(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 116438554 A
(43)申请公布日 2023.07.14

(21)申请号 202080107117.3
(22)申请日 2020.11.24
(85)PCT国际申请进入国家阶段日
2023.05.12
(86)PCT国际申请的申请数据
PCT/EP2020/083154 2020.11.24
(87)PCT国际申请的公布数据
WO2022/111789 EN 2022.06.02
(71)申请人 华为技术有限公司
地址 518129 广东省深圳市龙岗区坂田华为总部办公楼
(72)发明人 托马斯·凡内特 周雪冰
(74)专利代理机构 北京同立钧成知识产权代理有限公司 11205
专利代理师 杨文娟 黄健
(51)Int.Cl.
G06N 20/00(2019.01)
H04L 9/40(2022.01)

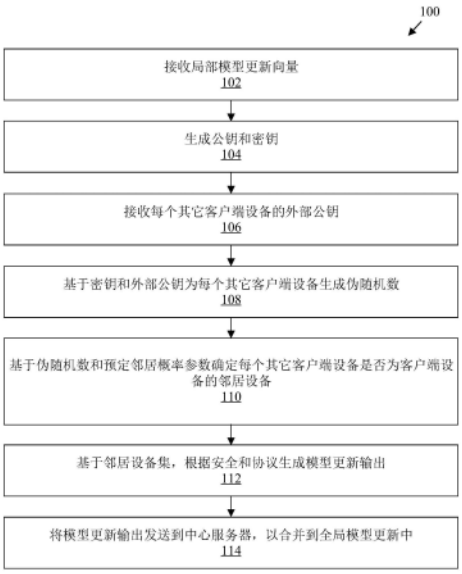
权利要求书4页 说明书21页 附图6页

(54)发明名称

利用随机安全平均的分布式训练

(57)摘要

一种用于网络的分布式机器学习的方法,包括:接收局部模型更新向量。所述方法还包括生成公钥和密钥。所述方法还包括接收每个其它客户端设备的外部公钥。所述方法还包括基于所述密钥和所述外部公钥生成伪随机数。所述方法还包括确定是否将每个其它客户端设备分配给所述客户端设备的邻居设备集。所述方法还包括:基于所述邻居设备集,根据安全和协议生成模型更新输出,并输出所述模型更新输出,用于将模型更新输出发送到中心服务器,以合并到全局模型更新中。所述方法增强了用户数据的隐私保护,同时解决了扩展问题,确保了高实际效用。



CN 116438554 A

1.一种用于网络(306)的分布式机器学习的方法(100),包括:

客户端设备(302A)的处理器(314)接收局部模型更新向量;

所述处理器(314)生成公钥和密钥,其中,所述公钥用于广播到所述网络(306)上的多个其它客户端设备(302B-302N);

所述处理器(314)接收每个所述其它客户端设备(302B-302N)的外部公钥;

对于每个其它客户端设备(302B-302N):

基于所述密钥和所述外部公钥生成伪随机数;

基于所述伪随机数和预定邻居概率参数确定是否将每个所述其它客户端设备(302B-302N)分配给所述客户端设备(302A)的邻居设备集;

所述处理器(314)基于所述邻居设备集,根据安全和协议(310)生成模型更新输出;

所述处理器(314)输出所述模型更新输出,用于将所述模型更新输出发送到中心服务器(304),以合并到全局模型更新(312)中。

2.根据权利要求1所述的方法(100),其中,所述邻居概率参数用于基于预定义值定义所述邻居设备集中的邻居的数量,其中,所述预定义值基于成功攻击的建模风险定义。

3.根据权利要求2所述的方法(100),其中,对于邻居概率参数p和预定义值r:

$f(n_h,p)>1-r$,并且对于精度δ,$f(n_h,p-\delta)<1-r$

其中:

$$\begin{cases} f(n_h,p)=1-\sum_{i=1}^{n_h-1} f(i,p)(1-p)^{i(n_h-i)}, & \text{对于} n_h>1 \\ f(1,p)=1 \end{cases}$$

其中:

$$n_h=\begin{cases} n-\left(\left\lceil\frac{n}{3}\right\rceil-1\right), & \text{对于主动安全} \\ n-2\left(\left\lceil\frac{n}{3}\right\rceil-1\right), & \text{对于被动安全} \end{cases}$$

4.根据上述权利要求中任一项所述的方法(100),其中,根据所述安全和协议(310)生成所述模型更新输出包括:为每个邻居设备生成一次性密码本,并将多个一次性密码本添加到所述局部模型更新向量中,其中,每个邻居设备的所述一次性密码本是基于从所述客户端设备的所述密钥和所述邻居设备的所述外部公钥导出的共享密钥生成的。

5.根据权利要求4所述的方法(100),其中,所述多个客户端设备(302)生成的所有一次性密码本的集合实质上总计为零。

6.根据权利要求1至3中任一项所述的方法(100),其中,根据所述安全和协议(310)生成所述模型更新输出包括:根据所述邻居设备的所述数量将所述局部模型向量更新拆分为多个部分,将所述多个部分发送到相应的邻居设备,从所述邻居设备接收多个外部部分,并将所述多个外部部分相加以形成所述模型更新输出。

7.根据上述权利要求中任一项所述的方法(100),其中,生成所述模型更新输出包括将局部生成的噪声信号添加到所述局部模型更新向量中,所述局部生成的噪声信号的分布是高斯或二项式的。

8.根据权利要求7所述的方法(100),其中,所述局部生成的噪声信号是以由从所述中

心服务器(304)接收的噪声参数定义的标准偏差生成的,其中,每个客户端设备(302A-302N)的所述噪声参数使得来自所述多个客户端设备(302)的对应的局部生成的噪声信号集总计为具有预定标准偏差的全局噪声。

9.根据上述权利要求中任一项所述的方法(100),还包括:将所述局部模型更新向量从浮点值的向量转换为整数向量。

10.根据上述权利要求中任一项所述的方法(100),还包括:向所述中心服务器(304)发送所述公钥,并从所述中心服务器(304)接收所述外部公钥。

11.一种计算机可读介质,用于存储指令,当所述指令被执行时,所述指令使客户端设备处理器(314)执行上述权利要求1至10中任一项所述的方法(100)。

12.一种客户端设备(302A),包括:

训练模块(318A),用于生成局部模型更新向量;

处理器(314),用于生成公钥和密钥;

收发器(316),用于:将所述公钥广播到所述网络(306)上的多个其它客户端设备(302B-302N),接收每个所述其它客户端设备(302B-302N)的外部公钥,并将模型更新输出发送到中心服务器(304),以合并到全局模型更新(312)中;

其中,所述处理器(314)还用于:

基于所述密钥和所述外部公钥,为每个其它客户端设备(302B-302N)生成伪随机数;

基于所述伪随机数和预定邻居概率参数确定是否将每个所述其它客户端设备(302B-302N)分配给所述客户端设备(302A)的邻居设备集;

基于所述邻居设备集,根据安全和协议(310)生成所述模型更新输出。

13.根据权利要求12所述的客户端设备(302A),其中,所述邻居概率参数用于基于预定义值定义所述邻居设备集中的邻居的数量,其中,所述预定义值基于成功攻击的建模风险定义。

14.根据权利要求13所述的客户端设备(302A),其中,对于邻居概率参数p和预定义值r:

$f(n_h,p)>1-r$,并且对于精度δ,$f(n_h,p-\delta)<1-r$

其中:

$$\begin{cases} f(n_h,p)=1-\sum_{i=1}^{n_h-1} f(i,p)(1-p)^{i(n_h-i)}, & \text{对于} n_h>1 \\ f(1,p)=1 \end{cases}$$

其中:

$$n_h=\begin{cases} n-\left(\left\lceil\frac{n}{3}\right\rceil-1\right), & \text{对于主动安全} \\ n-2\left(\left\lceil\frac{n}{3}\right\rceil-1\right), & \text{对于被动安全} \end{cases}$$

15.根据权利要求12至14中任一项所述的客户端设备(302A),其中,所述处理器(314)用于通过以下方式根据所述安全和协议(310)生成所述模型更新输出:为每个邻居设备生成一次性密码本,并将多个一次性密码本添加到所述局部模型更新向量中,其中,每个邻居设备的所述一次性密码本是基于从所述客户端设备的所述密钥和所述邻居设备的所述外

部公钥导出的共享密钥生成的。

16.根据权利要求15所述的客户端设备(302A),其中,由所述多个客户端设备(302)生成的所有一次性密码本的集合实质上总计为零。

17.根据权利要求12至14中任一项所述的客户端设备(302A),其中,所述处理器(314)用于通过以下方式根据所述安全和协议(310)生成所述模型更新输出:根据所述邻居设备的所述数量将所述局部模型向量更新拆分为多个部分,将所述多个部分发送到相应的邻居设备,从所述邻居设备接收外部部分,并将所述多个外部部分相加以形成所述模型更新输出。

18.根据权利要求12至17中任一项所述的客户端设备(302A),其中,所述处理器(314)用于通过将局部生成的噪声信号添加到所述局部模型更新向量中来生成所述模型更新输出,所述局部生成的噪声信号的分布是高斯或二项式的。

19.根据权利要求18所述的客户端设备(302A),其中,所述局部生成的噪声信号是以由从所述中心服务器(304)接收的噪声参数定义的标准偏差生成的,其中,每个客户端设备(302A-302N)的所述噪声参数使得来自所述多个客户端设备(302)的对应的差分私有噪声信号集总计为具有预定标准偏差的全局噪声。

20.根据权利要求12至19中任一项所述的客户端设备(302A),其中,所述处理器(314)还用于将所述局部模型更新向量从浮点值的向量转换为整数向量。

21.根据权利要求12至20中任一项所述的客户端设备(302A),其中,所述收发器(316)还用于向所述中心服务器(304)发送所述公钥,并从所述中心服务器(304)接收所述外部公钥。

22.一种用于网络(306)的分布式机器学习的方法(200),包括:

中心服务器(304)接收多个客户端设备(302)发送的多个模型更新输出;

所述中心服务器(304)基于所述多个模型更新输出确定模型更新的总和;

所述中心服务器(304)基于所述模型更新的所述总和更新全局模型;

所述中心服务器(304)向每个客户端设备(302A-302N)发送所述全局模型更新(312)。

23.根据权利要求22所述的方法(200),其中,更新所述全局模型包括将整数向量转换为浮点向量。

24.根据权利要求22或23所述的方法(200),还包括:

所述中心服务器(304)确定客户端设备已掉线;

所述中心服务器(304)基于添加到所述模型更新的总和中的局部噪声的预定方差值,将附加噪声添加到所述模型更新的总和中。

25.根据权利要求22至24中任一项所述的方法(200),还包括:执行客户端掉线恢复协议,包括:

所述中心服务器(304)从所述多个客户端设备(302)接收多个密钥部分,表示每个客户端设备(302)的密钥集,所述密钥集根据密钥共享协议拆分为多个密钥部分,分布在所述客户端设备中,并由每个客户端设备发送到所述中心服务器(304);

所述中心服务器(304)确定客户端设备已掉线;

所述中心服务器(304)组合接收到的对应于掉线客户端的多个密钥部分,以恢复对应于所述掉线客户端的密钥。

26.一种计算机可读介质,用于存储指令,当所述指令被执行时,所述指令使中心服务器处理器(322)执行权利要求22至25中任一项所述的方法(200)。

27.一种中心服务器(304),包括:

收发器(324),用于接收多个客户端设备(302)发送的多个模型更新输出,并向每个客户端设备(302A-302N)发送全局模型更新(312);

处理器(322),用于基于所述多个模型更新输出确定模型更新的总和,并基于所述模型更新的总和更新全局模型以生成所述全局模型更新(312)。

28.根据权利要求27所述的中心服务器(304),其中,更新所述全局模型包括将整数向量转换为浮点向量。

29.根据权利要求27或28所述的中心服务器(304),其中,所述处理器(322)还用于:

确定客户端设备已掉线;

基于添加到所述模型更新的总和中的局部噪声的预定方差值,将附加噪声添加到所述模型更新的总和中。

30.根据权利要求27至28中任一项所述的中心服务器(304),还用于:执行客户端掉线恢复协议,包括:

所述收发器(324)从所述多个客户端设备(302)接收多个密钥部分,表示每个客户端设备(302)的密钥集,所述密钥集根据密钥共享协议拆分为多个密钥部分,分布在所述客户端设备中,并由每个客户端设备发送到所述中心服务器(304);

所述处理器(322)确定客户端设备已掉线;

所述处理器(322)组合接收到的对应于掉线客户端的多个密钥部分,以恢复对应于所述掉线客户端的密钥。

利用随机安全平均的分布式训练

技术领域

[0001] 本公开大体上涉及数据安全和机器学习领域;更具体地,涉及利用随机安全平均协议的分布式机器学习(或训练)的方法和设备。

背景技术

[0002] 如今,众多服务使用机器学习(machine learning,ML)来提高其在日常生活中的可用性。通常,机器学习算法包括两个阶段:训练阶段和预测阶段。利用机器学习算法,开发了使用样本数据训练的模型。样本数据在训练阶段期间使用,以便在模型的预测阶段期间进行预测(或决策)。通常,当在模型的训练阶段和预测阶段期间使用客户的数据或个人数据时,如何实现隐私保护就成为一个主要的技术问题。

