

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2024 年 11 月 28 日 (28.11.2024)



(10) 国际公布号
WO 2024/239204 A1

- (51) 国际专利分类号:
G06F 16/176 (2019.01) **G06F 21/62** (2013.01)
- (21) 国际申请号: PCT/CN2023/095566
- (22) 国际申请日: 2023 年 5 月 22 日 (22.05.2023)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (71) 申请人: 浙江大学 (ZHEJIANG UNIVERSITY) [CN/CN]; 中国浙江省杭州市西湖区余杭塘路 866 号, Zhejiang 310058 (CN)。
- (72) 发明人: 刘健 (LIU, Jian); 中国浙江省杭州市西湖区余杭塘路 866 号, Zhejiang 310058 (CN)。李婧瑜 (LI, Jingyu); 中国浙江省杭州市西湖区余杭塘路 866 号, Zhejiang 310058 (CN)。吴狄 (WU, Di); 中国浙江省杭州市西湖区余杭塘路 866 号, Zhejiang

310058 (CN)。任奎 (REN, Kui); 中国浙江省杭州市西湖区余杭塘路 866 号, Zhejiang 310058 (CN)。

- (74) 代理人: 杭州求是专利事务所有限公司 (HANGZHOU QIUSHI PATENT OFFICE CO., LTD.); 中国浙江省杭州市西湖区古墩路 701 号绿城紫金广场 C 座 1506 室邱启旺, Zhejiang 310012 (CN)。
- (81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD,

(54) Title: BATCH PRIVACY INFORMATION RETRIEVAL METHOD AND DEVICE

(54) 发明名称: 一种批量隐私信息检索方法及装置

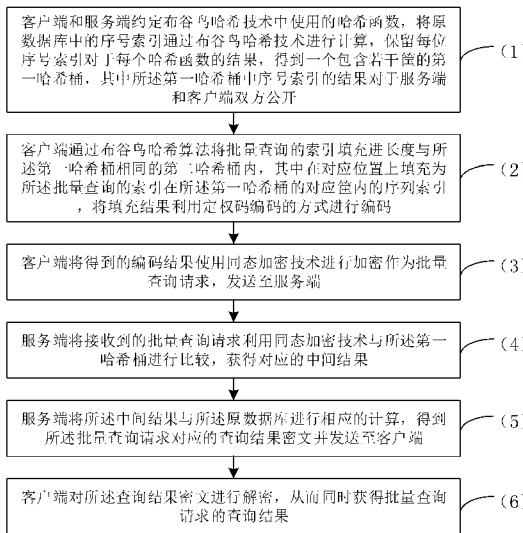


图 1

- (1) A client and a server agree on hash functions used in cuckoo hash technology, perform calculation on serial number indexes in an original database by means of cuckoo hash technology, keep a result of each serial number index for each hash function, and obtain a first hash bucket containing a plurality of baskets, wherein results of the serial number indexes in the first hash bucket are open to both the server and the client
- (2) The client fills a second hash bucket having the same length as the first hash bucket with a batch query index by means of a cuckoo hash algorithm, wherein the corresponding position is filled with a sequence index of the batch query index in the corresponding basket of the first hash bucket; and encode a filling result in a constant weight code encoding mode
- (3) The client encrypts an obtained encoding result by using homomorphic encryption technology to serve as a batch query request, and sends same to the server
- (4) The server compares the received batch query request with the first hash bucket by using the homomorphic encryption technology to obtain a corresponding intermediate result
- (5) The server performs corresponding calculation on the intermediate result and the original database to obtain a query result ciphertext corresponding to the batch query request, and sends the query result ciphertext to the client
- (6) The client decrypts the query result ciphertext so as to obtain query results of the batch query request at the same time

(57) Abstract: Disclosed in the present invention are a batch privacy information retrieval method and device. The method comprises: a client and a server agreeing on hash functions, performing calculation on serial number indexes in an original database by means of cuckoo hash technology, keeping a result of each serial number index for each hash function, and obtaining a first hash bucket containing a plurality of baskets; the client filling a second hash bucket having the same length as the first hash bucket with a batch query index by means of a cuckoo hash algorithm, and encoding a filling result in a constant weight code encoding mode; the client encrypting an obtained encoding result by using homomorphic encryption technology to serve as a batch query request; the server comparing the batch query request with the first hash bucket by using the homomorphic encryption technology to obtain a corresponding intermediate result, and performing corresponding calculation on the intermediate result and the original database to obtain a corresponding query result ciphertext; and the client decrypting the query result ciphertext so as to obtain query results of the batch query request.



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

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

本国际公布:
— 包括国际检索报告(条约第21条(3))。

(57) 摘要: 本发明公开了一种批量隐私信息检索方法及装置, 该方法包括: 客户端和服务端约定哈希函数, 将原数据库中的序号索引通过布谷鸟哈希技术进行计算, 保留每位序号索引对于每个哈希函数的结果, 得到一个包含若干筐的第一哈希桶; 客户端通过布谷鸟哈希算法将批量查询的索引填充进长度与第一哈希桶相同的第二哈希桶内, 将填充结果利用定权码编码的方式进行编码; 客户端将得到的编码结果使用同态加密技术进行加密作为批量查询请求; 服务端将批量查询请求利用同态加密技术与第一哈希桶进行比较, 获得对应的中间结果, 将中间结果与原数据库进行相应的计算, 得到对应的查询结果密文; 客户端对查询结果密文进行解密, 从而获得批量查询请求的查询结果。

一种批量隐私信息检索方法及装置

技术领域

本发明属于密码学应用领域，尤其涉及一种批量隐私信息检索方法及装置。

背景技术

互联网技术的飞速发展，使得云服务成为主流，大型服务商利用搭载在云服务器上的数据库为客户端提供服务，客户端通过访问云端的服务器获得信息，这一高效的服务模式下却存在隐私泄露风险，服务端能够获知用户端想要访问的具体内容，这在许多隐私敏感场景下是不被接受的，如医疗数据库等，基于隐私保护的信息检索方法能够避免这一安全问题。

