEL

(54) 发明名称: 一种用于加密模型的安全性评估的方法和装置



(57) Abstract: The present disclosure provides a security assessment scheme for an encryption model. The scheme comprises: a data demand party device sends a deployment request to a data source party device; the data source party device forwards a ciphertext model parameter and a public key in the deployment request to a security assessment party device; the security assessment party device computes a first ciphertext security assessment index by using the ciphertext model parameter and sends the first ciphertext security assessment index to the data demand party device; the data demand party device decrypts the first ciphertext security assessment index by using a private key to generate a plaintext security assessment index and forwards the plaintext security assessment index to the security assessment party device; the security assessment party device encrypts the plaintext safety assessment index by using the public key to generate a second ciphertext security assessment index, and compares the first ciphertext security assessment index with the second ciphertext security assessment index to determine consistency thereof to generate a corresponding model security assessment report; and the data source party device examines, according to the model security assessment report, whether to deploy an encrypted

ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US,
UZ, VC, VN, WS, ZA, ZM, ZW。

- (84) 指定国(除另有指明, 要求每一种可提供的地区
保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ,
NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM,
AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG,
CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU,
IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT,
RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI,
CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

- 包括国际检索报告(条约第21条(3))。

data model.

(57) 摘要: 本公开提供了一种用于加密模型的安全性评估方案, 所述方案包括: 数据需求方设备向数据源方设备发送部署请求; 数据源方设备将部署请求中的密文模型参数以及公钥转发给安全性评估方设备; 安全性评估方设备利用密文模型参数通过计算得到第一密文安全评估指标并将其发送给数据需求方设备; 数据需求方设备利用私钥解密第一密文安全评估指标以生成明文安全评估指标并将其转发给安全性评估方设备; 安全性评估方设备利用公钥对明文安全评估指标进行加密以生成第二密文安全评估指标, 比较第一密文安全评估指标与第二密文安全评估指标以确定它们的一致性来生成相应的模型安全评估报告, 数据源方设备根据模型安全评估报告审批是否部署加密的数据模型。

一种用于加密模型的安全性评估的方法和设备

技术领域

[01]本公开涉及加密模型安全性领域，具体而言，涉及一种加密模型安全部署评估方案。

背景技术

- 5 [02]随着互联网技术的不断发展和云计算概念的诞生，越来越多的公司和用户倾向于通过互联网来进行数据通信。由此，许多公司在自己的平台上建立起各种数据模型来处理来自各种源的各种类型的数据，所述处理包括：接收、计算、更新、存储、发送等各种数据操作。例如，一种被广泛使用的数据模型是机器学习模型，机器学习模型可通过模拟或实现人类的学习行为，来获取新的知识或技能，重新组织已有的知识结构，进而不断改善自身的性能。它是人工智能的核心，是计算机具有人工智能的关键技术，其应用遍及人工智能的各个领域。它主要使用数据的归纳、综合而不是演绎。数据模型可以分为未加密数据模型（明文数据模型）和加密数据模型（密文数据模型）。在部署数据模型时，需要评估模型的安全性。明文数据模型可以直接判断安全性，却不能保护数据需求方的隐私，而密文数据模型虽然能保护需求方的隐私，却使得要部署的数据源方难以直接评估模型的安全性。因此，存在对模型安全部署方案进行改进的需求，以便能够提供一种更加安全稳健的模型安全部署评估机制。
- 10
- 15

发明内容

- [03]本公开涉及一种加密模型安全部署评估方案，所述方案利用了同态加密技术实现对加密模型的安全性评估，因此，不存在泄露模型参数明文的风险。
- 20 [04]根据本公开的第一方面，提供了一种用于加密模型的安全性评估的数据需求方设备，所述数据需求方设备被配置为：向数据源方设备发送要将经同态加密的数据模型部署在所述数据源方设备上的部署请求，所述部署请求包括了基于所述数据模型的明文参数被同态加密的密文模型参数以及所述同态加密所使用的公钥；从安全性评估方设备接收第一密文安全评估指标；以及利用与所述公钥相对应的私钥解密接收到的第一密文安全评估指标以生成明文安全评估指标并将所述明文安全评估指标转发给所述安全性评估方设备。
- 25

[05]根据本公开的第二方面，提供了一种用于加密模型的安全性评估的数据源方设备，

所述数据源方设备被配置为：从数据需求方设备接收要将经同态加密的数据模型部署在所述数据源方设备上的部署请求，所述部署请求包括了被同态加密的密文模型参数以及所述同态加密所使用的公钥；将所述被同态加密的密文模型参数以及所述公钥转发给安全性评估方设备；以及从所述安全性评估方设备接收模型安全评估报告，并根据所述模型安全评估报告审批是否部署所述数据需求方设备提供的经同态加密的数据模型。

[06]根据本公开的第三方面，如第二方面所述的数据源方设备还被配置为：在将所述被同态加密的密文模型参数以及所述公钥转发给所述安全性评估方设备之前，打乱所述被同态加密的密文模型参数的顺序。

[07]根据本公开的第四方面，提供了一种用于加密模型的安全性评估的安全性评估方设备，所述安全性评估方设备被配置为：接收从数据源方设备转发的由数据需求方设备提供的被同态加密的密文模型参数以及公钥；利用所述被同态加密的密文模型参数通过计算得到第一密文安全评估指标并将所述第一密文安全评估指标发送给所述数据需求方设备；从所述数据需求方设备接收明文安全评估指标；利用由所述数据需求方设备提供的所述公钥对接收到的所述明文安全评估指标进行加密以生成第二密文安全评估指标；比较所述第一密文安全评估指标与所述第二密文安全评估指标以确定它们的一致性，并基于所述一致性生成相应的模型安全评估报告；以及将所述模型安全评估报告发送给所述数据源方设备。

[08]根据本公开的第五方面，提供了一种用于加密模型的安全性评估的方法，所述方法包括：数据需求方设备向数据源方设备发送要将经同态加密的数据模型部署在所述数据源方设备上的部署请求，所述部署请求包括了基于所述数据模型的明文参数被同态加密的密文模型参数以及所述同态加密所使用的公钥；所述数据源方设备将所述被同态加密的密文模型参数以及所述公钥转发给安全性评估方设备；所述安全性评估方设备利用接收到的所述被同态加密的密文模型参数通过计算得到第一密文安全评估指标，并将所述第一密文安全评估指标发送给所述数据需求方设备；所述数据需求方设备利用与所述公钥相对应的私钥解密接收到的所述第一密文安全评估指标以生成明文安全评估指标，并将所述明文安全评估指标转发给所述安全性评估方设备；所述安全性评估方设备利用由所述数据需求方设备提供的所述公钥对接收到的所述明文安全评估指标进行加密以生成第二密文安全评估指标，比较所述第一密文安全评估指标与所述第二密文安全评估指标以确定它们的一致性，并基于所述一致性生成相应的模型安全评估报告，再将所述模型安全评估报告发送给所述数据源方设备；以及所述数据源方设备根据接收到的所述模

型安全评估报告审批是否部署所述数据需求方设备提供的经同态加密的数据模型。

[09]根据本公开的第六方面，如第五方面所述的方法还包括：在所述数据需求方设备向所述数据源方设备发送所述部署请求之前，已经利用来自各个源的数据作为模型特征来训练所述数据模型，并通过所述模型参数进行同态加密得到加密数据模型。

5 [10]根据本公开的第七方面，如第五方面所述的方法还包括：当发生多次评估时，通过对多个所述第一密文安全评估指标和第二密文安全评估指标的比较结果进行综合考虑来生成综合的模型安全评估报告。

10 [11]根据本公开的第八方面，如第五方面所述的方法也可以在所述数据源方设备部署了所述经同态加密的数据模型之后被执行以用于评估所述数据需求方所提供的明文模型结果有没有反推所述数据源方的明细数据值的风险。

[12]根据本公开的第九方面，如第五方面所述的方法还包括：在所述解密步骤中，在将所述明文安全评估指标发送给所述安全评估方之前，所述数据需求方可以通过将所述解密出的方差 Var 与根据明文算出的 Var 比较来进行审核以避免泄露模型参数

15 [13]根据本公开的第十方面，如第五方面所述的方法还包括：随同所述公钥还提供了一个随机数以保证在相同随机数条件下相同公钥对同一个数据加密结果一致。

[14]根据本公开的第十一方面，提供了一种用于加密模型的安全性评估的数据需求方设备，所述数据需求方设备被配置为：向数据源方设备发送要将经同态加密的数据模型部署在所述数据源方设备上的部署请求，所述部署请求包括了基于所述数据模型的明文参数被同态加密的密文模型参数以及所述同态加密所使用的公钥；从所述数据源方设备接收第一密文安全评估指标；以及利用与所述公钥相对应的私钥解密接收到的第一密文安全评估指标以生成明文安全评估指标并将所述明文安全评估指标转发给所述数据源方设备。

20 [15]根据本公开的第十二方面，提供了一种用于加密模型的安全性评估的数据源方设备，所述数据源方设备被配置为：从数据需求方设备接收要将经同态加密的数据模型部署在所述数据源方设备上的部署请求，所述部署请求包括了被同态加密的密文模型参数以及所述同态加密所使用的公钥；利用所述被同态加密的密文模型参数通过计算得到第一密文安全评估指标并将所述第一密文安全评估指标发送给数据需求方设备；从所述数据需求方设备接收明文安全评估指标；利用由所述数据需求方设备提供的所述公钥对接收到的所述明文安全评估指标进行加密以生成第二密文安全评估指标；比较所述第一密文安