[0003] 目前,已经提出了一些方法来增强隐私,例如传统的联邦学习(federated learning,FL)。传统的联邦学习用作模型的分布式训练机制,而不强制客户(或客户端)在训练阶段期间上传其原始数据。因此,传统的联邦学习降低了总通信成本,同时部分增强了隐私。传统的联邦学习在一定程度上解决了以低成本不断更新模型的需要,并表现出通过使用客户的(或客户端的)局部数据局部训练模型的可能性。但是,传统的联邦学习存在隐私问题,例如,个人数据可以直接从传统服务器接收的模型更新中提取。此外,与其它客户端共享的模型(或训练后的模型)也可能泄露参与模型训练阶段的客户(或客户端)的个人信息。传统的联邦学习的隐私问题可以通过使用传统的安全和技术和传统的差分隐私(differential privacy,DP)技术在一定程度上解决。换句话说,通过使用传统的安全和技术或传统的差分隐私技术来执行模型(或机器学习模型)的分布式训练。传统的安全和技术是一种密码技术,该技术同时对多个客户(或客户端)的更新总和进行分布式和私有计算,并防止传统服务器观察任何单独的更新。因此,传统的安全和技术以分布式学习的方式保护单独的客户端的隐私。但是,当客户端数量增加时,传统的安全和技术会产生较高的计算成本,并且不会为最终训练的模型提供隐私保证。在传统的差分隐私技术中,传统的服务器在训练阶段的每次全局迭代之后添加噪声。因此,传统的差分隐私技术为最终训练的模型提供了一定量的正式隐私保证。但是,传统的差分隐私技术并不解决传统服务器有关的隐私问题。此外,传统的差分隐私技术需要大量的客户端,而传统的安全和技术在客户端数量增加时表现出扩展问题(即不能适当扩展,并且容易出错)。因此,当涉及大量客户端时,或当客户端(即客户端设备)数量增加时,存在对客户数据或个人数据的隐私保护不足的技术问题,扩展问题就会显现出来。

[0004] 因此,根据前面的讨论,需要克服与用于分布式机器学习的传统的技术有关的上述缺点。

发明内容

[0005] 本公开寻求提供用于分布式机器学习的方法和设备,这些方法和设备改进了用户数据或个人数据的隐私保护,具有高性能和高效用。本公开寻求提供一种方案,用于解决当

涉及大量客户端时或当客户端(即客户端设备)的数量增加导致扩展问题时,客户的数据或个人数据的隐私保护不足的现有问题。本公开的目的是提供一种至少部分地克服现有技术中遇到的问题的方案,并提供用于分布式机器学习的方法和设备,这些方法和设备改进了包括个人数据的用户数据的隐私保护,同时解决了扩展问题,并确保高性能和高实际效用。

[0006] 本公开的目的是通过所附独立权利要求中提供的方案实现的。本公开的有利实现方式在从属权利要求中进一步定义。

[0007] 在一个方面,本公开提供了一种用于网络的分布式机器学习的方法,包括:客户端设备的处理器接收局部模型更新向量。该方法还包括:处理器生成公钥和密钥,其中,公钥用于广播到网络上的多个其它客户端设备。该方法还包括处理器接收每个其它客户端设备的外部公钥。该方法还包括,对于每个其它客户端设备:基于密钥和外部公钥生成伪随机数;基于伪随机数和预定邻居概率参数确定是否将每个其它客户端设备分配给客户端设备的邻居设备集。该方法还包括:处理器基于邻居设备集,根据安全和协议生成模型更新输出;处理器输出模型更新输出,用于将模型更新输出发送到中心服务器,以合并到全局模型更新中。

[0008] 本公开的方法使用利用随机安全平均(random secure averaging,RdSA)协议的分布式机器学习。所公开的方法提供了客户端设备的(或用户的)数据(包括个人数据)的增强隐私保护、改进的性能和高效用。所公开的方法以协调和协同的方式使用安全和技术(或密码技术)和差分隐私技术的组合,以增强客户端设备的数据隐私。使用安全和协议的方法防止中心服务器从任何单独的客户端访问私有信息。有益的是,本公开的方法使用邻居选择算法。基于邻居选择算法,在随机安全平均(RdSA)协议的每次迭代中选择邻居集。通过使用邻居选择算法,在RdSA协议的每次迭代中,客户端设备的数量减少,这进一步解决了扩展问题。邻居选择算法的使用还支持使用差分隐私技术,以进一步增强客户端设备的数据的隐私保护。

[0009] 在一种实现方式中,邻居概率参数用于基于预定义值定义邻居设备集中的邻居的数量,其中,预定义值基于成功攻击的建模风险定义。

[0010] 所公开的方法提供了用户数据的增强隐私保护和改进的性能,即使在存在大量客户端设备的情况下也是如此。通过使用邻居选择算法,解决了传统安全和技术的扩展问题。通过使用邻居选择算法,客户端设备基于邻居概率参数选择自己的邻居。邻居的选择方式是,每对客户端设备就它们是否为邻居进行协商,并且任何其它客户端设备(例如中心服务器)对它们是否为邻居一无所知。当客户端设备的数量较多时,邻居选择算法仍然通过选择邻居客户端设备集,以较低的计算成本提供强大的隐私。此外,使用邻居选择算法使得能够使用差分隐私技术,以进一步增强客户端的数据隐私。

[0011] 在另一种实现方式中,对于邻居概率参数p和预定义值r:

[0012] $f(n_h,p)>1-r$,并且对于精度δ,$f(n_h,p-\delta)<1-r$

[0013] 其中:

[0014] $$\begin{cases} f(n_h,p)=1-\sum_{i=1}^{n_h-1} f(i,p)(1-p)^{i(n_h-i)}, & \text{对于} n_h>1 \\ f(1,p)=1 \end{cases}$$

[0015] 其中：

[0016] $$n_h = \begin{cases} n - \left(\left\lceil \frac{n}{3} \right\rceil - 1\right)，\text{对于主动安全} \\ n - 2\left(\left\lceil \frac{n}{3} \right\rceil - 1\right)，\text{对于被动安全} \end{cases}$$

[0017] 通过配置邻居概率参数p，所公开的方法可以与主动安全模型和被动安全模型一起使用。主动安全模型提供了针对非常强大(例如国家级或州级)攻击者或内部攻击者的保护。被动安全模型提供了防止数据泄露、远程黑客和遵守法规的保护。

[0018] 在另一种实现方式中，根据安全和协议生成模型更新输出包括：为每个邻居设备生成一次性密码本，并将多个一次性密码本添加到局部模型更新向量中，其中，每个邻居设备的一次性密码本是基于从客户端设备的密钥和邻居设备的外部公钥导出的共享密钥生成的。

[0019] 将多个一次性密码本添加到客户端设备的局部模型更新向量中，加密了客户端设备的密钥值。

[0020] 在另一种实现方式中，多个客户端设备生成的所有一次性密码本的集合实质上总计为零。

[0021] 多个客户端设备生成的所有一次性密码本的集合在中心服务器处相加，实质上为零，这意味着中心服务器获得了客户端的(未加密的)密钥的总和。

[0022] 在另一种实现方式中，根据安全和协议生成模型更新输出包括根据邻居设备的数量将局部模型向量更新拆分为多个部分，将多个部分发送到相应的邻居设备，从邻居设备接收外部部分，并将多个外部部分相加以形成模型更新输出。

[0023] 与扩展问题突出的传统安全和技术相比，根据邻居设备的数量将局部模型向量更新拆分为多个部分改善了客户端侧的密钥共享。

[0024] 在另一种实现方式中，生成模型更新输出包括将局部生成的噪声信号添加到局部模型更新向量中，局部生成的噪声信号的分布是高斯或二项式的。

[0025] 在差分隐私技术中，将局部生成的噪声信号添加到局部模型更新向量中，以提供改进的局部隐私，这意味着私有数据得到充分保护。中心服务器无法观察来自任何单独的客户端的任何私有信息。噪声信号是在局部(客户端侧)添加的，而不是在中心(服务器侧)添加的，因为局部隐私比中心隐私更可取。局部生成的二项式分布噪声信号用于提供局部隐私保护。

[0026] 在另一种实现方式中，局部生成的噪声信号是以由从中心服务器接收的噪声参数定义的标准偏差生成的，其中，每个客户端设备的噪声参数使得来自多个客户端设备的对应的局部生成的噪声信号集总计为具有预定标准偏差的全局噪声。

[0027] 通过生成具有中心服务器选择的标准偏差的局部生成的噪声信号，使全局噪声具有二项式分布。

[0028] 在另一种实现方式中，该方法还包括将局部模型更新向量从浮点值的向量转换为整数向量。

[0029] 将局部模型更新向量从浮点值的向量转换为(模)整数向量确保了客户端的安全性。

[0030] 在另一种实现方式中，该方法还包括向中心服务器发送公钥，并从中心服务器接

收外部公钥。

[0031] 公钥被传送到中心服务器，中心服务器进一步将公钥传送到其它客户端设备，以便客户端设备可以与其它客户端设备执行密钥协商并选择其邻居。

[0032] 在另一种实现方式中，一种计算机可读介质，用于存储指令，当这些指令被执行时，这些指令使客户端设备处理器执行该方法。

[0033] 客户端设备处理器实现了该方法的所有优点和效果。

[0034] 在另一方面，本公开提供了一种客户端设备，包括用于生成局部模型更新向量的训练模块。客户端设备还包括用于生成公钥和密钥的处理器。客户端设备还包括收发器，用于：将公钥广播到网络上的多个其它客户端设备，接收每个其它客户端设备的外部公钥，并将模型更新输出发送到中心服务器，以合并到全局模型更新中。客户端设备的处理器还用于基于密钥和外部公钥为每个其它客户端设备生成伪随机数。客户端设备的处理器还用于基于伪随机数和预定邻居概率参数确定是否将每个其它客户端设备分配给客户端设备的邻居设备集。客户端设备的处理器还用于基于邻居设备集，根据安全和协议生成模型更新输出。

[0035] 本公开的客户端设备表现出个人数据的增强的局部隐私。客户端设备使用邻居选择算法，该算法使RdSA协议的每次迭代中客户端设备的数量减少。因此，即使在存在大量客户端设备的情况下，客户端设备也表现出降低的计算成本。客户端设备使用安全和技术(或密码技术)，并防止中心服务器从任何单独的客户端访问私有信息。

[0036] 在一种实现方式中，邻居概率参数用于基于预定义值定义邻居设备集中的邻居的数量，其中，预定义值基于成功攻击的建模风险定义。

[0037] 客户端设备通过使用邻居选择算法中的邻居概率参数选择其邻居。邻居的选择方式是，每对客户端设备就它们是否为邻居进行协商，并且任何其它客户端设备(例如中心服务器)对它们是否为邻居一无所知。

[0038] 在另一种实现方式中，对于邻居概率参数p和预定义值r：

[0039] $f(n_h,p)>1-r$，并且对于精度δ，$f(n_h,p-\delta)<1-r$

[0040] 其中：

[0041] $$\begin{cases} f(n_h,p)=1-\sum_{i=1}^{n_h-1} f(i,p)(1-p)^{i(n_h-i)}, & \text{对于} n_h>1 \\ f(1,p)=1 \end{cases}$$

[0042] 其中：

[0043] $$n_h=\begin{cases} n-\left(\left\lceil\frac{n}{3}\right\rceil-1\right), & \text{对于主动安全} \\ n-2\left(\left\lceil\frac{n}{3}\right\rceil-1\right), & \text{对于被动安全} \end{cases}$$

[0044] 通过配置邻居概率参数p，客户端设备可以在主动安全模型中使用和用于被动安全模型。

[0045] 在另一种实现方式中，处理器还用于通过以下方式根据安全和协议生成模型更新输出：为每个邻居设备生成一次性密码本，并将多个一次性密码本添加到局部模型更新向量中，其中，每个邻居设备的一次性密码本是基于从客户端设备的密钥和邻居设备的外部

公钥导出的共享密钥生成的。

[0046] 将多个一次性密码本添加到客户端设备的局部模型更新向量中，加密了客户端设备的密钥值。

[0047] 在另一种实现方式中，多个客户端设备生成的所有一次性密码本的集合实质上总计为零。

[0048] 多个客户端设备生成的所有一次性密码本的集合在中心服务器处相加，实质上为零，这意味着中心服务器获得了客户端的(未加密的)密钥的总和。

[0049] 在另一种实现方式中，处理器还用于通过以下方式根据安全和协议生成模型更新输出：根据邻居设备的数量将局部模型向量更新拆分为多个部分，将多个部分发送到相应的邻居设备，从邻居设备接收外部部分，并将多个外部部分相加以形成模型更新输出。

[0050] 与扩展问题突出的传统安全和技术相比，根据邻居设备的数量将局部模型向量更新拆分为多个部分改善了客户端侧的密钥共享。

[0051] 在另一种实现方式中，处理器用于通过将局部生成的噪声信号添加到局部模型更新向量中来生成模型更新输出，局部生成的噪声信号的分布是高斯或二项式的。

[0052] 在差分隐私技术中，客户端设备将局部生成的噪声信号添加到局部模型更新向量中，以提供改进的局部隐私，这意味着私有数据得到充分保护。中心服务器无法观察来自任何单独的客户端的任何私有信息。噪声信号是在局部(客户端)添加的，而不是在中心(服务器侧)添加的，因为局部隐私比中心隐私更可取。局部生成的二项式分布噪声信号用于提供局部隐私保护。

[0053] 在另一种实现方式中，局部生成的噪声信号是以由从中心服务器接收的噪声参数定义的标准偏差生成的，其中，每个客户端设备的噪声参数使得来自多个客户端设备的对应的差分私有噪声信号集总计为具有预定标准偏差的全局噪声。

