(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 117688589 A
(43)申请公布日 2024.03.12

(21)申请号 202311796303.4

(22)申请日 2023.12.25

(71)申请人 北京富算科技有限公司
地址 100070 北京市丰台区南四环西路188号十六区18号楼1至15层101内7层701-8

(72)发明人 尤志强 赵东 陈立峰 王兆凯 蔡晓娟 杜吉锋 杨云波 卞阳 张伟奇

(74)专利代理机构 北京慧加伦知识产权代理有限公司 16035
专利代理师 李永敏

(51)Int.Cl.
G06F 21/60(2013.01)
G06N 20/00(2019.01)

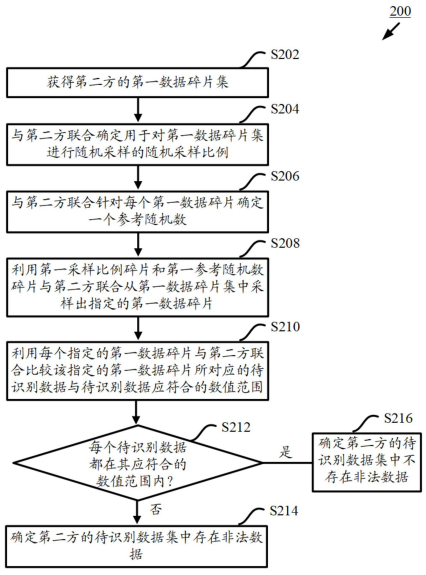
权利要求书4页 说明书14页 附图8页

(54)发明名称

第一方对第二方进行非法数据识别的方法及装置

(57)摘要

本公开的实施例提供一种第一方对第二方进行非法数据识别的方法及装置。该方法由第一方执行。该方法包括:获得第二方的第一数据碎片集;与第二方联合确定用于对第一数据碎片集进行随机采样的随机采样比例;与第二方联合针对每个第一数据碎片确定一个参考随机数,参考随机数符合均匀分布;利用第一采样比例碎片和第一参考随机数碎片与第二方联合从第一数据碎片集中采样出指定的第一数据碎片;利用每个指定的第一数据碎片与第二方联合比较该指定的第一数据碎片所对应的待识别数据与该待识别数据应符合的数值范围;以及响应于任一待识别数据不在其应符合的数值范围内,确定第二方的待识别数据集中存在非法数据。



CN 117688589 A

1.一种第一方对第二方进行非法数据识别的方法,所述方法由所述第一方执行,其特征在于,所述方法包括:

获得所述第二方的第一数据碎片集,其中,所述第二方的待识别数据集中的每个待识别数据被碎片化成第一数据碎片和第二数据碎片,所述第一数据碎片集包括所有第一数据碎片;

与所述第二方联合确定用于对所述第一数据碎片集进行随机采样的随机采样比例,其中,所述随机采样比例由第一采样比例碎片和第二采样比例碎片之和来确定,所述第一方持有所述第一采样比例碎片,所述第二方持有所述第二采样比例碎片;

与所述第二方联合针对每个第一数据碎片确定一个参考随机数,其中,所述参考随机数符合均匀分布,所述参考随机数由第一参考随机数碎片和第二参考随机数碎片之和来确定,所述第一方持有所述第一参考随机数碎片,所述第二方持有所述第二参考随机数碎片;

利用所述第一采样比例碎片和所述第一参考随机数碎片与所述第二方联合从所述第一数据碎片集中采样出指定的第一数据碎片;

利用每个指定的第一数据碎片与所述第二方联合比较该指定的第一数据碎片所对应的待识别数据与该待识别数据应符合的数值范围,其中,每个待识别数据应符合的数值范围是根据所述指定的第一数据碎片的属性特征来确定的;以及

响应于任一待识别数据不在其应符合的数值范围内,确定所述第二方的所述待识别数据集中存在非法数据。

2.根据权利要求1所述的方法,其特征在于,与所述第二方联合确定用于对所述第一数据碎片集进行随机采样的随机采样比例包括:

生成第一随机数,其中,所述第一随机数大于0且小于或者等于0.5;

将所述第一随机数碎片化成第一随机数碎片和第二随机数碎片;

接收来自所述第二方的第三随机数碎片,其中,所述第二方生成第二随机数并将所述第二随机数碎片化成第三随机数碎片和第四随机数碎片,所述第二随机数大于0且小于或者等于0.5;

将所述第一随机数碎片和所述第三随机数碎片相加以获得第一随机数碎片和;

生成第三随机数和第四随机数,其中,所述第三随机数大于或者等于0且小于1,所述第四随机数大于0且小于或者等于1,所述第三随机数小于所述第四随机数;

将所述第一随机数碎片和乘以所述第四随机数与所述第三随机数之差的积加上所述第三随机数以获得所述第一采样比例碎片;

向所述第二方发送所述第二随机数碎片、所述第三随机数和所述第四随机数,以便所述第二方将所述第二随机数碎片与所述第四随机数碎片之和乘以所述第四随机数与所述第三随机数之差的积加上所述第三随机数来获得所述第二采样比例碎片。

3.根据权利要求1所述的方法,其特征在于,与所述第二方联合针对每个第一数据碎片确定一个参考随机数包括:

生成第一随机数序列和第二随机数序列,其中,所述第一随机数序列和所述第二随机数序列中的随机数的数量等于所述第一数据碎片集中的第一数据碎片的数量,所述第一随机数序列和所述第二随机数序列中的每个随机数大于或者等于0且小于1;

将所述第一随机数序列碎片化成第一碎片序列和第二碎片序列;

将所述第二随机数序列碎片化成第三碎片序列和第四碎片序列；

接收来自所述第二方的第五碎片序列和第七碎片序列，其中，所述第二方生成第三随机数序列和第四随机数序列，将所述第三随机数序列碎片化成所述第五碎片序列和第六碎片序列，并且将所述第四随机数序列碎片化成所述第七碎片序列和第八碎片序列，所述第三随机数序列和所述第四随机数序列中的随机数数量等于所述第一数据碎片集中的第一数据碎片的数量，所述第三随机数序列和所述第四随机数序列中的每个随机数大于或者等于0且小于1；

向所述第二方发送所述第二碎片序列和所述第四碎片序列；

利用所述第一碎片序列和所述第五碎片序列，与所述第二方联合比较所述第一随机数序列中的每个随机数是否小于所述第三随机数序列中的相应随机数，其中，所述比较结果由第一布尔结果碎片和第二布尔结果碎片进行异或的结果来确定，所述第一方持有所述第一布尔结果碎片，所述第二方持有所述第二布尔结果碎片；

与所述第二方联合将所述第一布尔结果碎片和所述第二布尔结果碎片分别转化为第一算术结果碎片和第二算术结果碎片，其中，所述第一方持有所述第一算术结果碎片，所述第二方持有所述第二算术结果碎片；

根据所述第一算术结果碎片来生成第一乘法因子碎片，其中，所述第一乘法因子碎片中的每个元素等于1与所述第一算术结果碎片中的相应元素之差；

与所述第二方联合计算所述第四随机数序列与乘法因子之积以获得掩膜序列，其中，所述第二方根据所述第二算术结果碎片来生成第二乘法因子碎片，所述第二乘法因子碎片中的每个元素等于0与所述第二算术结果碎片中的相应元素之差，所述乘法因子等于所述第一乘法因子碎片与所述第二乘法因子碎片之和，所述掩膜序列等于第一掩膜碎片序列和第二掩膜碎片序列之和，所述第一方持有所述第一掩膜碎片序列，所述第二方持有所述第二掩膜碎片序列；

利用所述第三碎片序列、所述第一算术结果碎片和所述第一掩膜碎片序列，与所述第二方联合计算所述第二随机数序列乘以所述第一算术结果碎片与所述第二算术结果碎片之和的积加上所述掩膜序列以获得参考随机数序列，其中，所述参考随机数序列等于第一参考随机数碎片序列和第二参考随机数碎片序列之和，所述第一方持有所述第一参考随机数碎片序列，所述第二方持有所述第二参考随机数碎片序列，所述第一参考随机数碎片序列包括针对每个第一数据碎片的第一参考随机数碎片，所述第二参考随机数碎片序列包括针对每个第一数据碎片的第二参考随机数碎片。

4.根据权利要求3所述的方法，其特征在于，与所述第二方联合比较所述第一随机数序列中的每个随机数是否小于所述第三随机数序列中的相应随机数包括：

将所述第一碎片序列减去所述第五碎片序列以获得第九碎片序列；

获得第一布尔零碎片序列和第一算术零碎片序列，其中，所述第一布尔零碎片序列中的每个元素与第二布尔零碎片序列中的相应元素异或的结果为0，所述第一算术零碎片序列中的每个元素与第二算术零碎片序列中的相应元素相加的结果为0，所述第二方拥有所述第二布尔零碎片序列和所述第二算术零碎片序列；

计算所述第九碎片序列与所述第一算术零碎片序列之和与所述第一布尔零碎片序列异或的结果，以获得第一运算碎片序列；

与所述第二方联合利用所述第一方处的第一并行前缀加法器和所述第二方处的第二并行前缀加法器在所述第一方处获得第一符号位碎片序列并且在所述第二方处获得第二符号位碎片序列,其中,所述第一并行前缀加法器的输入为所述第一运算碎片序列和第三运算碎片序列,所述第二并行前缀加法器的输入为第二运算碎片序列和第四运算碎片序列,所述第二运算碎片序列由所述第二方计算第十碎片序列与所述第二算术零碎片序列之和来获得,所述第十碎片序列由所述第二方将所述第二碎片序列减去所述第六碎片序列来获得,所述第三运算碎片序列中的每个元素等于0,所述第四运算碎片序列等于所述第二布尔零碎片序列;

接收来自所述第二方的所述第二符号位碎片序列;

对所述第一符号位碎片序列与所述第二符号位碎片序列执行异或操作以获得比较值序列;

响应于所述比较值序列中的第一比较值为真,确定所述第一随机数序列中与所述第一比较值相对应的随机数小于所述第三随机数序列中与所述第一比较值相对应的随机数;以及

响应于所述比较值序列中的所述第一比较值不为真,确定所述第一随机数序列中与所述第一比较值相对应的随机数不小于所述第三随机数序列中与所述第一比较值相对应的随机数。

5.根据权利要求1所述的方法,其特征在于,所述指定的第一数据碎片对应的参考随机数小于所述随机采样比例。

6.根据权利要求5所述的方法,其特征在于,利用所述第一采样比例碎片和所述第一参考随机数碎片与所述第二方联合从所述第一数据碎片集中采样出指定的第一数据碎片包括:

将所述第一参考随机数碎片减去所述第一采样比例碎片以获得第一比较碎片值;

获得第一布尔零碎片和第一算术零碎片,其中,所述第一布尔零碎片与第二布尔零碎片异或的结果为0,所述第一算术零碎片与第二算术零碎片相加的结果为0,所述第二方拥有所述第二布尔零碎片和所述第二算术零碎片;

计算所述第一比较碎片值与所述第一算术零碎片之和与所述第一布尔零碎片异或的结果,以获得第一运算碎片;

由所述第一方和所述第二方联合利用所述第一方处的第一并行前缀加法器和所述第二方处的第二并行前缀加法器在所述第一方处获得第一符号位碎片并且在所述第二方处获得第二符号位碎片,其中,所述第一并行前缀加法器的输入为所述第一运算碎片和第三运算碎片,所述第二并行前缀加法器的输入为第二运算碎片和第四运算碎片,所述第二运算碎片由所述第二方计算第二比较碎片值与所述第二算术零碎片之和来获得,所述第二比较碎片值由所述第二方将所述第二参考随机数碎片减去所述第二采样比例碎片来获得,所述第三运算碎片等于0,所述第四运算碎片等于所述第二布尔零碎片;