全评估指标与所述第二密文安全评估指标以确定它们的一致性，并基于所述一致性生成相应的模型安全评估报告；以及根据所述模型安全评估报告审批是否部署所述数据需求方设备提供的经同态加密的数据模型。

5 [16]根据本公开的第十三方面，提供了一种用于加密模型的安全性评估的方法，所述方法包括：数据需求方设备向数据源方设备发送要将经同态加密的数据模型部署在所述数据源方设备上的部署请求，所述部署请求包括了基于所述数据模型的明文参数被同态加密的密文模型参数以及所述同态加密所使用的公钥；所述数据源方设备利用接收到的所述被同态加密的密文模型参数通过计算得到第一密文安全评估指标并将所述第一密文安全评估指标发送给所述数据需求方设备；所述数据需求方设备利用与所述公钥相对应的私钥解密接收到的所述第一密文安全评估指标以生成明文安全评估指标，并将所述明文安全评估指标转发给所述数据源方设备；所述数据源方设备利用由所述数据需求方设备提供的所述公钥对接收到的所述明文安全评估指标进行加密以生成第二密文安全评估指标，比较所述第一密文安全评估指标与所述第二密文安全评估指标以确定它们的一致性，并基于所述一致性生成相应的模型安全评估报告；以及根据所述模型安全评估报告审批是否部署所述数据需求方设备提供的经同态加密的数据模型。

[17]根据本公开的第十四方面，如第十三方面所述的方法还包括：在所述数据需求方设备向所述数据源方设备发送所述部署请求之前，已经利用来自各个源的数据作为模型特征来训练所述数据模型，并通过所述模型参数进行同态加密得到加密数据模型。

20 [18]根据本公开的第十五方面，如第十三方面所述的方法还包括：当发生多次评估时，通过对多个所述第一密文安全评估指标和第二密文安全评估指标的比较结果进行综合考虑来生成综合的模型安全评估报告。

[19]根据本公开的第十六方面，如第十三方面所述的方法也可以在所述数据源方设备部署了所述经同态加密的数据模型之后被执行以用于评估所述数据需求方所提供的明文模型结果有没有反推所述数据源方的明细数据值的风险。

25 [20]根据本公开的第十七方面，如第十三方面所述的方法包括：随同所述公钥还提供了一个随机数以保证在相同随机数条件下相同公钥对同一个数据加密结果一致。

[21]提供本概述以便以简化的形式介绍以下在详细描述中进一步描述的一些概念。本概述并不旨在标识所要求保护主题的关键特征或必要特征，也不旨在用于限制所要求保护主题的范围。

附图说明

[22]为了描述可获得本公开的上述和其它优点和特征的方式，将通过参考附图中示出的本公开的具体实施例来呈现以上简要描述的本公开的更具体描述。可以理解，这些附图只描绘了本公开的各典型实施例，并且因此不被认为是对其范围的限制，将通过使用附图并利用附加特征和细节来描述和解释本公开，在附图中：

[23]图 1 示出了本公开的一个示例性运行计算环境。

[24]图 2 是根据本公开的一个实施例的加密模型安全部署评估方案。

[25]图 3 根据本公开的一个实施例的加密模型安全部署评估方法的流程图。

[26]图 4 示出了本公开的另一个示例性运行计算环境。

[27]图 5 是根据本公开的另一个实施例的加密模型安全部署评估方案。

[28]图 6 根据本公开的另一个实施例的加密模型安全部署评估方法的流程图。

具体实施方式

[29]首先，为了解决在部署数据模型时的安全性和隐私之间的矛盾，在传统模型安全部署方案中，在部署加密数据模型之前，提供了可信评估方作为第三方评估机构来审批所述要部署的加密数据模型是否安全，它可以从数据需求方获取明文模型参数，然后，在基于该明文模型参数计算出模型安全性评估指标后，就向数据源方输出模型参数安全性评估报告。数据源方根据该安全性评估报告来判断模型安全性并决定是否将其部署在本地。

[30]但是，这种传统的评估方案存在很大的缺陷，即其方案是建立在假定可信评估方是完全可以信任的第三方的基础上，但在现实世界中，即使并非第三方刻意为之，但总是有各种入侵者试图利用各种漏洞或后门攻破可信第三方的安全壁垒来盗取各种有价值的信息。由于数据需求方将模型参数以明文方式传送给可信评估方以计算模型安全性评估指标，因此，一旦可信评估方被攻破（例如被植入木马程序），则所述明文模型参数很有可能被木马程序偷偷地传送给入侵者，从而导致数据需求方的模型明文在无人察觉的情况下被暴露给入侵者。

[31]或者，在另一种情况下，假如评估方实际上和数据源方形成共谋，以将数据需求方的模型参数明文泄露给数据源方（这种泄露对于正常状态下的可信评估方来说是绝对禁

止的，但不能排除可信评估方中的内部工作人员的私下共谋行为的存在可能性），则数据需求方的加密模型明文也会暴露给数据源方。无论是上述哪种情况，都将使得数据需求方的模型安全性受到严重的威胁。所以，传统模型安全部署方案并不能很好地解决模型安全部署的问题。

5 [32]而另一方面，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，它是一种基于数学难题的计算复杂性理论的密码学技术。对经过同态加密的数据进行处理得到一个输出，将这一输出进行解密，其结果与用同一方法处理未加密的原始数据得到的输出结果是一样的。
10 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。本质上，同态加密是指这样一种加密函数：对明文进行环上的加法和乘法运算再加密，与加密后对密文进行相应的运算，结果是等价的。正是同态加密这种特殊自然属性，使得要彻底解决部署数据模型时的安全性和隐私之间的矛盾变得可能。

15 [33]在具体介绍本公开的方案之前，首先对同态加密的具体算法加以说明，以帮助技术人员清楚认识到同态算法的所述特殊自然属性。

[34]具有同态性质的加密函数是指两个明文 a 、 b 满足 $\text{Dec}(\text{En}(a) \odot \text{En}(b)) = a \oplus b$ 的加密函数，其中 En 是加密运算， Dec 是解密运算， $\odot$ 、 $\oplus$ 分别对应明文和密文域上的运算。当 $\oplus$ 代表加法时，称该加密为加同态加密；当 $\odot$ 代表乘法时，称该加密为乘同态加密。

20 而全同态加密是指同时满足加同态和乘同态性质，可以进行任意多次加和乘运算的加密函数。用数学公式来表达，即 $\text{Dec}(f(\text{En}(m_1), \text{En}(m_2), \dots, \text{En}(m_k))) = f(m_1, m_2, \dots, m_k)$ ，或写成： $f(\text{En}(m_1), \text{En}(m_2), \dots, \text{En}(m_k)) = \text{En}(f(m_1, m_2, \dots, m_k))$ ，如果 f 是任意函数，称为全同态加密。

[35]为了便于理解，下面对最常用的同态加密算法加以说明。最常用的同态加密算法有
25 加法同态 Paillier 算法、乘法同态 Elgamal 算法、基于 LWE 的全同态算法。下面给出同态加密算法的定义。

[36]令加密运算为 $c = E(\text{PK}, m)$ ，解密运算为 $m = D(\text{SK}, d)$ 。其中， m 为明文， c 为密文， SK 为加密密钥， $E(\cdot)$ 为加密函数， $D(\cdot)$ 为解密函数。

[37]加法同态加密算法性质：可实现如下的加法同态运算，还可以利用多次加法同态运

算实现数乘同态运算。

[38]加法同态运算: $HAdd(c1, c2)$

[39]给定明文 $m1$ 和 $m2$ 的密文 $c1$ 和 $c2$, 即 $c1 = E(PK, m1)$, $c2 = E(PK, m2)$ 。对 $c1$ 和 $c2$ 计算可得到 $m1+m2$ 的密文, 即有下式成立:

$$5 \quad HAdd(c1, c2) = E(PK, m1+m2)$$

[40]数乘同态运算: $HSM(c1, m2)$

[41]给定明文 $m1$ 的密文 $c1$ 和明文 $m2$, 对 $c1$ 执行 $m2$ 次 $HAdd(,)$ 运算可以得到 $m1 \times m2$ 的密文。例如, 当 $m2=3$ 时, 有下式成立:

$$HSM(c1, m2) = HAdd(c1, HAdd(c1, HAdd(c1, c1))) = E(PK, m1 \times m2)$$

10 [42]乘法同态加密算法性质: 可实现如下的乘法同态运算:

[43]乘法同态运算: $HMul(c1, c2)$

[44]给定明文 $m1$ 和 $m2$ 的密文 $c1$ 和 $c2$, 即 $c1 = E(PK, m1)$, $c2 = E(PK, m2)$ 。对 $c1$ 和 $c2$ 计算可得到 $m1 \times m2$ 的密文, 即有下式成立:

$$HMul(c1, c2) = E(PK, m1 \times m2)。$$

15 [45]由此可见, 在上式中, 对明文进行环上的加法和乘法运算再加密, 与加密后对密文进行相应的运算, 其结果都是等价的。利用这个特殊自然属性, 有助于解决部署数据模型时的安全性和隐私之间的矛盾。上述算法公式都是同态加密领域的基本知识, 在此不再进一步详细描述。