[0054] 通过生成具有中心服务器选择的标准偏差的局部生成的噪声信号，使全局噪声具有二项式分布。

[0055] 在另一种实现方式中，处理器还用于将局部模型更新向量从浮点值的向量转换为整数向量。

[0056] 客户端设备将局部模型更新向量从浮点值的向量转换为(模)整数向量，以确保安全性。

[0057] 在另一种实现方式中，收发器还用于向中心服务器发送公钥，并从中心服务器接收外部公钥。

[0058] 客户端设备将公钥传送到中心服务器，中心服务器进一步将公钥传送到其它客户端设备，以便客户端设备可以与其它客户端设备执行密钥协商并选择其邻居。

[0059] 在另一方面，本公开提供了一种用于网络的分布式机器学习的方法，包括：中心服务器接收多个客户端设备发送的多个模型更新输出；该方法还包括：中心服务器基于多个模型更新输出确定模型更新的总和。该方法还包括：中心服务器基于模型更新的总和更新全局模型。该方法还包括：中心服务器向每个客户端设备发送全局模型更新。

[0060] 中心服务器对多个客户端设备发送的多个模型更新输出执行随机安全平均。基于随机安全平均，中心服务器确定进一步与多个客户端设备共享的全局模型更新。由于全局模型更新，所以多个客户端设备表现出提高的准确性和增强的个人数据隐私保护。

[0061] 在一种实现方式中,更新全局模型包括将整数向量转换为浮点向量。

[0062] 全局模型包括从整数向量到浮点向量的转换,以易于操作。

[0063] 在另一种实现方式中,该方法还包括中心服务器确定客户端设备已掉线。该方法还包括中心服务器基于添加到模型更新的总和中的局部噪声的预定方差值,将附加噪声添加到模型更新的总和中。

[0064] 中心服务器确定在执行RdSA协议期间是否有任何客户端设备掉线。在确定掉线之后,中心服务器通过将附加的噪声添加到模型更新的总和中来补偿丢失的噪声。这样,中心服务器也执行差分噪声恢复。

[0065] 在另一种实现方式中,该方法还包括执行客户端掉线恢复协议(client dropout recovery protocol),包括:中心服务器从多个客户端设备接收多个密钥部分,表示每个客户端设备的密钥集,该密钥集根据密钥共享协议拆分为多个密钥部分,分布在客户端设备中,并由每个客户端设备发送到中心服务器。该方法还包括中心服务器确定客户端设备已掉线。该方法还包括中心服务器组合接收到的对应于掉线客户端的多个密钥部分,以恢复对应于掉线客户端的密钥。

[0066] 中心服务器在客户端掉线的情况下通过使用密钥共享协议执行客户端掉线恢复。

[0067] 在另一种实现方式中,一种计算机可读介质,用于存储指令,当这些指令被执行时,这些指令使中心服务器处理器执行该方法。

[0068] 中心服务器处理器实现了该方法的所有优点和效果。

[0069] 在另一方面,本公开提供了一种中心服务器,该中心服务器包括收发器,该收发器用于接收多个客户端设备发送的多个模型更新输出,并向每个客户端设备发送全局模型更新。中心服务器还包括处理器,该处理器用于基于多个模型更新输出确定模型更新的总和,并基于模型更新的总和更新全局模型以生成全局模型更新。

[0070] 中心服务器对多个客户端设备发送的多个模型更新输出执行随机安全平均。基于随机安全平均,中心服务器确定进一步与多个客户端设备共享的全局模型更新。由于全局模型更新,所以多个客户端设备表现出提高的准确性和增强的个人数据隐私保护。

[0071] 在一种实现方式中,更新全局模型包括将整数向量转换为浮点向量。

[0072] 全局模型包括从整数向量到浮点向量的转换,以使中心服务器易于操作。

[0073] 在另一种实现方式中,处理器还用于确定客户端设备已掉线。处理器还用于基于添加到模型更新的总和中的局部噪声的预定方差值,将附加噪声添加到模型更新的总和中。

[0074] 中心服务器确定在执行RdSA协议期间是否有任何客户端设备掉线。在确定掉线之后,中心服务器通过将附加噪声添加到模型更新的总和中来补偿丢失的噪声。这样,中心服务器也执行差分噪声恢复。

[0075] 在另一种实现方式中,中心服务器还用于执行客户端掉线恢复协议,包括:收发器从多个客户端设备接收多个密钥部分,表示每个客户端设备的密钥集,该密钥集根据密钥共享协议拆分为多个密钥部分,分布在客户端设备中,并由每个客户端设备发送到中心服务器。中心服务器还用于执行客户端掉线恢复协议,包括:处理器确定客户端设备已掉线。中心服务器还用于执行客户端掉线恢复协议,包括:处理器组合接收到的对应于掉线客户端的多个密钥部分,以恢复对应于掉线客户端的密钥。

[0076] 中心服务器在客户端掉线的情况下通过使用密钥共享协议执行客户端掉线恢复。

[0077] 需要说明的是，本申请中描述的所有设备、元件、电路、单元和模块可以通过软件或硬件元件或其任何类型的组合实现。本申请中描述的各种实体执行的所有步骤和所描述的将由各种实体执行的功能旨在表明相应的实体适于或用于执行相应的步骤和功能。虽然在以下具体实施例的描述中，外部实体执行的具体功能或步骤没有在执行具体步骤或功能的实体的具体详述元件的描述中反映，但是技术人员应清楚，这些方法和功能可以通过相应的硬件或软件元件或其任何组合实现。应理解，可以对本公开的特征进行各种组合，这不会偏离所附权利要求书定义的本公开的范围。

[0078] 本公开的附加方面、优点、特征和目的从附图和结合以下所附权利要求书解释的说明性实现方式的详细描述中变得显而易见。

附图说明

[0079] 结合附图阅读，可以更好地理解以上概述以及以下说明性实施例的详细描述。为了说明本公开，在附图中示出了本公开的示例性结构。但是，本公开不限于本文公开的具体方法和工具。此外，本领域技术人员应理解，附图不是按比例绘制的。在可能的情况下，相同的元件用相同的数字表示。

[0080] 现在参考下图仅作为示例来描述本公开的实施例。

[0081] 图1是根据本公开的实施例的用于网络的分布式机器学习的方法的流程图。

[0082] 图2是根据本公开另一实施例的用于网络的分布式机器学习的方法的流程图。

[0083] 图3A是根据本公开的实施例的描述利用随机安全平均的分布式机器学习的网络环境图。

[0084] 图3B是根据本公开的实施例的客户端设备的各种示例性组件的框图。

[0085] 图3C是根据本公开的实施例的中心服务器的各种示例性组件的框图。

[0086] 图4是根据本公开的另一个实施例的描述利用随机安全平均的分布式机器学习的网络环境图。

[0087] 图5示出了根据本公开的实施例的分布式机器学习的示例性实现场景。

[0088] 在附图中，带下划线的数字用于表示带下划线的数字所在的项目或与带下划线的数字相邻的项目。不带下划线的数字是指由将不带下划线的数字与项目连接的线所标识的项目。当一个数字不带下划线并具有关联的箭头时，不带下划线的数字用于标识箭头指向的一般项目。

具体实施方式

[0089] 以下详细描述说明了本公开的实施例以及可以实现这些实施例的方式。尽管已经公开了实施本公开的一些模式，但本领域技术人员应认识到，也可以存在用于实施或实践本公开的其它实施例。

[0090] 图1是根据本公开的实施例的用于网络的分布式机器学习的方法的流程图。参考图1，示出了用于网络的分布式机器学习的方法100。方法100包括步骤102、104、106、108、110、112和114。在一种实现方式中，方法100由客户端设备执行，例如在图3A和图3B中详细描述。

[0091] 在步骤102中，方法100包括：客户端设备的处理器接收局部模型更新向量。客户端设备的处理器通过使用客户端设备的局部数据或原始数据更新局部模型更新向量。

[0092] 在步骤104中，方法100还包括：处理器生成公钥和密钥，其中，公钥用于广播到网络上的多个其它客户端设备。生成的公钥和客户端设备的密钥(或私钥)用于与另一个客户端设备执行密钥协商。

[0093] 在步骤106中，方法100还包括处理器接收每个其它客户端设备的外部公钥。客户端设备的处理器与网络上的多个其它客户端设备执行密钥协商，并基于接收到的外部公钥生成密钥对。密钥协商是通过使用密钥协商方案来执行的，例如椭圆曲线-迪菲-赫尔曼(elliptic-curve-diffie-hellman，ECDH)，该方案支持两个客户端设备在不安全的信道上建立共享密钥，每个客户端设备都具有椭圆曲线外部公钥-密钥对(也称为共享密钥)。例如，$s_{u,v}$是两个客户端(例如客户端u和客户端v)之间的共享密钥。共享密钥是通过将安全密钥导出函数(例如基于哈希的密钥导出函数(hash-based key derivation function，HKDF))应用于椭圆曲线外部公钥-密钥对(或共享密钥)来获得的。

[0094] 在步骤108中，方法100还包括，对于每个其它客户端设备：基于密钥和外部公钥生成伪随机数。基于密钥和外部公钥的椭圆曲线外部公钥-密钥对用于为每个其他客户端设备生成伪随机数。共享密钥$s_{u,v}$用作伪随机数生成器的种子，该伪随机数生成器是安全和确定性的。例如，$\left(rand_i^{u,v}\right)_{i\geq 0}$是由客户端对(u,v)生成的随机数序列。

[0095] 在步骤110中，方法100还包括，对于每个其它客户端设备：基于伪随机数和预定邻居概率参数确定是否将每个其它客户端设备分配给客户端设备的邻居设备集。例如，对于每个其它客户端设备(即客户端v)，客户端设备(即客户端u)生成伪随机数，例如$rand_0^{u,v}$，至少128位。客户端设备的(即客户端u的)邻居设备集是：

[0096] $$N_p(u)=\{v\left|rand_0^{u,v}\ mod\ 2^d<p2^d\right.\}$$

[0097] 其中：p是每个其它客户端设备的预定邻居概率参数，d是生成的伪随机数的比特大小。

[0098] 在步骤112中，方法100还包括：处理器基于邻居设备集，根据安全和协议生成模型更新输出。安全和协议用于通过使用加性密钥共享方案或基于一次性密码本(one-time-pad，OTP)的方案生成模型更新输出。

[0099] 在步骤114中，方法100还包括：处理器输出模型更新输出，用于将模型更新输出发送到中心服务器，以合并到全局模型更新中。客户端设备的处理器将局部训练的模型更新输出发送到中心服务器，以最终确定进一步与每个客户端设备共享的全局模型更新。

[0100] 根据实施例，邻居概率参数用于基于预定义值定义邻居设备集中的邻居的数量，其中，预定义值基于成功攻击的建模风险定义。对于N轮训练，基于成功攻击的建模风险的预定义值表示为r_{tot}。为每轮训练计算的预定义值是$r=1-\sqrt[N]{1-r_{tot}}$，使得$1-(1-r)^N=r_{tot}$。成功攻击的建模风险也被称为风险参数α，使得α∈[0,1]。风险参数α表示至少一个客户端的隐私得不到中心服务器保护的概率。风险参数α可以针对每轮训练计算，也可以针对N轮训练(或整个训练)计算。在一个示例中，如果每轮的α＝0.01，则对于每轮训练，至少一个客户端的隐私在中心服务器方面得不到保护的可能性为1％。在另一个示例中，如果100(即N＝100)轮训练的α＝0.1，则在训练期间的某个时候，至少一个客户的隐私得不到保护

的可能性为10％。风险参数α可以基于业务需要选择。较高的隐私风险将转化为更好的计算性能。通常，风险参数α选择与传统的差分隐私技术中使用的标准δ参数相同的数量级。

[0101] 根据实施例，对于邻居概率参数p和预定义值r：

[0102] $f(n_h,p)>1-r$，并且对于精度δ，$f(n_h,p-\delta)<1-r$

[0103] 其中：

[0104] $$\begin{cases} f(n_h,p)=1-\sum_{i=1}^{n_h-1} f(i,p)(1-p)^{i(n_h-i)}, & \text{对于} n_h>1 \\ f(1,p)=1 \end{cases}$$

[0105] 其中：

[0106] $$n_h=\begin{cases} n-\left(\left\lceil\frac{n}{3}\right\rceil-1\right), & \text{对于主动安全} \\ n-2\left(\left\lceil\frac{n}{3}\right\rceil-1\right), & \text{对于被动安全} \end{cases}$$

[0107] $f(n_h,p)$是n_h个顶点上的随机图连接的概率。方法100包括主动安全模型和被动安全模型。主动安全模型在恶意服务器尝试通过修改多个客户端发送的消息来恢复客户端的更新的情况下提供隐私保护。主动安全模型提供了适合高度机密数据(例如医疗数据或财务数据)的安全性。此外，主动安全模型还提供了针对非常强大(例如国家级或州级)攻击者或内部攻击者的保护。被动安全模型在诚实但好奇的服务器的情况下提供了隐私保护，在这种情况下，攻击者尝试通过侦听客户端数量与传统服务器之间的对话来恢复客户端的更新。被动安全模型提供了适合例如个人数据的安全性。附加地，被动安全模型提供了防止数据泄露、远程黑客和遵守法规的保护。