传统的基于隐私保护的信息检索方法一次只能够处理一个 query，而当用户同时有多个请求时，使用重复运行单词隐私信息检索方法无法获得更高的效率，其中包含大量的冗余步骤。使用批量隐私信息检索协议，其均摊计算开销远小于单词隐私信息查询方法，但对于以往的批量隐私信息检索方法，如 Hao Chen 等人提出的 SealPIR，其首次将布谷鸟哈希技术与批量查询相结合，但由于其使用了普通的编码特性，其在降低均摊计算开销的同时需要引入额外的通信开销。RA Mahdavi 等人首次提出将定权码编码应用于隐私信息检索方案，但其并未考虑批量检索的场景，且其实现方案没有充分利用同态加密的特性，计算效率无法应用于实际场景中。

发明内容

本申请实施例的目的是提供一种批量隐私信息检索方法及装置，以解决相关技术中存在的计算及通信开销过高实用性较低的技术问题，将定权码技术和布谷鸟哈希技术相结合，并针对同态加密计算特性优化，能够同时节省通信均摊开销及计算均摊开销，拥有更优的性能，能够更好的适用于实际场景。

根据本申请实施例的第一方面，提供一种批量隐私信息检索方法，包括：

客户端和服务端约定布谷鸟哈希技术中使用的哈希函数，将原数据库中的序号索引通过布谷鸟哈希技术进行计算，保留每位序号索引对于每个哈希函数的结果，得到一个包含若干筐的第一哈希桶，其中所述第一哈希桶中序号索引的结果对于服务端和客户端双方公开；

客户端通过布谷鸟哈希算法将批量查询的索引填充进长度与所述第一哈希桶相同的第二哈希桶内，其中在对应位置上填充为所述批量查询的索引在所述第一哈希桶的对应筐内的序列索引，将填充结果利用定权码编码的方式进行编码；

客户端将得到的编码结果使用同态加密技术进行加密作为批量查询请求，发送至服务端；

服务端将接收到的批量查询请求利用同态加密技术与所述第一哈希桶进行比较，获得对应的中间结果；

服务端将所述中间结果与所述原数据库进行相应的计算，得到所述批量查询请求对应的查询结果密文并发送至客户端；

客户端对所述查询结果密文进行解密，从而同时获得批量查询请求的查询结果。

进一步地，将填充结果利用定权码编码的方式进行编码，具体为：选择所述第一哈希桶中最大筐的长度 n 作为定权码的编码域，对应定权码的长度为 m ，选择定权码中 1 的数量 k ，将填充结果映射为定权码。

进一步地，客户端将得到的编码结果使用同态加密的 SIMD 技术进行加密。

进一步地，客户端将得到的编码结果使用同态加密的 SIMD 技术进行加密的过程具体为：对于长度为 B 的哈希桶，将哈希桶中 B 个对应第一位加密至同一个密文中，第二位加密至第二个密文中，依次类推，对应哈希桶中相应位置如为空，则默认定权码所有位为 0。

进一步地，服务端将接收到的批量查询请求利用同态加密技术与所述第一哈希桶进行比较，获得对应的中间结果，包括：

将所述第一哈希桶的每个筐内的每一个序列索引进行对应的定权码编码；

对于每个序列索引，选择对应当前定权码为 1 所在位为 1 所在的密文，并利用同态加密的 SIMD 技术计算选择出的所有的密文的乘积，得到对应的中间结果。

进一步地，服务端将所述中间结果与所述原数据库进行相应的计算，得到所述批量查询请求对应的查询结果密文，包括：

将所述中间结果与所述原数据库的对应位置相乘，以获得原数据库对应位置的密文；

将每个筐中的所有结果相加，形成查询结果密文。

进一步地，当原数据库对应位置的内容长度超过 SIMD 技术的明文域时，将该内容切片，分别查询后形成若干查询结果密文，分别发送至客户端。

根据本申请实施例的第二方面，提供一种批量隐私信息检索装置，包括：

布谷鸟哈希模块，用于客户端和服务端约定布谷鸟哈希技术中使用的哈希函数，将原数据库中的序号索引通过布谷鸟哈希技术进行计算，保留每位序号索引对于每个哈希函数的结果，得到一个包含若干筐的第一哈希桶，其中所述第一哈希桶中序号索引的结果对于服务端和客户端双方公开；

编码模块，用于客户端通过布谷鸟哈希算法将批量查询的索引填充进长度与所述第一哈希桶相同的第二哈希桶内，其中在对应位置上填充为所述批量查询的索引在所述第一哈希桶的对应筐内的序列索引，将填充结果利用定权码编码的方式进行编码；

加密模块,用于客户端将得到的编码结果使用同态加密技术进行加密作为批量查询请求,发送至服务端;

比较模块,用于服务端将接收到的批量查询请求利用同态加密技术与所述第一哈希桶进行比较,获得对应的中间结果;

计算模块,用于服务端将所述中间结果与所述原数据库进行相应的计算,得到所述批量查询请求对应的查询结果密文并发送至客户端;

解密模块,用于客户端对所述查询结果密文进行解密,从而同时获得批量查询请求的查询结果。

根据本申请实施例的第三方面,提供一种电子设备,包括:

一个或多个处理器;

存储器,用于存储一个或多个程序;

当所述一个或多个程序被所述一个或多个处理器执行,使得所述一个或多个处理器实现如第一方面所述的方法。

根据本申请实施例的第四方面,提供一种计算机可读存储介质,其上存储有计算机指令,该指令被处理器执行时实现如第一方面所述方法的步骤。

本申请的实施例提供的技术方案可以包括以下有益效果:

由上述实施例可知,本申请通过重新设计批量隐私信息检索具体查询方案,结合布谷鸟哈希、定权码编码及同态加密下 SIMD 技术,利用不同技术的优点结合具体场景设计优化查询算法,极大的降低了计算开销,减少批量隐私信息检索所需要的查询时间,使得查询方案能够更适用于实际场景,同时本方案下的批量隐私信息检索相比同类方案,能够极大的降低通信开销,大幅减小上传请求密文大小,具体的,针对百万条(2^{20} 个)数据(每个数据长度设为 256Bytes)的数据库在批量完成两千次查询时,其每次查询所需的均摊时间为 3.9 毫秒,相比以往批量查询在上传部分通信量减少约为 75 倍。本方案能够实现隐私保护下的批量信息检索服务,且通信开销及计算开销远低于之前的解决方案。