[46]在了解了同态加密的基本算法和含义之后, 为使本公开的目的、技术方案及优点更加清楚明白, 以下参照附图并例举实施例, 对本公开作进一步详细说明。

[47]首先, 在图 1 中示出了本公开的一个实施例的一个示例性运行计算环境。在图 1 所示的计算环境中, 包含了数据需求方设备 110 (简称为“数据需求方 110”)、安全性评估方设备 120 (简称为“安全性评估方 120”)以及数据源方设备 130 (简称为“数据源方 130”)这三方。所述数据需求方、安全性评估方以及数据源方可以由安全计算节点, 例如个人计算机、服务器、客户端、移动设备 (例如手机、个人数字助理、平板等) 以及各种计算设备等构成。在所述各方之间, 通过网络进行数据通信, 所述网络包括各种类型的有线和无线网络, 包括但不限于互联网、局域网、WIFI、WLAN、蜂窝通信网络 (GPRS、CDMA、2G/3G/4G/5G 蜂窝网络)、卫星通信网络等等。通过所述数据通信

网络，数据需求方、安全性评估方以及数据源方可以相互传送各种数据以供处理。在如图 2 所示的本公开的一个实施例中的加密模型安全部署评估方案中，就涉及在这三方之间的数据交换。而在另一个实施例中，所述方案可以仅仅涉及数据需求方 110 和数据源方 130 这两者，该方案在图 4 和 5 中有详细描述。

5 [48]在描述了图 1 的本公开的一个实施例的示例性运行计算环境之后，接着，结合所述运行计算环境来描述一种利用同态加密来更加安全稳健的模型安全部署评估机制。如图 2 所示，示出了一种在数据需求方 110、安全性评估方 120 以及数据源方 130 这三方之间实现的模型安全部署评估的整体方案。首先，数据需求方 110 向数据源方 130 发送要将数据模型（例如训练好的机器学习模型）经过同态加密后部署在数据源方 130 上的部署请求，在所述部署请求中包含了基于所述数据模型的明文参数被同态加密的密文模型参数以及加密所使用的公钥。在发起所述部署请求之前，数据需求方 110 已经利用来自各个源的数据作为模型特征来训练出数据模型，并用公钥对所述数据模型的模型参数进行同态加密后得到了加密数据模型（密文模型）。注意，在各实施例中，是以线性回归模型作为机器学习模型的一个示例来进行描述，但应该理解所述数据模型并不局限于线性回归模型，而是可以应用于各种类型的数据模型，例如模型结果为预测概率值的其他密文模型，如逻辑回归、回归决策树等数据模型。在本公开的上下文中，原始的数据模型也被称为“明文模型”，而经同态加密后的数据模型也被称为“密文模型”。在接收到来自数据需求方 110 的包含经同态加密的密文模型参数和公钥的请求之后，数据源方 130 将经同态加密的密文模型参数的顺序打乱，随后，将被打乱顺序的经同态加密的密文模型参数与所述公钥一起传送给安全性评估方 120 或直接将经同态加密的密文模型参数和公钥一起传送给安全性评估方 120。所述安全性评估方 120 在获取经同态加密的密文模型参数和公钥后，利用该经同态加密的密文模型参数通过计算得到第一密文安全评估指标（在本实施例中以方差计算为例），并将该第一密文安全评估指标传送给数据需求方 110。数据需求方 110 利用与所述公钥相对应的私钥解密接收到的第一密文安全评估指标以生成明文安全评估指标，接着将明文安全评估指标回传给安全评估方 120。安全评估方 120 利用先前由数据需求方 110 提供的公钥对接收到的明文安全评估指标进行加密以生成第二密文安全评估指标，并将所述第二密文安全评估指标与第一密文安全评估指标进行比较以确定它们的一致性。接着，安全评估方 120 根据所述一致性生成相应的模型安全评估报告并将其返回给数据源方 130 以供其参考。其中，当第一密文安全评估指标与第二密文安全评估指标一致时，可以提供所述密文模型是安全的模型安全评估报告，反之，当第一密文安全评估指标与第二密文安全评估指标不一致时，可以提供所述

密文模型是不安全(例如数据需求方 110 可能在解密过程中造假)的模型安全评估报告。在接收到来自安全评估方 120 的模型安全评估报告之后,数据源方 130 根据模型安全评估报告审批所述模型部署请求,即如果模型安全评估报告指示所述密文模型是安全的则批准部署所述密文模型,否则拒绝部署该密文模型并可向数据需求方 110 返回拒绝部署的提示。当审批批准后,在数据源方 130 上部署所述经加密的密文模型。至此,所述方案结束。

[49]在本公开的所述方案中,安全评估方只能拿到密文形式下的密文模型参数,因此,没有从其泄露出模型参数明文的风险;同时,数据需求方在将第一密文安全评估指标解密成明文安全评估指标之后,不是直接将明文安全评估指标发送给数据源方,而是先由安全评估方对数据需求方解密之前和之后的两份安全评估指标进行校验后生成模型安全评估报告,再将该报告返回给数据源方,从而防止了数据需求方在解密过程中造假(如果数据需求方提供的是加密密文指标则根据同态加密技术的特点,在安全评估方处对第一和第二这两份密文安全评估指标的比较就不可能一致)。由此可见,本公开的加密模型安全部署评估方案解决了传统模型安全部署方案中所存在的各种缺陷,提供了具有改进性能的更加安全稳健的模型安全部署评估机制。

[50]在了解了本公开的加密模型安全部署评估方案的概况之后,在图 3 中示出了根据本公开的一个实施例的利用同态加密的模型安全部署评估方法的流程图。所述流程图与图 2 所示的评估方案相对应。该流程图同样涉及在数据需求方 110、安全性评估方 120 以及数据源方 130 这三者之间的数据流动。

[51]首先,在步骤 310,数据需求方 110 生成同态加密的加密密钥(公钥 PK)和解密密钥(私钥 SK),对模型参数进行同态加密以生成加密数据模型,随后,向数据源方 130 发送要将经过同态加密后的数据模型部署在数据源方 130 上的部署请求,在所述部署请求中包含了基于所述数据模型的明文参数被同态加密的密文模型参数以及所述同态加密所使用的公钥 PK。具体的模型加密过程如下:

[52]算法输入: $k+1$ 个模型参数 $a_0, a_1, \dots, a_k$ 。

[53]算法过程: $c_0 = E(PK, a_0), c_1 = E(PK, a_1), \dots, c_k = E(PK, a_k)$ 。

[54]算法输出: $k+1$ 个加密模型参数 $c_0, c_1, \dots, c_k$ 。

[55]其中,应该理解上述算法示例仅仅是出于说明目的的简单加密过程的演示。实际上,根据非对称加密技术的特点,出于安全性考虑,采用非对称加密时,随同所述公钥还会

提供一个随机数给数据评估方,以保证在相同随机数条件下相同公钥对同一个数据加密结果一致。“随机数”是非对称加密的常用安全手段,与本公开本身的关系不大,因此,不再在此深入展开。

5 [56]随后,在步骤 320,在接收到所述部署请求之后,数据源方 130 将经同态加密的密文模型参数(即 $c_0, c_1, \dots, c_k$, 其顺序可以被打乱)和公钥(即 PK)一起传送给安全性评估方 120。

10 [57]接着,在步骤 330,所述安全性评估方 120 在获取经同态加密的密文模型参数和公钥后,利用该经同态加密的密文模型参数通过计算得到第一密文安全评估指标,并将该第一密文安全评估指标传送给数据需求方 110。在本公开中,“安全性评估指标”是指根据同态性质利用加密模型参数计算出的方差(var)或极差之类的统计指标。使用所述统计指标可以防止出现模型中某个特征的参数远大于其他特征的参数的情况。这种远大于其他特征的参数的参数很可能是用于反推其对应的原始值的“恶意”参数。因此,通过方差计算,可以使得可能的“恶意”参数彻底失去作用。在本实施例中通过对经同态加密的密文模型参数执行方差计算来作为计算安全性评估指标的一个示例,其算法如下:

15 [58]算法输入: $k+1$ 个加密模型参数 $c_0, c_1, \dots, c_k$ 。

[59]算法过程: 密文(样本)方差指标为:

$$\text{Var\_E} = \text{HAdd}(\text{HSM}(\text{HAdd}(\text{HMul}(c_0, c_0), \text{HMul}(c_1, c_1), \dots, \text{HMul}(c_k, c_k)), 1/k), \text{HSM}(\text{HMul}(\text{HAdd}(c_0, c_1, \dots, c_k), \text{HAdd}(c_0, c_1, \dots, c_k)), -1/(k*(k+1)))$$

20 [60]其中,如前所述, $\text{HAdd}()$ 表示加法同态运算, $\text{HSM}()$ 表示数乘同态运算,而 $\text{HMul}()$ 表示乘法同态运算。

[61]在步骤 340,数据需求方 110 利用与所述公钥相对应的私钥 SK 解密接收到的第一密文模型安全评估指标中的密文参数以生成明文安全评估指标($\text{Var} = \text{D}(\text{SK}, \text{Var\_E})$)并将该明文安全评估指标发给安全性评估方 120。