[0108] 根根据实施例，根据安全和协议生成模型更新输出包括：为每个邻居设备生成一次性密码本，并将多个一次性密码本添加到局部模型更新向量中，其中，每个邻居设备的一次性密码本是基于从客户端设备的密钥和邻居设备的外部公钥导出的共享密钥生成的。通常，在密码学中，一次性密码本是一种加密技术，使用客户端设备的密钥与邻居设备的外部公钥来对共享密钥进行加密。例如，对于每个其它客户端设备(即客户端v)，客户端设备(即客户端u)通过使用生成的伪随机数序列$\left(rand_i^{u,v}\right)_{i>0}$生成一次性密码本(OTP)，使得两个邻居设备使用相同的OTP。此后，为多个其它客户端设备生成的多个一次性密码本(即加密密钥)被添加到局部模型更新向量中。因此，为每个客户端设备的(未加密的)密钥获得总和。根据实施例，每个客户端的邻居数量通过以上所描述的邻居选择算法显著减少，这降低了与客户端设备上的聚合步骤关联的成本。

[0109] 根据实施例，由多个客户端设备生成的所有一次性密码本的集合实质上总计为零。每对邻居的生成的OTP相互抵消，因此，由多个客户端设备生成的所有一次性密码本的集合之和等于零。

[0110] 替代地，根据实施例，根据安全和协议生成模型更新输出包括根据邻居设备的数量将局部模型向量更新拆分为多个部分，将多个部分发送到相应的邻居设备，从邻居设备接收外部部分，并将多个外部部分相加以形成模型更新输出。安全和协议用于通过使用加性密钥共享方案(例如Shamir密钥共享(Shamir Secret Sharing，SSS))生成模型更新输出。在加性密钥共享方案中，局部模型更新向量被量化并视为固定长度整数的向量。客户端

设备将每个整数拆分为多个部分，每个邻居设备一个部分。多个部分被发送到相应的邻居设备。此后，从邻居设备接收外部部分，并添加外部部分，以形成模型更新输出。

[0111] 根据实施例，生成模型更新输出包括将局部生成的噪声信号添加到局部模型更新向量中，局部生成的噪声信号的分布是高斯或二项式的。局部生成的噪声信号具有高斯分布或二项式分布。二项式分布通常是首选的，具体是对于小字大小。

[0112] 根据实施例，局部生成的噪声信号是以由从中心服务器接收的噪声参数定义的标准偏差生成的，其中，每个客户端设备的噪声参数使得来自多个客户端设备的对应的局部生成的噪声信号集总计为具有预定标准偏差的全局噪声。来自多个客户端设备的对应的局部生成的噪声信号集在与全局噪声相加时，具有二项式分布和预定标准差(σ)。

[0113] 根据实施例，方法100还包括将局部模型更新向量从浮点值的向量转换为整数向量。通过量化过程，局部模型更新向量从浮点值的向量转换为整数(或模)向量。在量化过程中，使用了无偏、空间高效的算法。例如，对单独的模型参数要求边界支持高效地映射到整数空间。这些边界可以直接由中心服务器提供，也可以从差分隐私(DP)特定参数(例如模型更新截断边界S)推断。在求和来自多个其它客户端设备的加权更新时，量化必须考虑溢出的风险。具体地，多个其它客户端设备必须知道在此次迭代期间要使用的权重之和。

[0114] 根据实施例，方法100还包括向中心服务器发送公钥，并从中心服务器接收外部公钥。由客户端设备的处理器生成的公钥被传送到中心服务器。客户端设备的处理器用于从中心服务器接收每个其它客户端设备的外部公钥。

[0115] 根据实施例，一种计算机可读介质，用于存储指令，当这些指令被执行时，这些指令使客户端设备处理器执行方法100。客户端设备的处理器用于执行方法100。

[0116] 步骤102、104、106、108、110、112和114仅仅是说明性的，还可以提供其它替代方案，其中，添加一个或多个步骤、删除一个或多个步骤，或以不同的顺序提供一个或多个步骤，这不会偏离本文权利要求的范围。

[0117] 图2是根据本公开的另一个实施例的用于网络的分布式机器学习的方法的流程图。结合图1的元件描述图2。参考图2，示出了用于网络的分布式机器学习的方法200。方法200包括步骤202、204、206和208。在一种实现方式中，方法200由中心服务器执行，例如在图3A和图3C中详细描述。

[0118] 在步骤202中，方法200包括中心服务器接收多个客户端设备发送的多个模型更新输出。中心服务器接收对应于多个客户端设备的多个模型更新输出。多个客户端设备中的每个通过训练其相应的局部模型更新向量来生成多个模型更新输出之一。

[0119] 在步骤204中，方法200还包括：中心服务器基于多个模型更新输出确定模型更新的总和。中心服务器基于多个模型更新输出获得表示模型更新的总和的向量。

[0120] 在步骤206中，方法200还包括中心服务器基于模型更新的总和更新全局模型。中心服务器基于模型更新的总和确定与多个客户端设备中的每个共享的全局模型更新。

[0121] 在步骤208中，方法200还包括中心服务器向每个客户端设备发送全局模型更新。中心服务器将全局模型更新发送到每个客户端设备，用于下一次迭代。

[0122] 根据实施例，更新全局模型包括将整数向量转换为浮点向量。通过反量化过程从整数(模)向量转换为浮点向量，更新全局模型。从整数(模)向量到浮点向量的转换会产生噪声模型更新的总和。如果客户端设备在该过程期间已掉线，则中心服务器在执行反量化

过程时考虑丢失的噪声。

[0123] 根据实施例，该方法还包括中心服务器确定客户端设备已掉线。该方法还包括中心服务器基于添加到模型更新的总和中的局部噪声的预定方差值，将附加噪声添加到模型更新的总和中。在一种情况下，中心服务器用于在执行随机安全平均(RdSA)协议期间，多个客户端设备中的一个或多个已掉线的情况下进行噪声补偿。中心服务器用于基于添加到模型更新的总和中的局部噪声的预定方差值，将附加噪声添加到模型更新的总和中，例如在图3A中详细描述。

[0124] 根据实施例，该方法还包括执行客户端掉线恢复协议，包括：中心服务器从多个客户端设备接收多个密钥部分，表示每个客户端设备的密钥集，密钥集根据密钥共享协议拆分为多个密钥部分，分布在客户端设备中，并由每个客户端设备发送到中心服务器。该方法还包括中心服务器确定客户端设备已掉线。该方法还包括中心服务器组合接收到的对应于掉线客户端的多个密钥部分，以恢复对应于掉线客户端的密钥。中心服务器组合接收到的对应于掉线客户端的多个密钥部分，以导出在模型更新的总和中添加的随机噪声的值。该随机噪声的值的添加用以补偿由于掉线的客户端引起的模型更新的总和中丢失的噪声。例如，在图3A中进行了详细描述。

[0125] 根据实施例，一种计算机可读介质，用于存储指令，当这些指令被执行时，这些指令使中心服务器处理器执行方法200。中心服务器的处理器用于执行方法200。

[0126] 步骤202、204、206和208仅仅是说明性的，在不脱离本文权利要求的范围的情况下，还可以提供其它替代方案，其中，添加一个或多个步骤、删除一个或多个步骤，或以不同的顺序提供一个或多个步骤。

[0127] 图3A是根据本公开的实施例的描述利用随机安全平均协议的分布式机器学习的网络环境图。结合图1和图2的元件描述图3A。参考图3A，示出了包括多个客户端设备302、中心服务器304和网络306的系统300A。多个客户端设备302包括客户端设备302A和其它客户端设备302B-302N。系统300A描述了客户端设备302A执行的操作308A、308B、308C、308D、308E和308F的示例性序列。

[0128] 多个客户端设备302中的每个包括适当的逻辑、电路、接口和/或代码，这些逻辑、电路、接口和/或代码用于通过网络306与中心服务器304通信。多个客户端设备302中的每个还用于训练局部模型更新向量并计算模型更新输出。多个客户端设备302的示例可以包括但不限于用户设备、笔记本电脑、计算设备、包括便携式或非便携式电子设备的通信装置，或超级计算机。例如，在图3B中详细描述了客户端设备302A的各种示例性组件。

[0129] 中心服务器304包括用于经由网络306与多个客户端设备302通信的适当的逻辑、电路、接口或代码。中心服务器304还用于确定进一步与多个客户端设备302中的每个共享的全局模型更新，用于下一次迭代。中心服务器304的示例包括但不限于存储服务器、云服务器、Web服务器、应用服务器或其组合。根据实施例，中心服务器304包括能够增强信息以执行各种计算任务的物理或虚拟计算实体装置。在一个示例中，中心服务器304可以是单个硬件服务器。在另一个示例中，中心服务器304可以是并行或分布式架构中运行的多个硬件服务器。在一种实现方式中，中心服务器304可以包括存储器、处理器、网络接口等组件，以存储、处理或与多个客户端设备302共享信息。在另一种实现方式中，中心服务器304被实现为向多个客户端设备302或模块或装置提供各种服务(例如数据库服务)的计算机程序。例

如,在图3C中详细描述了中心服务器304的各种示例性组件。

[0130] 网络306包括介质(例如通信信道),多个客户端设备302通过该介质与中心服务器304通信,反之亦然。网络306可以是有线或无线通信网络。网络306的示例可以包括但不限于无线保真(Wireless Fidelity,Wi-Fi)网络、局域网(Local Area Network,LAN)、无线个人局域网(wireless personal area network,WPAN)、无线局域网(Wireless Local Area Network,WLAN)、无线广域网(wireless wide area network,WWAN)、云网络、长期演进(Long Term Evolution,LTE)网络、城域网(Metropolitan Area Network,MAN)或互联网。多个客户端设备302和中心服务器304可能用于根据各种有线和无线通信协议连接到网络306。这类有线和无线通信协议的示例可以包括但不限于传输控制协议和互联网协议(Transmission Control Protocol and Internet Protocol,TCP/IP)、用户数据报协议(User Datagram Protocol,UDP)、超文本传输协议(Hypertext Transfer Protocol,HTTP)、文件传输协议(File Transfer Protocol,FTP)、ZigBee、EDGE、红外(infrared,IR)、IEEE 802.11、802.16、长期演进(LTE)、光保真(Light Fidelity,Li-Fi)或其它蜂窝通信协议或蓝牙(Bluetooth,BT)通信协议,包括其变体。

[0131] 在操作中,多个客户端设备302中的客户端设备302A在一系列操作中执行局部训练。在操作308A中,客户端设备302A用于生成局部模型更新向量。客户端设备302A通过使用局部数据更新局部模型更新向量。

[0132] 在操作308B中,客户端设备302A还用于生成公钥和密钥。客户端设备302A生成公钥和密钥(或私钥),以便与其它客户端设备302B-302N执行密钥协商。

[0133] 在操作308C中,客户端设备302A还用于将公钥广播到网络306上的其它客户端设备302B-302N,接收每个其它客户端设备302B-302N的外部公钥,并将模型更新输出发送到中心服务器304,以合并到全局模型更新312中。客户端设备302A用于向网络306上的其它客户端设备302B-302N广播生成的公钥。在一种情况下,客户端设备302A不能直接与其它客户端设备302B-302N通信。在这种情况下,生成的公钥通过中心服务器304广播到其它客户端设备302B-302N。在主动安全模型(或恶意服务器)的另一种情况下,可信第三方用于初始建立客户端302中的每个之间的密钥协商。此后,客户端设备302A还用于接收每个其它客户端设备302B-302N的外部公钥,以基于接收到的外部公钥和生成的密钥生成密钥对。密钥协商是通过使用椭圆曲线-迪菲-赫尔曼(ECDH)密钥协商方案执行的,该方案支持两个客户端设备在不安全的信道上建立共享密钥,每个客户端设备都具有椭圆曲线外部公钥-密钥对(也称为共享密钥)。

[0134] 在操作308D中,客户端设备302A还用于基于密钥和外部公钥为每个其它客户端设备302B-302N生成伪随机数。建立共享密钥的两个客户端设备创建相同的伪随机数,用于指示两个客户端是彼此的邻居。

[0135] 在操作308E中,客户端设备302A还用于基于伪随机数和预定邻居概率参数确定是否将每个其它客户端设备302B-302N分配给客户端设备302A的邻居设备集。客户端设备302A的邻居设备集是:

[0136] $$N_p(u) = \{v | rand_0^{u,v} \bmod 2^d < p2^d\}$$

[0137] 其中:p是每个其它客户端设备302B-302N的预定邻居概率参数,d是生成的伪随机数的比特大小。

[0138] 在操作308F中，客户端设备302A还用于基于邻居设备集，根据安全和协议310生成模型更新输出。客户端设备302A通过将安全和协议310应用于邻居设备集来生成模型更新。安全和协议310通过加性密钥共享方案或基于一次性密码本(OTP)的方案被应用。

[0139] 根据实施例，邻居概率参数用于基于预定义值定义邻居设备集中的邻居的数量，其中，预定义值基于成功攻击的建模风险定义。邻居设备集中的邻居的数量是通过使用邻居选择算法计算的。邻居选择算法包括两个阶段，即设置阶段和在线阶段。在设置阶段，邻居设备集中邻居的数量的计算取决于N轮训练的风险参数r_{tot}和邻居概率参数p，例如，在图1中已经详细描述。在在线阶段，客户端设备302A获取其它客户端设备302B-302N之一的共享密钥。客户端设备302A用于从共享密钥中导出k个共享匀速比特b_i，然后，如果$\Sigma_i b_i 2^i < p2^k$，则将其它客户端设备302B-302N中的相应的一个添加到邻居列表中。