接收来自所述第二方的所述第二符号位碎片;

对所述第一符号位碎片与所述第二符号位碎片执行异或操作以获得第二比较值;

响应于所述第二比较值为真,确定与所述第二比较值相对应的第一数据碎片被采样;以及

响应于所述第二比较值不为真,确定与所述第二比较值相对应的第一数据碎片不被采样。

7.根据权利要求6所述的方法,其特征在于,利用所述第一采样比例碎片和所述第一参考随机数碎片与所述第二方联合从所述第一数据碎片集中采样出指定的第一数据碎片还包括:

将每个第一参考随机数碎片所对应的第二比较值转化为算数值,其中,为真的第二比较值被转化为1,不为真的第二比较值被转化为0;

将转化后的所有第二比较值按降序排列;以及

将值为1的所有第二比较值所对应的第一数据碎片确定为所述指定的第一数据碎片。

8.根据权利要求1所述的方法,其特征在于,利用每个指定的第一数据碎片与所述第二方联合比较该指定的第一数据碎片所对应的待识别数据与该待识别数据应符合的数值范围包括针对每个指定的第一数据碎片执行以下操作:

确定该指定的第一数据碎片的属性特征;

根据该指定的第一数据碎片的属性特征来确定该指定的第一数据碎片所对应的待识别数据应符合的数值范围的上限值和下限值;

与所述第二方联合比较该指定的第一数据碎片所对应的待识别数据是否小于所述上限值;

与所述第二方联合比较所述下限值是否小于该指定的第一数据碎片所对应的待识别数据;

响应于该指定的第一数据碎片所对应的待识别数据不小于所述上限值或者所述下限值不小于该指定的第一数据碎片所对应的待识别数据,确定该指定的第一数据碎片所对应的待识别数据不在其应符合的数值范围内。

9.一种第一方对第二方进行非法数据识别的装置,所述装置被布置在所述第一方处,其特征在于,所述装置包括:

至少一个处理器;以及

存储有计算机程序的至少一个存储器;

其中,当所述计算机程序由所述至少一个处理器执行时,使得所述装置执行根据权利要求1至8中任一项所述的方法的步骤。

10.一种存储有计算机程序的计算机可读存储介质,其特征在于,所述计算机程序在由处理器执行时实现根据权利要求1至8中任一项所述的方法的步骤。

第一方对第二方进行非法数据识别的方法及装置

技术领域

[0001] 本公开的实施例涉及数据处理技术领域,具体地,涉及第一方对第二方进行非法数据识别的方法及装置。

背景技术

[0002] 随着互联网的发展,各类政务主体、行业主体、公司主体、机构主体可经由互联网被关联起来。每个主体可被看作一个节点。在此类网络中,存在大量的用户节点,用户会上传大量的数据集作为多节点之间联合隐私计算任务所需的输入。这一类联合隐私计算任务可以包括联邦学习、联合计算、多方安全计算、隐私求交、全匿踪联邦学习等。

[0003] 隐私计算,因其特殊性,特别是针对纵向场景,也就是特征互补的场景,计算通信量往往会非常庞大。在处理亿级、十亿级,甚至百亿级的数据量时,通信和计算代价非常大,因此联合计算相关的用户参与者,对于计算任务的可靠性会非常关注,因为其直接与计算经济代价相关。

[0004] 影响计算成功或者有效性的因素有很多,比如硬件、网络等客观的环境条件。在这诸多因素中,数据质量是其中非常关键的因素。如果用户上传的数据合法性存在异常,那么往往得到的计算结果也是不可信、不可用的。因此在计算之前,提前识别非法数据是否存在,对于任务发起方来说,具有极大的性价比,可以在任务真实执行之前就决定是否发起昂贵的计算任务。并且,在对数据非法性校验的过程中,需要保证数据的安全性,不能造成原始数据和敏感数据的泄露。

发明内容

[0005] 本文中描述的实施例提供了一种第一方对第二方进行非法数据识别的方法、装置以及存储有计算机程序的计算机可读存储介质。

[0006] 根据本公开的第一方面,提供了一种第一方对第二方进行非法数据识别的方法。该方法由第一方执行。该方法包括:获得第二方的第一数据碎片集,其中,第二方的待识别数据集中的每个待识别数据被碎片化成第一数据碎片和第二数据碎片,第一数据碎片集包括所有第一数据碎片;与第二方联合确定用于对第一数据碎片集进行随机采样的随机采样比例,其中,随机采样比例由第一采样比例碎片和第二采样比例碎片之和来确定,第一方持有第一采样比例碎片,第二方持有第二采样比例碎片;与第二方联合针对每个第一数据碎片确定一个参考随机数,其中,参考随机数符合均匀分布,参考随机数由第一参考随机数碎片和第二参考随机数碎片之和来确定,第一方持有第一参考随机数碎片,第二方持有第二参考随机数碎片;利用第一采样比例碎片和第一参考随机数碎片与第二方联合从第一数据碎片集中采样出指定的第一数据碎片;利用每个指定的第一数据碎片与第二方联合比较该指定的第一数据碎片所对应的待识别数据与该待识别数据应符合的数值范围,其中,每个待识别数据应符合的数值范围是根据指定的第一数据碎片的属性特征来确定的;以及响应于任一待识别数据不在其应符合的数值范围内,确定第二方的待识别数据集中存在非法数

据。

[0007] 在本公开的一些实施例中,与第二方联合确定用于对第一数据碎片集进行随机采样的随机采样比例包括:生成第一随机数,其中,第一随机数大于0且小于或者等于0.5;将第一随机数碎片化成第一随机数碎片和第二随机数碎片;接收来自第二方的第三随机数碎片,其中,第二方生成第二随机数并将第二随机数碎片化成第三随机数碎片和第四随机数碎片,第二随机数大于0且小于或者等于0.5;将第一随机数碎片和第三随机数碎片相加以获得第一随机数碎片和;生成第三随机数和第四随机数,其中,第三随机数大于或者等于0且小于1,第四随机数大于0且小于或者等于1,第三随机数小于第四随机数;将第一随机数碎片和乘以第四随机数与第三随机数之差的积加上第三随机数以获得第一采样比例碎片;向第二方发送第二随机数碎片、第三随机数和第四随机数,以便第二方将第二随机数碎片与第四随机数碎片之和乘以第四随机数与第三随机数之差的积加上第三随机数来获得第二采样比例碎片。

[0008] 在本公开的一些实施例中,与第二方联合针对每个第一数据碎片确定一个参考随机数包括:生成第一随机数序列和第二随机数序列,其中,第一随机数序列和第二随机数序列中的随机数的数量等于第一数据碎片集中的第一数据碎片的数量,第一随机数序列和第二随机数序列中的每个随机数大于或者等于0且小于1;将第一随机数序列碎片化成第一碎片序列和第二碎片序列;将第二随机数序列碎片化成第三碎片序列和第四碎片序列;接收来自第二方的第五碎片序列和第七碎片序列,其中,第二方生成第三随机数序列和第四随机数序列,将第三随机数序列碎片化成第五碎片序列和第六碎片序列,并且将第四随机数序列碎片化成第七碎片序列和第八碎片序列,第三随机数序列和第四随机数序列中的随机数数量等于第一数据碎片集中的第一数据碎片的数量,第三随机数序列和第四随机数序列中的每个随机数大于或者等于0且小于1;向第二方发送第二碎片序列和第四碎片序列;利用第一碎片序列和第五碎片序列,与第二方联合比较第一随机数序列中的每个随机数是否小于第三随机数序列中的相应随机数,其中,比较结果由第一布尔结果碎片和第二布尔结果碎片进行异或的结果来确定,第一方持有第一布尔结果碎片,第二方持有第二布尔结果碎片;与第二方联合将第一布尔结果碎片和第二布尔结果碎片分别转化为第一算术结果碎片和第二算术结果碎片,其中,第一方持有第一算术结果碎片,第二方持有第二算术结果碎片;根据第一算术结果碎片来生成第一乘法因子碎片,其中,第一乘法因子碎片中的每个元素等于1与第一算术结果碎片中的相应元素之差;与第二方联合计算第四随机数序列与乘法因子之积以获得掩膜序列,其中,第二方根据第二算术结果碎片来生成第二乘法因子碎片,第二乘法因子碎片中的每个元素等于0与第二算术结果碎片中的相应元素之差,乘法因子等于第一乘法因子碎片与第二乘法因子碎片之和,掩膜序列等于第一掩膜碎片序列和第二掩膜碎片序列之和,第一方持有第一掩膜碎片序列,第二方持有第二掩膜碎片序列;利用第三碎片序列、第一算术结果碎片和第一掩膜碎片序列,与第二方联合计算第二随机数序列乘以第一算术结果碎片与第二算术结果碎片之和的积加上掩膜序列以获得参考随机数序列,其中,参考随机数序列等于第一参考随机数碎片序列和第二参考随机数碎片序列之和,第一方持有第一参考随机数碎片序列,第二方持有第二参考随机数碎片序列,第一参考随机数碎片序列包括针对每个第一数据碎片的第一参考随机数碎片,第二参考随机数碎片序列包括针对每个第一数据碎片的第二参考随机数碎片。

[0009] 在本公开的一些实施例中,与第二方联合比较第一随机数序列中的每个随机数是否小于第三随机数序列中的相应随机数包括:将第一碎片序列减去第五碎片序列以获得第九碎片序列;获得第一布尔零碎片序列和第一算术零碎片序列,其中,第一布尔零碎片序列中的每个元素与第二布尔零碎片序列中的相应元素异或的结果为0,第一算术零碎片序列中的每个元素与第二算术零碎片序列中的相应元素相加的结果为0,第二方拥有第二布尔零碎片序列和第二算术零碎片序列;计算第九碎片序列与第一算术零碎片序列之和与第一布尔零碎片序列异或的结果,以获得第一运算碎片序列;与第二方联合利用第一方处的第一并行前缀加法器和第二方处的第二并行前缀加法器在第一方处获得第一符号位碎片序列并且在第二方处获得第二符号位碎片序列,其中,第一并行前缀加法器的输入为第一运算碎片序列和第三运算碎片序列,第二并行前缀加法器的输入为第二运算碎片序列和第四运算碎片序列,第二运算碎片序列由第二方计算第十碎片序列与第二算术零碎片序列之和来获得,第十碎片序列由第二方将第二碎片序列减去第六碎片序列来获得,第三运算碎片序列中的每个元素等于0,第四运算碎片序列等于第二布尔零碎片序列;接收来自第二方的第二符号位碎片序列;对第一符号位碎片序列与第二符号位碎片序列执行异或操作以获得比较值序列;响应于比较值序列中的第一比较值为真,确定第一随机数序列中与第一比较值相对应的随机数小于第三随机数序列中与第一比较值相对应的随机数;以及响应于比较值序列中的第一比较值不为真,确定第一随机数序列中与第一比较值相对应的随机数不小于第三随机数序列中与第一比较值相对应的随机数。