25 [62]在步骤 350,安全评估方 120 利用由数据需求方 110 提供的公钥 PK 对接收到的明文安全评估指标进行加密以生成第二密文安全评估指标:

$$\text{Var\_E2} = \text{E}(\text{PK}, \text{Var})$$

[63]其中,如前所述,Var 表示方差, $\text{E}()$ 为加密函数, $\text{D}()$ 为解密函数,SK 为私钥,而 PK 为公钥。

[64]在步骤 360, 安全评估方 120 将 (步骤 330 生成的) 所述第一密文安全评估指标与第二密文安全评估指标进行比较以确定它们的一致性, 即将 Var_E 与 Var_E2 进行比较。其中, 如果数据需求方 110 是意图通过造假模型参数来恶意套取数据源方的明细数据, 则在数据需求方 110 用私钥 SK 解密时造假的模型参数必然导致所生成的明文安全评估指标的明文参数发生变化, 进而导致在安全评估方 120 处基于所述明文参数加密生成的 Var_E2 与 Var_E 不一致。这种不一致性, 揭示了数据需求方 110 在步骤 340 解密时进行了造假。

[65]在步骤 370, 安全评估方 120 基于所述一致性生成相应的模型安全评估报告并将其返回给数据源方 130 以供其参考。其中, 当第一密文安全评估指标与第二密文安全评估指标一致时, 可以提供所述数据模型是安全的模型安全评估报告, 反之, 当第一密文安全评估指标与第二密文安全评估指标不一致时, 可以提供所述数据模型是不安全 (例如数据需求方 110 可能在解密过程中造假) 的模型安全评估报告。

[66]最后, 在步骤 380, 数据源方 130 根据接收到的模型安全评估报告, 审批所述模型部署请求, 即如果模型安全评估报告指示所述加密密文模型是安全的则批准部署所述密文模型, 否则拒绝部署所述密文模型并可向数据需求方 110 返回拒绝部署的提示。当批准部署后, 在数据源方 130 上部署所述加密的密文模型。

[67]至此, 所述评估方法的整个流程结束。需要理解的是, 本公开所涉及的同态加密和解密算法利用了常用的同态加密解密算法中的一些参数、函数和算式。这些参数、函数和算式都是同态加密的基本常识, 因此, 不再在此过多描述。上述示例算法仅仅是为了说明性目的提供的, 并不是要把本公开的同态加密、解密以及计算局限于上述算法。实际上, 技术人员可以根据实际需要选择其他任何合适的同态加密和解密和计算算法。

[68]在一个改进实施例中, 在步骤 330 中基于经同态加密的密文模型参数生成方差 Var_E 之后, 还可以将该方差与设定的一个阈值进行比较, 如果所述方差超过该阈值则判定为不安全, 因为所述模型参数可能是用来攻击模型的造假参数。

[69]在一个改进的实施例中, 提供了一种综合的安全等级评估报告, 即在该报告的生成过程中, 可以存在多次评估, 进而同时提供了多个第一安全评估指标和第二密文安全评估指标, 通过对这些第一密文安全评估指标和第二密文安全评估指标的比较结果进行综合考虑 (例如评分) 之后, 就能获得更加准确的综合的安全等级评估报告。具体而言, 数据源方 130 可以要求数据需求方 110 在不同的时间多次发送部署请求, 并将每次部署请求中的加密模型参数和公钥发送给安全评估方 120。同样, 安全评估方 120 也在每次

对加密模型参数进行计算生成第一安全评估指标后将该第一安全评估指标发送给数据需求方 110，并在接收到经数据需求方 110 解密后的明文安全评估指标后使用公钥对所述明文安全评估指标进行加密以生成相应的第二密文安全评估指标并与相应的第一安全评估指标进行比较。随后，综合每次比较后的结果再生成综合的安全等级评估报告。

- 5 所述综合可以包括当多次比较结果中有一次结果超过阈值就给出不安全的评估，或者对所有结果进行加权平均后获得一个总的安全评分分数加入报告中以反馈给数据源方 130，等等。在另一个改进的实施例中，在步骤 340 的数据需求方 110 解密接收到的密文模型安全评估指标中的密文参数以生成明文安全评估指标的步骤中，在将明文安全评估指标发送给安全评估方 120 之前，数据需求方可以通过将所述解密出的方差 Var 与根据明文算出的 Var 比较来进行审核，以确定是否有人将一个加密的模型参数变型后假冒成 Var_E 让数据需求方 110 解密。通过所述审核，可以发现这种假冒行为以避免泄露模型参数。

- [70]在其它实施例中，所述模型安全性评估方案还可以用于在加密数据模型被部署到数据源方之后的模型结果评估中，以评估数据需求方 110 所提供的明文模型结果有没有反推数据源方的明细数据值的风险。在数据源方存储有大量有价值的数据，因此，如果数据需求方 110 所提供的数据模型所提供的明文模型结果可以轻易推算出数据源方输入的明细数据值，则说明数据需求方 110 可能设定了特定的模型参数来进行反推。因此，可以使用本公开的模型安全性评估方案对已部署的加密数据模型进行审计以确定是否存在反推风险。由于其过程类似，不再在此详细陈述。

- 20 [71]另外，如前所述，在非对称加密过程中，随同所述公钥还会提供一个随机数，以保证在相同随机数条件下相同公钥对同一个数据加密结果一致。因此，在上述涉及公钥的步骤中，实际上还包含了随机数的处理。但所述“随机数”是非对称加密的常用安全手段

- [72] 上面主要描述的是在数据需求方、安全性评估方以及数据源方这三方之间的模型安全部署评估机制。而根据本公开的另一个实施例，也提供了一种利用同态加密来更加安全稳健的模型安全部署评估机制。如图 4 所示，示出了一种在无需安全性评估方的情况下，在数据需求方 110 以及数据源方 130 这两方之间实现的模型安全部署评估的示例性运行环境。在该示例性运行环境中，数据需求方 110 以及数据源方 130 通过网络直接双向通信，而不需要安全性评估方的参与。

[73]而在图 5 中示出了根据本公开的另一个实施例的加密模型安全部署评估方案。

- 30 [74]首先，数据需求方 110 向数据源方 130 发送要将数据模型（例如训练好的机器学习

模型) 经过同态加密后部署在数据源方 130 上的部署请求, 在所述部署请求中包含了基于所述数据模型的明文参数被同态加密的密文模型参数以及加密所使用的公钥。在接收到来自数据需求方 110 的包含经同态加密的密文模型参数和公钥的请求之后, 数据源方 130 利用该经同态加密的密文模型参数通过计算得到第一密文安全评估指标, 并将该第一密文安全评估指标传送给数据需求方 110。数据需求方 110 利用与所述公钥相对应的私钥解密接收到的密文模型安全评估指标以生成明文安全评估指标, 接着将明文安全评估指标回传给数据源方 130。数据源方 130 利用先前由数据需求方 110 提供的公钥对接收到的明文安全评估指标进行加密以生成第二密文安全评估指标, 并将所述第二密文安全评估指标与第一密文安全评估指标进行比较以确定它们的一致性。接着, 数据源方 130 根据所述一致性生成相应的模型安全评估报告。其中, 当第一密文安全评估指标与第二密文安全评估指标一致时, 可以提供所述数据模型是安全的模型安全评估报告, 反之, 当第一密文安全评估指标与第二密文安全评估指标不一致时, 可以提供所述数据模型是不安全 (例如数据需求方 110 可能在解密过程中造假) 的模型安全评估报告。然后, 数据源方 130 根据模型安全评估报告审批所述模型部署请求, 即如果模型安全评估报告指示所述加密密文模型是安全的则批准部署所述密文模型, 否则拒绝部署所述密文模型并可向数据需求方 110 返回拒绝部署的提示。当审批批准后, 在数据源方 130 上部署所述加密的密文模型。至此, 所述方案结束。

[75] 所述模型安全评估报告的生成方便了对其进行存档以备将来使用, 也即如果所述模型安全评估报告显示来自所述数据需求方 110 的数据模型是不安全的, 则在今后再收到来自同一数据需求方 110 的后续部署请求后, 可以直接查阅所存储的先前生成的与该数据需求方 110 相关联的模型安全评估报告来直接拒绝来自该数据需求方 110 的后续部署请求以节省资源。

[76] 因此, 在一个实施例中, 也可以跳过所述安全评估报告生成的步骤, 而直接根据第一密文安全评估指标与第二密文安全评估指标的比较结果来审批所述模型部署请求, 即如果第一密文安全评估指标与第二密文安全评估指标的比较结果为一一致, 则批准部署所述密文模型, 反之, 如果比较结果不一致, 则拒绝部署所述密文模型。

[77] 在本公开的所述方案中, 数据源方 130 只能拿到密文形式下的模型参数, 因此, 没有泄露模型参数明文的风险; 同时数据需求方 110 在对第一密文安全评估指标解密之后, 由数据源方 130 通过利用公钥对明文安全评估指标进行加密以生成第二密文安全评估指标并对数据需求方解密前和后的这两份安全评估指标进行校验后生成模型安全评估报

告,从而防止了数据需求方在解密过程中造假(如果提供的是造假指标则根据同态加密技术的特点,这两份密文安全评估指标不可能一致)。由此可见,本公开的加密模型安全部署评估方案解决了传统模型安全部署方案中所存在的各种缺陷,提供了具有改进性能的更加安全稳健的模型安全部署评估机制。

5 [78]在了解了本公开的另一个实施例的加密模型安全部署评估方案的概况之后,在图 6 中示出了根据本公开的该另一个实施例的利用同态加密的模型安全部署评估方法的流程图。所述流程图涉及在数据需求方 110 以及数据源方 130 这两者之间的数据流动。其中与图 3 相类似的内容、定义和描述不再详述。