[0140] 根据实施例，对于邻居概率参数p和预定义值r：

[0141] $f(n_h,p) > 1-r$，并且对于精度δ，$f(n_h,p-\delta) < 1-r$

[0142] 其中：

[0143] $$\begin{cases} f(n_h,p) = 1 - \sum_{i=1}^{n_h-1} f(i,p)(1-p)^{i(n_h-i)}, & \text{对于} n_h > 1 \\ f(1,p) = 1 \end{cases}$$

[0144] 其中：

[0145] $$n_h = \begin{cases} n - \left(\left\lceil \frac{n}{3} \right\rceil - 1\right), & \text{对于主动安全} \\ n - 2\left(\left\lceil \frac{n}{3} \right\rceil - 1\right), & \text{对于被动安全} \end{cases}$$

[0146] 客户端设备302A根据需要显示主动安全模型以及被动安全模型。例如，在图1中已经详细描述了主动安全模型和被动安全模型。

[0147] 根据实施例，客户端设备302A还用于通过以下方式根据安全和协议310生成模型更新输出：为每个邻居设备生成一次性密码本，并将多个一次性密码本添加到局部模型更新向量中，其中，每个邻居设备的一次性密码本是基于从客户端设备302A的密钥和邻居设备的外部公钥导出的共享密钥生成的。在邻居选择之后，通过使用例如在图1中详细描述的基于OTP的方案来应用安全和协议310。在为每个邻居设备生成OTP之后，客户端设备302A将多个一次性密码本添加到局部模型更新向量中，并生成模型更新输出。

[0148] 根据实施例，由多个客户端设备302生成的所有一次性密码本的集合实质上总计为零。为每对邻居生成的OTP相互抵消，因此，由多个客户端设备生成的所有一次性密码本的集合之和等于零。

[0149] 替代地，根据实施例，客户端设备302A还用于通过以下方式根据安全和协议310生成模型更新输出：根据邻居设备的数量将局部模型向量更新拆分为多个部分，将多个部分发送到相应的邻居设备，从邻居设备接收外部部分，并将多个外部部分相加以形成模型更新输出。在邻居选择之后，安全和协议310通过加性密钥共享方案被应用，例如Shamir密钥共享(SSS)方案。在加性密钥共享方案中，局部模型更新向量被量化并被视为固定长度模整数的向量。客户端设备302A将每个整数拆分为多个部分，每个邻居设备一个部分。多个部分被发送到相应的邻居设备302B-302N。此后，从邻居设备302B-302N接收外部部分，并添加外

部部分，以形成模型更新输出。

[0150] 根据实施例，客户端设备302A还用于通过将局部生成的噪声信号添加到局部模型更新向量中来生成模型更新输出，局部生成的噪声信号的分布是高斯或二项式的。局部生成的噪声信号具有高斯分布或二项式分布。二项式分布通常是首选的，具体是对于小字大小。

[0151] 根据实施例，局部生成的噪声信号是以由从中心服务器304接收的噪声参数定义的标准偏差生成的，其中，每个客户端设备302A-302N的噪声参数使得来自多个客户端设备302的对应的差分私有噪声信号集总计为具有预定标准偏差的全局噪声。中心服务器304用于确定噪声参数(例如噪声乘数z)和模型更新截断边界S。局部生成的噪声信号通过以下步骤从噪声乘数导出：在第一步骤中，计算标准偏差，该标准偏差被添加到和中，为$\sigma=zS$。在第二步骤中，选择噪声拆分策略，以将局部生成的噪声信号拆分到多个客户端302。噪声拆分策略被提供标准差σ和多个客户端302的权重集$\{w_u\}_u$，并返回局部噪声集$\{\sigma_u\}_u$，使得$\sum_u \sigma_u^2 = \sigma^2$。可以使用三种不同的噪声拆分策略，策略1：$\sigma_u = \sqrt{\frac{w_u}{\sum_u w_u}}\sigma$，策略2：$\sigma_u = \frac{w_u}{\sqrt{\sum w_u^2}}\sigma$，策略3：$\sigma_u = \frac{\sigma}{\sqrt{n}}$。为了生成高斯分布局部生成的噪声信号，客户端设备302A局部噪声为$\mathcal{N}(0,\sigma_u^2)$，具有正态分布、均值0和方差$\sigma_u^2$。为了生成二项式局部生成的分布式噪声信号，遵循四个步骤。这些步骤是：(a)设置$N=2^{\text{wordsize}}$，(b)设置量化参数$k = \frac{2\sigma}{\sqrt{N}}$，(c)设置$N_u = \left\lfloor \left(\frac{2\sigma_u}{k}\right)^2 \right\rfloor$或$\left\lceil \left(\frac{2\sigma_u}{k}\right)^2 \right\rceil$，使得$\Sigma_u N_u = N$，(d)计算遵循以下二项式机制的局部噪声：$\left(Binomial\left(N_u, \frac{1}{2}\right) - \frac{N_u}{2}\right)k$。这样，来自多个客户端设备302的对应的局部生成的噪声信号集在与全局噪声相加时，具有二项式分布和预定标准差(σ)。

[0152] 根据实施例，客户端设备302A还用于将局部模型更新向量从浮点值的向量转换为整数向量。通过量化过程，局部模型更新向量从浮点值的向量转换为整数(或模)向量。在量化过程中，使用了无偏、空间高效的算法。例如，对单独的模型参数要求边界支持高效地映射到整数空间。这些边界可以直接提供或从差分隐私参数S推断。量化过程在求和来自多个其它客户端设备302B-302N的加权更新时必须考虑溢出的风险。具体地，多个其它客户端设备302B-302N必须知道在此次迭代期间要使用的权重之和。

[0153] 根据实施例，客户端设备302A还用于向中心服务器304发送公钥，并从中心服务器304接收外部公钥。例如，在一种情况下，客户端设备302A不能直接与多个其它客户端设备302B-302N通信。然后，在这种情况下，客户端设备302A将生成的公钥传送到中心服务器304，中心服务器304进一步将生成的公钥传送到多个其它客户端设备302B-302N。在将生成的公钥传送到中心服务器304之后，客户端设备302A用于从中心服务器304接收外部公钥。

[0154] 因此，多个客户端设备302中的每个训练局部模型更新向量并生成模型更新输出。多个客户端设备302中的每个缩放局部模型更新向量，量化局部模型更新向量，然后将具有二项式分布的校准噪声添加到局部模型更新向量。校准噪声基于隐私要求和所使用的安全模型(即主动安全模型或被动安全模型)的类型。之后，多个客户端设备302中的每个被赋予风险参数和所选择的安全模型，然后多个客户端设备302中的每个通过使用邻居选择算法

导出邻居概率参数p。邻居选择算法基于随机图的连通性概念。邻居选择之后，每对客户端设备共享一个密钥。共享密钥用于导出共享随机性。使用邻居概率参数p(在共享随机性上)，每对客户端设备选择为邻居。每对邻居生成相同的伪随机数，并将其添加到相应的密钥值(具有相反符号)中，以生成模型更新输出。之后，多个客户端设备302中的每个将生成的模型更新输出发送到中心服务器304。

[0155] 中心服务器304用于接收多个客户端设备302发送的多个模型更新输出，并将全局模型更新312发送到每个客户端设备302A-302N。中心服务器304通过使用多个客户端设备302发送的多个模型更新输出来确定全局模型更新312。中心服务器304还用于将全局模型更新312发送到每个客户端设备302A-302N，用于下一次迭代。

[0156] 中心服务器304还用于基于多个模型更新输出确定模型更新的总和，并基于模型更新的总和更新全局模型以生成全局模型更新312。中心服务器304基于多个模型更新输出获得表示模型更新的总和的向量。中心服务器304基于模型更新的总和确定与多个客户端设备302中的每个共享的全局模型更新312。

[0157] 根据实施例，更新全局模型包括将整数向量转换为浮点向量。通过反量化过程从整数(模)向量转换为浮点向量，更新全局模型。从整数(模)向量到浮点向量的转换会产生噪声模型更新的总和。如果客户端设备在该过程期间已掉线，则中心服务器304在执行反量化过程时考虑。

[0158] 根据实施例，中心服务器304还用于确定客户端设备已掉线。中心服务器304还用于基于添加到模型更新的总和中的局部噪声的预定方差值，将附加噪声添加到模型更新的总和中。在一种情况下，中心服务器304用于在执行随机安全平均(RdSA)协议期间，多个客户端设备302中的一个或多个已掉线的情况下进行噪声补偿。在这种客户端掉线的情况下，模型更新的总和中的总噪声小于所需的值。因此，中心服务器304用于基于添加到模型更新的总和中的局部噪声的预定方差值，将附加噪声添加到模型更新的总和中。在一个示例中，如果将高斯分布局部噪声添加到模型更新的总和中，则预定方差为$\sum_{u\in\mathcal{A}}\sigma_u^2$，其中，$\mathcal{A}$是其模型更新输出包括在模型更新的总和中的客户端集。在另一个示例中，如果将二项式分布局部噪声添加到模型更新的总和中，则预定方差为$\sum_{u\in\mathcal{A}}N_u k^2/4$，其中，$\mathcal{A}$是其模型更新输出包括在模型更新的总和中的客户端集。根据局部噪声的分布，附加噪声添加到模型更新的总和中。在一种实现方式中，如果使用高斯分布，则将具有平均值0和标准偏差$\sigma_d=\sqrt{\sigma^2-\sum_{u\in\mathcal{A}}\sigma_u^2}$的高斯噪声(即附加噪声)$\mathcal{N}(0,\sigma_d^2)$添加到模型更新的总和中。在另一种实现方式中，如果使用二项式分布，则将二项式噪声$\left(Binomial\left(N_d,\frac{1}{2}\right)-\frac{N_d}{2}\right)k$(其中，$N_d=N-\sum_{u\in\mathcal{A}}N_u$)添加到模型更新的总和中。

[0159] 根据实施例，中心服务器304还用于执行客户端掉线恢复协议，包括：从多个客户端设备302接收多个密钥部分，表示每个客户端设备302A-302N的密钥集，该密钥集根据密钥共享协议拆分为多个密钥部分，分布在客户端设备中，并由每个客户端设备302A-302N发送到中心服务器304。中心服务器304还用于确定客户端设备已掉线。中心服务器304还用于组合接收到的对应于掉线客户端的多个密钥部分，以恢复对应于掉线客户端的密钥。在一种实现方式中，为了确定客户端设备已掉线，中心服务器304用于组合接收到的对应于掉线

客户端的多个密钥部分,以恢复对应于掉线客户端的密钥。在中心服务器304确定客户端设备已掉线之后,中心服务器304组合接收到的对应于掉线客户端的多个密钥部分,以导出在模型更新的总和中添加的随机噪声的值。由于掉线客户端而添加的该随机噪声的值从模型更新的总和中删除。在另一种实现方式中,中心服务器304用于组合接收到的对应于非掉线客户端的多个密钥部分,以导出在模型更新的总和中添加的局部随机噪声的值。由于非掉线客户端而添加的该局部随机噪声的值从模型更新的总和中删除。这样,中心服务器304获得非掉线的模型更新的总和。

[0160] 因此,中心服务器304通过使用从多个客户端设备302接收的模型更新输出来确定全局模型更新312。之后,中心服务器304与多个客户端设备302共享全局模型更新312,以便为下一次迭代做准备。

[0161] 系统300A基于图论和随机图的连通性属性。在系统300A中,中心服务器304学习至少n/3个私有模型更新的总和,除非概率等于风险参数α。系统300A提供了客户端设备的数据(包括个人数据)的增强隐私保护、改进的性能和高效用。系统300A使用安全和技术(或密码技术)和差分隐私技术的组合。安全和技术(或密码技术)防止中心服务器304从任何单独的客户端访问私有信息。在差分隐私技术中,噪声是在局部(在客户端侧)而不是在中心(在中心服务器侧)添加的,这提供了强大的隐私保护。

[0162] 图3B是根据本公开的实施例的客户端设备的各种示例性组件的框图。结合图1、图2和图3A的元件描述图3B。参考图3B,示出了(图3A的)客户端设备302A的框图300B,包括处理器314、网络接口316、存储器318和输入/输出(input/output,I/O)组件320。存储器318还包括训练模块318A。

[0163] 处理器314包括用于生成公钥和密钥的适当的逻辑、电路或接口。在一种实现方式中,处理器314用于执行存储在存储器318中的指令。在一个示例中,处理器314可以是通用处理器。处理器314的其它示例可以包括但不限于微处理器、微控制器、复杂指令集计算(complex instruction set computing,CISC)处理器、专用集成电路(application-specific integrated circuit,ASIC)处理器、精简指令集(reduced instruction set,RISC)处理器、超长指令字(very long instruction word,VLIW)处理器、中央处理单元(central processing unit,CPU)、状态机、数据处理单元和其它处理器或控制电路。此外,处理器314可以是指一个或多个单独的处理器、处理设备或作为机器的一部分的处理单元,例如客户端设备302A。

[0164] 网络接口316包括用于执行以下操作的适当的逻辑、电路或接口:将公钥广播到网络306上的多个其它客户端设备302B-302N,接收每个其它客户端设备302B-302N的外部公钥,并将模型更新输出发送到中心服务器304,以合并到全局模型更新312中。网络接口316的示例可以包括但不限于天线、射频(radio frequency,RF)收发器、一个或多个放大器、数字信号处理器或用户识别模块(subscriber identity module,SIM)卡。