[0010] 在本公开的一些实施例中,指定的第一数据碎片对应的参考随机数小于随机采样比例。

[0011] 在本公开的一些实施例中,利用第一采样比例碎片和第一参考随机数碎片与第二方联合从第一数据碎片集中采样出指定的第一数据碎片包括:将第一参考随机数碎片减去第一采样比例碎片以获得第一比较碎片值;获得第一布尔零碎片和第一算术零碎片,其中,第一布尔零碎片与第二布尔零碎片异或的结果为0,第一算术零碎片与第二算术零碎片相加的结果为0,第二方拥有第二布尔零碎片和第二算术零碎片;计算第一比较碎片值与第一算术零碎片之和与第一布尔零碎片异或的结果,以获得第一运算碎片;由第一方和第二方联合利用第一方处的第一并行前缀加法器和第二方处的第二并行前缀加法器在第一方处获得第一符号位碎片并且在第二方处获得第二符号位碎片,其中,第一并行前缀加法器的输入为第一运算碎片和第三运算碎片,第二并行前缀加法器的输入为第二运算碎片和第四运算碎片,第二运算碎片由第二方计算第二比较碎片值与第二算术零碎片之和来获得,第二比较碎片值由第二方将第二参考随机数碎片减去第二采样比例碎片来获得,第三运算碎片等于0,第四运算碎片等于第二布尔零碎片;接收来自第二方的第二符号位碎片;对第一符号位碎片与第二符号位碎片执行异或操作以获得第二比较值;响应于第二比较值为真,确定与第二比较值相对应的第一数据碎片被采样;以及响应于第二比较值不为真,确定与第二比较值相对应的第一数据碎片不被采样。

[0012] 在本公开的一些实施例中,利用第一采样比例碎片和第一参考随机数碎片与第二方联合从第一数据碎片集中采样出指定的第一数据碎片还包括:将每个第一参考随机数碎片所对应的第二比较值转化为算数值,其中,为真的第二比较值被转化为1,不为真的第二比较值被转化为0;将转化后的所有第二比较值按降序排列;以及将值为1的所有第二比较

值所对应的第一数据碎片确定为指定的第一数据碎片。

[0013] 在本公开的一些实施例中,利用每个指定的第一数据碎片与第二方联合比较该指定的第一数据碎片所对应的待识别数据与该待识别数据应符合的数值范围包括针对每个指定的第一数据碎片执行以下操作:确定该指定的第一数据碎片的属性特征;根据该指定的第一数据碎片的属性特征来确定该指定的第一数据碎片所对应的待识别数据应符合的数值范围的上限值和下限值;与第二方联合比较该指定的第一数据碎片所对应的待识别数据是否小于上限值;与第二方联合比较下限值是否小于该指定的第一数据碎片所对应的待识别数据;响应于该指定的第一数据碎片所对应的待识别数据不小于上限值或者下限值不小于该指定的第一数据碎片所对应的待识别数据,确定该指定的第一数据碎片所对应的待识别数据不在其应符合的数值范围内。

[0014] 根据本公开的第二方面,提供了一种第一方对第二方进行非法数据识别的装置。装置被布置在第一方处。该装置包括至少一个处理器;以及存储有计算机程序的至少一个存储器。当计算机程序由至少一个处理器执行时,使得装置执行根据本公开的第一方面所述的方法的步骤。

[0015] 根据本公开的第三方面,提供了一种存储有计算机程序的计算机可读存储介质,其中,计算机程序在由处理器执行时实现根据本公开的第一方面所述的方法的步骤。

附图说明

[0016] 为了更清楚地说明本公开的实施例的技术方案,下面将对实施例的附图进行简要说明,应当知道,以下描述的附图仅仅涉及本公开的一些实施例,而非对本公开的限制,其中:

[0017] 图1是数联网的示意性拓扑图;

[0018] 图2是根据本公开的实施例的第一方对第二方进行非法数据识别的方法的示意性流程图;

[0019] 图3是根据本公开的实施例的确定用于对第一数据碎片集进行随机采样的随机采样比例的步骤的示意性流程图和信令方案;

[0020] 图4是根据本公开的实施例的针对每个第一数据碎片确定一个参考随机数的步骤的示意性流程图和信令方案;

[0021] 图5是根据本公开的实施例的确定第一值是否小于第二值的步骤的示意性流程图和信令方案;

[0022] 图6是图5中的动作511的示意性流程图和信令方案;

[0023] 图7是图6中的动作603的示意性流程图和信令方案;

[0024] 图8是图6中的动作604和605的示意性流程图和信令方案;

[0025] 图9是根据本公开的实施例的第一方对第二方进行非法数据识别的装置的示意性框图。

[0026] 需要注意的是,附图中的元素是示意性的,没有按比例绘制。

具体实施方式

[0027] 为了使本公开的实施例的目的、技术方案和优点更加清楚,下面将结合附图,对本

公开的实施例的技术方案进行清楚、完整的描述。显然,所描述的实施例是本公开的一部分实施例,而不是全部的实施例。基于所描述的本公开的实施例,本领域技术人员在无需创造性劳动的前提下所获得的所有其它实施例,也都属于本公开保护的范围。

[0028] 除非另外定义,否则在此使用的所有术语(包括技术和科学术语)具有与本公开主题所属领域的技术人员所通常理解的相同含义。进一步将理解的是,诸如在通常使用的词典中定义的那些的术语应解释为具有与说明书上下文和相关技术中它们的含义一致的含义,并且将不以理想化或过于正式的形式来解释,除非在此另外明确定义。另外,诸如“第一”和“第二”的术语仅用于将一个部件(或部件的一部分)与另一个部件(或部件的另一部分)区分开。

[0029] 如上所述,各行各业中的各类政务主体、行业主体、公司主体、机构主体可组合成网络,而此类网络中的主体数量可能非常大。数联网就是此类网络的一个示例。数联网作为开放性的网络体系,可以支持各类主体通过被允许的方式接入数联网,接入数联网后就形成了相应数据节点。数据节点之间存在联合建模、联合统计等多方协同计算的业务需求。成熟的数联网中数据节点数量级可以达到万、十万甚至是百万量级,规模跨越多个数量级。

[0030] 图1示出数联网的示意性拓扑图。数联网可包括多个子网10。每个子网10包括枢纽节点11和与枢纽节点直接连接的多个参与节点12。该多个子网10中的枢纽节点11相互直接连接。枢纽节点11与枢纽节点11之间可以通过专网进行互联。枢纽节点11承担对参与节点12进行信息聚合、寻址导航等功能。参与节点12可以是各类政务主体、行业主体、公司主体、机构主体等。直接连接到同一个枢纽节点11的参与节点12通过该枢纽节点11进行通信。直接连接到不同枢纽节点11的参与节点12通过它们各自直接连接的枢纽节点11进行通信。也就是说,参与节点12只与其直接连接的枢纽节点11直接通信,枢纽节点11之间可直接通信,而参与节点12之间需经由相应的枢纽节点11进行通信。

[0031] 在实践中,参与节点12之间可能需要联合执行计算任务,并且联合计算任务所涉及的数据量可能是非常庞大的。本公开提出在正式执行联合计算任务之前联合计算任务的一方(在上下文中称为“第一方”)对联合计算任务的其它方(在上下文中称为“第二方”)进行非法数据识别。如果识别到第二方的非法数据,第一方可不与第二方联合执行任务,以避免经济和时间损失。

[0032] 图2示出根据本公开的实施例的第一方对第二方进行非法数据识别的方法的示意性流程图。非法数据在上下文中指的是数值不在其应符合的数值范围内的数据。例如,等于负数的年龄是非法数据。

[0033] 在图2的框S202处,第一方获得第二方的第一数据碎片集。第一数据碎片集来自第二方。第二方的待识别数据集中的每个待识别数据被碎片化成第一数据碎片和第二数据碎片。第一数据碎片集包括所有第一数据碎片。第一方从第一数据碎片无法还原出第二方的待识别数据。

[0034] 在框S204处,第一方与第二方联合确定用于对第一数据碎片集进行随机采样的随机采样比例。随机采样比例由第一采样比例碎片和第二采样比例碎片之和来确定。第一方持有第一采样比例碎片。第二方持有第二采样比例碎片。图3示出根据本公开的实施例的确定用于对第一数据碎片集进行随机采样的随机采样比例的步骤的示意性流程图和信令方案。

[0035] 在图3的示例中,由第一方P1在动作301生成第一随机数r1并将第一随机数r1碎片化成第一随机数碎片r1_0和第二随机数碎片r1_1。其中,第一随机数r1大于0且小于或者等于0.5。由第二方P2在动作302生成第二随机数r2并将第二随机数r2碎片化成第三随机数碎片r2_0和第四随机数碎片r2_1。其中,第二随机数r2大于0且小于或者等于0.5。由第一方P1在动作303向第二方P2发送第二随机数碎片r1_1。由第二方P2在动作304向第一方P1发送第三随机数碎片r2_0。由第一方P1在动作305将第一随机数碎片r1_0和第三随机数碎片r2_0相加以获得第一随机数碎片和rs_0。由第二方P2在动作306将第二随机数碎片r1_1和第四随机数碎片r2_1相加以获得第二随机数碎片和rs_1。由第一方P1在动作307生成第三随机数start和第四随机数end。其中,第三随机数start大于或者等于0且小于1。第四随机数end大于0且小于或者等于1。第三随机数start小于第四随机数end。在一个示例中,可生成范围在0至1之间的2个随机数,将数值较小的随机数称为第三随机数start,将数值较大的随机数称为第四随机数end。由第一方P1在动作308向第二方P2发送第三随机数start和第四随机数end。由第一方P1在动作311将第一随机数碎片和rs_0乘以第四随机数end与第三随机数start之差的积加上第三随机数start以获得第一采样比例碎片P_0,即P_0＝start+rs_0×(end-start)。由第二方P2在动作312将第二随机数碎片和rs_1(即,第二随机数碎片r1_1与第四随机数碎片r2_1之和)乘以第四随机数end与第三随机数start之差的积加上第三随机数start来获得第二采样比例碎片P_1,即P_1＝start+rs_1×(end-start)。

[0036] 在框S206处,第一方与第二方联合针对每个第一数据碎片确定一个参考随机数。参考随机数符合均匀分布。参考随机数由第一参考随机数碎片和第二参考随机数碎片之和来确定。第一方持有第一参考随机数碎片。第二方持有第二参考随机数碎片。图4示出根据本公开的实施例的针对每个第一数据碎片确定一个参考随机数的步骤的示意性流程图和信令方案。

[0037] 在图4的示例中,由第一方P1在动作401生成第一随机数序列r11和第二随机数序列r12,将第一随机数序列r11碎片化成第一碎片序列r11_0和第二碎片序列r11_1(r11＝r11_0+r11_1),并将第二随机数序列r12碎片化成第三碎片序列r12_0和第四碎片序列r12_1(r12＝r12_0+r12_1)。第一随机数序列r11和第二随机数序列r12分别包括多个随机数。第一随机数序列r11和第二随机数序列r12中的随机数的数量等于第一数据碎片集中的第一数据碎片的数量。第一随机数序列r11和第二随机数序列r12中的每个随机数大于或者等于0且小于1。第一碎片序列r11_0、第二碎片序列r11_1、第三碎片序列r12_0和第四碎片序列r12_1的大小与第一随机数序列r11的大小相同。这里的大小指的是所包含的元素数量。假设第一随机数序列r11为[0.5,0.8,0.1],则可将第一随机数序列r11碎片化成第一碎片序列r11_0＝[0.4,0.1,0.5]和第二碎片序列r11_1＝[0.1,0.7,-0.4]。