10 [79]首先,在步骤 610,数据需求方 110 生成同态加密的加密密钥(公钥 PK)和解密密钥(私钥 SK),对模型参数进行同态加密以生成加密数据模型,随后,向数据源方 130 发送要将经过同态加密后的数据模型部署在数据源方 130 上的部署请求,在所述部署请求中包含了基于所述数据模型的明文参数被同态加密的密文模型参数以及加密所使用的公钥 PK。具体的加密过程如下:

[80]算法输入: $k+1$ 个模型参数 $a_0, a_1, \dots, a_k$ 。

15 [81]算法过程: $c_0=E(PK, a_0), c_1=E(PK, a_1), \dots, c_k=E(PK, a_k)$ 。

[82]算法输出: $k+1$ 个加密模型参数 $c_0, c_1, \dots, c_k$ 。

[83]随后,在步骤 620,所述数据源方 130 在接收到所述部署请求之后,利用该经同态加密的密文模型参数通过计算得到第一密文安全评估指标,并将该第一密文安全评估指标传送给数据需求方 110。所述计算指标可以是例如根据同态性质利用加密模型参数计算极差、方差类型之类的统计指标,以防止模型中某个特征的参数远大于其他特征的参数。这种远大于其他特征的参数的参数很可能是用于反推其对应的原始值的“恶意”参数。以方差为例,其算法如下:

20 [84]算法输入: $k+1$ 个加密模型参数 $c_0, c_1, \dots, c_k$ 。

[85]算法过程: 密文(样本)方差指标为:

25 [86]
$$\text{Var\_E} = \text{HAdd}(\text{HSM}(\text{HAdd}(\text{HMul}(c_0, c_0), \text{HMul}(c_1, c_1), \dots, \text{HMul}(c_k, c_k)), 1/k), \text{HSM}(\text{HMul}(\text{HAdd}(c_0, c_1, \dots, c_k), \text{HAdd}(c_0, c_1, \dots, c_k)), -1/(k*(k+1))))$$

[87]在步骤 630,数据需求方 110 利用与所述公钥相对应的私钥 SK 解密接收到的第一密文模型安全评估指标中的密文参数以生成明文安全评估指标($\text{Var}=\text{D}(\text{SK}, \text{Var\_E})$)并将该明文安全评估指标返回给数据源方 130。

[88]在步骤 640, 所述数据源方 130 利用由数据需求方 110 提供的公钥 PK 对接收到的明文安全评估指标进行加密以生成第二密文安全评估指标 $Var\_E2=E(PK, Var)$ 。

[89]在步骤 650, 数据源方 130 将 (在步骤 620 生成的) 所述第一密文安全评估指标与第二密文安全评估指标进行比较以确定它们的一致性, 即将 Var_E 与 Var_E2 进行比较。

5 其中, 如果数据需求方 110 是意图通过造假模型参数来恶意套取数据源方的明细数据, 则在解密时进行的造假必然导致明文安全评估指标的明文参数发生变化, 进而导致基于所述明文参数生成的 Var_E2 与先前计算的 Var_E 不一致。这种不一致性, 揭示了数据需求方 110 在解密时进行了造假。

10 [90]在步骤 660, 数据源方 130 基于所述一致性生成相应的模型安全评估报告。其中, 当第一密文安全评估指标与第二密文安全评估指标一致时, 可以提供所述数据模型是安全的模型安全评估报告, 反之, 当第一密文安全评估指标与第二密文安全评估指标不一致时, 可以提供所述数据模型是不安全 (例如数据需求方 110 可能在解密过程中造假) 的模型安全评估报告。

15 [91]最后, 在步骤 670, 数据源方 130 根据所述模型安全评估报告, 审批所述模型部署请求, 即如果模型安全评估报告指示所述加密密文模型是安全的则批准部署所述密文模型, 否则拒绝部署所述密文模型并可向数据需求方 110 返回拒绝部署的提示。当审批批准后, 在数据源方 130 上部署所述加密的密文模型。

20 [92]在一个改进的实施例, 提供了一种综合的安全等级评估报告, 即在该报告的生成过程中, 可以存在多次评估, 进而同时提供了多个第一安全评估指标和第二密文安全评估指标, 通过对这些第一密文安全评估指标和第二密文安全评估指标的多次比较结果进行综合考虑 (例如进行加权求和, 过滤最大和最小差异等等), 就能获得更加准确的综合的安全等级评估报告。具体而言, 数据源方 130 可以要求数据需求方 110 在不同的时间多次发送部署请求, 并在每次接收到部署请求时, 都对加密模型参数进行计算生成第一安全评估指标以发送给数据需求方 110, 并在接收到经数据需求方 110 解密后的明文安全评估指标后使用公钥对所述明文安全评估指标进行加密以生成相应的第二密文安全评估指标并与相应的第一安全评估指标进行比较。随后, 数据源方 130 综合每次比较后的结果再生成综合的安全等级评估报告。所述综合可以包括当多次比较结果中有一次结果超过阈值就给出不安全的评估, 或者对所有结果进行加权平均后获得一个总的安全评分分数加入报告中以反馈安全性, 等等。

30 [93]在另一个改进的实施例, 在步骤 340 的数据需求方 110 解密接收到的密文模型安

全评估指标中的密文参数以生成明文安全评估指标的步骤中，在将明文安全评估指标发送给安全评估方 120 之前，数据需求方可以通过将所述解密出的方差 Var 与根据明文算出的 Var 比较来进行审核，以确定是否是否有人将一个加密得分模型参数变型后假冒成 Var_E 让数据需求方 110 解密。通过所述审核，可以发现这种假冒行为以避免泄露模型参数。

[94]另外，如前所述，在非对称加密过程中，随同所述公钥还会提供一个随机数，以保证在相同随机数条件下相同公钥对同一个数据加密结果一致。因此，在上述涉及公钥的步骤中，实际上还包含了随机数的处理。但所述“随机数”是非对称加密的常用安全手段。

[95]在上述另一个实施例中，数据源方可以取代安全评估方来完成第一和第二密钥安全指标的生成、比较以及模型安全评估报告的生成。这样，进一步避免了评估方的存在可能带来的各种安全性隐患，使得所述安全评估报告更加可信。因此，其也能解决传统模型安全部署方案中所存在的各种缺陷，并提供具有改进性能的更加安全稳健的模型安全部署评估机制。

[96]虽然以上描述了不同的实施例，但应当理解的是它们只是作为示例而非限制。（诸）相关领域的技术人员将领会，在不偏离如所附权利要求书所定义的本公开的精神和范围的情况下，可以在形式和细节方面进行各种修改。因此，此处所公开的本公开的宽度和范围不应被上述所公开的示例性实施例所限制，而应当仅根据所附权利要求书及其等同替换来定义。

权利要求书

1. 一种用于加密模型的安全性评估的数据需求方设备, 其特征在于, 所述数据需求方设备被配置为:

向数据源方设备发送要将经同态加密的数据模型部署在所述数据源方设备上的部署请求, 所述部署请求包括了基于所述数据模型的明文参数被同态加密的密文模型参数以及所述同态加密所使用的公钥;

从安全性评估方设备接收第一密文安全评估指标; 以及

利用与所述公钥相对应的私钥解密接收到的所述第一密文安全评估指标以生成明文安全评估指标并将所述明文安全评估指标转发给所述安全性评估方设备。

2. 一种用于加密模型的安全性评估的数据源方设备, 其特征在于, 所述数据源方设备被配置为:

从数据需求方设备接收要将经同态加密的数据模型部署在所述数据源方设备上的部署请求, 所述部署请求包括了被同态加密的密文模型参数以及所述同态加密所使用的公钥;

将所述被同态加密的密文模型参数以及所述公钥转发给安全性评估方设备; 以及

从所述安全性评估方设备接收模型安全评估报告, 并根据所述模型安全评估报告审批是否部署所述数据需求方设备提供的所述经同态加密的数据模型。

3. 如权利要求 2 所述的所述数据源方设备, 其特征在于, 所述数据源方设备还被配置为:

在将所述被同态加密的密文模型参数以及所述公钥转发给所述安全性评估方设备之前, 打乱所述被同态加密的密文模型参数的顺序。

4. 一种用于加密模型的安全性评估的安全性评估方设备, 其特征在于, 所述安全性评估方设备被配置为:

接收从数据源方设备转发的由数据需求方设备提供的被同态加密的密文模型参数以及公钥;

利用所述被同态加密的密文模型参数通过计算得到第一密文安全评估指标并将所述第一密文安全评估指标发送给所述数据需求方设备;

从所述数据需求方设备接收明文安全评估指标;

利用由所述数据需求方设备提供的所述公钥对接收到的所述明文安全评估指标进行加密以生成第二密文安全评估指标;

比较所述第一密文安全评估指标与所述第二密文安全评估指标以确定它们的一致

性，并基于所述一致性生成相应的模型安全评估报告；以及

将所述模型安全评估报告发送给所述数据源方设备。

5. 一种用于加密模型的安全性评估的方法，其特征在于，所述方法包括：

数据需求方设备向数据源方设备发送要将经同态加密的数据模型部署在所述数据

5 源方设备上的部署请求，所述部署请求包括了基于所述数据模型的明文参数被同态加密的密文模型参数以及所述同态加密所使用的公钥；

所述数据源方设备将所述被同态加密的密文模型参数以及所述公钥转发给安全性评估方设备；

10 所述安全性评估方设备利用接收到的所述被同态加密的密文模型参数通过计算得到第一密文安全评估指标，并将所述第一密文安全评估指标发送给所述数据需求方设备；

所述数据需求方设备利用与所述公钥相对应的私钥解密接收到的所述第一密文安全评估指标以生成明文安全评估指标，并将所述明文安全评估指标转发给所述安全性评估方设备；

15 所述安全性评估方设备利用由所述数据需求方设备提供的所述公钥对接收到的所述明文安全评估指标进行加密以生成第二密文安全评估指标，比较所述第一密文安全评估指标与所述第二密文安全评估指标以确定它们的一致性，并基于所述一致性生成相应的模型安全评估报告，再将所述模型安全评估报告发送给所述数据源方设备；以及

