(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 118590236 A
(43)申请公布日 2024.09.03

(21)申请号 202411073968.7

(22)申请日 2024.08.07

(71)申请人 富算科技(上海)有限公司
地址 200135 上海市浦东新区自由贸易试验区浦东大道1200号2层A区

(72)发明人 尤志强 赵东 陈立峰

(74)专利代理机构 上海弼兴律师事务所 31283
专利代理师 陈臻晔 罗朗

(51)Int.Cl.
H04L 9/08(2006.01)
H04L 9/00(2022.01)

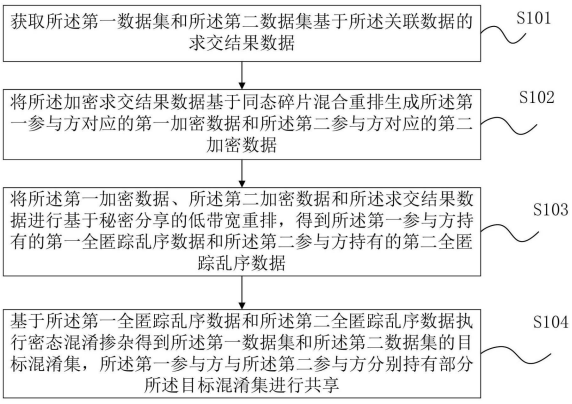
权利要求书3页 说明书13页 附图5页

(54)发明名称

全匿踪混淆求交数据的共享方法、系统、设备及介质

(57)摘要

本公开提供了全匿踪混淆求交数据的共享方法、系统、设备及介质。该共享方法包括:获取第一数据集和第二数据集的加密求交结果数据;基于同态碎片混合重排生成第一参与方对应的第一加密数据和第二参与方对应的第二加密数据;将第一加密数据、第二加密数据和加密求交结果数据进行基于秘密分享的低带宽重排,得到第一参与方持有的第一全匿踪乱序数据和第二参与方持有的第二全匿踪乱序数据;基于第一全匿踪乱序数据和第二全匿踪乱序数据执行密态混淆掺杂得到第一数据集和第二数据集的目标混淆集,第一参与方与第二参与方分别持有目标混淆集的其中一份碎片数据。基于全匿踪混淆求交处理,保护参与双方的交集和非交集不泄露,数据处理过程计算量级低。



CN 118590236 A

1.一种全匿踪混淆求交数据的共享方法,其特征在于,第一参与方持有第一数据集,第二参与方持有第二数据集,所述第一数据集与所述第二数据集均包括同一目标标识的关联数据,所述方法包括:

获取所述第一数据集和所述第二数据集基于所述关联数据的加密求交结果数据;

将所述加密求交结果数据基于同态碎片混合重排生成所述第一参与方对应的第一加密数据和所述第二参与方对应的第二加密数据;

将所述第一加密数据、所述第二加密数据和所述加密求交结果数据进行基于秘密分享的低带宽重排,得到所述第一参与方持有的第一全匿踪乱序数据和所述第二参与方持有的第二全匿踪乱序数据;

其中,所述基于秘密分享的低带宽重排与所述同态碎片混合重排的重排顺序一致;

基于所述第一全匿踪乱序数据和所述第二全匿踪乱序数据执行密态混淆掺杂得到所述第一数据集和所述第二数据集的目标混淆集,所述第一参与方与所述第二参与方分别持有部分所述目标混淆集进行共享。

2.根据权利要求1所述的全匿踪混淆求交数据的共享方法,其特征在于,所述基于同态碎片混合重排生成所述第一参与方对应的第一加密数据和所述第二参与方对应的第二加密数据的步骤包括:

随机生成所述第一参与方的第一混淆矩阵GB_g和所述第二参与方的第二混淆矩阵GB_h,所述第一混淆矩阵GB_g与所述第二混淆矩阵GB_h的大小一致,且混淆矩阵集合$GB = GB_g + GB_h$;

将所述第一混淆矩阵GB_g与所述第二混淆矩阵GB_h基于同态加密计算和乱序处理,得到所述第一参与方的第一秘密共享数据Z_g和所述第一参与方执行的第一乱序索引$shuffle_index_g$,以及所述第二参与方的第二秘密共享数据Z_h和所述第二参与方执行的第二乱序索引$shuffle_index_h$;

其中,所述第一秘密共享数据Z_g与所述第二秘密共享数据Z_h之和,等于所述混淆矩阵集合GB基于所述第一乱序索引$shuffle_index_g$和所述第二乱序索引$shuffle_index_h$进行乱序处理的结果;

将所述第一混淆矩阵GB_g、所述第一秘密共享数据Z_g和所述第一乱序索引$shuffle_index_g$作为所述第一加密数据,所述第二混淆矩阵GB_h、所述第二秘密共享数据Z_h和所述第二乱序索引$shuffle_index_h$作为所述第二加密数据。

3.根据权利要求1所述的全匿踪混淆求交数据的共享方法,其特征在于,所述将所述第一加密数据、所述第二加密数据和所述加密求交结果数据进行基于秘密分享的低带宽重排,得到所述第一参与方持有的第一全匿踪乱序数据和所述第二参与方持有的第二全匿踪乱序数据的步骤包括:

将所述加密求交结果数据分为所述第一参与方持有的第一交集数据和所述第二参与方持有的第二交集数据;

生成所述第一参与方持有的第三混淆矩阵和所述第二参与方持有的第四混淆矩阵,所述第三混淆矩阵和所述第四混淆矩阵的大小一致;

所述第一参与方基于持有的所述第一交集数据、所述第一加密数据和所述第三混淆矩

阵计算得到第一碎片数据,并将所述第一碎片数据发送至所述第二参与方;

所述第二参与方基于所述第一碎片数据、所述第四混淆矩阵和所述第二加密数据计算得到第二碎片数据,并将所述第二碎片数据发送至所述第一参与方;

所述第一参与方基于所第二碎片数据计算得到所述第一全匿踪乱序数据,所述第二参与方基于所述第四混淆矩阵和所述第二加密数据计算得到所述第二全匿踪乱序数据。

4.根据权利要求1所述的全匿踪混淆求交数据的共享方法,其特征在于,所述获取所述第一数据集和所述第二数据集基于所述关联数据的加密求交结果数据的步骤包括:

获取所述第一数据集和所述第二数据集;

其中,所述第一数据集包括第一样本特征,所述第二数据集包括第二样本特征;

将所述第一数据集和所述第二数据集进行全匿踪安全求交,得到指示结果数据和特征结果数据;

所述指示结果数据表征对应数据是否属于所述关联数据,所述特征结果数据为所述第一样本特征与所述第二样本特征的对齐结果;

将所述指示结果数据和所述特征结果数据拼接得到所述加密求交结果数据。

5.根据权利要求4所述的全匿踪混淆求交数据的共享方法,其特征在于,所述将所述指示结果数据和所述特征结果数据拼接得到所述加密求交结果数据的步骤包括:

将所述特征结果数据分为所述第一参与方持有的第一特征结果数据和所述第二参与方持有的第二特征结果数据;

其中,所述第一特征结果数据为$M \times P$大小的矩阵,所述第二特征结果数据为$M \times Q$大小的矩阵;

将所述第一特征结果数据、所述第二特征结果数据和所述指示结果数据横向拼接得到大小为$M \times N$的矩阵作为所述加密求交结果数据;

其中,$N=P+Q+1$,M、N、P和Q是正整数。

6.根据权利要求4所述的全匿踪混淆求交数据的共享方法,其特征在于,所述目标混淆集包括所述指示结果数据,所述共享方法还包括:

将所述目标混淆集中的所述指示结果数据恢复得到所述目标混淆集中数据对应的指示结果;

所述第一参与方基于所述指示结果对持有的部分所述目标混淆集进行筛选,得到所述第一数据集中属于所述关联数据的第一目标样本;

获取所述第一目标样本对应的目标索引,并将所述目标索引发送至所述第二参与方;

所述第二参与方基于所述目标索引对持有的另一部分所述目标混淆集进行筛选,得到所述第二数据集中与所述第一目标样本对应的第二目标样本。

7.根据权利要求6所述的全匿踪混淆求交数据的共享方法,其特征在于,所述第一数据集为用户消费数据,所述第二数据集为用户财产数据,所述目标标识为用户身份信息,所述共享方法还包括:

基于预设筛选信息对所述第一目标样本和所述第二目标样本筛选,得到所述用户身份信息对应的目标消费数据和目标财产数据。

8.一种全匿踪混淆求交数据的共享系统,其特征在于,第一参与方持有第一数据集,第二参与方持有第二数据集,所述第一数据集与所述第二数据集均包括同一目标标识的关联

数据,所述共享系统包括数据求交模块、同态重排模块、全匿踪乱序模块和密态混淆掺杂模块;

所述数据求交模块,用于获取所述第一数据集和所述第二数据集基于所述关联数据的加密求交结果数据;

所述同态重排模块,用于将所述加密求交结果数据基于同态碎片混合重排生成所述第一参与方对应的第一加密数据和所述第二参与方对应的第二加密数据;

所述全匿踪乱序模块,用于将所述第一加密数据、所述第二加密数据和所述加密求交结果数据进行基于秘密分享的低带宽重排,得到所述第一参与方持有的第一全匿踪乱序数据和所述第二参与方持有的第二全匿踪乱序数据;其中,所述基于秘密分享的低带宽重排与所述同态碎片混合重排的重排顺序一致;

所述密态混淆掺杂模块,用于基于所述第一全匿踪乱序数据和所述第二全匿踪乱序数据执行密态混淆掺杂得到所述第一数据集和所述第二数据集的目标混淆集,所述第一参与方与所述第二参与方分别持有部分所述目标混淆集进行共享。