[0038] 第二方P2在动作402生成第三随机数序列r21和第四随机数序列r22,将第三随机数序列r21碎片化成第五碎片序列r21_0和第六碎片序列r21_1(r21＝r21_0+r21_1),并且将第四随机数序列r22碎片化成第七碎片序列r22_0和第八碎片序列r22_1(r22＝r22_0+r22_1)。第三随机数序列r21和第四随机数序列r22分别包括多个随机数。第三随机数序列r21和第四随机数序列r22中的随机数数量等于第一数据碎片集中的第一数据碎片的数量。第三随机数序列r21和第四随机数序列r22中的每个随机数大于或者等于0且小于1。第五碎片序列r21_0、第六碎片序列r21_1、第七碎片序列r22_0和第八碎片序列r22_1的大小与第

一随机数序列r11的大小相同。

[0039] 由第一方P1在动作403向第二方P2发送第二碎片序列r11_1和第四碎片序列r12_1。由第二方P2在动作404向第一方P1发送第五碎片序列r21_0和第七碎片序列r22_0。

[0040] 由第一方P1在动作405利用第一碎片序列r11_0和第五碎片序列r21_0,与第二方P2联合比较第一随机数序列r11中的每个随机数是否小于第三随机数序列r21中的相应随机数。如图4所示,在动作405中,第二方P2利用第二碎片序列r11_1和第六碎片序列r21_1来联合比较。动作405的具体操作过程将在下文结合图5至图8进行介绍。动作405的比较结果Con由第一布尔结果碎片Con_0和第二布尔结果碎片Con_1进行异或的结果来确定。$Con=Con\_0 \oplus Con\_1$。第一方P1持有第一布尔结果碎片Con_0。第二方P2持有第二布尔结果碎片Con_1。如果r11中的某个随机数小于r21中的相应随机数,则Con中的相应元素为真,否则为假。

[0041] 由第一方P1在动作406与第二方P2联合将第一布尔结果碎片Con_0和第二布尔结果碎片Con_1分别转化为第一算术结果碎片sel_0和第二算术结果碎片sel_1。第一方持有第一算术结果碎片sel_0,第二方持有第二算术结果碎片sel_1。第一算术结果碎片sel_0和第二算术结果碎片sel_1之和sel(sel＝sel_0+sel_1)是比较结果Con所对应的算术值。在比较结果Con为真的情况下,sel＝1。在比较结果Con不为真的情况下,sel＝0。该动作相当于将布尔值Con转化为算术值sel。

[0042] 由第一方P1在动作407根据第一算术结果碎片sel_0来生成第一乘法因子碎片sb_0。第一乘法因子碎片sb_0中的每个元素等于1与第一算术结果碎片sel_0中的相应元素之差(sb_0＝1-sel_0)。在这里,相应元素指的是位置相同的元素。例如,第一算术结果碎片sel_0中的第i个元素是第一乘法因子碎片sb_0中的第i个元素的相应元素。由第二方P2在动作408根据第二算术结果碎片sel_1来生成第二乘法因子碎片sb_1。第二乘法因子碎片sb_1中的每个元素等于0与第二算术结果碎片sel_1中的相应元素之差。或者可以认为第二乘法因子碎片sb_1中的每个元素等于第二算术结果碎片sel_1中的相应元素的相反数。sb_1＝-sel_1。

[0043] 由第一方P1在动作409与第二方P2联合计算第四随机数序列r22与乘法因子sb之积以获得掩膜序列r22_mask。其中,乘法因子sb等于第一乘法因子碎片sb_0与第二乘法因子碎片sb_1之和(sb＝sb_0+sb_1)。掩膜序列r22_mask等于第一掩膜碎片序列r22_mask_0和第二掩膜碎片序列r22_mask_1之和。第一方P1持有第一掩膜碎片序列r22_mask_0。第二方P2持有第二掩膜碎片序列r22_mask_1。在图4中,<>表示碎片态。<r22_mask>表示掩膜序列r22_mask的碎片态。<r22>表示第四随机数序列r22的碎片态。<sb>表示乘法因子sb的碎片态。<r22_mask>＝<r22>⊙<sb>表示r22_mask＝r22⊙sb的操作是在碎片态下完成的,不会泄露任何一方的原始数据。⊙表示哈达玛积。

[0044] 由第一方P1在动作410利用第三碎片序列r12_0、第一算术结果碎片sel_0和第一掩膜碎片序列r22_mask_0,与第二方P2联合计算第二随机数序列r12乘以sel(第一算术结果碎片sel_0与第二算术结果碎片sel_1之和)的积加上掩膜序列r22_mask以获得参考随机数序列R。R＝r12⊙sel+r22_mask。如图4所示,在动作410中,第二方P2利用第四碎片序列r12_1、第二算术结果碎片sel_1和第二掩膜碎片序列r22_mask_1来联合计算。在动作410计算出的参考随机数序列R等于第一参考随机数碎片序列R0和第二参考随机数碎片序列R1之

和(R＝R0+R1)。第一方P1持有第一参考随机数碎片序列R0,第二方P2持有第二参考随机数碎片序列R1。第一参考随机数碎片序列R0包括针对每个第一数据碎片的第一参考随机数碎片。第二参考随机数碎片序列R1包括针对每个第一数据碎片的第二参考随机数碎片。<R>＝<r12>⊙<sel>+<r22_mask>表示R＝r12⊙sel+r22_mask的操作是在碎片态下完成的,不会泄露任何一方的原始数据。

[0045] 图5示出根据本公开的实施例的确定第一值是否小于第二值的步骤的示意性流程图和信令方案。由第一方P1在动作501将第一值x碎片化为第一碎片值x1和第二碎片值x2(x＝x1+x2)并在动作503向第二方P2发送第二碎片值x2。由第二方P2在动作502将第二值y碎片化为第三碎片值y1和第四碎片值y2(y＝y1+y2)并在动作504向第一方P1发送第三碎片值y1。在动作505,由第一方P1将第一碎片值x1减去第三碎片值y1以获得第五碎片值z1,即z1＝x1-y1。在动作506,由第二方P2将第二碎片值x2减去第四碎片值y2以获得第六碎片值z2,即z2＝x2-y2。

[0046] 可由第一方P1和第二方P2中的一者生成第一布尔零碎片a1、第二布尔零碎片a2、第一算术零碎片b1、第二算术零碎片b2。其中,第一布尔零碎片a1与第二布尔零碎片a2异或的结果为$0\ (a1\oplus a2=0)$,第一算术零碎片b1与第二算术零碎片b2相加的结果为0(b1+b2＝0)。在动作507,将第一布尔零碎片a1和第一算术零碎片b1分配给第一方P1。在动作508,将第二布尔零碎片a2和第二算术零碎片b2分配给第二方P2。

[0047] 在动作509,由第一方P1计算第五碎片值z1与第一算术零碎片b1之和与第一布尔零碎片a1异或的结果,以获得第一运算碎片op11,即,$op11=(z1+b1)\oplus a1$。第一方P1还持有第三运算碎片op21,其中,op21＝0。

[0048] 在动作510,由第二方P2计算第六碎片值z2与第二算术零碎片b2之和,以获得第二运算碎片op22,即,op22＝z2+b2。第二方P2还持有第四运算碎片op12,其中,op12＝a2。

[0049] 在动作511,由第一方P1和第二方P2联合利用第一方P1处的第一并行前缀加法器和第二方P2处的第二并行前缀加法器在第一方P1处获得第一符号位碎片B1并且在第二方P2处获得第二符号位碎片B2。第一并行前缀加法器的输入为第一运算碎片op11和第三运算碎片op21。第二并行前缀加法器的输入为第二运算碎片op22和第四运算碎片op12。

[0050] 在由第一方P1来确定比较结果的示例中,第二方P2在动作513向第一方P1发送第二符号位碎片B2。由第一方P1在动作514对第一符号位碎片B1与第二符号位碎片B2执行异或操作以获得比较值。如果比较值为真,则确定第一值小于第二值。如果比较值不为真,则确定第一值不小于第二值。类似地,也可以由第二方P2来确定比较结果。

[0051] 图6示出图5中的动作511的具体过程。在动作603,由第一方P1根据第一运算碎片op11和第三运算碎片op21并且由第二方P2根据第二运算碎片op22和第四运算碎片op12来共同生成第一中间碎片G1和第二中间碎片G2。

[0052] 图7示出由第一方P1和第二方P2联合执行的与运算的示意性流程图和信令方案。在图7中以第一方P1拥有第一输入碎片W1和第二输入碎片V1且第二方P2拥有第三输入碎片W2和第四输入碎片V2为例来进行说明。当图6中的动作603使用图7所示的方案时,第一运算碎片op11相当于第一输入碎片W1,第三运算碎片op21相当于第二输入碎片V1,第二运算碎片op22相当于第三输入碎片W2,第四运算碎片op12相当于第四输入碎片V2。

[0053] 下面描述图7所示的过程。

[0054] 第一方P1在动作701获得三元组碎片矩阵<R1,S1,T1>,第二方P2在动作702获得三元组碎片矩阵<R2,S2,T2>。其中,$(R1 \oplus R2) \& (S1 \oplus S2) = (T1 \oplus T2)$。

[0055] 第一方P1在动作703对W1和R1执行异或操作以获得第三中间碎片D1,对V1和S1执行异或操作以获得第四中间碎片E1。第二方P2在动作704对W2和R2执行异或操作以获得第五中间碎片D2,对V2和S2执行异或操作以获得第六中间碎片E2。

[0056] 第二方P2在动作705向第一方P1发送D2和E2。第一方P1在动作706向第二方P2发送D1和E1。第一方P1在动作707对D1和D2执行异或操作以获得第一合成碎片D,对E1和E2执行异或操作以获得第二合成碎片E。类似的,第二方P2在动作708对D1和D2执行异或操作以获得第一合成碎片D,对E1和E2执行异或操作以获得第二合成碎片E。

[0057] 第一方P1在动作709计算第一输出碎片$O1=T1 \oplus (R1\&E) \oplus (S1\&D)$

[0058] $\oplus (E\&D)$。第二方P2在动作710计算第二输出碎片$O2=T2 \oplus (R2\&E) \oplus (S2\&D)$。当图6中的动作603使用图7所示的方案时,第一中间碎片G1相当于第一输出碎片O1,第二中间碎片G2相当于第二输出碎片O2。

[0059] 回到图6,第一方P1在动作604根据第五中间碎片$p1(p1=op11 \oplus op21)$对G1的每一位进行逐位循环计算。第二方P2在动作605根据第六中间碎片$p2(p2=op12 \oplus op22)$对G2的每一位进行逐位循环计算。图8示出图6中的动作604和605的示意性流程图和信令方案。

[0060] 第一方P1在动作801对G1执行左移2^i位的操作以得到第一临时碎片G11,即G11=G1<<2^i。第二方P2在动作802对G2执行左移2^i位的操作以得到第二临时碎片G12,即G12=G2<<2^i。其中,i表示当前循环的索引。

[0061] 在动作803处,由第一方P1和第二方P2联合执行图7所示的与运算。G11相当于第一输入碎片W1,p1相当于第二输入碎片V1,G12相当于第三输入碎片W2,p2相当于第四输入碎片V2。经过动作803的操作,第一方P1获得第七中间碎片F1,第二方P2获得第八中间碎片F2。F1相当于第一输出碎片O1,F2相当于第二输出碎片O2。

[0062] 第一方P1在动作804对p1执行左移2^i位的操作以获得第三临时碎片p11(即p11=p1<<2^i),然后再将p11更新为p11与kmask异或的结果(即,$p11=p11 \oplus kmask$)。其中,kmask是大小与op11相同的矩阵且其每一个元素值均为2^i-1。

[0063] 第二方P2在动作805对p2执行左移2^i位的操作以获得第四临时碎片p12(即p12=p2<<2^i),然后再将p12更新为p12与kmask异或的结果(即,