20 所述数据源方设备根据接收到的所述模型安全评估报告审批是否部署所述数据需求方设备提供的经同态加密的数据模型。

6. 如权利要求 5 所述的方法，其特征在于，所述方法还包括：

在所述数据需求方设备向所述数据源方设备发送所述部署请求之前，已经利用来自各个源的数据作为模型特征来训练所述数据模型，并通过对所述模型参数进行同态加密得到加密数据模型。

25 7. 如权利要求 5 所述的方法，其特征在于，所述方法还包括：

当发生多次评估时，通过对多个所述第一密文安全评估指标和第二密文安全评估指标的比较结果进行综合考虑来生成综合的模型安全评估报告。

8. 如权利要求 5 所述的方法，其特征在于，所述方法也可以在所述数据源方设备部署了所述经同态加密的数据模型之后被执行以用于评估数据需求方所提供的明文模型
30 结果有没有反推所述数据源方的明细数据值的风险。

9. 如权利要求 5 所述的方法，其特征在于，在所述解密步骤中，在将所述明文安全

评估指标发送给所述安全评估方之前，所述数据需求方可以通过将所述解密出的方差 Var 与根据明文算出的 Var 比较来进行审核以避免泄露模型参数。

10. 如权利要求 5 所述的方法，其特征在于，随同所述公钥还提供了一个随机数以保证在相同随机数条件下相同公钥对同一个数据加密结果一致。

5 11. 一种用于加密模型的安全性评估的数据需求方设备，其特征在于，所述数据需求方设备被配置为：

向数据源方设备发送要将经同态加密的数据模型部署在所述数据源方设备上的部署请求，所述部署请求包括了基于所述数据模型的明文参数被同态加密的密文模型参数以及所述同态加密所使用的公钥；

10 从所述数据源方设备接收第一密文安全评估指标；以及

利用与所述公钥相对应的私钥解密接收到的第一密文安全评估指标以生成明文安全评估指标并将所述明文安全评估指标转发给所述数据源方设备。

12. 一种用于加密模型的安全性评估的数据源方设备，其特征在于，所述数据源方设备被配置为：

15 从数据需求方设备接收要将经同态加密的数据模型部署在所述数据源方设备上的部署请求，所述部署请求包括了被同态加密的密文模型参数以及所述同态加密所使用的公钥；

利用所述被同态加密的密文模型参数通过计算得到第一密文安全评估指标并将所述第一密文安全评估指标发送给数据需求方设备；

20 从所述数据需求方设备接收明文安全评估指标；

利用由所述数据需求方设备提供的所述公钥对接收到的所述明文安全评估指标进行加密以生成第二密文安全评估指标；

比较所述第一密文安全评估指标与所述第二密文安全评估指标以确定它们的一致性，并基于所述一致性生成相应的模型安全评估报告；以及

25 根据所述模型安全评估报告审批是否部署所述数据需求方设备提供的经同态加密的数据模型。

13. 一种用于加密模型的安全性评估的方法，其特征在于，所述方法包括：

数据需求方设备向数据源方设备发送要将经同态加密的数据模型部署在所述数据源方设备上的部署请求，所述部署请求包括了基于所述数据模型的明文参数被同态加密的密文模型参数以及所述同态加密所使用的公钥；

30 所述数据源方设备利用接收到的所述被同态加密的密文模型参数通过计算得到第

一密文安全评估指标并将所述第一密文安全评估指标发送给所述数据需求方设备;

所述数据需求方设备利用与所述公钥相对应的私钥解密接收到的所述第一密文安全评估指标以生成明文安全评估指标,并将所述明文安全评估指标转发给所述数据源方设备;

5 所述数据源方设备利用由所述数据需求方设备提供的所述公钥对接收到的所述明文安全评估指标进行加密以生成第二密文安全评估指标,比较所述第一密文安全评估指标与所述第二密文安全评估指标以确定它们的一致性,并基于所述一致性生成相应的模型安全评估报告;以及根据所述模型安全评估报告审批是否部署所述数据需求方设备提供的经同态加密的数据模型。

10 14.如权利要求 13 所述的方法,其特征在于,所述方法还包括:

在所述数据需求方设备向所述数据源方设备发送所述部署请求之前,已经利用来自各个源的数据作为模型特征来训练所述数据模型,并通过对所述模型参数进行同态加密得到加密数据模型。

15.如权利要求 13 所述的方法,其特征在于,所述方法还包括:

15 当发生多次评估时,通过对多个所述第一密文安全评估指标和第二密文安全评估指标的比较结果进行综合考虑来生成综合的模型安全评估报告。

16. 如权利要求 13 所述的方法,其特征在于,所述方法也可以在所述数据源方设备部署了所述经同态加密的数据模型之后被执行以用于评估数据需求方所提供的明文模型结果有没有反推所述数据源方的明细数据值的风险。