9.一种电子设备,包括存储器、处理器及存储在存储器上并用于在处理器上运行的计算机程序,其特征在于,所述处理器执行所述计算机程序时实现权利要求1至7中任一项所述的全匿踪混淆求交数据的共享方法。

10.一种计算机可读存储介质,其上存储有计算机程序,其特征在于,所述计算机程序被处理器执行时实现权利要求1至7中任一项所述的全匿踪混淆求交数据的共享方法。

11.一种计算机程序产品,包括计算机程序,其特征在于,所述计算机程序被处理器执行时实现如权利要求1-7中任一项所述的全匿踪混淆求交数据的共享方法。

全匿踪混淆求交数据的共享方法、系统、设备及介质

技术领域

[0001] 本公开涉及数据处理技术领域,尤其是涉及全匿踪混淆求交数据的共享方法、系统、设备及介质。

背景技术

[0002] 流量是商业运营的基础之一,大量的流量可以转化为广告收入、产品销售额等商业价值,因此流量对于机构发展非常重要。为了获客,机构愿意花费很大的代价进行广告投放,但收效一般。银行、保险、证券之间尝试以基金或者保险代销、资金存管等业务进行一定程度的互相合作,但是由于机构之间的数据割裂,难以形成符合期望的规模。

[0003] 现有解决数据割裂的数据处理手段,主要用于保护参与双方的非交集ID不泄露,对交集部分的ID默认双方知晓,但这样的数据处理方法无法应对差分攻击或者信息反推等数据攻击,容易暴露出个体的私有信息。另一方面,现有的数据处理手段要同时保护参与双方的非交集ID和交集ID不泄露,所需的计算量庞大,计算收益低。

发明内容

[0004] 本公开要解决的技术问题是为了克服现有技术中无法同时保护参与双方的非交集ID和交集ID不泄露,和应对差分攻击或者信息反推的缺陷,提供一种全匿踪混淆求交数据的共享方法、系统、设备、介质及计算机程序产品。

[0005] 本公开是通过下述技术方案来解决上述技术问题:

[0006] 第一方面,提供一种全匿踪混淆求交数据的共享方法,第一参与方持有第一数据集,第二参与方持有第二数据集,所述第一数据集与所述第二数据集均包括同一目标标识的关联数据,所述方法包括:

[0007] 获取所述第一数据集和所述第二数据集基于所述关联数据的加密求交结果数据;

[0008] 将所述加密求交结果数据基于同态碎片混合重排生成所述第一参与方对应的第一加密数据和所述第二参与方对应的第二加密数据;

[0009] 将所述第一加密数据、所述第二加密数据和所述加密求交结果数据进行基于秘密分享的低带宽重排,得到所述第一参与方持有的第一全匿踪乱序数据和所述第二参与方持有的第二全匿踪乱序数据;

[0010] 其中,所述基于秘密分享的低带宽重排与所述同态碎片混合重排的重排顺序一致;

[0011] 基于所述第一全匿踪乱序数据和所述第二全匿踪乱序数据执行密态混淆掺杂得到所述第一数据集和所述第二数据集的目标混淆集,所述第一参与方与所述第二参与方分别持有部分所述目标混淆集进行共享。

[0012] 较佳地,所述基于同态碎片混合重排生成所述第一参与方对应的第一加密数据和所述第二参与方对应的第二加密数据的步骤包括:

[0013] 随机生成所述第一参与方的第一混淆矩阵GB_g和所述第二参与方的第二混淆矩阵

GB_h，所述第一混淆矩阵GB_g与所述第二混淆矩阵GB_h的大小一致，且混淆矩阵集合$GB = GB_g + GB_h$；

[0014] 将所述第一混淆矩阵GB_g与所述第二混淆矩阵GB_h基于同态加密计算和乱序处理，得到所述第一参与方的第一秘密共享数据Z_g和所述第一参与方执行的第一乱序索引$shuffle_index_g$，以及所述第二参与方的第二秘密共享数据Z_h和所述第二参与方执行的第二乱序索引$shuffle_index_h$；

[0015] 其中，所述第一秘密共享数据Z_g与所述第二秘密共享数据Z_h之和，等于所述混淆矩阵集合GB基于所述第一乱序索引$shuffle_index_g$和所述第二乱序索引$shuffle_index_h$进行乱序处理的结果；

[0016] 将所述第一混淆矩阵GB_g、所述第一秘密共享数据Z_g和所述第一乱序索引$shuffle_index_g$作为所述第一加密数据，所述第二混淆矩阵GB_h、所述第二秘密共享数据Z_h和所述第二乱序索引$shuffle_index_h$作为所述第二加密数据。

[0017] 较佳地，所述将所述第一加密数据、所述第二加密数据和所述加密求交结果数据进行基于秘密分享的低带宽重排，得到所述第一参与方持有的第一全匿踪乱序数据和所述第二参与方持有的第二全匿踪乱序数据的步骤包括：

[0018] 将所述加密求交结果数据分为所述第一参与方持有的第一交集数据和所述第二参与方持有的第二交集数据；

[0019] 生成所述第一参与方持有的第三混淆矩阵和所述第二参与方持有的第四混淆矩阵，所述第三混淆矩阵和所述第四混淆矩阵的大小一致；

[0020] 所述第一参与方基于持有的所述第一交集数据、所述第一加密数据和所述第三混淆矩阵计算得到第一碎片数据，并将所述第一碎片数据发送至所述第二参与方；

[0021] 所述第二参与方基于所述第一碎片数据、所述第四混淆矩阵和所述第二加密数据计算得到第二碎片数据，并将所述第二碎片数据发送至所述第一参与方；

[0022] 所述第一参与方基于所第二碎片数据计算得到所述第一全匿踪乱序数据，所述第二参与方基于所述第四混淆矩阵和所述第二加密数据计算得到所述第二全匿踪乱序数据。

[0023] 较佳地，所述获取所述第一数据集和所述第二数据集基于所述关联数据的加密求交结果数据的步骤包括：

[0024] 获取所述第一数据集和所述第二数据集；

[0025] 其中，所述第一数据集包括第一样本特征，所述第二数据集包括第二样本特征；

[0026] 将所述第一数据集和所述第二数据集进行全匿踪安全求交，得到指示结果数据和特征结果数据；

[0027] 所述指示结果数据表征对应数据是否属于所述关联数据，所述特征结果数据为所述第一样本特征与所述第二样本特征的对齐结果；

[0028] 将所述指示结果数据和所述特征结果数据拼接得到所述加密求交结果数据。

[0029] 较佳地，所述将所述指示结果数据和所述特征结果数据拼接得到所述加密求交结果数据的步骤包括：

[0030] 将所述特征结果数据分为所述第一参与方持有的第一特征结果数据和所述第二参与方持有的第二特征结果数据；

[0031] 其中,所述第一特征结果数据为M×P大小的矩阵,所述第二特征结果数据为M×Q大小的矩阵;

[0032] 将所述第一特征结果数据、所述第二特征结果数据和所述指示结果数据横向拼接得到大小为M×N的矩阵作为所述加密求交结果数据;

[0033] 其中,N=P+Q+1,M、N、P和Q是正整数。

[0034] 较佳地,所述目标混淆集包括所述指示结果数据,所述共享方法还包括:

[0035] 将所述目标混淆集中的所述指示结果数据恢复得到所述目标混淆集中数据对应的指示结果;

[0036] 所述第一参与方基于所述指示结果对持有的部分所述目标混淆集进行筛选,得到所述第一数据集中属于所述关联数据的第一目标样本;

[0037] 获取所述第一目标样本对应的目标索引,并将所述目标索引发送至所述第二参与方;

[0038] 所述第二参与方基于所述目标索引对持有的另一部分所述目标混淆集进行筛选,得到述第二数据集中与所述第一目标样本对应的第二目标样本。

[0039] 较佳地,所述第一数据集为用户消费数据,所述第二数据集为用户财产数据,所述目标标识为用户身份信息,所述共享方法还包括:

[0040] 基于预设筛选信息对所述第一目标样本和所述第二目标样本筛选,得到所述用户身份信息对应的目标消费数据和目标财产数据。

[0041] 第二方面,提供一种全匿踪混淆求交数据的共享系统,第一参与方持有第一数据集,第二参与方持有第二数据集,所述第一数据集与所述第二数据集均包括同一目标标识的关联数据,所述共享系统包括数据求交模块、同态重排模块、全匿踪乱序模块和密态混淆掺杂模块;

[0042] 所述数据求交模块,用于获取所述第一数据集和所述第二数据集基于所述关联数据的加密求交结果数据;

[0043] 所述同态重排模块,用于将所述加密求交结果数据基于同态碎片混合重排生成所述第一参与方对应的第一加密数据和所述第二参与方对应的第二加密数据;

[0044] 所述全匿踪乱序模块,用于将所述第一加密数据、所述第二加密数据和所述加密求交结果数据进行基于秘密分享的低带宽重排,得到所述第一参与方持有的第一全匿踪乱序数据和所述第二参与方持有的第二全匿踪乱序数据;其中,所述基于秘密分享的低带宽重排与所述同态碎片混合重排的重排顺序一致;

[0045] 所述密态混淆掺杂模块,用于基于所述第一全匿踪乱序数据和所述第二全匿踪乱序数据执行密态混淆掺杂得到所述第一数据集和所述第二数据集的目标混淆集,所述第一参与方与所述第二参与方分别持有部分所述目标混淆集进行共享。

[0046] 第三方面,提供一种电子设备,包括存储器、处理器及存储在存储器上并用于在处理器上运行的计算机程序,所述处理器执行所述计算机程序时实现第一方面所述的全匿踪混淆求交数据的共享方法。