[0064] $p12=p12 \oplus kmask$)。

[0065] 在动作806处,由第一方P1和第二方P2联合执行图7所示的与运算。p1相当于第一输入碎片W1,p11相当于第二输入碎片V1,p2相当于第三输入碎片W2,p12相当于第四输入碎片V2。经过动作806的操作,第一方P1获得更新后的p1,第二方P2获得更新后的p2。更新后的p1相当于第一输出碎片O1,更新后的p2相当于第二输出碎片O2。更新后的p1会被代入下一循环的动作803处。更新后的p2也会被代入下一循环的动作803处。

[0066] 第一方P1在动作807对G1和F1执行异或操作以获得更新后的G1(即，$G1=G1\oplus F1$)。更新后的G1会被代入下一循环的动作801处。第二方P2在动作808对G2和F2执行异或操作以获得更新后的G2(即，$G2=G2\oplus F2$)。更新后的G2会被代入下一循环的动作802处。

[0067] 再次回到图6，第一方P1在动作606对G1左移1位以获得第九中间碎片C1(即，C1＝G1<<1)。第二方P2在动作607对G2左移1位以获得第十中间碎片C2(即，C2＝G2<<1)。第一方P1在动作608对p1和C1执行异或操作以获得第十一中间碎片Z1(即，$Z1=p1\oplus C1$)。第二方P2在动作609对p2和C2执行异或操作以获得第十二中间碎片Z2(即，$Z2=p2\oplus C2$)。

[0068] 第一方P1在动作610对Z1和mask执行按位与操作以获得更新后的Z1(即，Z1＝Z1&mask)。其中，mask＝0x1<<n-1，n表示第一值x的位数。第二方P2在动作611对Z2和mask执行按位与操作以获得更新后的Z2(即，Z2＝Z2&mask)。其中，mask＝0x1<<n-1，n表示第二值y的位数。

[0069] 第一方P1在动作612将Z1转换成布尔类型以获得第一符号位碎片B1。

[0070] 第二方P2在动作613将Z2转换成布尔类型以获得第二符号位碎片B2。

[0071] 在上述过程中，由于第一方P1没有获得第二值的完整信息，而第二方P2也没有获得第一值的完整信息，因此该计算过程是安全的，不会泄露任何原始信息。

[0072] 在本公开的一些实施例中，在图4的动作405中，在第一方P1与第二方P2联合比较第一随机数序列r11中的每个随机数是否小于第三随机数序列r21中的相应随机数的过程中，可使用图5至图8所示的方法来执行比较操作。可将第一碎片序列r11_0中的每个随机数当成图5中的x1，可将第五碎片序列r21_0中的每个随机数当成图5中的y1，可将第二碎片序列r11_1中的每个随机数当成图5中的x2，可将第六碎片序列r21_1中的每个随机数当成图5中的y2。然后从图5的动作505和506开始执行后续操作。

[0073] 可替代地，也可以将图5中x和y看成是序列。那么在将图5的方案应用于动作405的情况下，第一方P1将第一碎片序列r11_0减去第五碎片序列r21_0以获得第九碎片序列。第一方P1获得第一布尔零碎片序列和第一算术零碎片序列。其中，第一布尔零碎片序列中的每个元素与第二布尔零碎片序列中的相应元素异或的结果为0。第一算术零碎片序列中的每个元素与第二算术零碎片序列中的相应元素相加的结果为0。第二方P2拥有第二布尔零碎片序列和第二算术零碎片序列。第一方P1计算第九碎片序列与第一算术零碎片序列之和与第一布尔零碎片序列异或的结果，以获得第一运算碎片序列。第一方P1与第二方P2联合利用第一方P1处的第一并行前缀加法器和第二方P2处的第二并行前缀加法器在第一方P1处获得第一符号位碎片序列并且在第二方P2处获得第二符号位碎片序列。其中，第一并行前缀加法器的输入为第一运算碎片序列和第三运算碎片序列。第二并行前缀加法器的输入为第二运算碎片序列和第四运算碎片序列。第二运算碎片序列由第二方P2计算第十碎片序列与第二算术零碎片序列之和来获得。第十碎片序列由第二方P2将第二碎片序列r11_1减去第六碎片序列r21_1来获得。第三运算碎片序列中的每个元素等于0。第四运算碎片序列等于第二布尔零碎片序列。第一方P1接收来自第二方P2的第二符号位碎片序列。第一方P1对第一符号位碎片序列与第二符号位碎片序列执行异或操作以获得比较值序列。如果比较值序列中的第一比较值为真，第一方P1确定第一随机数序列r11中与第一比较值相对应的

随机数小于第三随机数序列r21中与第一比较值相对应的随机数。如果比较值序列中的第一比较值不为真,第一方P1确定第一随机数序列r11中与第一比较值相对应的随机数不小于第三随机数序列r21中与第一比较值相对应的随机数。在这里,第一比较值指的是比较值序列中的任一个比较值。

[0074] 回到图2,在框S208处,第一方利用第一采样比例碎片和第一参考随机数碎片与第二方联合从第一数据碎片集中采样出指定的第一数据碎片。在本公开的一些实施例中,指定的第一数据碎片对应的参考随机数小于在框S204处确定的随机采样比例。可将每个第一参考随机数碎片对应的参考随机数与随机采样比例进行比较,如果该参考随机数小于随机采样比例,则该参考随机数对应的第一数据碎片被确定为该指定的第一数据碎片。由于参考随机数符合均匀分布,因此采样出的第一数据碎片的数量与第一数据碎片集的大小之比可等于随机采样比例。

[0075] 在本公开的一些实施例中,在第一方利用第一采样比例碎片和第一参考随机数碎片与第二方联合从第一数据碎片集中采样出指定的第一数据碎片的过程中,可使用图5至图8所示的方法来比较第一数据碎片对应的参考随机数是否小于随机采样比例。可将第一参考随机数碎片当成图5中的x1,可将第一采样比例碎片P_0当成图5中的y1,可将第二参考随机数碎片当成图5中的x2,可将第二采样比例碎片P_1当成图5中的y2。然后从图5的动作505和506开始执行后续操作。

[0076] 具体地,第一方P1将第一参考随机数碎片减去第一采样比例碎片P_0以获得第一比较碎片值(相当于图5中的z1)。第一方P1获得第一布尔零碎片(相当于图5中的a1)和第一算术零碎片(相当于图5中的b1)。第一布尔零碎片与第二布尔零碎片(相当于图5中的a2)异或的结果为0,第一算术零碎片与第二算术零碎片(相当于图5中的b2)相加的结果为0。第二方P2拥有第二布尔零碎片和第二算术零碎片。第一方P1计算第一比较碎片值与第一算术零碎片之和与第一布尔零碎片异或的结果,以获得第一运算碎片(相当于图5中的op11)。由第一方P1和第二方P2联合利用第一方P1处的第一并行前缀加法器和第二方P2处的第二并行前缀加法器在第一方P1处获得第一符号位碎片(相当于图5中的B1)并且在第二方P2处获得第二符号位碎片(相当于图5中的B2)。其中,第一并行前缀加法器的输入为第一运算碎片和第三运算碎片(相当于图5中的op21)。第二并行前缀加法器的输入为第二运算碎片(相当于图5中的op22)和第四运算碎片(相当于图5中的op12)。第二运算碎片由第二方P2计算第二比较碎片值(相当于图5中的z2)与第二算术零碎片之和来获得。第二比较碎片值由第二方P2将第二参考随机数碎片减去第二采样比例碎片P_1来获得。第三运算碎片等于0。第四运算碎片等于第二布尔零碎片。第一方P1接收来自第二方P2的第二符号位碎片。第一方P1对第一符号位碎片与第二符号位碎片执行异或操作以获得第二比较值。如果第二比较值为真,第一方P1确定与第二比较值相对应的第一数据碎片被采样。如果第二比较值不为真,第一方P1确定与第二比较值相对应的第一数据碎片不被采样。

[0077] 在本公开的一些实施例中,在第一方P1利用第一采样比例碎片和第一参考随机数碎片与第二方P2联合从第一数据碎片集中采样出指定的第一数据碎片的过程中,可将每个第一参考随机数碎片所对应的第二比较值转化为算数值。为真的第二比较值被转化为1,不为真的第二比较值被转化为0。将转化后的所有第二比较值按降序排列。将值为1的所有第二比较值所对应的第一数据碎片确定为指定的第一数据碎片。这样一旦发现某个第二比较

值为0,则该第二比较值之后的第二比较值都为0,不需要再继续处理。而该第二比较值之前的所有第二比较值对应的第一数据碎片即为全部的指定的第一数据碎片。通过上述方式可以快速的确定出所有指定的第一数据碎片。

[0078] 在图2的框S210处,第一方利用每个指定的第一数据碎片与第二方联合比较该指定的第一数据碎片所对应的待识别数据与该待识别数据应符合的数值范围。其中,每个待识别数据应符合的数值范围是根据指定的第一数据碎片的属性特征(也是对应的待识别数据的属性特征)来确定的。例如,如果待识别数据是年龄,则年龄应符合的数值范围是大于0且小于150(在这里,150可以是根据当前已知的最高在世人口来确定的年龄值)。如果待识别数据是薪水,则薪水应符合的数值范围是大于0(不需要检测上限值,或者可认为上限值是无穷大)。

[0079] 在本公开的一些实施例中,在第一方利用每个指定的第一数据碎片与第二方联合比较该指定的第一数据碎片所对应的待识别数据与该待识别数据应符合的数值范围的过程中,第一方可针对每个指定的第一数据碎片执行以下操作:确定该指定的第一数据碎片的属性特征(即,待识别数据的属性特征);根据该指定的第一数据碎片的属性特征来确定该指定的第一数据碎片所对应的待识别数据应符合的数值范围的上限值和下限值;与第二方联合比较该指定的第一数据碎片所对应的待识别数据是否小于上限值;与第二方联合比较下限值是否小于该指定的第一数据碎片所对应的待识别数据;响应于该指定的第一数据碎片所对应的待识别数据不小于上限值或者下限值不小于该指定的第一数据碎片所对应的待识别数据,确定该指定的第一数据碎片所对应的待识别数据不在其应符合的数值范围内。

[0080] 在这里,“比较该指定的第一数据碎片所对应的待识别数据是否小于上限值”以及“比较下限值是否小于该指定的第一数据碎片所对应的待识别数据”可使用图5至图8所示的方法来执行比较操作,在此不再赘述。

[0081] 在框S212处,第一方确定是否每个待识别数据都在其应符合的数值范围内。如果任一待识别数据不在其应符合的数值范围内(在框S212处为“否”),则在框S214处第一方确定第二方的待识别数据集中存在非法数据。如果每个待识别数据都在其应符合的数值范围内(在框S212处为“是”),则在框S216处第一方确定第二方的待识别数据集中不存在非法数据。

[0082] 在图3至图8的流程图中动作编号的顺序不用于限定动作执行的先后顺序。除了具有输入输出关系(或者因果关系)的动作必须具有先后顺序之外,其他动作可以并行地执行,或者按照除图示之外的其他顺序来执行。

[0083] 图9示出根据本公开的实施例的第一方对第二方进行非法数据识别的装置900的示意性框图。装置900位于第一方处。如图9所示,该装置900可包括处理器910和存储有计算机程序的存储器920。当计算机程序由处理器910执行时,使得装置900可执行如图2所示的方法200的步骤。在一个示例中,装置900可以是计算机设备或云计算节点。装置900可获得第二方的第一数据碎片集。其中,第二方的待识别数据集中的每个待识别数据被碎片化成第一数据碎片和第二数据碎片。第一数据碎片集包括所有第一数据碎片。装置900可与第二方联合确定用于对第一数据碎片集进行随机采样的随机采样比例。随机采样比例由第一采样比例碎片和第二采样比例碎片之和来确定。第一方持有第一采样比例碎片。第二方持有