20 17. 如权利要求 13 所述的方法,其特征在于,随同所述公钥还提供了一个随机数以保证在相同随机数条件下相同公钥对同一个数据加密结果一致。

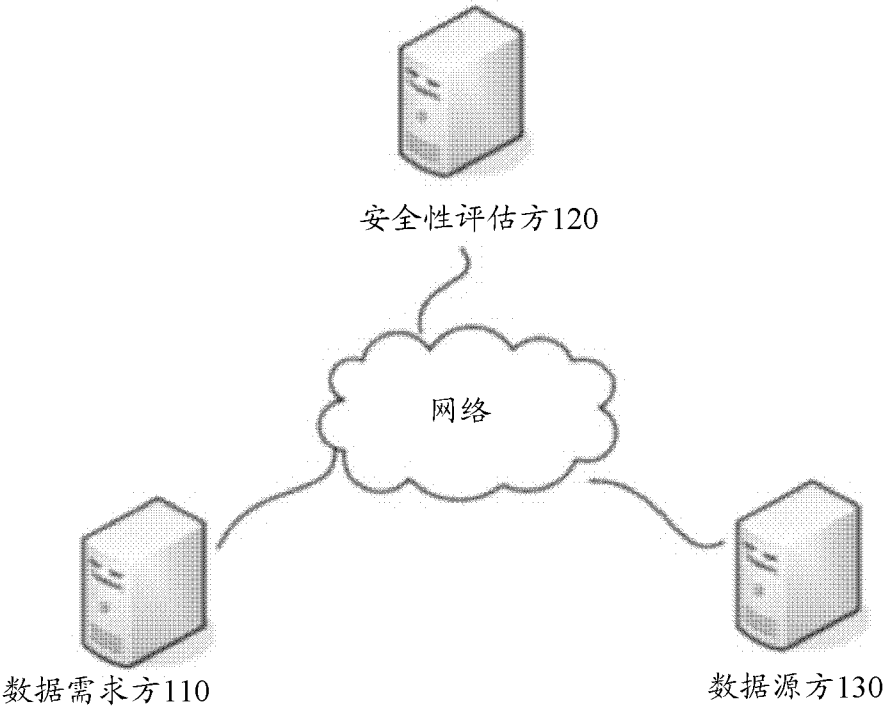


图 1

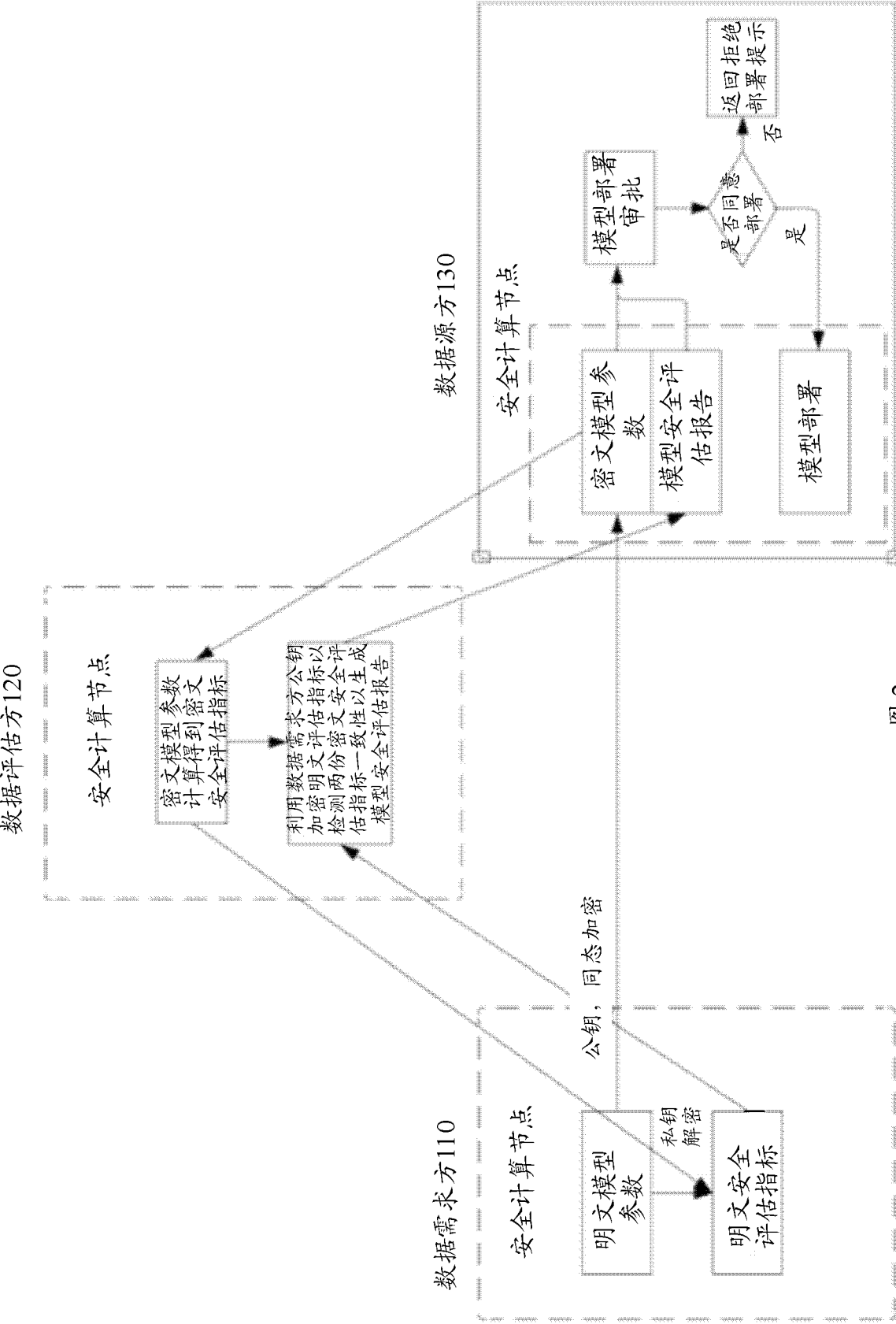


图 2

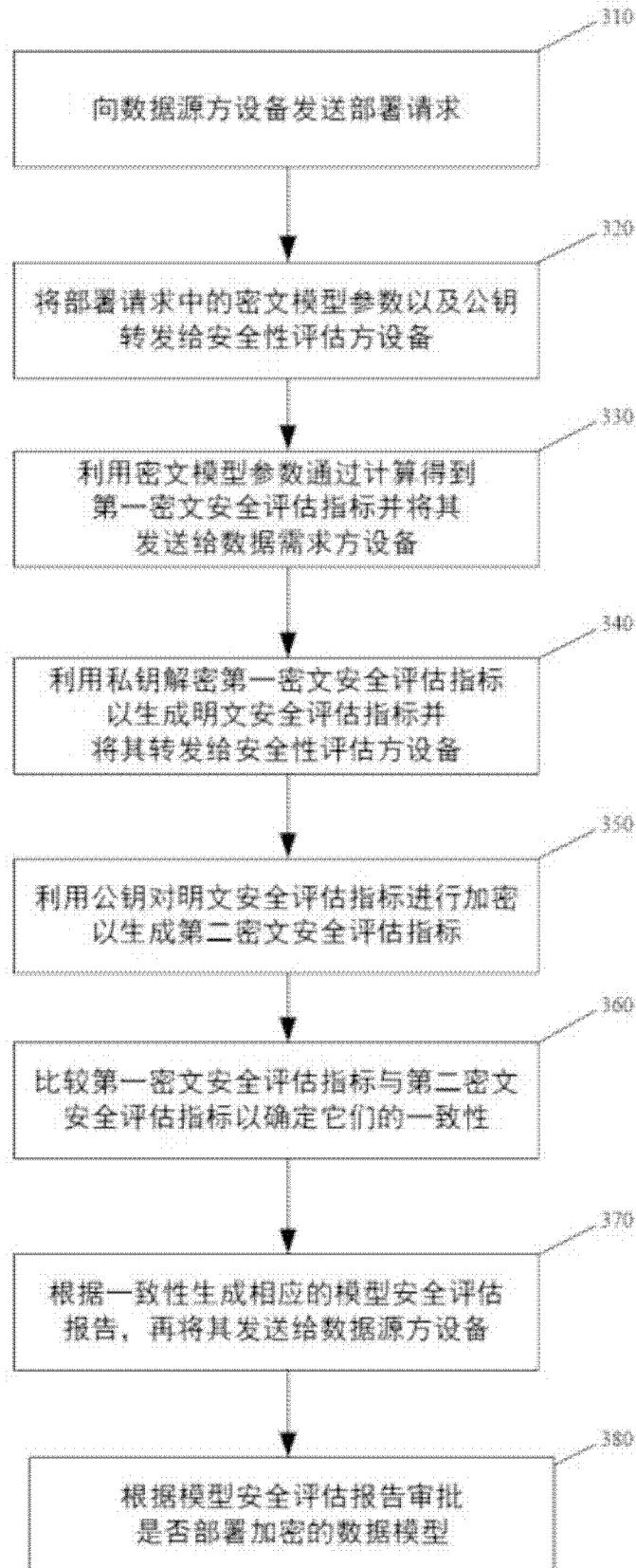


图 3



图 4

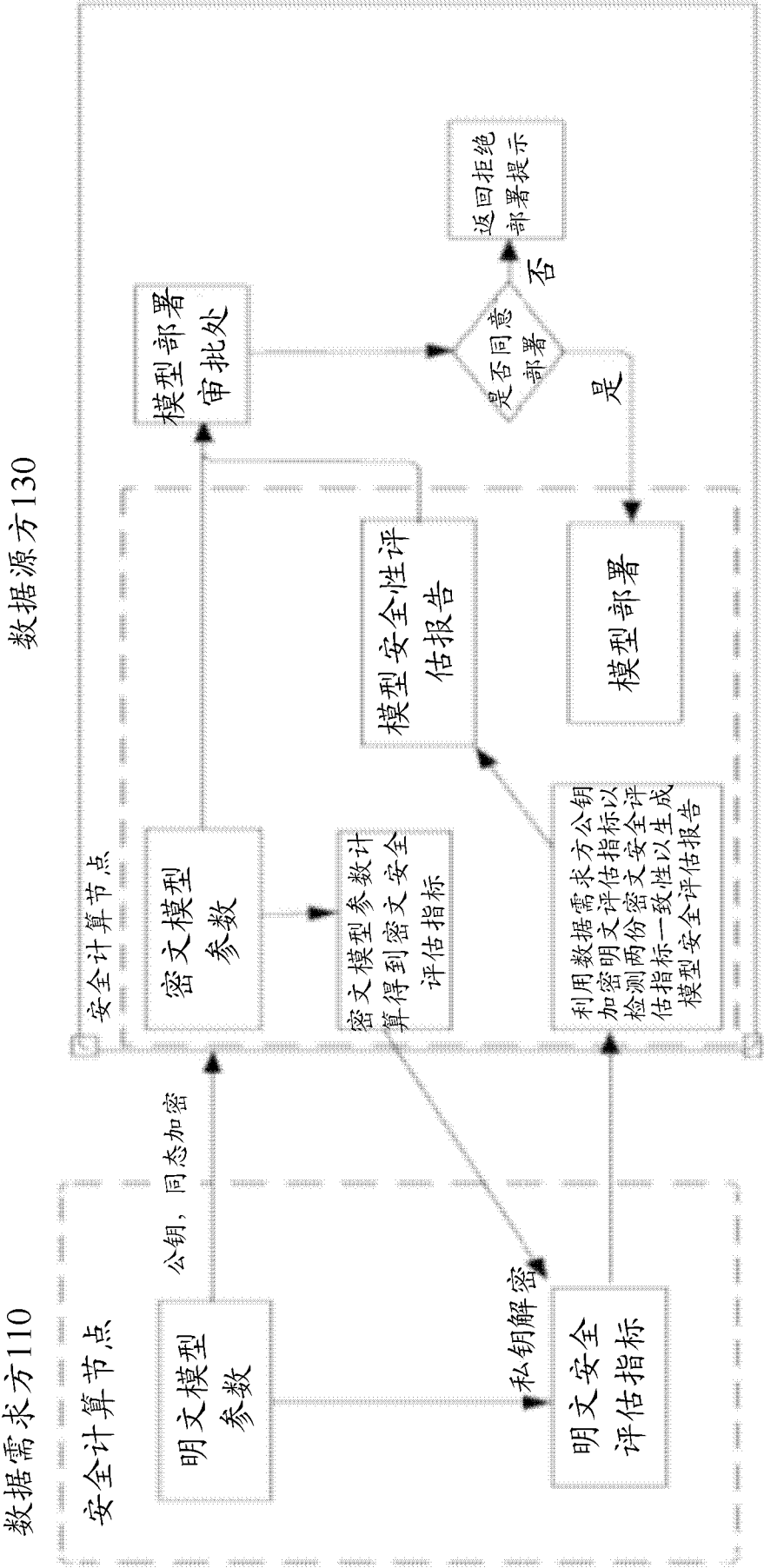


图 5

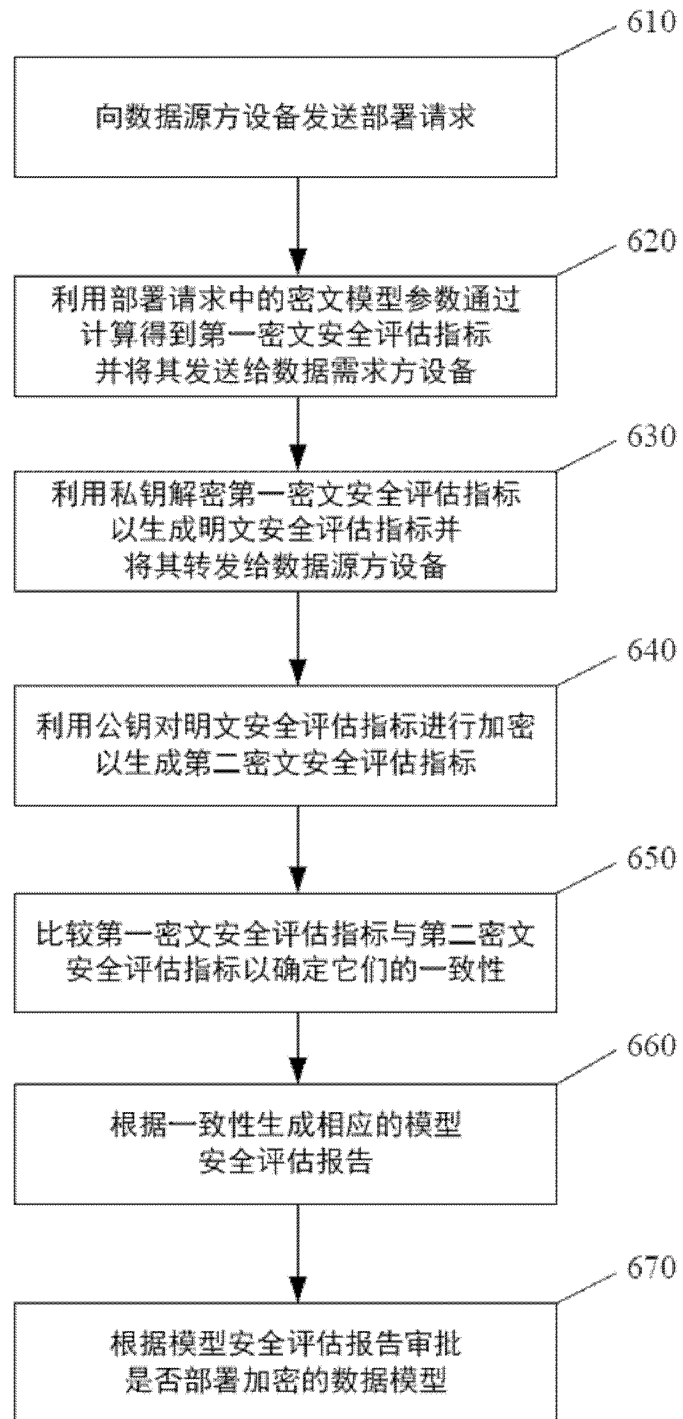


图 6

A. CLASSIFICATION OF SUBJECT MATTER H04L 12/24(2006.01)i; H04L 9/00(2006.01)i; H04L 9/08(2006.01)i According to International Patent Classification (IPC) or to both national classification and IPC				
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) H04L Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) CNABS; CNTXT; VEN; CNKI; USTXT; EPTXT; WOTXT: 安全, 评估, 测评, 评测, 同态, 加密, 公钥, 需求, 请求, 源, 提供, security, assessment, evaluate, homomorphic, encrypt, public key, request, source, provide				
C. DOCUMENTS CONSIDERED TO BE RELEVANT				
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.		
PX	CN 110113203 A (ALIBABA GROUP HOLDING LIMITED) 09 August 2019 (2019-08-09) claims 1-17, and description, paragraphs [0003]-[0102]	1-17		
Y	CN 109190876 A (ALIBABA GROUP HOLDING LIMITED) 11 January 2019 (2019-01-11) description, paragraphs [0036]-[0090]	2, 3		
Y	CN 109033854 A (ALIBABA GROUP HOLDING LIMITED) 18 December 2018 (2018-12-18) description, paragraphs [0031]-[0086]	2, 3		
A	CN 106910057 A (ALIBABA GROUP HOLDING LIMITED) 30 June 2017 (2017-06-30) entire document	1-17		
A	US 2007140479 A1 (MICROSOFT CORPORATION) 21 June 2007 (2007-06-21) entire document	1-17		
☐ Further documents are listed in the continuation of Box C. ☒ See patent family annex.				
<table style="width: 100%; border: none;"> <tr> <td style="width: 50%; vertical-align: top; border: none;"> * Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed </td> <td style="width: 50%; vertical-align: top; border: none;"> "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family </td> </tr> </table>			* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family			
Date of the actual completion of the international search 14 March 2020		Date of mailing of the international search report 14 April 2020		
Name and mailing address of the ISA/CN China National Intellectual Property Administration (ISA/CN) No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088 China Facsimile No. (86-10)62019451		Authorized officer Telephone No.		

INTERNATIONAL SEA^A
Information on patent family members

PCT/CN2020/071605

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	110113203	A	09 August 2019	None			
CN	109190876	A	11 January 2019	None			
CN	109033854	A	18 December 2018	WO	2020015478	A1	23 January 2020
CN	106910057	A	30 June 2017	None			
US	2007140479	A1	21 June 2007	US	7856100	B2	21 December 2010