[0047] 第四方面,提供一种计算机可读存储介质,其上存储有计算机程序,所述计算机程序被处理器执行时实现第一方面所述的全匿踪混淆求交数据的共享方法。

[0048] 第五方面,提供一种计算机程序产品,包括计算机程序,所述计算机程序被处理器

执行时实现如第一方面所述的全匿踪混淆求交数据的共享方法。

[0049] 在符合本领域常识的基础上,上述各优选条件,可任意组合,即得本公开各较佳实例。

[0050] 本公开的积极进步效果在于:基于全匿踪混淆求交的数据处理流程,达到保护参与双方的交集和非交集不泄露,防止数据被差分攻击破解,且数据处理中无需个体信息使用授权,数据处理过程计算量级低。

附图说明

[0051] 图1为本公开一示例性实施例提供的一种全匿踪混淆求交数据的共享方法的第一流程图;

[0052] 图2为本公开一示例性实施例提供的一种全匿踪混淆求交数据的共享方法中步骤S101的流程图;

[0053] 图3为本公开一示例性实施例提供的一种全匿踪混淆求交数据的共享方法中步骤S103的流程图;

[0054] 图4为本公开一示例性实施例提供的一种全匿踪混淆求交数据的共享方法的第二流程图;

[0055] 图5为本公开一示例性实施例提供的一种全匿踪混淆求交数据的共享方法中交集数据示意图;

[0056] 图6为本公开一示例性实施例提供的一种全匿踪混淆求交数据的共享系统的模块示意图;

[0057] 图7为本公开一示例性实施例提供的电子设备的硬件结构示意图。

具体实施方式

[0058] 下面通过实施例的方式进一步说明本公开,但并不因此将本公开限制在所述的实施例范围之中。

[0059] 本公开实施例中采用诸如“第一”、“第二”的前缀词,仅仅为了区分不同的描述对象,对被描述对象的位置、顺序、优先级、数量或内容等没有限定作用。本公开实施例中对序数词等用于区分描述对象的前缀词的使用不对所描述对象构成限制,对所描述对象的陈述参见权利要求或实施例中上下文的描述,不应因为使用这种前缀词而构成多余的限制。此外,在本实施例的描述中,除非另有说明,“多个”的含义是两个或两个以上。

[0060] 本公开的技术方案中,所涉及的用户个人信息的收集、存储、使用、加工、传输、提供和公开等处理,均符合相关法律法规的规定,且不违背公序良俗。

[0061] 本公开实施例中,同态碎片混合重排(Shuffling with Homomorphic Encryption and Secret Sharing),是一种利用同态加密技术对秘密分享数据碎片进行加密处理,并在加密状态下对这些碎片进行打乱顺序的数据保护技术。

[0062] 基于秘密分享的低带宽重排((Low Bandwidth-friendly Shuffling Based on Secret Sharing),是一种通过秘密分享技术,将数据切分成两份碎片并分发给各个参与方,然后利用辅助重排因子信息对这些数据碎片进行打乱顺序的数据保护技术。

[0063] 实施例1

[0064] 本实施例,如图1所示,提供一种全匿踪混淆求交数据的共享方法,第一参与方持有第一数据集,第二参与方持有第二数据集,所述第一数据集与所述第二数据集均包括同一目标标识的关联数据,所述方法包括:

[0065] S101、获取所述第一数据集和所述第二数据集基于所述关联数据的求加密交结果数据;

[0066] S102、将所述加密求交结果数据基于同态碎片混合重排生成所述第一参与方对应的第一加密数据和所述第二参与方对应的第二加密数据;

[0067] S103、将所述第一加密数据、所述第二加密数据和所述加密求交结果数据进行基于秘密分享的低带宽重排,得到所述第一参与方持有的第一全匿踪乱序数据和所述第二参与方持有的第二全匿踪乱序数据;

[0068] 其中,所述基于秘密分享的低带宽重排与所述同态碎片混合重排的重排顺序一致;

[0069] S104、基于所述第一全匿踪乱序数据和所述第二全匿踪乱序数据执行密态混淆掺杂得到所述第一数据集和所述第二数据集的目标混淆集,所述第一参与方与所述第二参与方分别持有部分所述目标混淆集进行共享。

[0070] 作为一种可实现的方式,如图2所示,步骤S101包括:

[0071] S1011、获取所述第一数据集和所述第二数据集;

[0072] 其中,所述第一数据集包括第一样本特征,所述第二数据集包括第二样本特征;

[0073] S1012、将所述第一数据集和所述第二数据集进行全匿踪安全求交,得到指示结果数据和特征结果数据;

[0074] S1013、所述指示结果数据表征对应数据是否属于所述关联数据,所述特征结果数据为所述第一样本特征与所述第二样本特征的对齐结果;

[0075] S1014、将所述指示结果数据和所述特征结果数据拼接得到所述加密求交结果数据。

[0076] 作为一种可实现的方式,步骤包S1014包括:

[0077] 将所述特征结果数据分为所述第一参与方持有的第一特征结果数据和所述第二参与方持有的第二特征结果数据;

[0078] 其中,所述第一特征结果数据为$M \times P$大小的矩阵,所述第二特征结果数据为$M \times Q$大小的矩阵;

[0079] 将所述第一特征结果数据、所述第二特征结果数据和所述指示结果数据横向拼接得到大小为$M \times N$的矩阵作为所述加密求交结果数据;

[0080] 其中,$N=P+Q+1$,M、N、P和Q是正整数。

[0081] 在本方案中,全匿踪指数据处理过程中不暴露包括交集样本和交集量级的交集信息,也不暴露非交集信息,同时无法再数据中定位个人粒度的信息。

[0082] 在步骤S101中,通过将第一数据集和第二数据集进行基于电路的安全求交(circuit-based private set intersection,Circuit-PSI),用于计算第一参与方与第二参与方所述持有的第一数据集和第二数据集中与同一目标标识相关联的关联数据。其中,第一数据集和第二数据集中可以包括电话号码、身份证号、银行卡号等唯一性标识作为目标标识,第一数据集和第二数据集中分别包括对应的第一特征结果数据和第二特征结果数

据,比如年龄、性别、存款、消费行为等。安全求交输出的结果包含两部分,第一部分为表征该集合是否为交集的指示结果数据<B>,该指示结果数据为碎片态数据,对应的明文结果为1或者0,其中1表示交集,0表示非交集。另一部分为第一数据集和第二数据集特征对齐之后的特征结果数据<Fx>和<Fy>。在计算输出的时候会将<B>、<Fx>、<Fy>三种数据进行分布式横向拼接,得到特征大宽表数据<FF>,其中这三种数据均为碎片态数据。第一参与方持有第一交集数据FF0,第二参与方持有第二交集数据FF1,由于第一参与方与第二参与方是对等关系,也可以第一参与方持有第二交集数据FF1,第二参与方持有第一交集数据FF0,FF1+FF0=<FF>。<FF>的样本量级被缩小至交集大小*(1+P)的样本量,为后续的群体特征计算提供了客观的计算收益。

[0083] 作为一种可实现的方式,步骤S102包括:

[0084] 随机生成所述第一参与方的第一混淆矩阵GB_g和所述第二参与方的第二混淆矩阵GB_h,所述第一混淆矩阵GB_g与所述第二混淆矩阵GB_h的大小一致,且混淆矩阵集合$GB = GB_g + GB_h$;

[0085] 将所述第一混淆矩阵GB_g与所述第二混淆矩阵GB_h基于同态加密计算和乱序处理,得到所述第一参与方的第一秘密共享数据Z_g和所述第一参与方执行的第一乱序索引$shuffle_index_g$,以及所述第二参与方的第二秘密共享数据Z_h和所述第二参与方执行的第二乱序索引$shuffle_index_h$;

[0086] 其中,所述第一秘密共享数据Z_g与所述第二秘密共享数据Z_h之和,等于所述混淆矩阵集合GB基于所述第一乱序索引$shuffle_index_g$和所述第二乱序索引$shuffle_index_h$进行乱序处理的结果;

[0087] 将所述第一混淆矩阵GB_g、所述第一秘密共享数据Z_g和所述第一乱序索引$shuffle_index_g$作为所述第一加密数据,所述第二混淆矩阵GB_h、所述第二秘密共享数据Z_h和所述第二乱序索引$shuffle_index_h$作为所述第二加密数据。

[0088] 在本方案中,以第一参与方作为Receiver方,第二参与方为Sender方,随机生成(m*1)的第一混淆矩阵GB_g和第二混淆矩阵GB_h,即混淆矩阵$GB = GB_g + GB_h$,且各自生成对应的第一参与方持有的同态第一加密公钥PK_g、第一加密私钥SK_g,第二参与方持有的同态第二加密公钥PK_h、第二加密私钥SK_h。

[0089] 具体的,进行两轮乱序(Shuffle)流程。

[0090] 第一轮乱序重排流程中的第一步,首先将秘密共享碎片转化为同态加密,即Receiver方使用第一加密公钥PK_g加密GB_g得到第一加密混淆矩阵$[GB_g]$,并发送给sender方;

[0091] 第二步中,Sender方计算加密混淆矩阵$[GB] = [GB_g] + GB_h$,第二参与方在本地对加密混淆矩阵$[GB]$执行乱序shuffle得到乱序加密混淆矩阵$[GB]*$,并保存打乱的第二乱序索引$shuffle_index_h$,同时生成第二随机矩阵R_h,执行第一加密混淆矩阵$[GBR] = [GB]* + R_h$,并发送给Receiver方;

[0092] 第三步中,Receiver方使用第一加密私钥SK_g解密第一加密混淆矩阵$[GBR]$得到第一混淆矩阵GBR,同时生成第一随机矩阵R_g,执行重排混淆矩阵$GBR_hR_g = GBR - R_g$,并发送

给Sender方;并执行第一临时数据$PQ_g = R_g$;