应当理解的是,以上的一般描述和后文的细节描述仅是示例性和解释性的,并不能限制本申请。

附图说明

此处的附图被并入说明书中并构成本说明书的一部分,示出了符合本申请的实施例,并与说明书一起用于解释本申请的原理。

图 1 是根据一示例性实施例示出的一种批量隐私信息检索方法的流程图。

图 2 是根据一示例性实施例示出的基于所有序号保留哈希函数结果的示意图。

图 3 是根据一示例性实施例示出的客户端发出请求的加密过程示意图。

图 4 是根据一示例性实施例示出的遍历筐中序号索引时针对每个索引的的算法示意图。

图 5 是根据一示例性实施例示出的利用中间结果返回数据库内容的算法示意图。

图 6 是根据一示例性实施例示出的一种批量隐私信息检索装置的框图。

图 7 是根据一示例性实施例示出的一种电子设备的示意图。

具体实施方式

这里将详细地对示例性实施例进行说明，其示例表示在附图中。下面的描述涉及附图时，除非另有表示，不同附图中的相同数字表示相同或相似的要素。以下示例性实施例中所描述的实施方式并不代表与本发明相一致的所有实施方式。

在本申请使用的术语是仅仅出于描述特定实施例的目的，而非旨在限制本申请。在本申请和所附权利要求书中所使用的单数形式的“一种”、“所述”和“该”也旨在包括多数形式，除非上下文清楚地表示其他含义。还应当理解，本文中使用的术语“和/或”是指并包含一个或多个相关联的列出项目的任何或所有可能组合。

应当理解，尽管在本申请可能采用术语第一、第二、第三等来描述各种信息，但这些信息不应限于这些术语。这些术语仅用来将同一类型的信息彼此区分开。例如，在不脱离本申请范围的情况下，第一信息也可以被称为第二信息，类似地，第二信息也可以被称为第一信息。取决于语境，如在此所使用的词语“如果”可以被解释成为“在……时”或“当……时”或“响应于确定”。

图 1 是根据一示例性实施例示出的一种批量隐私信息检索方法的流程图，如图 1 所示，该方法应用于终端中，可以包括以下步骤：

(1) 客户端和服务端约定布谷鸟哈希技术中使用的哈希函数，将原数据库中的序号索引通过布谷鸟哈希技术进行计算，保留每位序号索引对于每个哈希函数的结果，得到一个包含若干筐的第一哈希桶，其中所述第一哈希桶中序号索引的结果对于服务端和客户端双方公开；

(2) 客户端通过布谷鸟哈希算法将批量查询的索引填充进长度与所述第一哈希桶相同的第二哈希桶内，其中在对应位置上填充为所述批量查询的索引在所述第一哈希桶的对应筐内的序列索引，将填充结果利用定权码编码的方式进行编码；

(3) 客户端将得到的编码结果使用同态加密技术进行加密作为批量查询请求，发送至服务端；

(4) 服务端将接收到的批量查询请求利用同态加密技术与所述第一哈希桶进行比较，获得对应的中间结果；

(5) 服务端将所述中间结果与所述原数据库进行相应的计算，得到所述批量查询请求对

应的查询结果密文并发送至客户端；

(6) 客户端对所述查询结果密文进行解密，从而同时获得批量查询请求的查询结果。

在具体实施中，本申请可应用于如医疗数据库等场景，访问者（客户端）期望通过查询某一系列药品的详细说明如对应治疗病症、禁忌反应、使用说明、库存等，同时不期望该医疗数据库拥有者（服务端）知道访问者对哪一系药品感兴趣，如服务端知道当前查询用户所感兴趣的药品列表，即可推断出客户端当前可能患有某种疾病等，如乙肝等，用户本身的身体情况应当为敏感隐私信息，不期望暴露给服务端，在无意中的泄露可能导致带来不必要的麻烦，敏感信息泄露的可能性也会使部分患者担心泄露风险从而放弃查询功能，错失部分有效治疗信息。此外，本协议也能够应用于其它隐私计算协议中的基本模块，如用于单次隐私查询的离线加速部分或隐私机器学习中的批量查表操作等。以下结合药品信息查询进行描述。

(1) 客户端和服务端约定布谷鸟哈希技术中使用的哈希函数，将原数据库中的序号索引过布谷鸟哈希算法进行计算，保留每位序号索引对于每个哈希函数的结果，得到一个包含若干筐的第一哈希桶，其中所述第一哈希桶中序号索引的结果对于服务端和客户端双方是公开的，示例如图 2 所示。

所述原数据库中的序号索引对服务端和客户端双方均是公开的，序号索引对应的元素对服务端公开、对客户端不公开。在具体实施中，服务端拥有医疗数据库，以药品数据库为例，数据库中包含药品的详细说明如治疗病症、禁忌反应、使用说明、库存等，客户端想要了解一系列药品的具体信息，同时不想暴露自己对这系列药品感兴趣，因为此类信息会间接泄露客户端的身体情况，客户端能够通过所需药品编号查询数据库了解药品的详细说明，且同时支持查询多个药品，查询后，服务端无法判断客户端关心哪些药品的信息，从而无法根据查询推断出用户的身体情况等有效信息，客户端在进行一些敏感类疾病的药物查询时不必担心查询过程中个人身体情况信息泄露。布谷鸟哈希技术中包括多个不同的哈希函数，这里使用的布谷鸟哈希技术以及具体的哈希函数对于服务端和客户端都是公开的，本方法通常选择哈希函数数量为 3。

具体地，对于每个序号索引，依序计算并保留 3 个哈希函数的结果，利用布谷鸟哈希算法将计算的结果均分在大小为 B 的第一哈希桶中，所述第一哈希桶包括若干筐，药品数据库中每个序号索引在每个筐中的位置对于服务端和客户端都是公开的。

对于哈希桶中的每个位置，有多个序号索引经过哈希函数运算后会被填入，假设哈希桶的长度为 B，假设哈希桶大小其相当于将原数据库复制三倍后，利用布谷鸟哈希算法将复制后的数据库均分在 B 个位置。以便于步骤 (2) 中客户端在查询的时候，能够在每个筐内进行查询，此时每个筐的长度远小于分隔之前，这对于隐私信息检索过程是非常友好的。且运