[0165] 存储器318包括用于存储可由处理器314执行的指令合适的逻辑、电路或接口。存储器318的实现示例可以包括但不限于电可擦除可编程只读存储器(Electrically Erasable Programmable Read-Only Memory,EEPROM)、随机存取存储器(Random Access Memory,RAM)、只读存储器(Read Only Memory,ROM)、硬盘驱动器(Hard Disk Drive,HDD)、闪存、固态硬盘(Solid-State Drive,SSD)或CPU高速缓冲存储器。存储器318可以存储操作

系统或其它程序产品(包括一种或多种操作算法),以操作客户端设备302A。

[0166] 在示例性实现方式中,训练模块318A用于生成局部模型更新向量。训练模块318A对应于通过使用局部模型更新向量训练的机器学习模型。训练模块318A在客户端设备302A上局部训练。在训练期间,客户端设备302A使用安全和协议310,该协议防止中心服务器304观察来自训练模块318A的任何私有信息。附加地,客户端设备302A向训练模块318A添加局部差分隐私(DP)噪声,以进一步增强隐私保护。在训练之后,客户端设备302A与中心服务器304共享局部训练的模型。类似地,中心服务器304相对于多个其它客户端设备302B-302N中的每个接收局部训练的模型。在接收到局部训练的模块之后,中心服务器304计算与多个客户端设备302中的每个共享的全局模型更新。全局模型更新在更准确的预测或决策、个人数据的隐私保护以及即使在存在大量客户端设备的情况下更低的计算成本方面表现出改进的性能。在一种实现方式中,训练模块318A(可以包括一个或多个软件模块)可能被实现为客户端设备302A中的单独电路。替代地,在另一实现方式中,训练模块318A被实现为另一个电路的一部分,以执行各种操作。

[0167] 输入/输出(I/O)组件320是指可以从用户(例如客户端设备302A)接收输入并向用户(即客户端设备302A)提供输出的输入和输出组件(或设备)。I/O组件320可以通信地耦合到处理器314。输入组件的示例可以包括但不限于触摸屏,例如显示设备的触摸屏,麦克风、运动传感器、光传感器、专用硬件输入单元(例如按钮)和扩展坞。输出组件的示例包括显示设备和扬声器。

[0168] 在操作中,训练模块318A用于生成局部模型更新向量。处理器314用于生成公钥和密钥。收发器316(或网络接口)用于:将公钥广播到网络306上的多个其它客户端设备302B-302N,接收每个其它客户端设备302B-302N的外部公钥,并将模型更新输出发送到中心服务器304,以合并到全局模型更新312中。处理器314还用于基于密钥和外部公钥,为每个其它客户端设备302B-302N生成伪随机数。处理器314还用于基于伪随机数和预定邻居概率参数确定是否将每个其它客户端设备302B-302N分配给客户端设备302A的邻居设备集。处理器314还用于基于邻居设备集,根据安全和协议生成模型更新输出。在一种实现方式中,处理器314还用于通过以下方式根据安全和协议生成模型更新输出:为每个邻居设备生成一次性密码本,并将多个一次性密码本添加到局部模型更新向量中,其中,每个邻居设备的一次性密码本是基于从客户端设备302A的密钥和邻居设备的外部公钥导出的共享密钥生成的。多个客户端设备302生成的所有一次性密码本的集合实质上总计为零。在另一种实现方式中,处理器314还用于通过以下方式根据安全和协议生成模型更新输出:根据邻居设备的数量将局部模型向量更新拆分为多个部分,将多个部分发送到相应的邻居设备,从邻居设备接收外部部分,并将多个外部部分相加以形成模型更新输出。根据实施例,处理器314还用于通过将局部生成的噪声信号添加到局部模型更新向量中来生成模型更新输出,其中,局部生成的噪声信号的分布是高斯或二项式的。处理器314还用于将局部模型更新向量从浮点值的向量转换为整数向量。图1的方法100中公开的各种实施例、操作和变型比照适用于客户端设备302A和处理器314。

[0169] 图3C是根据本公开的实施例的中心服务器的各种示例性组件的框图。结合图1、图2、图3A和图3B的元件描述图3C。参考图3C,示出了(图3A的)中心服务器304的框图300C,包括处理器322、网络接口324和存储器326。

[0170] 处理器322包括用于执行以下操作的适当的逻辑、电路或接口:基于多个模型更新输出确定模型更新的总和,并基于模型更新的总和更新全局模型以生成全局模型更新。在一种实现方式中,处理器322用于执行存储在存储器326中的指令。在一个示例中,处理器322可以是通用处理器。处理器322的其它示例可以包括但不限于微处理器、微控制器、复杂指令集计算(CISC)处理器、专用集成电路(ASIC)处理器、精简指令集(RISC)处理器、超长指令字(very long instruction word,VLIW)处理器、中央处理单元(central processing unit,CPU)、状态机、数据处理单元和其它处理器或控制电路。此外,处理器322可以是指一个或多个单独的处理器、处理设备或作为机器的一部分的处理单元,例如中心服务器304。

[0171] 网络接口324包括用于执行以下操作的适当的逻辑、电路或接口:接收多个客户端设备302发送的多个模型更新输出,并将全局模型更新312发送到每个客户端设备302A-302N。网络接口324的示例可以包括但不限于天线、射频(RF)收发器、一个或多个放大器、数字信号处理器或用户识别模块(SIM)卡。

[0172] 存储器326包括用于存储可由处理器322执行的指令合适的逻辑、电路或接口。存储器326的实现示例可以包括但不限于电可擦除可编程只读存储器(EEPROM)、随机存取存储器(RAM)、只读存储器(ROM)、硬盘驱动器(HDD)、闪存、固态硬盘(SSD)或CPU高速缓冲存储器。存储器326可以存储操作系统或其它程序产品(包括一种或多种操作算法),以操作中心服务器304。

[0173] 在操作中,收发器324(或网络接口)用于接收多个客户端设备302发送的多个模型更新输出,并向每个客户端设备302A-302N发送全局模型更新312。处理器322用于基于多个模型更新输出确定模型更新的总和,并基于模型更新的总和更新全局模型以生成全局模型更新312。

[0174] 根据实施例,处理器322还用于确定客户端设备已掉线,并基于添加到模型更新的总和中的局部噪声的预定方差值将附加噪声添加到模型更新的总和中。中心服务器304还用于执行客户端掉线恢复协议,包括:收发器324从多个客户端设备302接收多个密钥部分,表示每个客户端设备302的密钥集,该密钥集根据密钥共享协议拆分为多个密钥部分,分布在客户端设备中,并由每个客户端设备302发送到中心服务器304。中心服务器304还用于执行客户端掉线恢复协议,包括:处理器322确定客户端设备已掉线,处理器322组合接收到的对应于掉线客户端的多个密钥部分,以恢复对应于该掉线客户端的密钥。图2的方法200中公开的各种实施例、操作和变型比照适用于中心服务器304和处理器322。

[0175] 图4是根据本公开的另一个实施例的描述分布式机器学习随机安全平均的网络环境图。结合图1、图2、图3A、图3B和图3C的元件描述图4。参考图4,示出了更详细地描述了利用随机安全平均(RdSA)的分布式机器学习的实现方式的系统400。系统400描述了客户端设备302A执行的操作402、402A、402B、402C、402D、402E、402F、402G、402H和402I的示例性序列。还示出了中心服务器304执行的操作404A、404B、404C、404D、404E、404F和406的示例性序列。

[0176] 系统400包括图3A的多个客户端设备302和中心服务器304。关于客户端设备302A和多个其它客户端设备302B-302N解释客户端侧分布式学习或客户端侧随机安全平均。类似地,关于中心服务器304解释中心服务器侧分布式学习或中心服务器侧随机安全平均。

[0177] 在操作中,多个客户端设备302中的客户端设备302A在一系列操作中执行全局模

型的局部训练。全局模型由中心服务器304与多个客户端设备302中的每个共享。此外，中心服务器304为标准分布式学习参数选择值，例如局部轮训周期的数量、学习速率或迭代总数N。中心服务器304与多个客户端设备302共享选择的标准分布式学习参数。

[0178] 在操作402中，客户端设备302通过使用局部数据或原始数据开始局部模型更新向量的局部训练。

[0179] 在操作402A中，客户端设备302对中心服务器304共享的参数执行选择。所选择的参数用于局部模型更新向量的局部训练。

[0180] 在操作402B中，客户端设备302A还用于生成公钥和密钥。客户端设备302A生成公钥和密钥(或私钥)，以便与其它客户端设备302B-302N执行密钥协商。在一种情况下，客户端设备302A不能直接与多个其它客户端设备302B-302N通信。在这种情况下，客户端设备302A将生成的公钥传送到中心服务器304。中心服务器304与多个客户端设备302中的每个共享生成的公钥。因此，作为回报，每个客户端设备302从中心服务器304接收外部公钥。

[0181] 在操作402C中，客户端设备302A还用于基于生成的密钥和接收到的外部公钥生成密钥对。生成的密钥对用于与多个其它客户端设备302B-302N执行密钥协商。密钥协商是通过使用椭圆曲线-迪菲-赫尔曼(ECDH)密钥协商方案执行的，该方案支持两个客户端设备在不安全的信道上建立共享密钥，每个客户端设备都具有椭圆曲线外部公钥-密钥对(也称为共享密钥)。

[0182] 在操作402D中，客户端设备302A还用于通过使用邻居选择算法来选择其邻居，例如，在图1和图3A中已经进行了详细描述。建立共享密钥的两个客户端设备创建相同的伪随机数，用于指示两个客户端是彼此的邻居。

[0183] 在操作402E中，执行密钥共享以便拆分客户端设备302A的私钥。客户端设备302A的两个私钥在操作402E中被拆分，例如客户端设备302A的密钥，从该密钥导出共享密钥，用于为每个邻居设备生成一次性密码本，以及用于在进一步操作中添加随机噪声的客户端设备302A的个人种子。

[0184] 在操作402F中，客户端设备302A还用于将局部生成的差分隐私(DP)噪声信号添加到局部模型更新向量中，其中，局部生成的DP噪声信号的分布是高斯或二项式的。局部生成的噪声信号具有高斯分布或二项式分布。噪声参数由中心服务器304选择并与多个客户端设备302中的每个共享，用于生成DP噪声信号。

[0185] 在操作402G中，在将局部生成的差分隐私(DP)噪声信号添加到局部模型更新向量中之后，客户端设备302A还用于执行量化过程。在量化过程中，局部模型更新向量从浮点值的向量转换为整数(或模)向量。

[0186] 在操作402H中，客户端设备302A还用于通过为每个邻居设备生成一次性密码本并将生成的一次性密码本添加到量化模型更新向量中，应用安全和协议310。每个邻居设备的一次性密码本是基于从客户端设备302A的密钥和邻居设备的外部公钥导出的共享密钥生成的。此外，客户端设备302A还用于基于其个人种子计算局部随机噪声，并将计算的局部随机噪声添加到量化模型更新向量中。

[0187] 在操作402I中，客户端设备302A用于将每个其它客户端的一个部分传送到中心服务器304。在一个示例中，如果另一个客户端设备在执行RdSA协议期间掉线，则客户端设备302A用于将其它客户端设备的密钥的部分传送到中心服务器304，以便导出其它客户端设

备的共享OTP。在另一个示例中，如果其它客户端设备在执行RdSA协议期间掉线，则客户端设备302A用于将其它客户端设备的个人种子的部分传送到中心服务器304，以导出多个其它客户端设备302B-302N的局部随机噪声。

[0188] 操作402A至操作402G是执行客户端侧随机安全平均所必需的。对于多个客户端设备302中的每个，操作402A至402I以相同的顺序执行。通过执行来自402A、402B、402C、402D、402E、402F和402G的操作，多个客户端设备302中的每个用于分别在训练局部模型更新向量之后确定多个模型更新输出。多个客户端设备302中的每个与中心服务器304共享多个模型更新输出。

[0189] 类似地，中心服务器304用于执行一系列操作以确定全局模型更新406。在操作404A中，中心服务器304用于将客户端设备302A的生成的公钥广播到多个其它设备302B-302N。

[0190] 在操作404B中，中心服务器304用于将客户端设备302A的多个部分分配给多个其它客户端设备302B-302N。

[0191] 在操作404C中，中心服务器304用于接收多个客户端设备302发送的多个模型更新输出。中心服务器304还用于基于多个模型更新输出确定模型更新的总和。中心服务器304基于多个模型更新输出获得表示模型更新的总和的向量。中心服务器304基于模型更新的总和确定与多个客户端设备302中的每个共享的全局模型更新406。

[0192] 在操作404D中，中心服务器304用于执行掉线恢复。例如，图3A中已经详细描述了掉线恢复的两种方法。

[0193] 在操作404E中，中心服务器304用于执行将整数向量转换为浮点向量的反量化过程。从整数(模)向量到浮点向量的转换会产生噪声模型更新的总和。

[0194] 在操作404F中，中心服务器304用于在RdSA协议执行期间一个或多个客户端设备掉线的情况下补偿差分隐私噪声。中心服务器304还用于基于添加到模型更新的总和中的局部噪声的预定方差值，将附加噪声添加到模型更新的总和中。

[0195] 在执行来自404A、404B、404C、404D、404E和404F的操作之后，中心服务器304确定全局模型更新406。全局模型更新406传送给多个客户端设备302中的每个，以便为下一次迭代做准备。404A、404B、404C、404D、404E和404F的操作顺序指示中心服务器侧随机安全平均(RdSA)。