[0093] 第四步中:Sender方接收重排混淆矩阵GBR_hR_g,去除噪声R_h,执行第二临时数据$PQ_h = GBR_hR_g - R_h$, $PQ = PQ_g + PQ_h = GB*$,得到一次乱序混淆矩阵$GB*$。

[0094] 第二轮乱序重排流程中的第一步,Sender方使用第二加密公钥PK_h加密第二临时数据PQ_h得到加密临时数据$[PQ_h]$,并发送给Receiver方;

[0095] 第二步中,Receiver方计算加密临时数据$[PQ] = [PQ_h] + PQ_g$,在第一参与方的本地对加密临时数据$[PQ]$执行乱序shuffle得到乱序加密临时数据$[PQ] *= [GB] **$, $[GB] **$为二次乱序加密混淆矩阵,并保存打乱的第一乱序索引$shuffle_index_g$,同时生成第三随机矩阵R_g',执行第三加密临时数据$[PQR] = [PQ] * + R_g'$,并发送给Sender方;

[0096] 第三步中,Sender方使用第二加密私钥SK_h解密第一加密临时数据$[PQR]$得到第三临时数据PQR,同时生成第四随机矩阵R_h',执行重排混淆临时数据$PQR_gR_h = PQR - R_h'$,并发送给Receiver方;并执行第二输出数据$Z_h = R_h'$;

[0097] 第四步中,Receiver方接收重排混淆临时数据PQR_gR_h,去除噪声R_g',执行第一输出数据$Z_g = PQR_gR_h - R_g'$;得到输出数据$Z = Z_g + Z_h = PQ *= GB **$, $GB **$为二次乱序混淆矩阵。

[0098] 以一个具体的模拟示例来解释:

[0099] GB_g:[0, 88, 96, -17, 53]、Z_g:[31, -67, -4, 41, 17]、$shuffle\_index_g$:[4, 0, 2, 1, 3];

[0100] GB_h:[-42, -79, 66, -89, -22]、Z_h:[-137, 229, 35, -32, -59]、$shuffle\_index_h$:[4, 2, 0, 3, 1];

[0101] $GB = GB_g + GB_h$:[-42, 9, 162, -106, 31];

[0102] $Z = Z_g + Z_h$:[-106, 162, 31, 9, -42];

[0103] $Z = GB[shuffle\_index_h][shuffle\_index_g] = GB **$;

[0104] GB经过第一轮shuffle得到$GB *$[31, -42, 162, 9, -106],经过第二轮shuffle得到$GB **$[-106, 162, 31, 9, -42];

[0105] 即$Z = GB **$。

[0106] 在本方案中,通过步骤S102对秘密共享分片数据,结合半同态加密算法,相对于纯粹使用多方安全计算技术,可以极大降低计算的通信次数,并且只针对单列的混淆矩阵进行密文加法计算,其通信量以及计算耗时都较低。不依赖于具体的实际业务数据,可以在线下提前做好计算,在计算实际业务数据时直接使用,节省这一步骤的在线计算耗时,通过这种调度策略的优化,更进一步提升本乱序方案的执行效率。

[0107] 作为一种可实现的方式,如图3所示,步骤S103包括:

[0108] S1031、将所述加密求交结果数据分为所述第一参与方持有的第一交集数据FF0和所述第二参与方持有的第二交集数据FF1;

[0109] S1032、生成所述第一参与方持有的第三混淆矩阵T_g和所述第二参与方持有的第四混淆矩阵T_h,所述第三混淆矩阵T_g和所述第四混淆矩阵T_h的大小一致;

[0110] S1033、所述第一参与方基于持有的所述第一交集数据FF0、所述第一加密数据和

所述第三混淆矩阵T_g计算得到第一碎片数据S_0,并将所述第一碎片数据S_0发送至所述第二参与方;

[0111] S1034、所述第二参与方基于所述第一碎片数据S_0、所述第四混淆矩阵T_h和所述第二加密数据计算得到第二碎片数据S_1,并将所述第二碎片数据S_1发送至所述第一参与方;

[0112] S1035、所述第一参与方基于所第二碎片数据S_1计算得到所述第一全匿踪乱序数据Z_g,所述第二参与方基于所述第四混淆矩阵T_h和所述第二加密数据计算得到所述第二全匿踪乱序数据Z_h。

[0113] 在本方案中,仍以第一参与方作为Receiver方,第二参与方为Sender方,进行一轮乱序重排的流程;

[0114] 在第一轮的第一步中,Receiver方随机生成(1*n)第三混淆矩阵T_g,发送第一碎片数据 $s_0 = FF_0 - T_g + GB_g$给到Sender方;其中, GB_g为第一加密数据中的第一混淆矩阵。同时第三混淆矩阵T_g即为第三临时数据P_0;

[0115] 第二步中,Sender方接收碎片数据s_0,计算第四临时数据$P_1 = s_0 + FF_1 + GB_h = FF_0 - T_g + GB_g + FF_1 + GB_h = FF + GB - T_g$;其中,Sender方通过$P_1 = FF + GB - T_g$无法反推$FF$;

[0116] 第三步中,Sender方使用第二加密数据中的第二乱序索引$shuffle_index_h$执行在第二参与方的本地乱序, $P_1 = P_1[shuffle\_index_h] = FF* + GB* - T_g$;同时Receiver方的混淆矩阵无需同步乱序,即$P_0 = T_g$;

[0117] 第四步中,Sender方随机生成大小为(1*n)的第四混淆矩阵T_h,发送第二碎片数据$s_1 = P_1 - T_h$给到sender方,同时第四混淆矩阵T_h即为Q_1;

[0118] 第五步中,Receiver方接收第二碎片数据s_1,计算

$Q_0 = s_1 + P_0 = FF* + GB* - T_g - T_h + T_g = FF* + GB* - T_h$;Receiver方通过

$Q_0 = FF* + GB* - R_h*$无法反推FF;

[0119] 第六步中,Receiver方使用第一乱序索引$shuffle_index_g$执行第一参与方的本地乱序, $Q_0 = Q_0[shuffle\_index_g] = FF** + GB** - T_h$;同时Sender方的混淆矩阵无需同步进行乱序,即$Q_1 = T_h$;

[0120] 第七步中,Receiver方计算第一全匿踪乱序数据

$Z_0 = Q_0 - Z_g = FF** + GB** - T_h - Z_g$;Sender方计算第二全匿踪乱序数据

$Z_1 = Q_1 - Z_h = T_h - Z_h$;即

$Z_0 + Z_1 = Q_0 + Q_1 - (Z_g + Z_h) = (FF** + GB** - T_h) + T_h - GB** = FF**$。

[0121] 在本方案中,基于秘密分享的低带宽重排阶段只涉及混淆的数值类数据,不涉及到同态加密的数据,因此其通信量较小,通常是64bit的数值数据。由于在同态碎片重排阶段已经预先计算了在基于秘密分享的低带宽重排阶段所需的中间依赖数据,因此本阶段仅需要执行两次通信,且通信量很小,即可输出最终的原始分片数据< FF>的密态重排结果<Z>,双方中的任意一方都无法单独反推原始样本顺序。

[0122] 在步骤S104中,针对步骤S103输出的密态重排结果<Z>,基于其中的指示结果数据<B>,执行密态混淆,生成最终的目标混淆集<PSIDiffusion>。其中,得到混淆集合保留的样

本的指示结果数据<B'>,该样本的指示结果数据为碎片态数据,对应的明文结果为1或者0,其中1表示混淆集合中选择的特定的数据(真实交集+一部分掺杂非交集),0表示一定是非交集。由于已经经过全匿踪乱序,因此无法推知非交集对应的原始样本是哪一条。

[0123] 作为一种可实现的方式,如图4所示,所述目标混淆集包括所述指示结果数据,所述共享方法还包括:

[0124] S105、将所述目标混淆集中的所述指示结果数据恢复得到所述目标混淆集中数据对应的指示结果;

[0125] S106、所述第一参与方基于所述指示结果对持有的部分所述目标混淆集进行筛选,得到所述第一数据集中属于所述关联数据的第一目标样本;

[0126] S107、获取所述第一目标样本对应的目标索引,并将所述目标索引发送至所述第二参与方;

[0127] S108、所述第二参与方基于所述目标索引对持有的另一部分所述目标混淆集进行筛选,得到述第二数据集中与所述第一目标样本对应的第二目标样本。

[0128] 在本方案中,根据输出的目标混淆集,得到第一参与方持有的部分目标混淆集和第二参与方持有的另一部分目标混淆集,对其中混淆集合保留结果的指标结果数据进行结果恢复,第一参与方保留其中指示结果为1的样本,并同步保留下来的样本索引给到S方进行相应的筛选,以得到最终的结果。

[0129] 作为一种可实现的方式,所述第一数据集为用户消费数据,所述第二数据集为用户财产数据,所述目标标识为用户身份信息,所述共享方法还包括:

[0130] 基于预设筛选信息对所述第一目标样本和所述第二目标样本筛选,得到所述用户身份信息对应的目标消费数据和目标财产数据。

[0131] 在本方案中,在高密业务计算场景中,用户希望在全流程保护原始数据(包括交集数据及非交集数据)的前提下,完成业务相关的数据指标的计算,帮助营销运营、广告投放等业务筛选出目标投放人群。

[0132] 以商场近一周信用卡使用次数大于5次的用户群筛选需求为例。涉及的业务主体为银行侧卡运营部门作为第一参与方以及商场侧作为第二参与方。第一参与方希望通过挖掘出在商场一周内刷卡大于5次的用户群(Gm)的特性,进一步找出相似特性的但没有达到有效次数刷卡记录的信用卡用户群(Gn),进行优惠券刺激营销,激活这批不活跃的信用卡用户群体。同时,商家侧的第二参与方不想泄露这一批高价值用户的身份信息,一方面避免竞对竞争,另一方面是满足数据保护合规层面的要求。