行完所有哈希函数后的结果对于服务端和客户端双方都是公开的，即客户端能够知道每个元素在每个筐中的位置。

(2) 客户端通过布谷鸟哈希算法将批量查询的索引填充进长度与所述第一哈希桶相同的第二哈希桶内，其中在对应位置上填充为所述批量查询的索引在所述第一哈希桶的对应筐内的序列索引，将填充结果利用定权码编码的方式进行编码；

具体地，本实施例中批量查询的索引为所需药品编号，本步骤的目的是将所有的序号索引都填入一个第二哈希桶中，其中第二哈希桶长度为 B ，则利用布谷鸟哈希算法，在一定失败率的前提下，能够将 L 个想要查询的请求利用该算法填充至 B 个位置中，当满足 $B=1.5L$ 时，失败率为 $(1e-40)$ ，可以忽略不记。因此，本申请在具体实施中优选地预先确定每次批量查询的个数 L ，从而确定哈希桶的长度，进而开展步骤 (1) ~ (6) 的具体过程。由于布谷鸟哈希算法的特性，第二哈希桶中每个索引被填充的位置，一定是布谷鸟哈希函数结果，即一定会出现在对应第一哈希桶相应位置所在筐中，此时只需使用框内的列索引在筐中选取元素即可。

需要说明的是，由于步骤 (1) 获得的结果对于双方是公开的，且保留了所有哈希函数的结果，其相当于将数据库在重复 3 倍后重新划分在 B 个筐中。客户端对于索引运行普通的布谷鸟哈希算法，即对于每个索引只会保留其中一个哈希结果，遇见哈希冲突问题时，会使用解决冲突的策略重新计算冲突结果的哈希值，每个索引在布谷鸟哈希算法后在哈希桶中只会出现一次当客户端将所有需要查询的索引填充进长度为 B 的第二哈希桶内后，相当于在每一个筐中取回了一个数据，而每个筐的大小远小于原数据库的大小，将利用布谷鸟哈希后每个需要查询的序号映射到对应的位置，即获得元素所在的筐，将对应的序号索引替换为所在筐需要查询的对应索引。

具体地，选择所述第一哈希桶中最大筐的长度 n 作为定权码的编码域，对应定权码的长度为 m ，选择定权码中 1 的数量为 k ，通常本方法推荐 k 为 2 时能获得计算量和通信量的平衡，将每一个由 (2.2) 得到的具体筐内索引使用定权码进行编码，以下伪代码提供了一种稳定的从普通序号到定权码的映射方式，定权码长度 m 满足 $m \leq O(\sqrt[k]{k!n} + k)$ ，远小于普通自然数序号编码方式：

输入: $i \in [n], m, k \in \mathbb{N}$ with $\binom{m}{k} \geq n$
 输出: $x \in CW(m, k)$
 1: $j := i$
 2: $l := k$
 3: $x := 0^m$
 4: **for** $m' := m - 1, \dots, 0$ **do**
 5: **if** $j \geq \binom{m'}{l}$ **then**
 6: $x[m'] := 1$
 7: $j := j - \binom{m'}{l}$
 8: $l := l - 1$
 9: **end if**
 10: **if** $l = 0$ **then**
 11: **break**
 12: **end if**
 13: **end for**
 14: **返回** x

本算法输入为序号索引 i ，为 $1 - n$ 间的整数，并输入定权码长度 m ，确定定权码中 1 的数量。输出定权码的编码结果。初始化变量 j 等于 i ，变量 l 为 k ，使用循环对 m 位定权码的每一位进行运算， m' 为当前计算位，当 j 结果大于等于 $\binom{m'}{l}$ 时，则令输出的定权码中当前的计算位为 1，同时减小 l 的值表示仍有 l 个 1 未出现，当 l 为 0 时退出循环，表示定权码剩余位置中不再有 1 出现。

需要说明的是，序号映射至定权码的方法可以随机定义，只要服务端和客户端实现约定同一种映射即可。在另一实施例中，如图 3 中所示，采取随机映射实现序号至定权码的映射，当客户端与服务端使用相同的从普通序号到定权码的映射时，不会影响本方法的其它步骤，实际中 n 远大于 k ，使用定权码编码长度小于普通序号编码。

(3) 客户端将得到的编码结果使用同态加密技术进行加密作为批量查询请求，发送至服务端；

使用同态加密中的 SIMD(Single Instruction Multiple Data, 单指令流多数据流)技术对定权码结果进行加密，具体加密方法为：将哈希桶中 B 个对应第一位加密至同一个密文中，第二位加密至第二个密文中，依次类推，对应哈希桶中相应位置如为空，则默认定权码所有位为 0。对于 SIMD 技术，一次支持将哈希桶中的 N 个定权码的对应位加密至同一个密文中，其中 N 取决于同态加密选择的多项式次数，对于服务端和客户端是公开的，为方便表示，图中选择 $N = B$ ，示例见图 3，在实际应用场景中 N 取值一般为 4096、8192 或 16384。

使用同态加密中的对称加密算法进行加密，其相对于公钥加密而言，能够减小密文的大小，将加密后的结果全部发送给服务端，密文个数为 $m \cdot B / N$ 。

(4) 服务端将接收到的批量查询请求利用同态加密技术与所述第一哈希桶进行比较，获得对应的中间结果；如图 4 所示，可以包括以下过程：

(4.1) 将所述第一哈希桶的每个筐内的每一个序列索引进行对应的定权码编码。

具体地,采用与步骤(2)中相同的定权码映射方法,服务端与客户端事先约定使用相同的同态加密算法。

(4.2) 对于每个序列索引,选择对应当前定权码为对应位为1所在的密文,并利用同态加密的 SIMD 技术计算选择出的所有的密文的乘积,得到对应的中间结果,其中同态加密算法支持密文乘操作,且 SIMD 技术能够同时对 N 个筐执行相同的计算操作。

经过(4.2)后,若当前处理的定权码,等于对应筐中所取序号索引的定权码,则乘积结果对应位置为加密的1,否则为0,遍历所有定权码,存储所有的乘积结果。

(5) 服务端将所述中间结果与所述原数据库进行相应的计算,得到所述批量查询请求对应的查询结果密文并发送至客户端,如图5所示。

(5.1) 将所述中间结果与所述原数据库的对应位置相乘,以获得原数据库对应位置的密文;