第二采样比例碎片。装置900可与第二方联合针对每个第一数据碎片确定一个参考随机数。参考随机数符合均匀分布。参考随机数由第一参考随机数碎片和第二参考随机数碎片之和来确定。第一方持有第一参考随机数碎片。第二方持有第二参考随机数碎片。装置900可利用第一采样比例碎片和第一参考随机数碎片与第二方联合从第一数据碎片集中采样出指定的第一数据碎片。装置900可利用每个指定的第一数据碎片与第二方联合比较该指定的第一数据碎片所对应的待识别数据与该待识别数据应符合的数值范围。每个待识别数据应符合的数值范围是根据指定的第一数据碎片的属性特征来确定的。响应于任一待识别数据不在其应符合的数值范围内,装置900可确定第二方的待识别数据集中存在非法数据。

[0084] 在本公开的实施例中,处理器910可以是例如中央处理单元(CPU)、微处理器、数字信号处理器(DSP)、基于多核的处理器架构的处理器等。存储器920可以是使用数据存储技术实现的任何类型的存储器,包括但不限于随机存取存储器、只读存储器、基于半导体的存储器、闪存、磁盘存储器等。

[0085] 此外,在本公开的实施例中,装置900也可包括输入设备930,例如键盘、鼠标等,用于输入开始执行方法200的指令。另外,装置900还可包括输出设备940,例如显示器等,用于输出识别结果。

[0086] 在本公开的其它实施例中,还提供了一种存储有计算机程序的计算机可读存储介质,其中,计算机程序在由处理器执行时能够实现如图2所示的方法的步骤。

[0087] 综上所述,根据本公开的实施例的第一方对第二方进行非法数据识别的方法和装置通过安全的随机采样方式从第二方的待识别数据中采样一部分数据,并通过多方安全计算的方式来对该部分数据进行非法数据识别的操作,能够以安全高效的方式进行非法数据识别。而待识别数据中的哪些数据被采样是通过以多方安全计算的方式计算随机采样比例和针对每个待识别数据的参考随机数,并根据随机采样比例和参考随机数来确定的。在整个识别过程中,随机采样比例、针对每个待识别数据的参考随机数、所采样的待识别数据都以碎片态的形式参与计算,因此不会泄露任一方的原始数据。采样一部分数据进行非法数据识别的方式能够以低代价的计算耗损,完成非法数据的探查识别,给联邦学习、联合统计等任务的发起方以可靠的数字依据。

[0088] 附图中的流程图和框图显示了根据本公开的多个实施例的装置和方法的可能实现的体系架构、功能和操作。在这点上,流程图或框图中的每个方框可以代表一个模块、程序段或指令的一部分,所述模块、程序段或指令的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。在有些作为替换的实现中,方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如,两个连续的方框实际上可以基本并行地执行,它们有时也可以按相反的顺序执行,这依所涉及的功能而定。也要注意的是,框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合,可以用执行规定的功能或动作的专用的基于硬件的系统来实现,或者可以用专用硬件与计算机指令的组合来实现。

[0089] 除非上下文中另外明确地指出,否则在本文和所附权利要求中所使用的词语的单数形式包括复数,反之亦然。因而,当提及单数时,通常包括相应术语的复数。相似地,措辞“包含”和“包括”将解释为包含在内而不是独占性地。同样地,术语“包括”和“或”应当解释为包括在内的,除非本文中明确禁止这样的解释。在本文中使用术语“示例”之处,特别是当其位于一组术语之后时,所述“示例”仅仅是示例性的和阐述性的,且不应当被认为是独占

性的或广泛性的。

[0090] 适应性的进一步的方面和范围从本文中提供的描述变得明显。应当理解,本申请的各个方面可以单独或者与一个或多个其它方面组合实施。还应当理解,本文中的描述和特定实施例旨在仅说明的目的并不旨在限制本申请的范围。

[0091] 以上对本公开的若干实施例进行了详细描述,但显然,本领域技术人员可以在不脱离本公开的精神和范围的情况下对本公开的实施例进行各种修改和变型。本公开的保护范围由所附的权利要求限定。

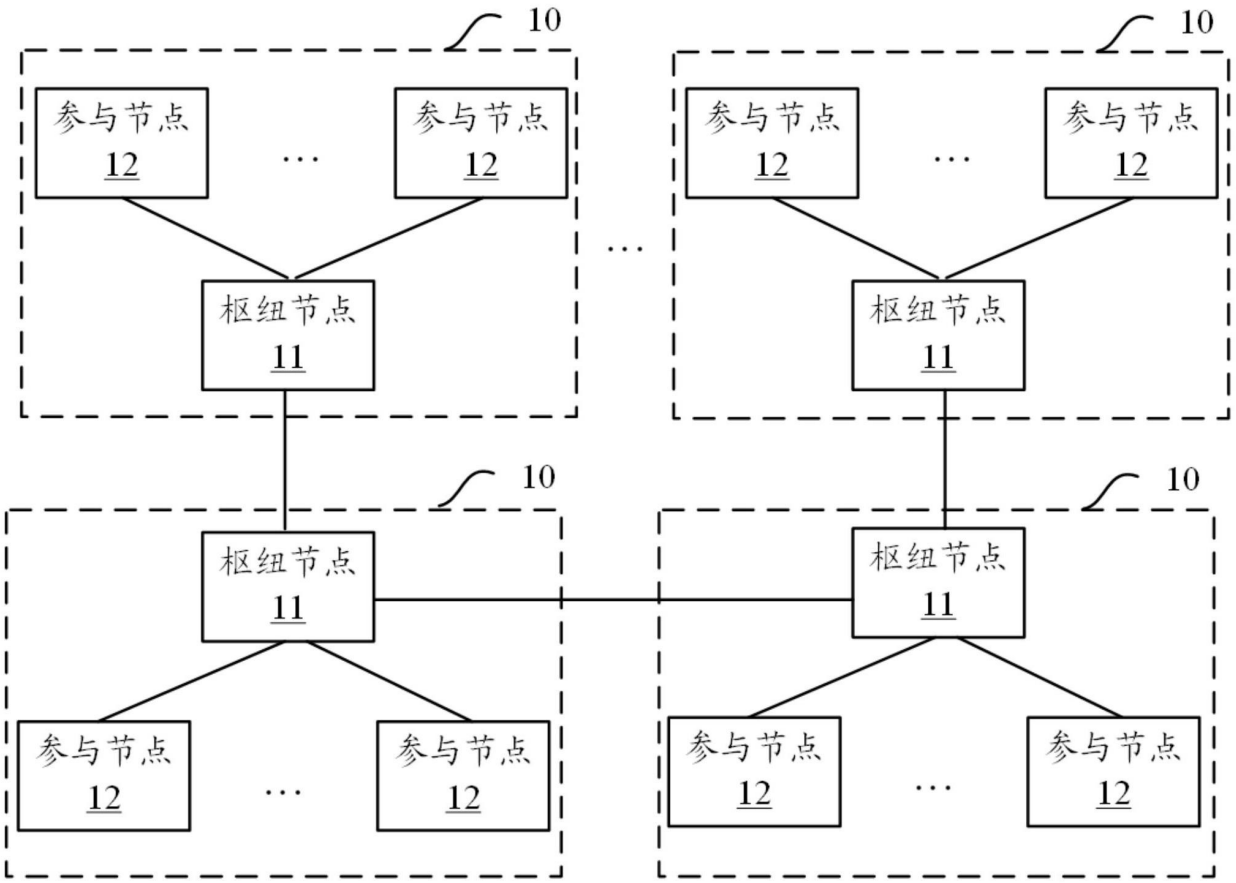

10
参与节点
12
参与节点
12
枢纽节点
11
10
参与节点
12
参与节点
12
枢纽节点
11
10
枢纽节点
11
参与节点
12
参与节点
12
10
枢纽节点
11
参与节点
12
参与节点
12

图1

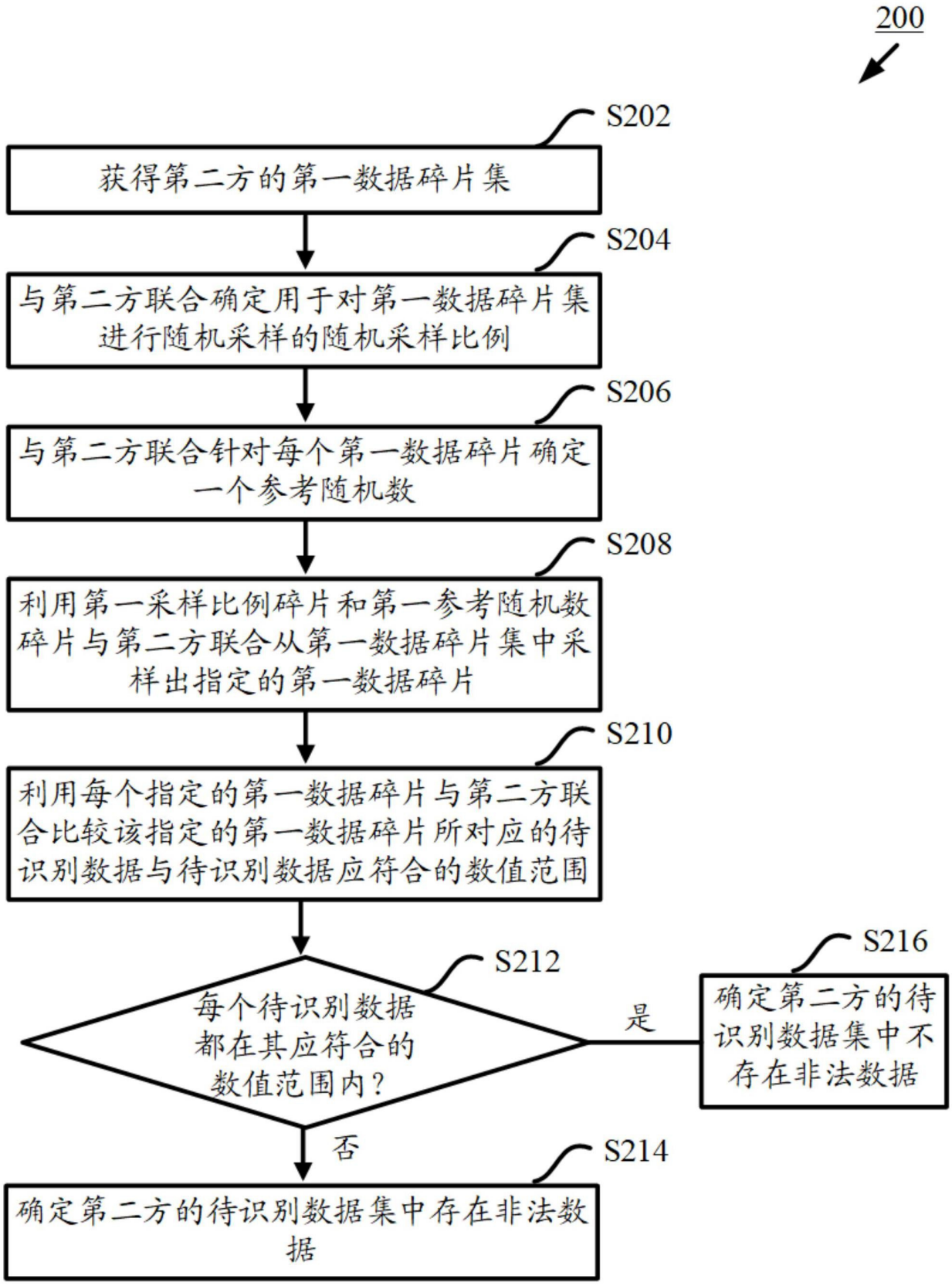

200
S202
获得第二方的第一数据碎片集
S204
与第二方联合确定用于对第一数据碎片集进行随机采样的随机采样比例
S206
与第二方联合针对每个第一数据碎片确定一个参考随机数
S208
利用第一采样比例碎片和第一参考随机数碎片与第二方联合从第一数据碎片集中采样出指定的第一数据碎片
S210
利用每个指定的第一数据碎片与第二方联合比较该指定的第一数据碎片所对应的待识别数据与待识别数据应符合的数值范围
S212
每个待识别数据都在其应符合的数值范围内？
是
S216
确定第二方的待识别数据集中不存在非法数据
否
S214
确定第二方的待识别数据集中存在非法数据