[0133] 如图5所示,在作为第一参与方的银行用户作为第二参与方的商场用户执行全匿踪求交算法,计算得到特征大宽表数据<FF>,第一参与方持有第一交集数据FF0,第一交集数据FF0包括如图中第一栏从左至右依次包括交集标识B、近一周信用卡刷卡次数、近一周消费金额、年龄、性别、月收入、存款金额、婚姻、职业等级、社保缴存、公积金和爱好等级等数据。第二参与方持有第二交集数据FF1,第二交集数据FF1包括与第一交集数据同态分布的如图中第一栏从左至右依次包括交集标识B、近一周信用卡刷卡次数、近一周消费金额、年龄、性别、月收入、存款金额、婚姻、职业等级、社保缴存、公积金和爱好等级等数据。<B>表示“是否为交集”的指示结果数据,<1>表示“是交集”,<0>表示“不是交集”。经过全匿踪求交算法之后得到的指示碎片以及属性碎片的大宽表FF的分片数据,就进入到全匿踪乱序。

[0134] 后续的同态碎片混合重排和基于秘密分享的低带宽重排过程已经在前述步骤S102和S103中详细说明执行过程,输出的结果进入到混淆掺杂后得到目标混淆集。

[0135] 最终对筛选出的目标混淆数据,执行按预设筛选条件的密态计算,比如在第一部分中列举的任务描述,需要刻画出“近1周商场消费刷卡大于5次”的用户群体的特征。那么需要对B侧特征进行条件过滤筛选出消费刷卡大于5次,也就是基于第二参与方持有的部分目标混淆集输出结果,统计出刷卡这一维特征中,次数大于5次的人群,这样我们就利用B方特征,在全匿踪混淆求交碎片态数据上,筛选出了银行想要的目标人群G。基于目标人群G,统计A方银行侧的特征,比如近30天、近3个月、近6个月的银行存款资产水平,该群体的年龄分布特点,银行侧流水金额数据分布等目标群体G的群体特征信息。据此圈选出需要定向激励的不活跃的信用卡人群,激活类似特征的信用卡存量用户,达到有效运营的目标。以上的群体特征统计,都是在碎片态下,执行多方安全计算得到,因此,并不会暴露交集或者非交集信息,真正实现全流程全匿踪的数据处理目的。

[0136] 本方案中,实现交集数据以及非交集数据的全流程隐私保护,并且无法通过差分攻击或者信息反推来定位出个体的私有信息,完成A方与B方的业务需求,且在计算性能上可以满足真实生产环境的性能要求。

[0137] 本实施例提供的全匿踪混淆求交数据的共享方法,基于全匿踪混淆求交的数据处理流程,达到保护参与双方的交集和非交集不泄露,防止数据被差分攻击破解,且数据处理中无需个体信息使用授权,数据处理过程计算量级低。

[0138] 实施例2

[0139] 本实施例提供一种全匿踪混淆求交数据的共享系统100,第一参与方持有第一数据集,第二参与方持有第二数据集,所述第一数据集与所述第二数据集均包括同一目标标识的关联数据,如图6所示,所述共享系统包括数据求交模块101、同态重排模块102、全匿踪乱序模块103和密态混淆掺杂模块104;

[0140] 所述数据求交模块,用于获取所述第一数据集和所述第二数据集基于所述关联数据的加密求交结果数据;

[0141] 所述同态重排模块,用于将所述加密求交结果数据基于同态碎片混合重排生成所述第一参与方对应的第一加密数据和所述第二参与方对应的第二加密数据;

[0142] 所述全匿踪乱序模块,用于将所述第一加密数据、所述第二加密数据和所述加密求交结果数据进行基于秘密分享的低带宽重排,得到所述第一参与方持有的第一全匿踪乱序数据和所述第二参与方持有的第二全匿踪乱序数据;其中,所述基于秘密分享的低带宽重排与所述同态碎片混合重排的重排顺序一致;

[0143] 所述密态混淆掺杂模块,用于基于所述第一全匿踪乱序数据和所述第二全匿踪乱序数据执行密态混淆掺杂得到所述第一数据集和所述第二数据集的目标混淆集,所述第一参与方与所述第二参与方分别持有部分所述目标混淆集进行共享。

[0144] 作为一种可实现的方式,所述同态重排模块102包括第一混淆矩阵生成单元、同态乱序单元和加密数据输出单元;

[0145] 所述第一混淆矩阵生成单元,用于随机生成所述第一参与方的第一混淆矩阵GB_g和所述第二参与方的第二混淆矩阵GB_h,所述第一混淆矩阵GB_g与所述第二混淆矩阵GB_h的大小一致,且混淆矩阵集合$GB = GB_g + GB_h$;

[0146] 所述同态乱序单元,用于将所述第一混淆矩阵GB_g与所述第二混淆矩阵GB_h基于同态加密计算和乱序处理,得到所述第一参与方的第一秘密共享数据Z_g和所述第一参与方执行的第一乱序索引$shuffle_index_g$,以及所述第二参与方的第二秘密共享数据Z_h和所述第二参与方执行的第二乱序索引$shuffle_index_h$;

[0147] 其中,所述第一秘密共享数据Z_g与所述第二秘密共享数据Z_h之和,等于所述混淆矩阵集合GB基于所述第一乱序索引$shuffle_index_g$和所述第二乱序索引$shuffle_index_h$进行乱序处理的结果;

[0148] 所述加密数据输出单元,用于将所述第一混淆矩阵GB_g、所述第一秘密共享数据Z_g和所述第一乱序索引$shuffle_index_g$作为所述第一加密数据,所述第二混淆矩阵GB_h、所述第二秘密共享数据Z_h和所述第二乱序索引$shuffle_index_h$作为所述第二加密数据。

[0149] 作为一种可实现的方式,所述全匿踪乱序模块103包括交集数据分配单元、第二混淆矩阵生成单元、第一碎片计算单元、第二碎片计算单元和全匿踪乱序数据输出单元;

[0150] 所述交集数据分配单元,用于将所述加密求交结果数据分为所述第一参与方持有的第一交集数据和所述第二参与方持有的第二交集数据;

[0151] 所述第二混淆矩阵生成单元,用于生成所述第一参与方持有的第三混淆矩阵和所述第二参与方持有的第四混淆矩阵,所述第三混淆矩阵和所述第四混淆矩阵的大小一致;

[0152] 所述第一碎片计算单元,用于所述第一参与方基于持有的所述第一交集数据、所述第一加密数据和所述第三混淆矩阵计算得到第一碎片数据,并将所述第一碎片数据发送至所述第二参与方;

[0153] 所述第二碎片计算单元,用于所述第二参与方基于所述第一碎片数据、所述第四混淆矩阵和所述第二加密数据计算得到第二碎片数据,并将所述第二碎片数据发送至所述第一参与方;

[0154] 所述全匿踪乱序数据输出单元,用于所述第一参与方基于所第二碎片数据计算得到所述第一全匿踪乱序数据,所述第二参与方基于所述第四混淆矩阵和所述第二加密数据计算得到所述第二全匿踪乱序数据。

[0155] 作为一种可实现的方式,所述交集数据分配单元包括数据获取子模块、安全求交子模块和拼接子模块;

[0156] 所述数据获取子模块,用于获取所述第一数据集和所述第二数据集;

[0157] 其中,所述第一数据集包括第一样本特征,所述第二数据集包括第二样本特征;

[0158] 所述安全求交子模块,用于将所述第一数据集和所述第二数据集进行全匿踪安全求交,得到指示结果数据和特征结果数据;

[0159] 指示结果数据表征对应数据是否属于所述关联数据,所述特征结果数据为所述第一样本特征与所述第二样本特征的对齐结果;

[0160] 所述拼接子模块,用于将所述指示结果数据和所述特征结果数据拼接得到所述加密求交结果数据。

[0161] 作为一种可实现的方式,所述拼接子模块还用于将所述特征结果数据分为所述第一参与方持有的第一特征结果数据和所述第二参与方持有的第二特征结果数据;

[0162] 其中,所述第一特征结果数据为M×P大小的矩阵,所述第二特征结果数据为M×Q

大小的矩阵；

[0163] 所述拼接子模块还用于将所述第一特征结果数据、所述第二特征结果数据和所述指示结果数据横向拼接得到大小为M×N的矩阵作为所述加密求交结果数据；

[0164] 其中，N=P+Q+1，M、N、P和Q是正整数。

[0165] 作为一种可实现的方式，所述目标混淆集包括所述指示结果数据，所述共享系统还包括过滤模块和筛选模块；

[0166] 所述过滤模块，用于将所述目标混淆集中的所述指示结果数据恢复得到所述目标混淆集中数据对应的指示结果；

[0167] 所述筛选模块，用于所述第一参与方基于所述指示结果对持有的部分所述目标混淆集进行筛选，得到所述第一数据集中属于所述关联数据的第一目标样本；

[0168] 所述筛选模块还用于获取所述第一目标样本对应的目标索引，并将所述目标索引发送至所述第二参与方；

[0169] 所述筛选模块还用于所述第二参与方基于所述目标索引对持有的另一部分所述目标混淆集进行筛选，得到述第二数据集中与所述第一目标样本对应的第二目标样本。

[0170] 作为一种可实现的方式，所述第一数据集为用户消费数据，所述第二数据集为用户财产数据，所述目标标识为用户身份信息；

[0171] 所述筛选模块还用于基于预设筛选信息对所述第一目标样本和所述第二目标样本筛选，得到所述用户身份信息对应的目标消费数据和目标财产数据。

[0172] 对于系统实施例而言，由于其基本对应于方法实施例，所以相关之处参见方法实施例的部分说明即可。以上所描述的系统实施例仅仅是示意性的，其中所述作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部模块来实现本公开方案的目的。