(5.2) 将每个筐中的所有列的结果相累加,形成查询结果密文,减少回传通信量。由于对于每个筐内,经过4.3的操作,最多只有一个加密位置为1,所以将每个筐内位置相加时,不会影响5.1获得对应数据库索取位置的内容。

在具体实施中,当数据库内容长度超过 SIMD 技术的明文域时,将数据库内容切片,分别查询后传回,其中步骤4获得的结果可以复用。

如对通信量有更高的要求,如数据库信息长度较大时,则可以通过重塑数据库和请求明文打包方式结合旋转操作重新组合回传结果,进一步减少通信及计算开销。当 B 小于 N 时,则结果仍有压缩空间,具体表示为:

可将(2)中每个筐中的 n 个元素进行拆分,拆分为 N/B 块,纵向拼接,数据库重塑为 N 个筐,每个筐最大长度为 $n*B/N$,此时对应的编码长度 m 相应减小,进一步减少计算过程中的计算量。由于返回的密文中的 N/B 个筐的结果,来自于原始的同一个筐中,即只会出现一个密文对应的明文是有效信息,其余均是 0,此时若原始数据库信息长度长于 SIMD 技术的单个明文域时,会进行切片并获得多个查询结果,通过将每个查询结果旋转 B 相应的倍数后,其中 SIMD 技术支持对加密后的密文旋转任意槽个数,对于结果进行累加,则最多每 N/B 个密文能够累加在一起,且不会影响最终的查询结果。对于放回通信量,能够缩小 N/B 倍。

本申请提出的旋转及合并算法,能够进一步压缩返回结果密文大小,更适用于实际中的网络环境,具体地,针对百万条(2^{20} 个)数据(每个数据长度设为 20KB)的数据库在批量完成两千次查询时,其每次查询所需的均摊时间为 0.27 秒,相比以往批量查询在上传部分通信量减少约为 75 倍,下行通信量与其他同类方案持平(若不进行旋转,该场景下的下行通信

量会出现膨胀)。

(6) 客户端对所述查询结果密文进行解密, 从而同时获得批量查询请求的查询结果。

具体地, 将回传得到的密文利用在加密时生成的私钥利用同态加密原理进行解密操作, 同时利用中国剩余定理进行解码, 此时客户端能够得到所查询的对应信息。如果使用了旋转操作重新组合回传结果, 由于客户端已知服务端的旋转算法, 即服务端知晓对应信息出现的具体位置, 通过重新组合排列后即可获得所需的药品信息。

与前述的批量隐私信息检索方法的实施例相对应, 本申请还提供了批量隐私信息检索装置的实施例。

图 6 是根据一示例性实施例示出的一种批量隐私信息检索装置框图。参照图 6, 该装置可以包括:

布谷鸟哈希模块 21, 用于客户端和服务端约定布谷鸟哈希技术中使用的哈希函数, 将原数据库中的序号索引通过布谷鸟哈希技术进行计算, 保留每位序号索引对于每个哈希函数的结果, 得到一个包含若干筐的第一哈希桶, 其中所述第一哈希桶中序号索引的结果对于服务端和客户端双方公开;

编码模块 22, 用于客户端通过布谷鸟哈希算法将批量查询的索引填充进长度与所述第一哈希桶相同的第二哈希桶内, 其中在对应位置上填充为所述批量查询的索引在所述第一哈希桶的对应筐内的序列索引, 将填充结果利用定权码编码的方式进行编码;

加密模块 23, 用于客户端将得到的编码结果使用同态加密技术进行加密作为批量查询请求, 发送至服务端;

比较模块 24, 用于服务端将接收到的批量查询请求利用同态加密技术与所述第一哈希桶进行比较, 获得对应的中间结果;

计算模块 25, 用于服务端将所述中间结果与所述原数据库进行相应的计算, 得到所述批量查询请求对应的查询结果密文并发送至客户端;

解密模块 26, 用于客户端对所述查询结果密文进行解密, 从而同时获得批量查询请求的查询结果。

关于上述实施例中的装置, 其中各个模块执行操作的具体方式已经在有关该方法的实施例中进行了详细描述, 此处将不做详细阐述说明。

对于装置实施例而言, 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。以上所描述的装置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, 作为单元显示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申请方案的目的。本领域普通技术人员在不付出创造性劳动的情况下，即可以理解并实施。

相应的，本申请还提供一种电子设备，包括：一个或多个处理器；存储器，用于存储一个或多个程序；当所述一个或多个程序被所述一个或多个处理器执行，使得所述一个或多个处理器实现如上述的批量隐私信息检索方法。如图 7 所示，为本发明实施例提供的一种批量隐私信息检索方法所在任意具备数据处理能力的设备的一种硬件结构图，除了图 7 所示的处理器、内存以及网络接口之外，实施例中装置所在的任意具备数据处理能力的设备通常根据该任意具备数据处理能力的设备的实际功能，还可以包括其他硬件，对此不再赘述。

相应的，本申请还提供一种计算机可读存储介质，其上存储有计算机指令，该指令被处理器执行时实现如上述的批量隐私信息检索方法。所述计算机可读存储介质可以是前述任一实施例所述的任意具备数据处理能力的设备的内部存储单元，例如硬盘或内存。所述计算机可读存储介质也可以是外部存储设备，例如所述设备上配备的插接式硬盘、智能存储卡（Smart Media Card，SMC）、SD 卡、闪存卡（Flash Card）等。进一步的，所述计算机可读存储介质还可以既包括任意具备数据处理能力的设备的内部存储单元也包括外部存储设备。所述计算机可读存储介质用于存储所述计算机程序以及所述任意具备数据处理能力的设备所需的其他程序和数据，还可以用于暂时地存储已经输出或者将要输出的数据。

本领域技术人员在考虑说明书及实践这里公开的内容后，将容易想到本申请的其它实施方案。本申请旨在涵盖本申请的任何变型、用途或者适应性变化，这些变型、用途或者适应性变化遵循本申请的一般性原理并包括本申请未公开的本技术领域中的公知常识或惯用技术手段。

应当理解的是，本申请并不局限于上面已经描述并在附图中示出的精确结构，并且可以在不脱离其范围进行各种修改和改变。

权利要求书

1. 一种批量隐私信息检索方法，其特征在于，包括：

客户端和服务端约定布谷鸟哈希技术中使用的哈希函数，将原数据库中的序号索引通过布谷鸟哈希技术进行计算，保留每位序号索引对于每个哈希函数的结果，得到一个包含若干筐的第一哈希桶，其中所述第一哈希桶中序号索引的结果对于服务端和客户端双方公开；