图2

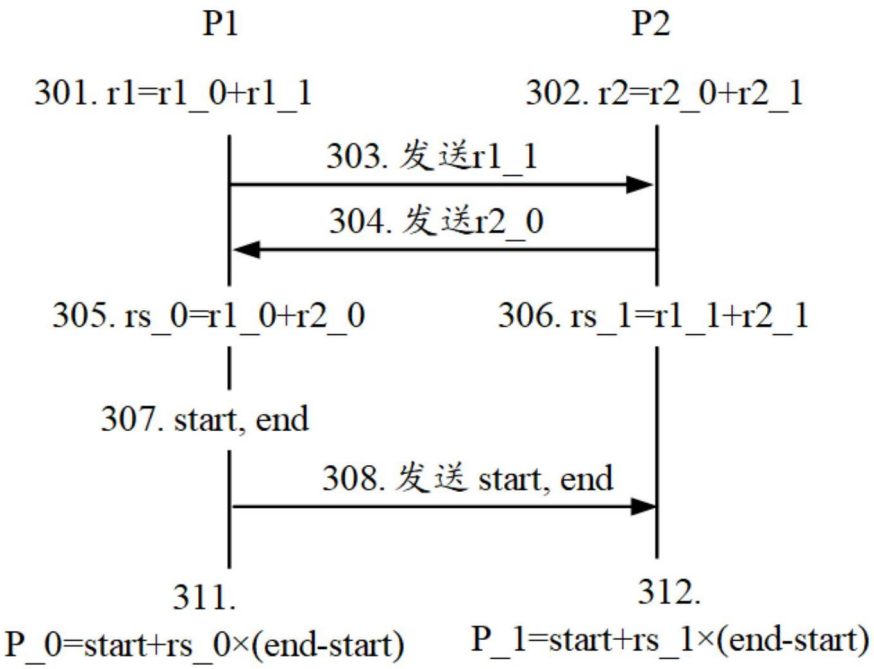

P1
P2
301. r1=r1_0+r1_1
302. r2=r2_0+r2_1
303. 发送r1_1
304. 发送r2_0
305. rs_0=r1_0+r2_0
306. rs_1=r1_1+r2_1
307. start, end
308. 发送 start, end
311.
P_0=start+rs_0×(end-start)
312.
P_1=start+rs_1×(end-start)

图3

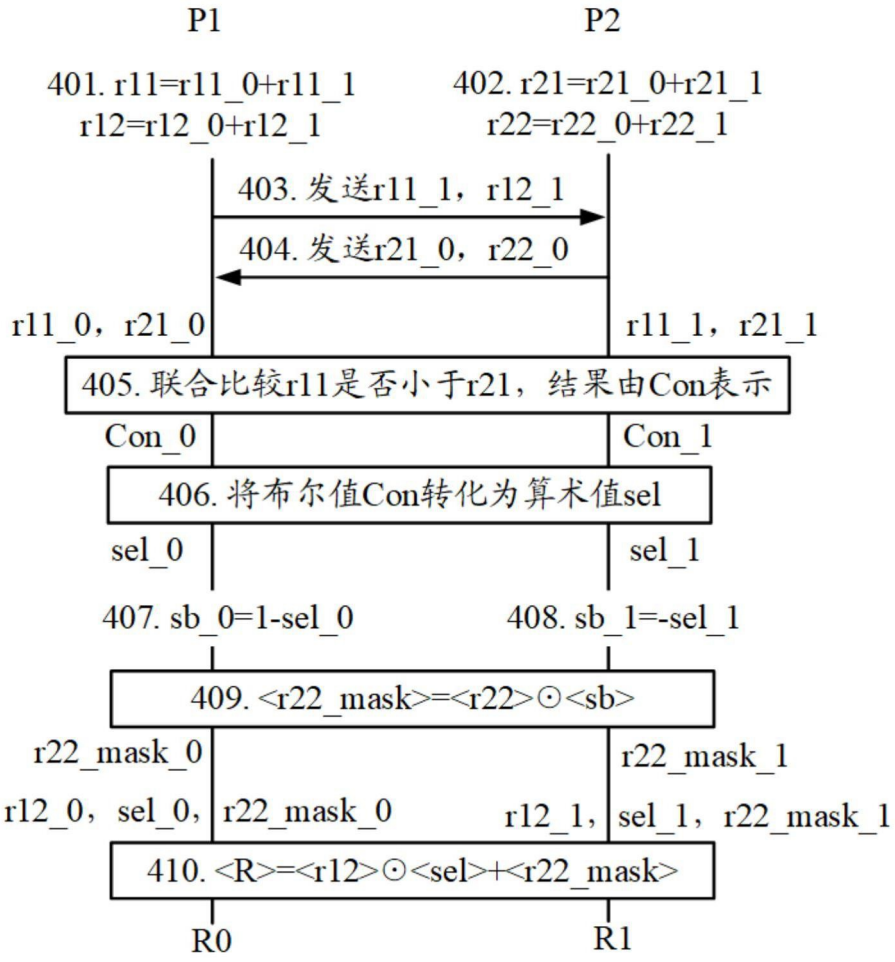

P1
P2
401. r11=r11_0+r11_1
r12=r12_0+r12_1
402. r21=r21_0+r21_1
r22=r22_0+r22_1
403. 发送r11_1，r12_1
404. 发送r21_0，r22_0
r11_0，r21_0
r11_1，r21_1
405. 联合比较r11是否小于r21，结果由Con表示
Con_0
Con_1
406. 将布尔值Con转化为算术值sel
sel_0
sel_1
407. sb_0=1-sel_0
408. sb_1=-sel_1
409. <r22_mask>=<r22>⊙<sb>
r22_mask_0
r22_mask_1
r12_0，sel_0，r22_mask_0
r12_1，sel_1，r22_mask_1
410. <R>=<r12>⊙<sel>+<r22_mask>
R0
R1

图4

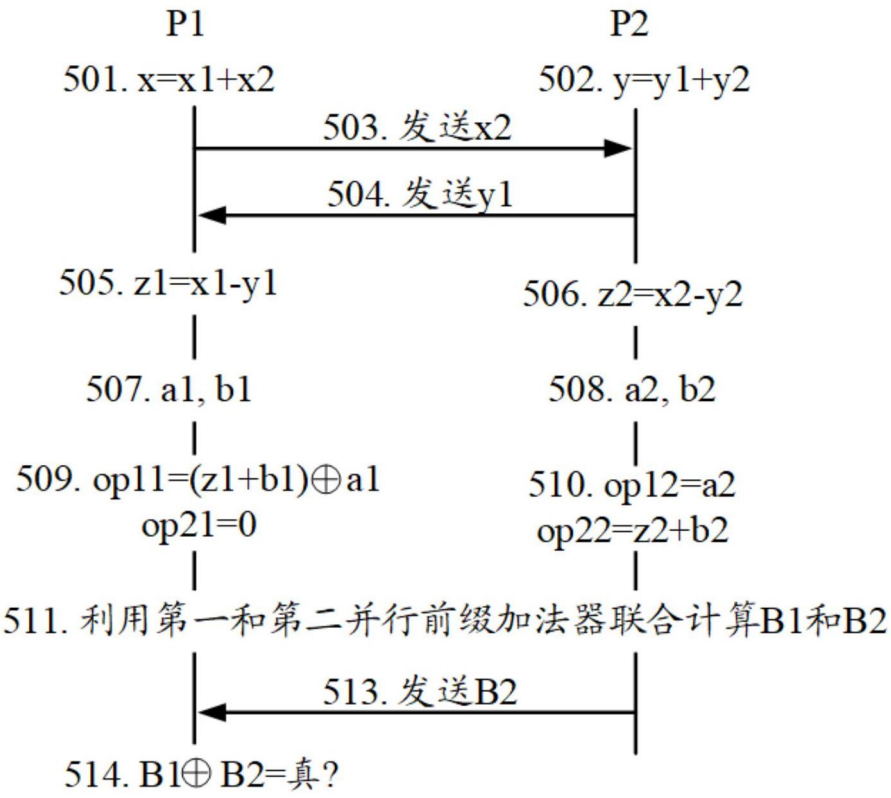

P1
P2
501. x=x1+x2
502. y=y1+y2
503. 发送x2
504. 发送y1
505. z1=x1-y1
506. z2=x2-y2
507. a1, b1
508. a2, b2
509. op11=(z1+b1)⊕a1
op21=0
510. op12=a2
op22=z2+b2
511. 利用第一和第二并行前缀加法器联合计算B1和B2
513. 发送B2
514. B1⊕B2=真?