[0196] 图5示出了根据本公开的实施例的分布式机器学习的示例性实现场景。结合图1、图2、图3A、图3B、图3C和图4的元件描述图5。参考图5，示出了包括多个客户端设备502、中心服务器504和网络506的系统500。多个客户端设备502包括客户端设备502A和其它客户端设备502B-502N。客户端设备502A使用视频推荐工具508A(例如机器学习模型)。类似地，其它客户端设备502B-502N使用视频推荐工具508B-508N。

[0197] 多个客户端设备502、中心服务器504和网络506分别对应于图3A的多个客户端设备302、中心服务器304和网络306。

[0198] 客户端设备502A使用视频推荐工具508A，视频推荐工具508A使用预测模型。视频推荐工具508A在客户端设备502A上局部训练，因为系统500使用利用随机安全平均(RdSA)协议的分布式机器学习。类似地，其它客户端设备502B-502N局部训练它们相应的视频推荐工具508B-508N。多个客户端设备502中的每个与中心服务器504共享其相应的局部训练视

频推荐工具。此后，RdSA协议用于在中心服务器504为N(数百至数千)个客户端设备计算私有全局模型更新。在中心服务器504处计算的私有全局模型更新与多个客户端设备502中的每个共享，多个客户端设备502受益于视频推荐工具的提高的精度以及隐私保护。

[0199] 在另一种实现场景中，多个客户端设备502可以对应于医院或实验室使用的多个计算设备。例如，客户端设备502A对应于医院使用的计算设备，客户端设备502B对应于实验室或组织使用的另一个计算设备，等等。多个客户端设备502中的每个使用从图像训练的训练模型以及医生手动提供的注释，该注释用于使用医学图像检测疾病。在许多司法管辖区，这种医学图像模型不能在不同的医院或实验室之间共享。因此，这种模型的局部训练是在医院、实验室或组织使用的相应计算设备上进行的。在局部训练之后，在医院或实验室或组织中使用的多个计算设备中的每个与中心服务器504共享其局部训练的模型。在该示例性场景中，中心服务器504可以通过使用RdSA协议计算全局模型更新，并与医院、实验室或组织中使用的多个计算设备中的每个共享计算的全局模型更新，该医院、实验室或组织正在准确地检测疾病，同时保持改进的隐私和个人数据保护(可以提供强大的数据隐私保护的正式保证)，并避免扩展问题。

[0200] 可以对上文描述的本公开的实施例进行修改，这不会偏离所附权利要求所定义的本公开范围。如“包括”、“结合”、“具有”、“是”等用于描述和要求保护本公开的表述旨在以非排他的方式解释，即支持未明确描述的项目、组件或元件也存在。对单数的引用也应解释为与复数有关。本文使用的词语“示例性”表示“作为一个示例、实例或说明”。任何被描述为“示例性”的实施例不一定解释为比其它实施例更优选或更有利，和/或排除其它实施例的特征的结合。本文使用的词语“可选地”表示“在一些实施例中提供而在其它实施例中没有提供”。应理解，为了清楚起见而在单独实施例的上下文中描述的本公开的一些特征还可以在单个实施例中组合提供。相反，为简洁起见而在单个实施例的上下文中描述的本公开的各个特征也可以单独提供、以任何适当的组合提供，或适合于本公开的任何其它描述的实施例。

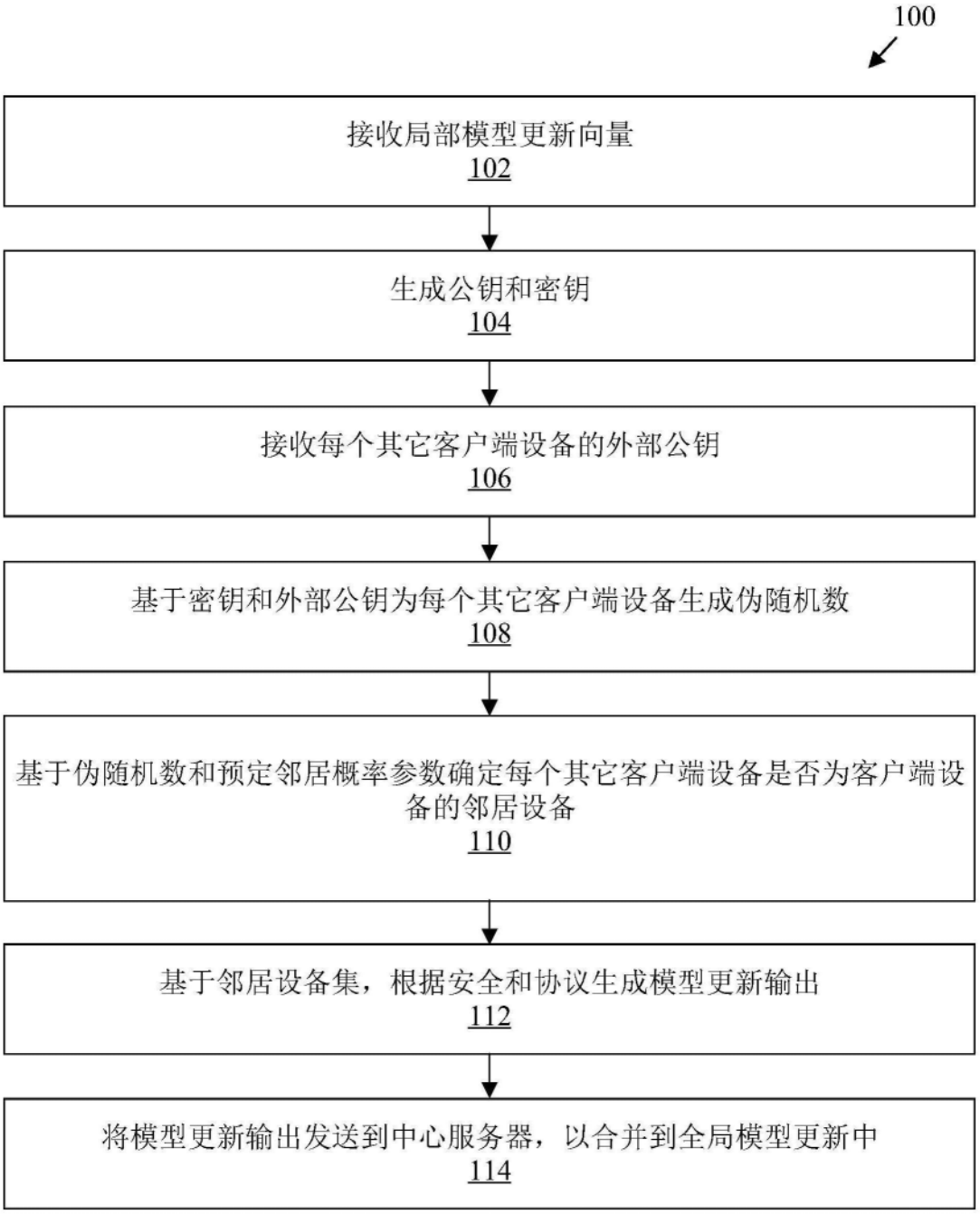

100
接收局部模型更新向量
102
生成公钥和密钥
104
接收每个其它客户端设备的外部公钥
106
基于密钥和外部公钥为每个其它客户端设备生成伪随机数
108
基于伪随机数和预定邻居概率参数确定每个其它客户端设备是否为客户端设备的邻居设备
110
基于邻居设备集，根据安全和协议生成模型更新输出
112
将模型更新输出发送到中心服务器，以合并到全局模型更新中
114

图1

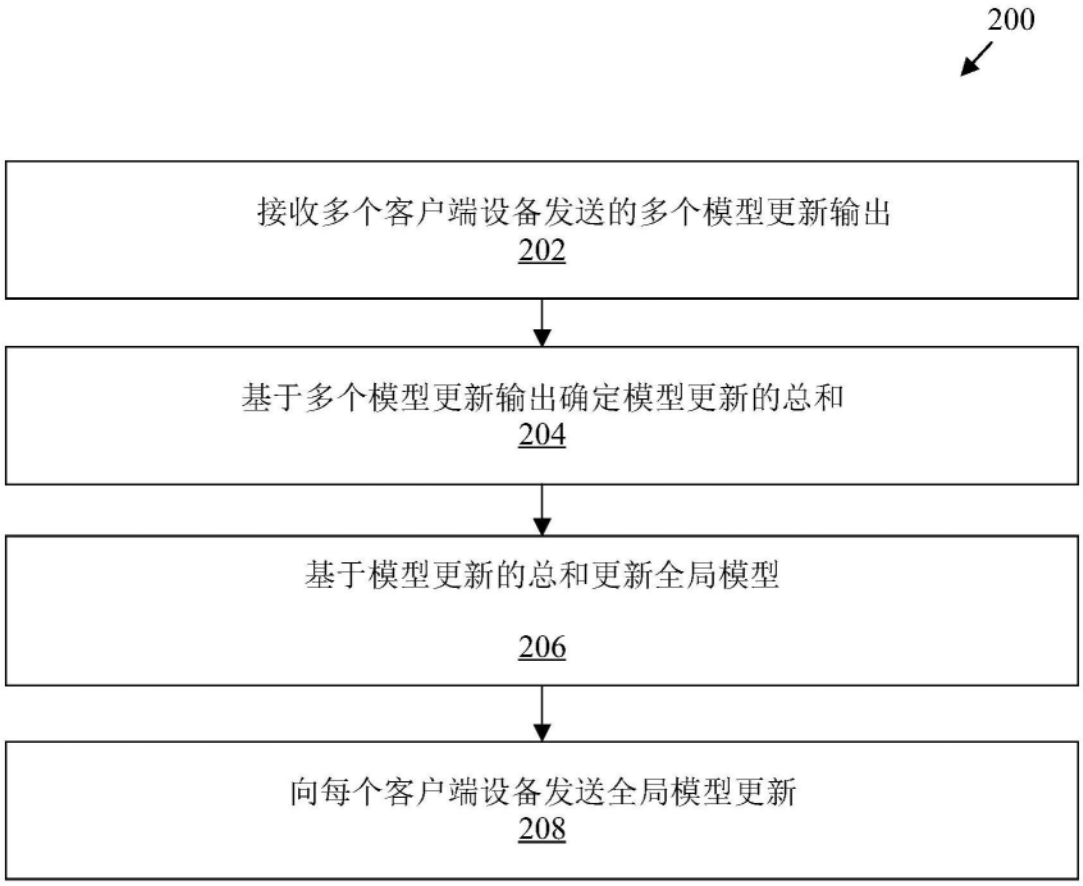

200
接收多个客户端设备发送的多个模型更新输出
202
基于多个模型更新输出确定模型更新的总和
204
基于模型更新的总和更新全局模型
206
向每个客户端设备发送全局模型更新
208

图2

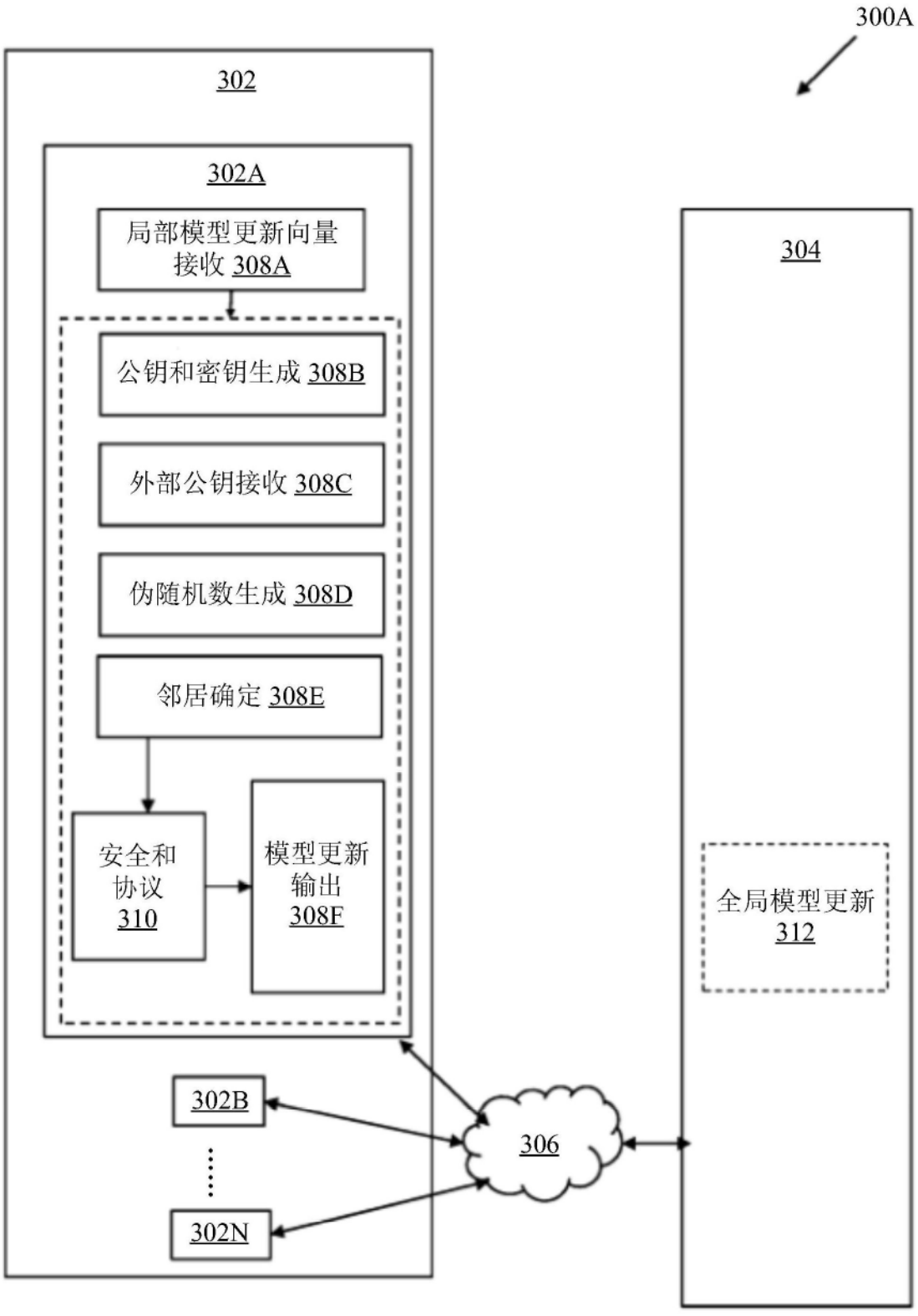

300A
302
302A
局部模型更新向量接收 308A
公钥和密钥生成 308B
外部公钥接收 308C
伪随机数生成 308D
邻居确定 308E
安全和协议 310
模型更新输出 308F
302B
302N
306
304
全局模型更新 312