[0173] 本实施例提供的全匿踪混淆求交数据的共享系统，基于全匿踪混淆求交的数据处理流程，达到保护参与双方的交集和非交集不泄露，防止数据被差分攻击破解，且数据处理中无需个体信息使用授权，数据处理过程计算量级低

[0174] 实施例3

[0175] 图7为本公开一示例实施例示出的一种电子设备的结构示意图，电子设备包括存储器、处理器及存储在存储器上并用于在处理器上运行的计算机程序，所述处理器执行所述计算机程序时实现上述任一实施例所述的全匿踪混淆求交数据的共享方法。图7显示的电子设备90仅仅是一个示例，不应对本公开实施例的功能和使用范围带来任何限制。

[0176] 如图7所示，电子设备90可以以通用计算设备的形式表现，例如其可以为服务器设备。电子设备90的组件可以包括但不限于：上述至少一个处理器91、上述至少一个存储器92、连接不同系统组件(包括存储器92和处理器91)的总线93。

[0177] 总线93包括数据总线、地址总线和控制总线。

[0178] 存储器92可以包括易失性存储器，例如随机存取存储器(RAM)921和/或高速缓存存储器922，还可以进一步包括只读存储器(ROM)923。

[0179] 存储器92还可以包括具有一组(至少一个)程序模块924的程序工具925(或实用工具)，这样的程序模块924包括但不限于：操作系统、一个或者多个应用程序、其它程序模块

以及程序数据,这些示例中的每一个或某种组合中可能包括网络环境的实现。

[0180] 处理器91通过运行存储在存储器92中的计算机程序,从而执行各种功能应用以及数据处理,例如上述任一实施例所提供的全匿踪混淆求交数据的共享方法。

[0181] 电子设备90也可以与一个或多个外部设备94(例如键盘、指向设备等)通信。这种通信可以通过输入/输出(I/O)接口95进行。并且,电子设备90还可以通过网络适配器96与一个或者多个网络(例如局域网(LAN),广域网(WAN)和/或公共网络,例如因特网)通信。如图所示,网络适配器96通过总线93与电子设备90的其它模块通信。应当明白,尽管图中未示出,可以结合电子设备90使用其它硬件和/或软件模块,包括但不限于:微代码、设备驱动器、冗余处理器、外部磁盘驱动阵列、RAID(磁盘阵列)系统、磁带驱动器以及数据备份存储系统等。

[0182] 应当注意,尽管在上文详细描述中提及了电子设备的若干单元/模块或子单元/模块,但是这种划分仅仅是示例性的并非强制性的。实际上,根据本公开的实施方式,上文描述的两个或更多单元/模块的特征和功能可以在一个单元/模块中具体化。反之,上文描述的一个单元/模块的特征和功能可以进一步划分为由多个单元/模块来具体化。

[0183] 实施例4

[0184] 本公开实施例还提供一种计算机可读存储介质,其上存储有计算机程序,所述程序被处理器执行时实现上述任一实施例所提供的全匿踪混淆求交数据的共享方法。

[0185] 其中,可读存储介质可以采用的更具体可以包括但不限于:便携式盘、硬盘、随机存取存储器、只读存储器、可擦拭可编程只读存储器、光存储器件、磁存储器件或上述的任意合适的组合。

[0186] 实施例5

[0187] 本公开实施例还提供一种计算机程序产品,包括计算机程序,所述计算机程序被处理器执行时实现上述任一项所述的全匿踪混淆求交数据的共享方法。

[0188] 其中,可以以一种或多种程序设计语言的任意组合来编写用于执行本公开的计算机程序产品的程序代码,所述程序代码可以完全地在用户设备上执行、部分地在用户设备上执行、作为一个独立的软件包执行、部分在用户设备上部分在远程设备上执行或完全在远程设备上执行。

[0189] 虽然以上描述了本公开的具体实施方式,但是本领域的技术人员应当理解,这仅是举例说明,本公开的保护范围是由所附权利要求书限定的。本领域的技术人员在不背离本公开的原理和实质的前提下,可以对这些实施方式做出多种变更或修改,但这些变更和修改均落入本公开的保护范围。

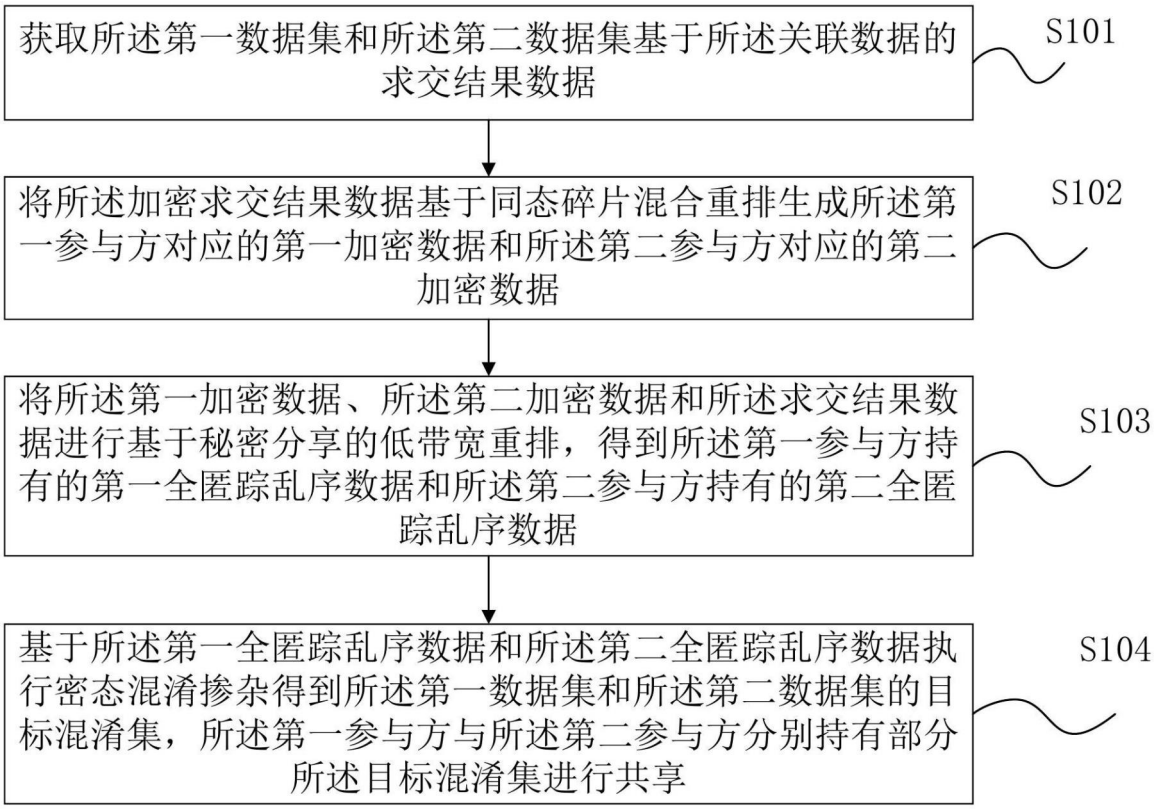

获取所述第一数据集和所述第二数据集基于所述关联数据的求交结果数据
S101
将所述加密求交结果数据基于同态碎片混合重排生成所述第一参与方对应的第一加密数据和所述第二参与方对应的第二加密数据
S102
将所述第一加密数据、所述第二加密数据和所述求交结果数据进行基于秘密分享的低带宽重排，得到所述第一参与方持有的第一全匿踪乱序数据和所述第二参与方持有的第二全匿踪乱序数据
S103
基于所述第一全匿踪乱序数据和所述第二全匿踪乱序数据执行密态混淆掺杂得到所述第一数据集和所述第二数据集的目标混淆集，所述第一参与方与所述第二参与方分别持有部分所述目标混淆集进行共享
S104

图 1

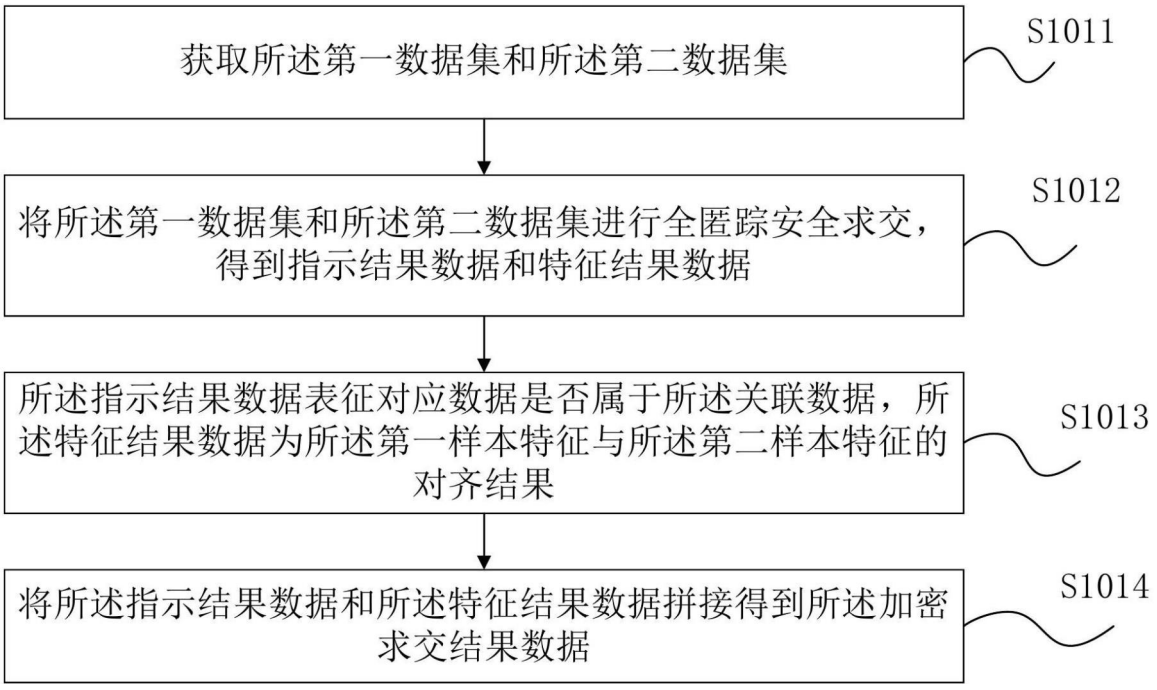

获取所述第一数据集和所述第二数据集
S1011
将所述第一数据集和所述第二数据集进行全匿踪安全求交，得到指示结果数据和特征结果数据
S1012
所述指示结果数据表征对应数据是否属于所述关联数据，所述特征结果数据为所述第一样本特征与所述第二样本特征的对齐结果
S1013
将所述指示结果数据和所述特征结果数据拼接得到所述加密求交结果数据
S1014