A. 主题的分类 H04L 12/24(2006.01)i; H04L 9/00(2006.01)i; H04L 9/08(2006.01)i 按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类																				
B. 检索领域 检索的最低限度文献(标明分类系统和分类号) H04L 包含在检索领域中的除最低限度文献以外的检索文献 在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用)) CNABS;CNTXT;VEN;CNKI;USTXT;EPTXT;WOTXT;安全, 评估, 测评, 评测, 同态, 加密, 公钥, 需求, 请求, 源, 提供, security, assessment, evaluate, homomorphic, encrypt, public key, request, source, provide																				
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>PX</td> <td>CN 110113203 A (阿里巴巴集团控股有限公司) 2019年 8月 9日 (2019 - 08 - 09) 权利要求1-17, 说明书[0003]-[0102]段</td> <td>1-17</td> </tr> <tr> <td>Y</td> <td>CN 109190876 A (阿里巴巴集团控股有限公司) 2019年 1月 11日 (2019 - 01 - 11) 说明书[0036]-[0090]段</td> <td>2-3</td> </tr> <tr> <td>Y</td> <td>CN 109033854 A (阿里巴巴集团控股有限公司) 2018年 12月 18日 (2018 - 12 - 18) 说明书[0031]-[0086]段</td> <td>2-3</td> </tr> <tr> <td>A</td> <td>CN 106910057 A (阿里巴巴集团控股有限公司) 2017年 6月 30日 (2017 - 06 - 30) 全文</td> <td>1-17</td> </tr> <tr> <td>A</td> <td>US 2007140479 A1 (MICROSOFT CORP) 2007年 6月 21日 (2007 - 06 - 21) 全文</td> <td>1-17</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	PX	CN 110113203 A (阿里巴巴集团控股有限公司) 2019年 8月 9日 (2019 - 08 - 09) 权利要求1-17, 说明书[0003]-[0102]段	1-17	Y	CN 109190876 A (阿里巴巴集团控股有限公司) 2019年 1月 11日 (2019 - 01 - 11) 说明书[0036]-[0090]段	2-3	Y	CN 109033854 A (阿里巴巴集团控股有限公司) 2018年 12月 18日 (2018 - 12 - 18) 说明书[0031]-[0086]段	2-3	A	CN 106910057 A (阿里巴巴集团控股有限公司) 2017年 6月 30日 (2017 - 06 - 30) 全文	1-17	A	US 2007140479 A1 (MICROSOFT CORP) 2007年 6月 21日 (2007 - 06 - 21) 全文	1-17
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求																		
PX	CN 110113203 A (阿里巴巴集团控股有限公司) 2019年 8月 9日 (2019 - 08 - 09) 权利要求1-17, 说明书[0003]-[0102]段	1-17																		
Y	CN 109190876 A (阿里巴巴集团控股有限公司) 2019年 1月 11日 (2019 - 01 - 11) 说明书[0036]-[0090]段	2-3																		
Y	CN 109033854 A (阿里巴巴集团控股有限公司) 2018年 12月 18日 (2018 - 12 - 18) 说明书[0031]-[0086]段	2-3																		
A	CN 106910057 A (阿里巴巴集团控股有限公司) 2017年 6月 30日 (2017 - 06 - 30) 全文	1-17																		
A	US 2007140479 A1 (MICROSOFT CORP) 2007年 6月 21日 (2007 - 06 - 21) 全文	1-17																		
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。																				
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件																				
国际检索实际完成的日期 2020年 3月 14日		国际检索报告邮寄日期 2020年 4月 14日																		
ISA/CN的名称和邮寄地址 中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451		授权官员 薛乐梅 电话号码 86-(20)-28950448																		

国际检索报
关于同族专利的信息

PCT/CN2020/071605

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	110113203	A	2019年 8月 9日	无			
CN	109190876	A	2019年 1月 11日	无			
CN	109033854	A	2018年 12月 18日	W0	2020015478	A1	2020年 1月 23日
CN	106910057	A	2017年 6月 30日	无			
US	2007140479	A1	2007年 6月 21日	US	7856100	B2	2010年 12月 21日