客户端通过布谷鸟哈希算法将批量查询的索引填充进长度与所述第一哈希桶相同的第二哈希桶内，其中在对应位置上填充为所述批量查询的索引在所述第一哈希桶的对应筐内的序列索引，将填充结果利用定权码编码的方式进行编码；

客户端将得到的编码结果使用同态加密技术进行加密作为批量查询请求，发送至服务端；

服务端将接收到的批量查询请求利用同态加密技术与所述第一哈希桶进行比较，获得对应的中间结果；

服务端将所述中间结果与所述原数据库进行相应的计算，得到所述批量查询请求对应的查询结果密文并发送至客户端；

客户端对所述查询结果密文进行解密，从而同时获得批量查询请求的查询结果。

2. 根据权利要求1所述的方法，其特征在于，将填充结果利用定权码编码的方式进行编码，具体为：选择所述第一哈希桶中最大筐的长度 n 作为定权码的编码域，对应定权码的长度为 m ，选择定权码中1的数量 k ，将填充结果映射为定权码。

3. 根据权利要求1所述的方法，其特征在于，客户端将得到的编码结果使用同态加密的SIMD技术进行加密。

4. 根据权利要求3所述的方法，其特征在于，客户端将得到的编码结果使用同态加密的SIMD技术进行加密的过程具体为：对于长度为 B 的哈希桶，将哈希桶中 B 个对应第一位加密至同一个密文中，第二位加密至第二个密文中，依次类推，对应哈希桶中相应位置如为空，则默认定权码所有位为0。

5. 根据权利要求1所述的方法，其特征在于，服务端将接收到的批量查询请求利用同态加密技术与所述第一哈希桶进行比较，获得对应的中间结果，包括：

将所述第一哈希桶的每个筐内的每一个序列索引进行对应的定权码编码；

对于每个序列索引，选择对应当前定权码为对应位为1所在的密文，并利用同态加密的SIMD技术计算选择出的所有的密文的乘积，得到对应的中间结果。

6. 根据权利要求1所述的方法，其特征在于，服务端将所述中间结果与所述原数据库进行相应的计算，得到所述批量查询请求对应的查询结果密文，包括：

将所述中间结果与所述原数据库的对应位置相乘，以获得原数据库对应位置的密文；

将每个筐中的所有结果相加，形成查询结果密文。

7. 根据权利要求 6 所述的方法，其特征在于，当原数据库对应位置的内容长度超过 SIMD 技术的明文域时，将该内容切片，分别查询后形成若干查询结果密文，分别发送至客户端。

8. 一种批量隐私信息检索装置，其特征在于，包括：

布谷鸟哈希模块，用于客户端和服务端约定布谷鸟哈希技术中使用的哈希函数，将原数据库中的序号索引通过布谷鸟哈希技术进行计算，保留每位序号索引对于每个哈希函数的结果，得到一个包含若干筐的第一哈希桶，其中所述第一哈希桶中序号索引的结果对于服务端和客户端双方公开；

编码模块，用于客户端通过布谷鸟哈希算法将批量查询的索引填充进长度与所述第一哈希桶相同的第二哈希桶内，其中在对应位置上填充为所述批量查询的索引在所述第一哈希桶的对应筐内的序列索引，将填充结果利用定权码编码的方式进行编码；