图3A

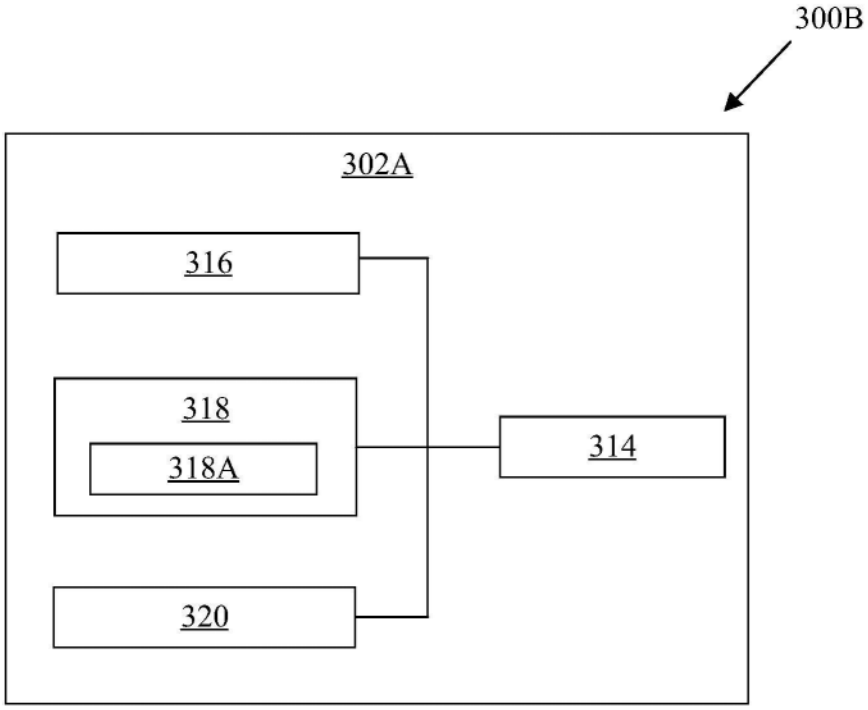


图3B

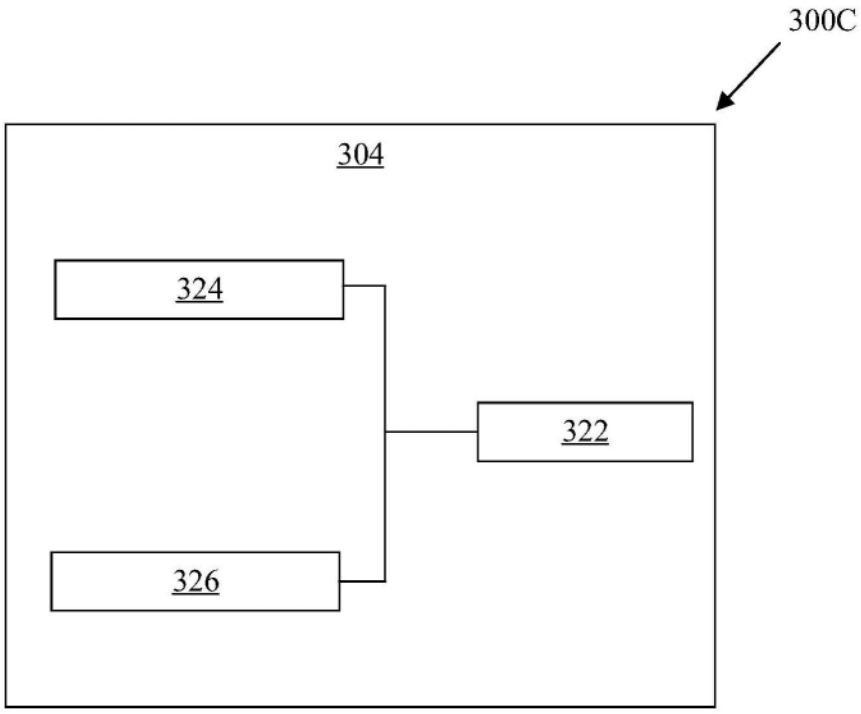


图3C

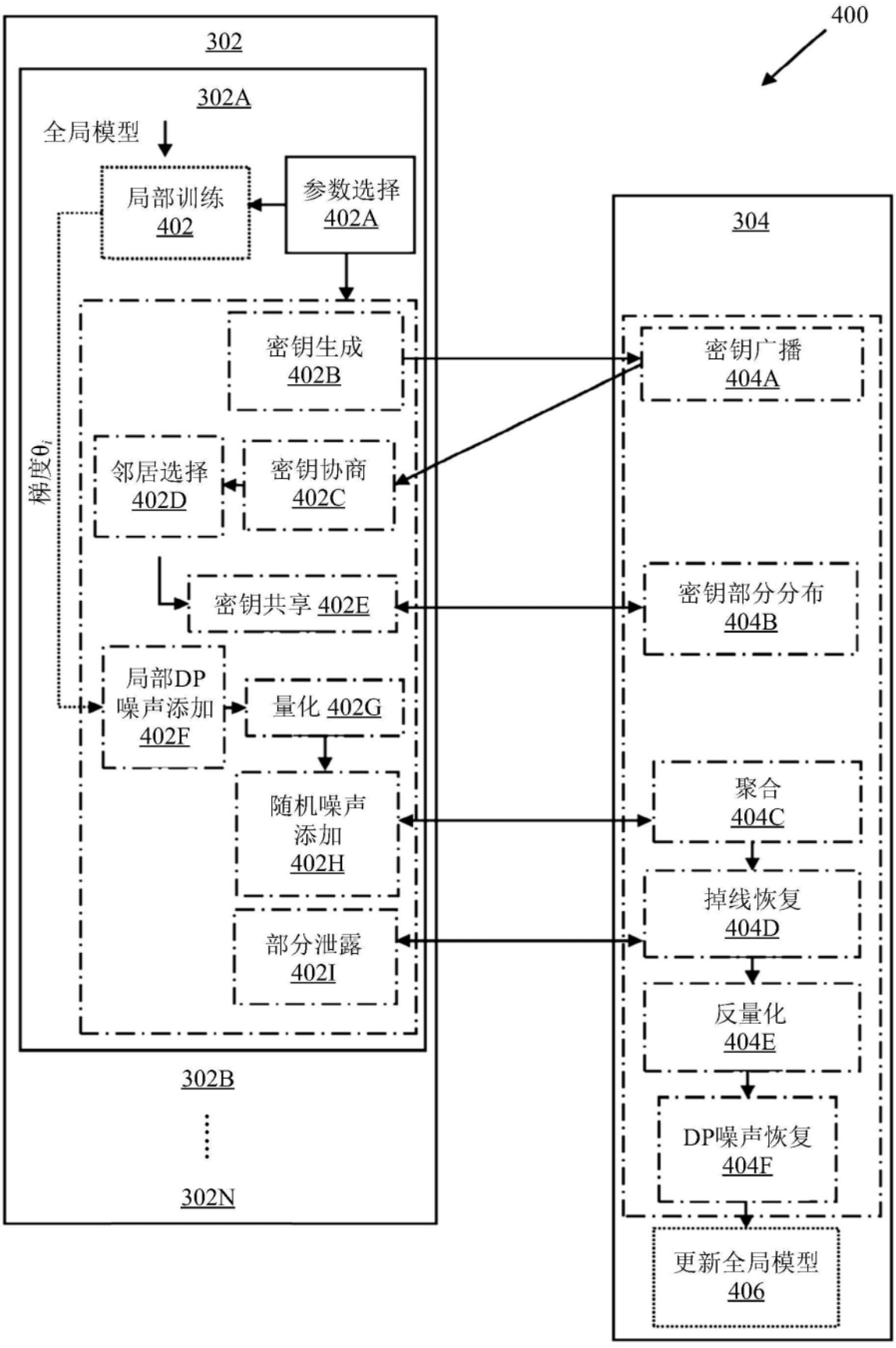

400
302
302A
全局模型
局部训练
402
参数选择
402A
密钥生成
402B
邻居选择
402D
密钥协商
402C
密钥共享 402E
局部DP
噪声添加
402F
量化 402G
随机噪声
添加
402H
部分泄露
402I
梯度θ_i
302B
302N
304
密钥广播
404A
密钥部分分布
404B
聚合
404C
掉线恢复
404D
反量化
404E
DP噪声恢复
404F
更新全局模型
406

图4

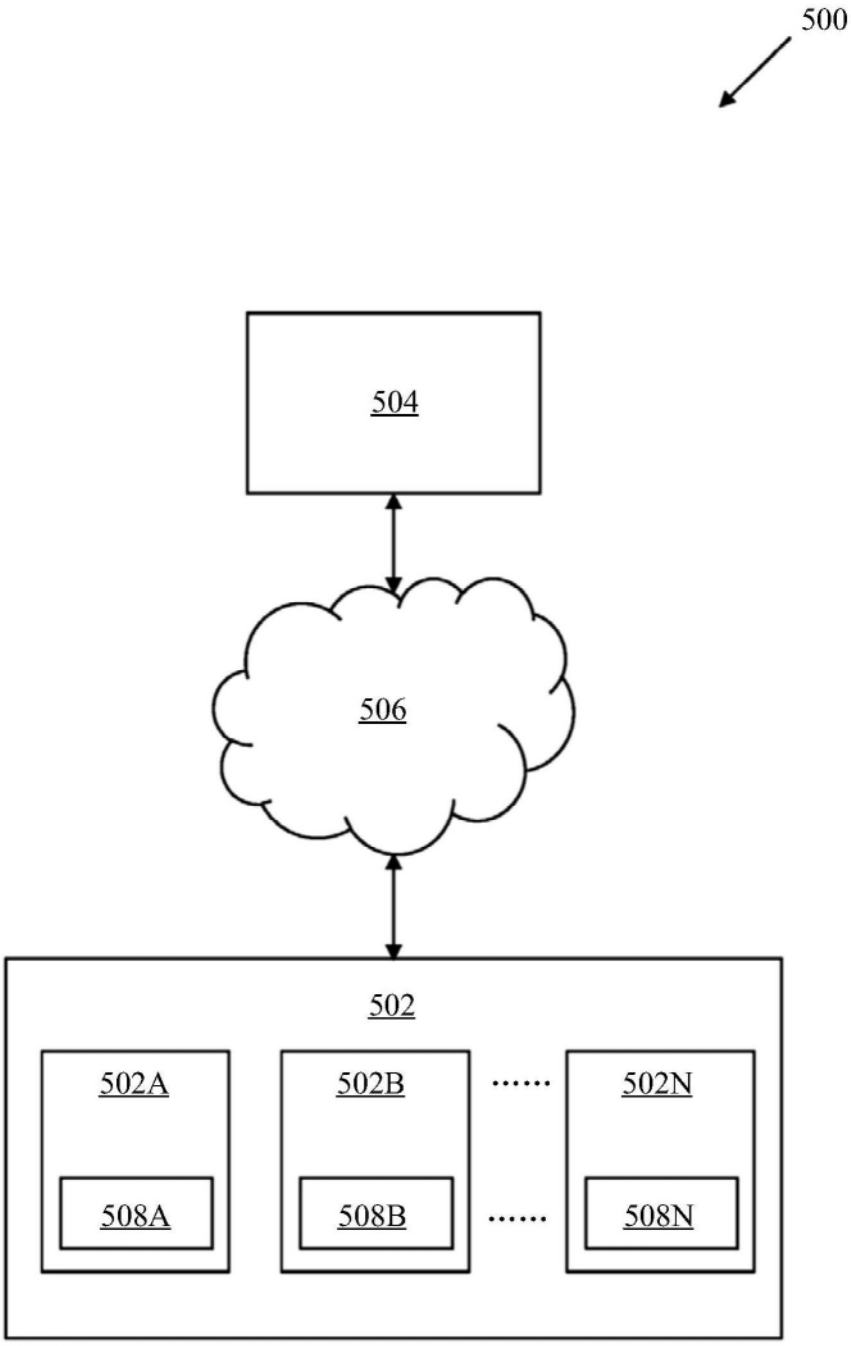


图5