图 2

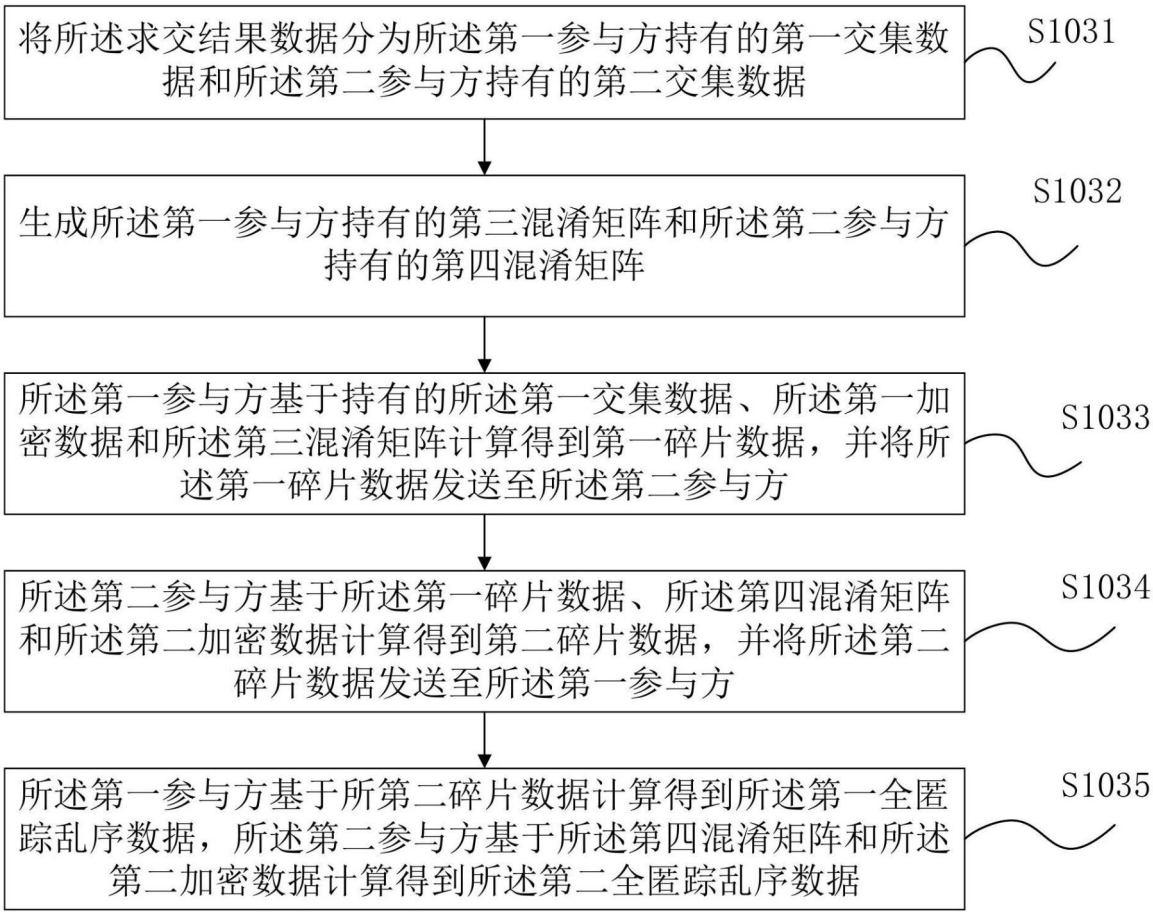

将所述求交结果数据分为所述第一参与方持有的第一交集数据和所述第二参与方持有的第二交集数据
S1031
生成所述第一参与方持有的第三混淆矩阵和所述第二参与方持有的第四混淆矩阵
S1032
所述第一参与方基于持有的所述第一交集数据、所述第一加密数据和所述第三混淆矩阵计算得到第一碎片数据，并将所述第一碎片数据发送至所述第二参与方
S1033
所述第二参与方基于所述第一碎片数据、所述第四混淆矩阵和所述第二加密数据计算得到第二碎片数据，并将所述第二碎片数据发送至所述第一参与方
S1034
所述第一参与方基于所第二碎片数据计算得到所述第一全匿踪乱序数据，所述第二参与方基于所述第四混淆矩阵和所述第二加密数据计算得到所述第二全匿踪乱序数据
S1035

图 3

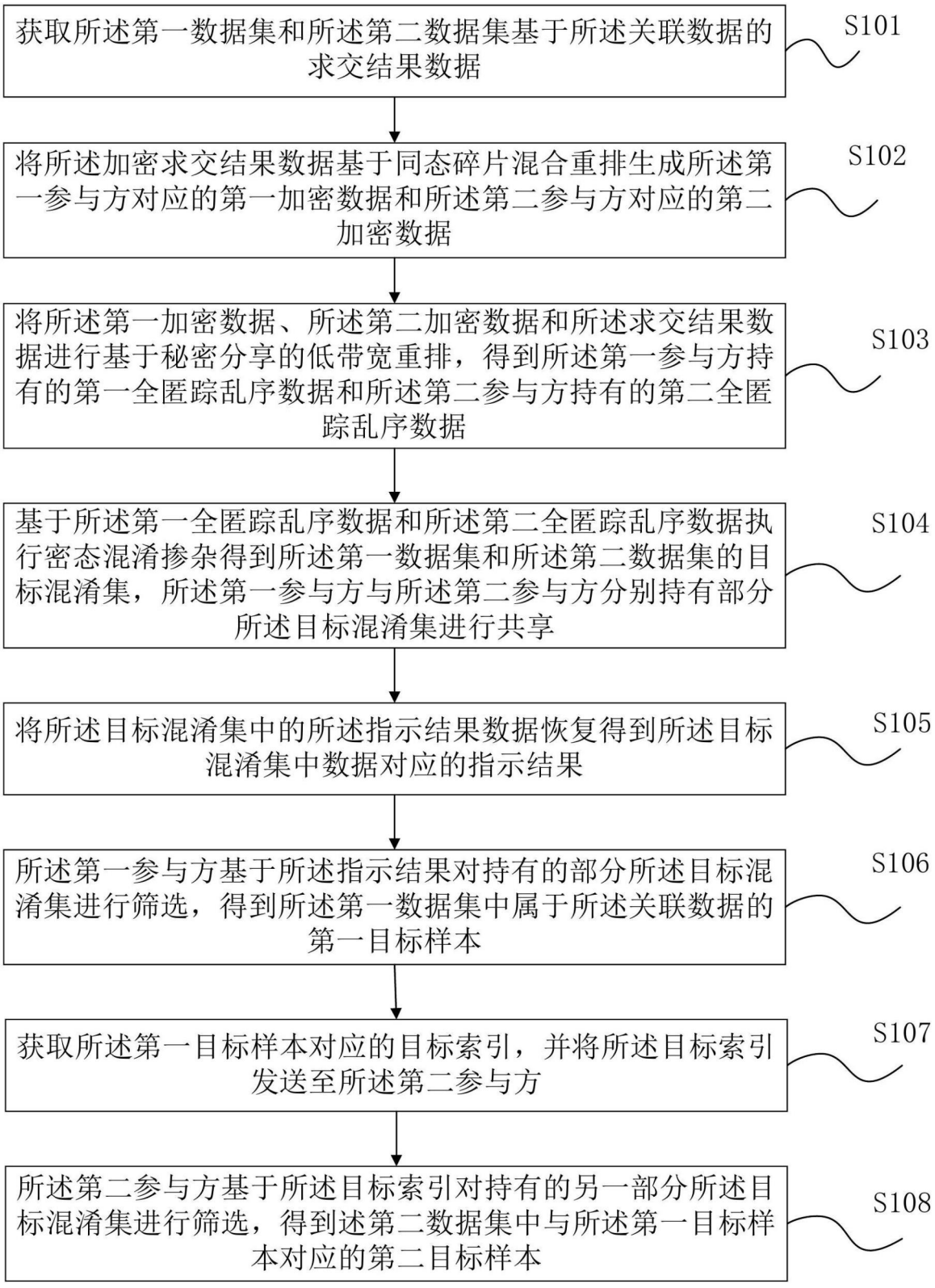

获取所述第一数据集和所述第二数据集基于所述关联数据的求交结果数据
S101
将所述加密求交结果数据基于同态碎片混合重排生成所述第一参与方对应的第一加密数据和所述第二参与方对应的第二加密数据
S102
将所述第一加密数据、所述第二加密数据和所述求交结果数据进行基于秘密分享的低带宽重排，得到所述第一参与方持有的第一全匿踪乱序数据和所述第二参与方持有的第二全匿踪乱序数据
S103
基于所述第一全匿踪乱序数据和所述第二全匿踪乱序数据执行密态混淆掺杂得到所述第一数据集和所述第二数据集的目标混淆集，所述第一参与方与所述第二参与方分别持有部分所述目标混淆集进行共享
S104
将所述目标混淆集中的所述指示结果数据恢复得到所述目标混淆集中数据对应的指示结果
S105
所述第一参与方基于所述指示结果对持有的部分所述目标混淆集进行筛选，得到所述第一数据集中属于所述关联数据的第一目标样本
S106
获取所述第一目标样本对应的目标索引，并将所述目标索引发送至所述第二参与方
S107
所述第二参与方基于所述目标索引对持有的另一部分所述目标混淆集进行筛选，得到述第二数据集中与所述第一目标样本对应的第二目标样本
S108