加密模块，用于客户端将得到的编码结果使用同态加密技术进行加密作为批量查询请求，发送至服务端；

比较模块，用于服务端将接收到的批量查询请求利用同态加密技术与所述第一哈希桶进行比较，获得对应的中间结果；

计算模块，用于服务端将所述中间结果与所述原数据库进行相应的计算，得到所述批量查询请求对应的查询结果密文并发送至客户端；

解密模块，用于客户端对所述查询结果密文进行解密，从而同时获得批量查询请求的查询结果。

9. 一种电子设备，其特征在于，包括：

一个或多个处理器；

存储器，用于存储一个或多个程序；

当所述一个或多个程序被所述一个或多个处理器执行，使得所述一个或多个处理器实现如权利要求 1-7 任一项所述的方法。

10. 一种计算机可读存储介质，其上存储有计算机指令，其特征在于，该指令被处理器执行时实现如权利要求 1-7 中任一项所述方法的步骤。

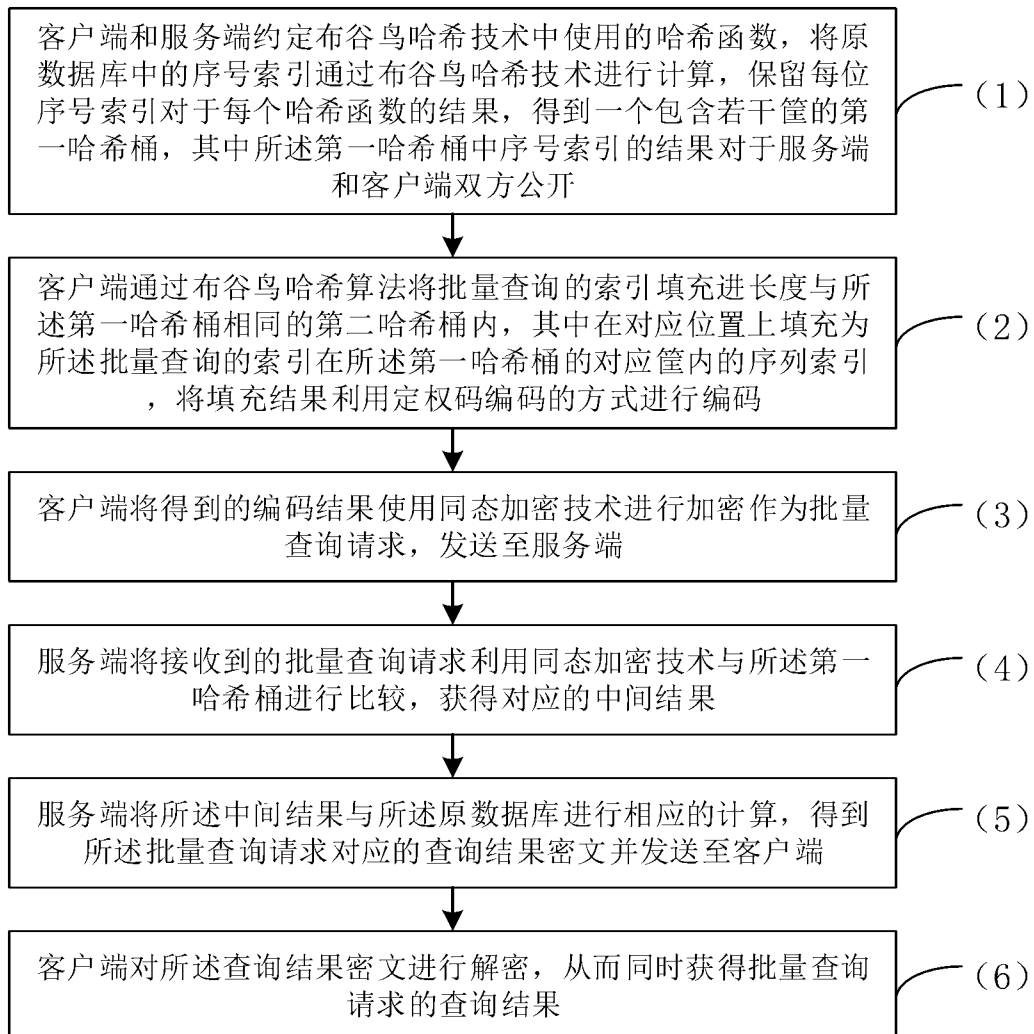


图 1

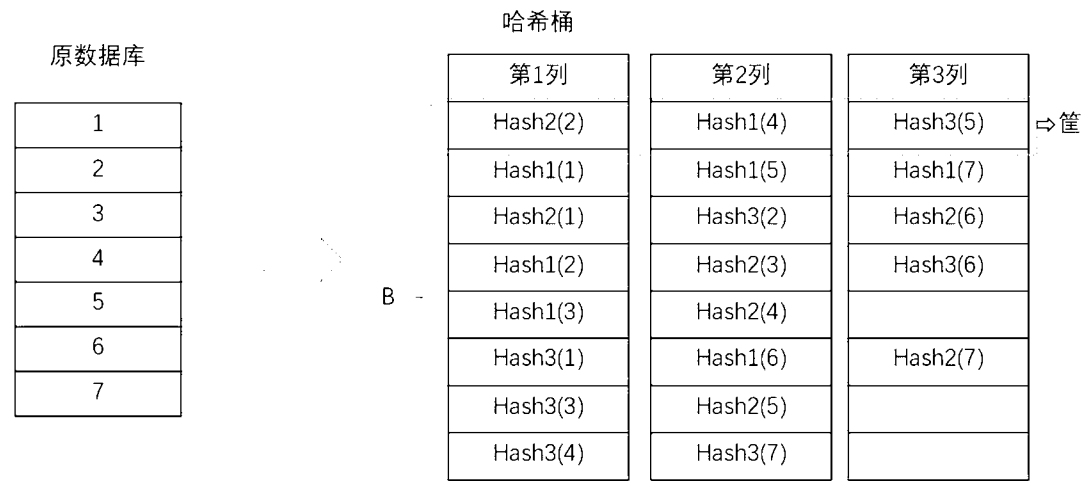


图 2

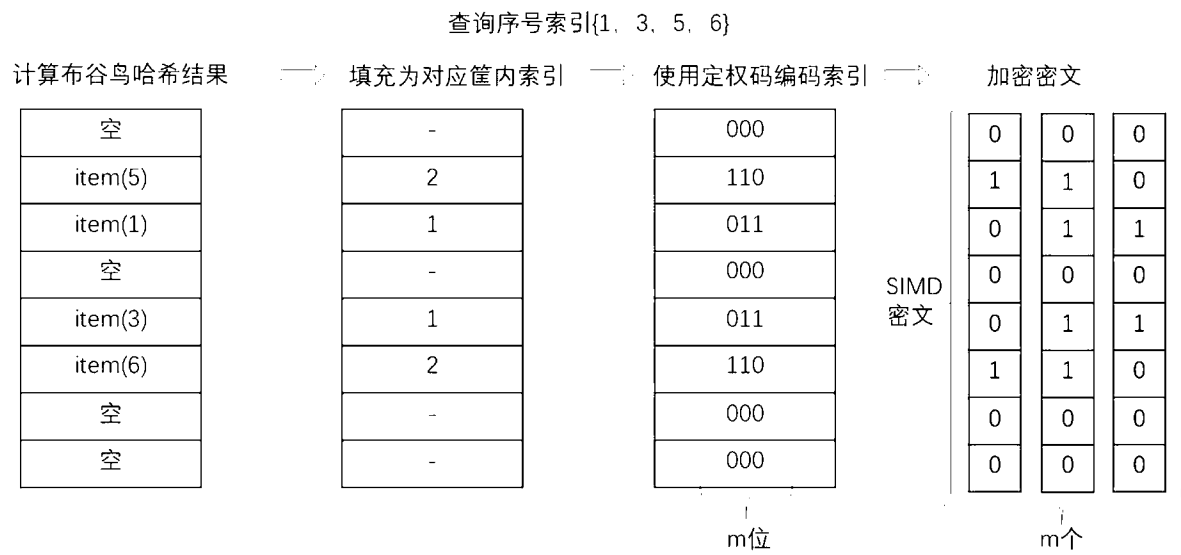


图 3

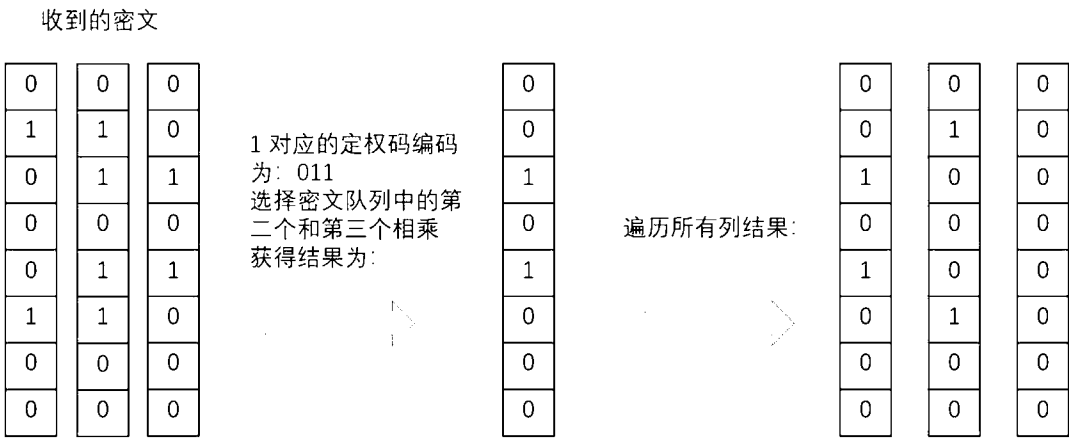


图 4



图 5

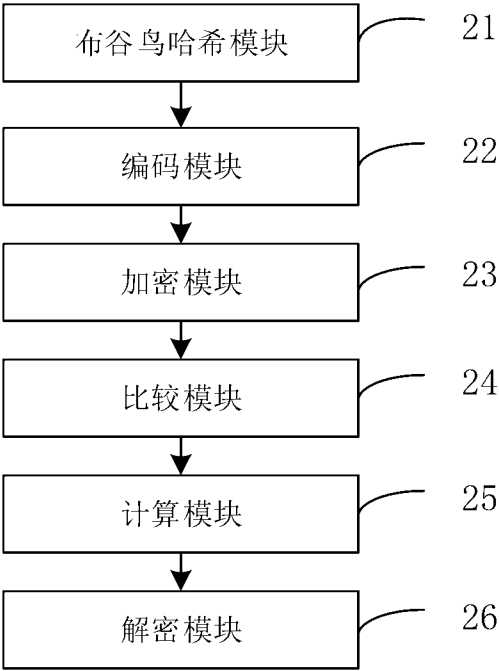


图 6

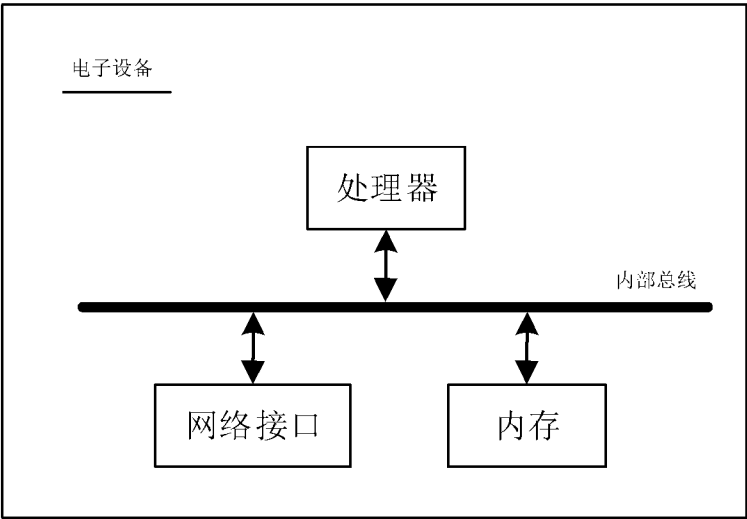


图 7

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2023/095566

A. CLASSIFICATION OF SUBJECT MATTER

G06F16/176(2019.01)i; G06F21/62(2013.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC: G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNKI, CNTXT, VEN, EPTXT, USTXT, ENTXT, IEEE: 隐私, 批量, 哈希, 布谷鸟, 编码, 索引, private, batch, hash, buckoo, coding, index

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 114661680 A (LANXIANG ZHILIAN (HANGZHOU) TECHNOLOGY CO., LTD.) 24 June 2022 (2022-06-24) description, paragraphs [0030]-[0046], and figures 1-2	1-10
A	CN 113343305 A (CMG FINTECH CO., LTD.) 03 September 2021 (2021-09-03) entire document	1-10
A	CN 115529118 A (SHANDONG UNIVERSITY) 27 December 2022 (2022-12-27) entire document	1-10
A	US 2019026491 A1 (MEDIASIFT LIMITED) 24 January 2019 (2019-01-24) entire document	1-10



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“D” document cited by the applicant in the international application

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

13 July 2023