图5

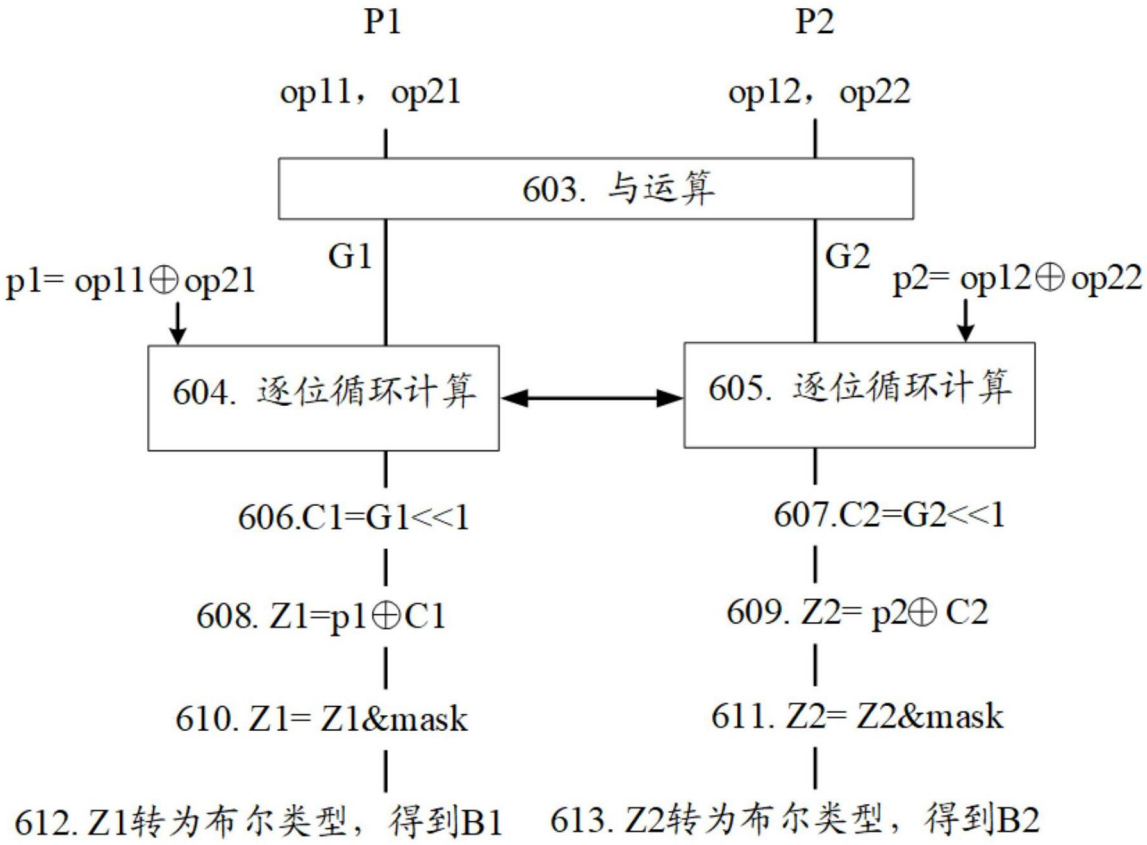

P1
P2
op11，op21
op12，op22
603. 与运算
G1
G2
p1= op11⊕op21
p2= op12⊕op22
604. 逐位循环计算
605. 逐位循环计算
606.C1=G1<<1
607.C2=G2<<1
608. Z1=p1⊕C1
609. Z2= p2⊕ C2
610. Z1= Z1&mask
611. Z2= Z2&mask
612. Z1转为布尔类型，得到B1
613. Z2转为布尔类型，得到B2

图6

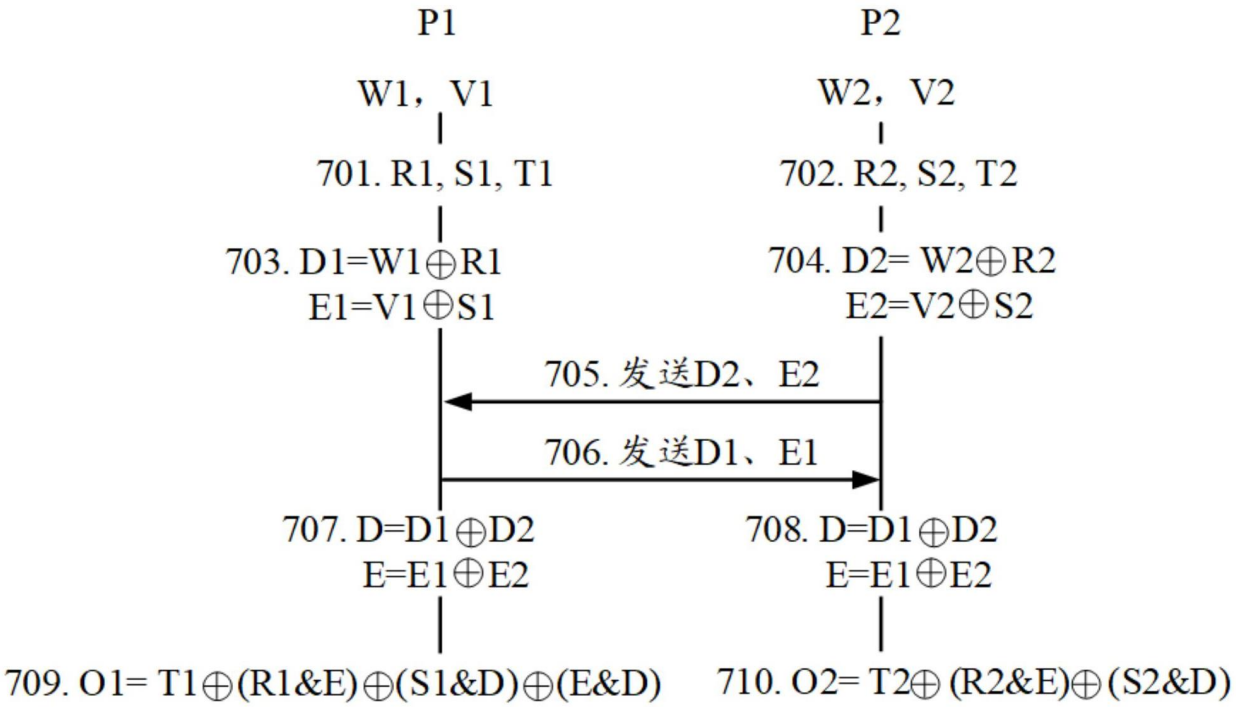

P1
P2
W1，V1
W2，V2
701. R1, S1, T1
702. R2, S2, T2
703. D1=W1⊕R1
E1=V1⊕S1
704. D2= W2⊕R2
E2=V2⊕S2
705. 发送D2、E2
706. 发送D1、E1
707. D=D1⊕D2
E=E1⊕E2
708. D=D1⊕D2
E=E1⊕E2
709. O1= T1⊕(R1&E)⊕(S1&D)⊕(E&D)
710. O2= T2⊕ (R2&E)⊕(S2&D)

图7

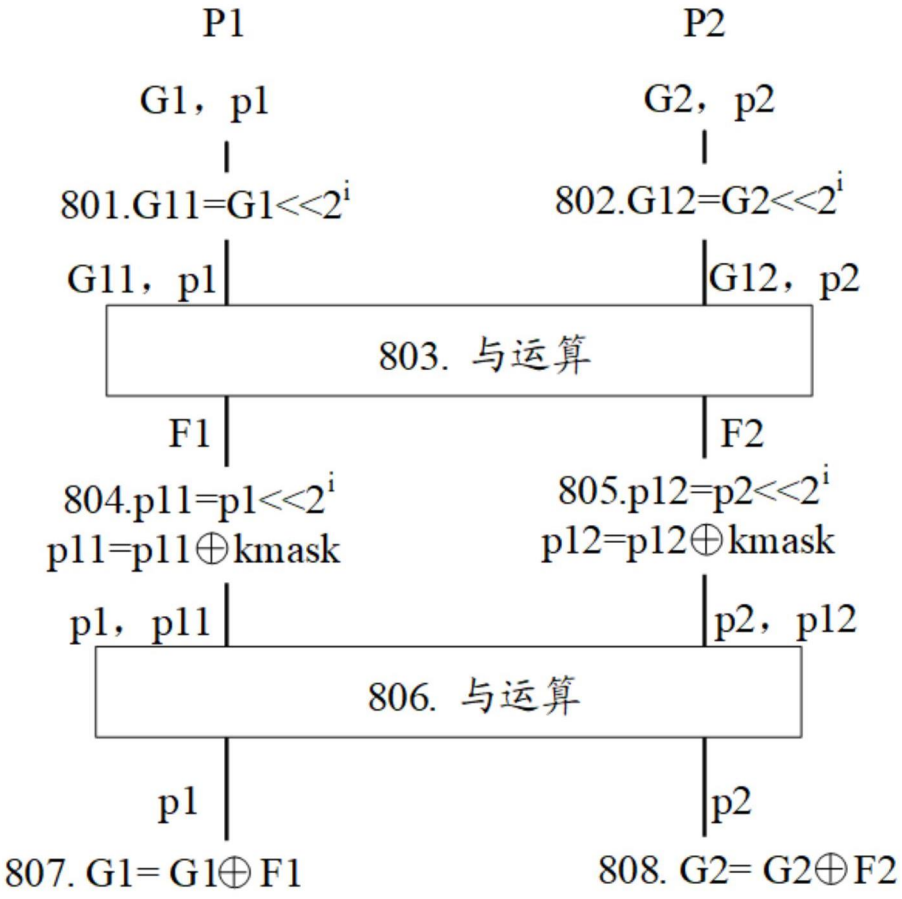

P1
P2
G1，p1
G2，p2
801.G11=G1<<2^i
802.G12=G2<<2^i
G11，p1
G12，p2
803. 与运算
F1
F2
804.p11=p1<<2^i
p11=p11⊕kmask
805.p12=p2<<2^i
p12=p12⊕kmask
p1，p11
p2，p12
806. 与运算
p1
p2
807. G1= G1⊕ F1
808. G2= G2⊕F2

图8

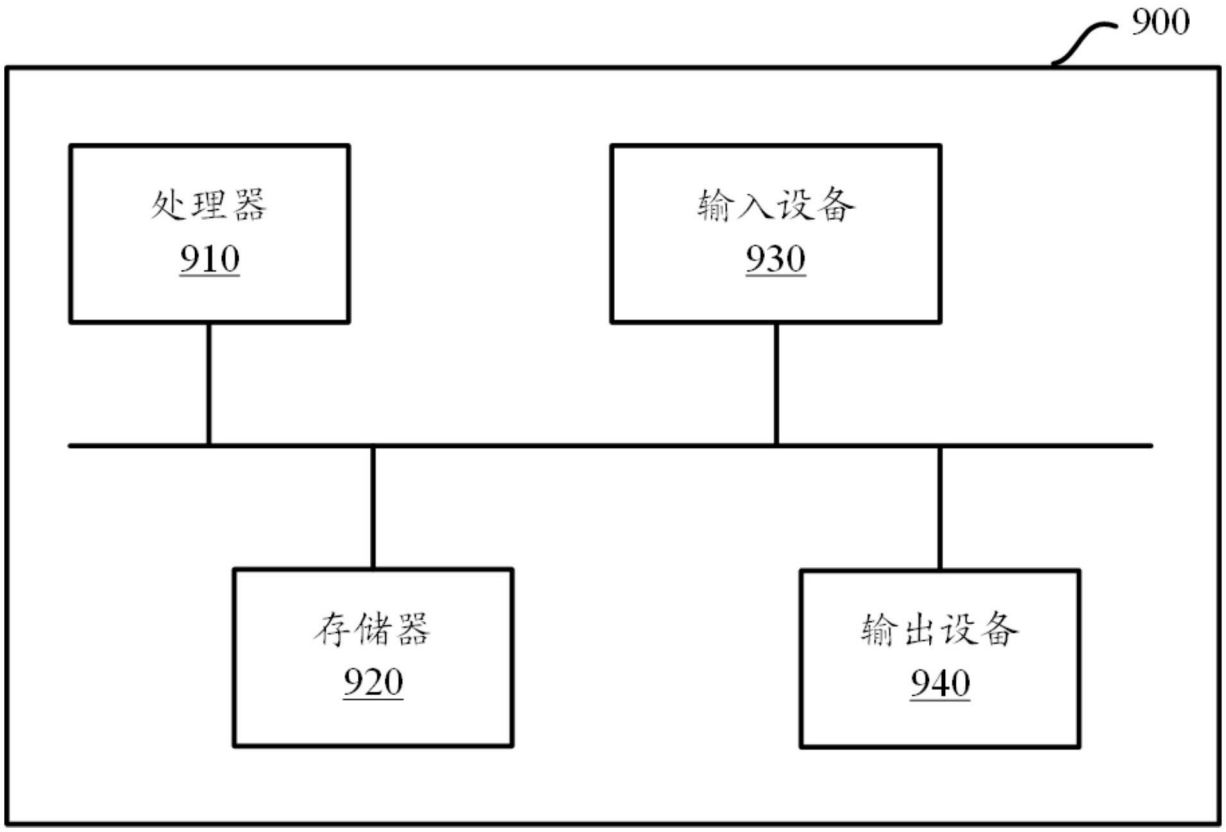


图9