图 4

第一参与方持有的第一交集数据FF0

交集标识B	近一周信用卡刷卡次数	近一周消费金额	年龄	性别	月收入	存款金额	婚姻	职业等级	社保缴存	公积金	爱好等级
<1>_0	<6>_0	<3000>_0	<32>_0	<0>_0	<6000>_0	<120000>_0	<1>_0	<2>_0	<5000>_0	<7000>_0	<1>_0
<0>_0	<2>_0	<450>_0	<age_r1>_0	<ge_r1>_0	<sa_r1>_0	<sav_r1>_0	<ma_r1>_0	<job_r1>_0	<ss_r1>_0	<af_r1>_0	<ho_r1>_0
<1>_0	<9>_0	<5000>_0	<29>_0	<1>_0	<8000>_0	<100000>_0	<0>_0	<5>_0	<4000>_0	<6000>_0	<3>_0
<0>_0	<3>_0	<15>_0	<age_r2>_0	<ge_r2>_0	<sa_r2>_0	<sav_r2>_0	<ma_r2>_0	<job_r2>_0	<ss_r2>_0	<af_r2>_0	<ho_r2>_0
<0>_0	<cu_r1>_0	<sh_r1>_0	<30>_0	<1>_0	<9000>_0	<150000>_0	<1>_0	<5>_0	<8000>_0	<4000>_0	<1>_0
<0>_0	<1>_0	<100>_0	<age_r3>_0	<ge_r3>_0	<sa_r3>_0	<sav_r3>_0	<ma_r3>_0	<job_r3>_0	<ss_r3>_0	<af_r3>_0	<ho_r3>_0

第二参与方持有的第二交集数据FF1

交集标识B	近一周信用卡刷卡次数	近一周消费金额	年龄	性别	月收入	存款金额	婚姻	职业等级	社保缴存	公积金	爱好等级
<1>_1	<6>_1	<3000>_1	<32>_1	<0>_1	<6000>_1	<120000>_1	<1>_1	<2>_1	<5000>_1	<7000>_1	<1>_1
<0>_1	<2>_1	<450>_1	<age_r1>_1	<ge_r1>_1	<sa_r1>_1	<sav_r1>_1	<ma_r1>_1	<job_r1>_1	<ss_r1>_1	<af_r1>_1	<ho_r1>_1
<1>_1	<9>_1	<5000>_1	<29>_1	<1>_1	<8000>_1	<100000>_1	<0>_1	<5>_1	<4000>_1	<6000>_1	<3>_1
<0>_1	<3>_1	<15>_1	<age_r2>_1	<ge_r2>_1	<sa_r2>_1	<sav_r2>_1	<ma_r2>_1	<job_r2>_1	<ss_r2>_1	<af_r2>_1	<ho_r2>_1
<0>_1	<cu_r1>_1	<sh_r1>_1	<30>_1	<1>_1	<9000>_1	<150000>_1	<1>_1	<5>_1	<8000>_1	<4000>_1	<1>_1
<0>_1	<1>_1	<100>_1	<age_r3>_1	<ge_r3>_1	<sa_r3>_1	<sav_r3>_1	<ma_r3>_1	<job_r3>_1	<ss_r3>_1	<af_r3>_1	<ho_r3>_1

图 5

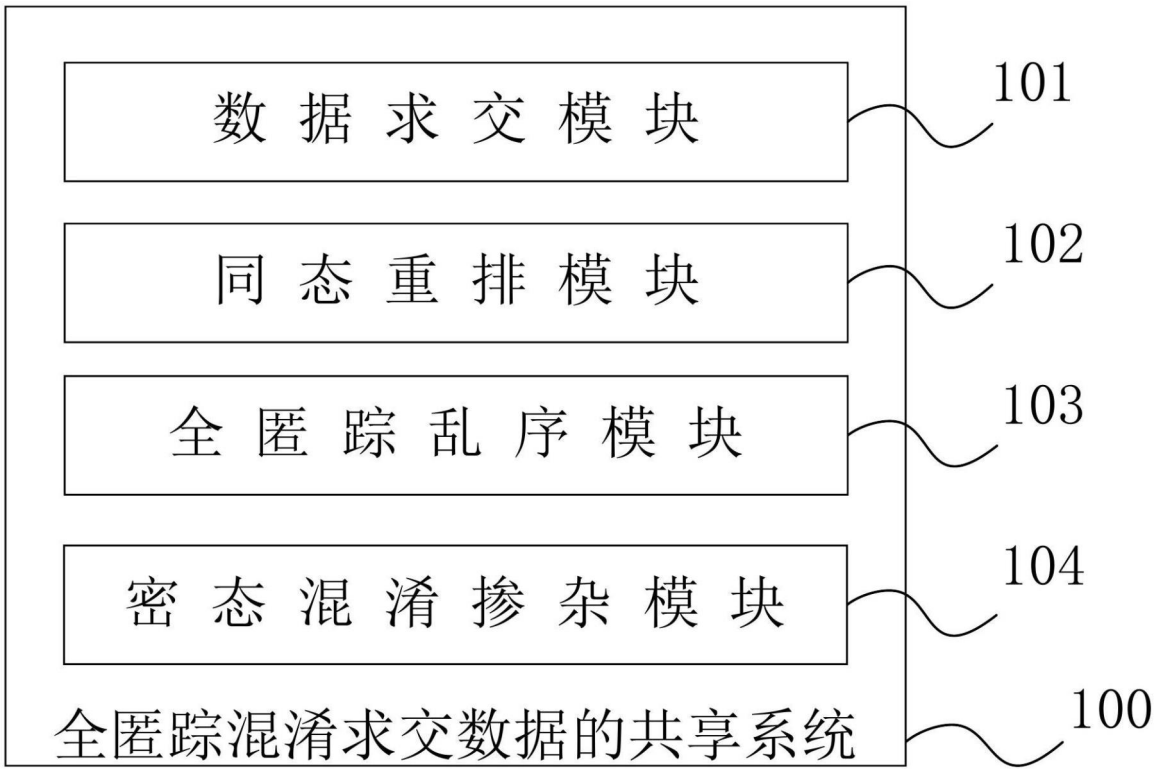

数 据 求 交 模 块
101
同 态 重 排 模 块
102
全 匿 踪 乱 序 模 块
103
密 态 混 淆 掺 杂 模 块
104
全匿踪混淆求交数据的共享系统
100

图 6

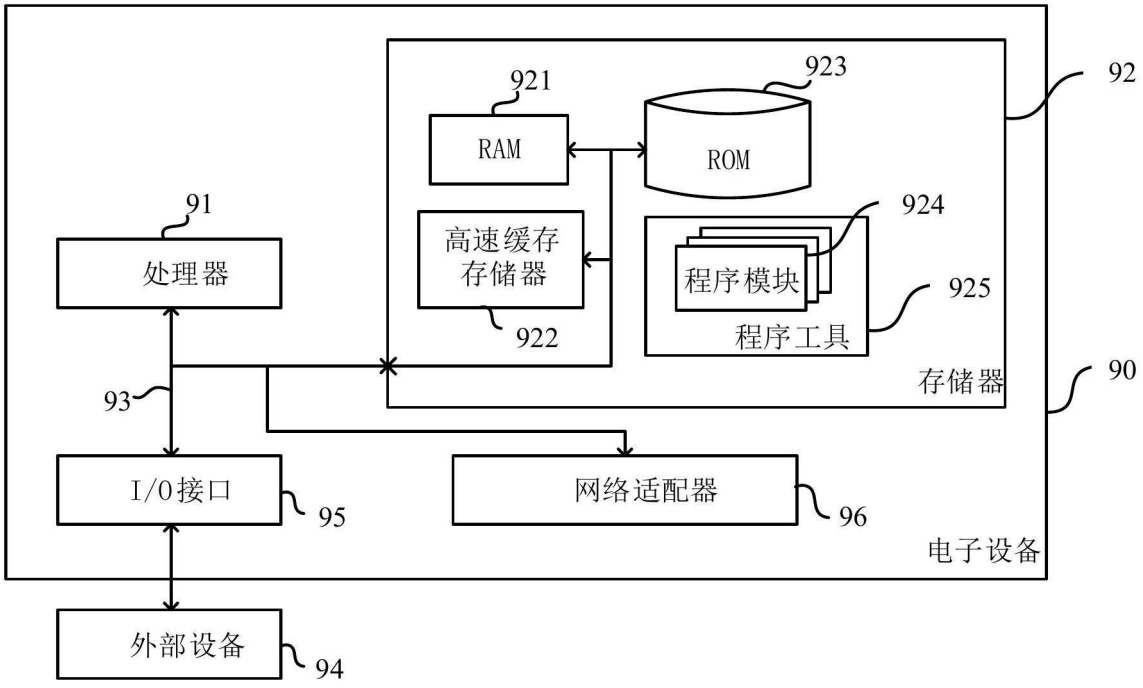

921
RAM
923
ROM
92
91
处理器
高速缓存
存储器
922
924
程序模块
925
程序工具
存储器
90
93
I/O接口
95
网络适配器
96
电子设备
外部设备
94

图 7