Date of mailing of the international search report

25 December 2023

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
China No. 6, Xitucheng Road, Jimenqiao, Haidian District,
Beijing 100088

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2023/095566

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN	114661680	A	24 June 2022	None	
CN	113343305	A	03 September 2021	None	
CN	115529118	A	27 December 2022	None	
US	2019026491	A1	24 January 2019	None	

A. 主题的分类 G06F16/176(2019.01)i; G06F21/62(2013.01)i 按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类		
B. 检索领域 检索的最低限度文献(标明分类系统和分类号) IPC: G06F 包含在检索领域中的除最低限度文献以外的检索文献 在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用)) CNKI, CNTXT, VEN, EPTXT, USTXT, ENTXT, IEEE: 隐私, 批量, 哈希, 布谷鸟, 编码, 索引, private, batch, hash, buckoo, coding, index		
C. 相关文件		
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
A	CN 114661680 A (蓝象智联(杭州)科技有限公司) 2022年6月24日 (2022 - 06 - 24) 说明书第[0030]-[0046]段, 附图1-2	1-10
A	CN 113343305 A (招商局金融科技有限公司) 2021年9月3日 (2021 - 09 - 03) 全文	1-10
A	CN 115529118 A (山东大学) 2022年12月27日 (2022 - 12 - 27) 全文	1-10
A	US 2019026491 A1 (MEDIASIFT LIMITED) 2019年1月24日 (2019 - 01 - 24) 全文	1-10
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。		
★ 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “D” 申请人在国际申请中引证的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件		
国际检索实际完成的日期 2023年7月13日	国际检索报告邮寄日期 2023年12月25日	
ISA/CN的名称和邮寄地址 中国国家知识产权局 中国北京市海淀区蓟门桥西土城路6号 100088	授权官员 白桦 电话号码 (+86) 010-53961425	

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2023/095566

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	114661680	A	2022年6月24日	无	
CN	113343305	A	2021年9月3日	无	
CN	115529118	A	2022年12月27日	无	
US	2019026491	A1	2019年1月24日	无	