(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 101957902 B
(45)授权公告日 2014.03.26

(21)申请号 200910164561.4
(22)申请日 2009.07.20
(73)专利权人 日电(中国)有限公司
地址 100007 北京市东城区东四十条甲22号南新仓国际大厦B栋12层1222室
(72)发明人 曾珂 福岛俊一
(74)专利代理机构 中科专利商标代理有限责任公司 11021
代理人 王波波
(51)Int.Cl.
G06F 21/62(2013.01)
G06F 17/30(2006.01)

(56)对比文件
US 6052466 A,2000.04.18, 全文.
US 5677952 A,1997.10.14, 全文.
CN 1858747 A,2006.11.08, 全文.

审查员 赵洋

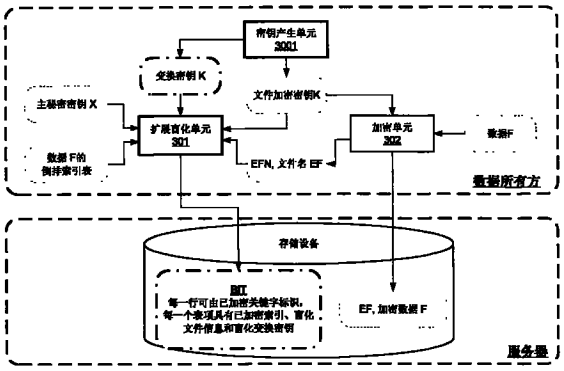
权利要求书6页 说明书16页 附图5页

(54)发明名称

扩展盲化倒排索引表产生方法和设备、联合关键字搜索方法和设备

(57)摘要

本发明提出了一种产生扩展盲化倒排索引表的设备，包括：键控行标识符产生器，用于针对倒排索引表的每一行，根据行标识符和第一密钥，产生第一和第二键控行标识符 $KRID_1$ 和 $KRID_2$；键控数据项标识符产生器，用于针对该行中的每一个数据项标识符，根据数据项标识符和第二密钥，产生键控数据项标识符 KFN；盲化文件信息项产生器，用于针对该行中的每一个数据项标识符，根据第三密钥、第一和第二键控行标识符 $KRID_1$ 和 $KRID_2$、键控数据项标识符 KFN 以及生成子，产生盲化文件信息项；以及排列器，用于通过排列行标识符和盲化文件信息项，形成扩展盲化倒排索引表的一行。本发明还提出一种产生扩展盲化倒排索引表的方法。可选地，本发明还提出一种发布扩展盲化倒排索引表的联合关键字搜索权限的方法和设备以及在扩展盲化倒排索引表中执行联合搜索的方法和设备。



CN 101957902 B

1. 一种产生扩展盲化倒排索引表的设备，包括：

键控行标识符产生器，用于针对倒排索引表的每一行，根据行标识符和第一密钥，产生第一和第二键控行标识符 $KRID_1$ 和 $KRID_2$；

键控数据项标识符产生器，用于针对该行中的每一个数据项标识符，根据数据项标识符和第二密钥，产生键控数据项标识符 KFN；

盲化文件信息项产生器，用于针对该行中的每一个数据项标识符，根据第三密钥、第一和第二键控行标识符 $KRID_1$ 和 $KRID_2$、键控数据项标识符 KFN 以及生成子，产生盲化文件信息项；以及

排列器，用于通过排列行标识符和盲化文件信息项，形成扩展盲化倒排索引表的一行。

2. 根据权利要求 1 所述的设备，还包括：

文件加密器，用于针对每一个数据项标识符，选择文件加密密钥，并利用所选的文件加密密钥来对与每一个数据项标识符相对应的文件进行加密，以获得已加密文件数据。

3. 根据权利要求 1 所述的设备，还包括：

加密索引产生器，用于针对该行中的每一个数据项标识符，选择变换密钥，并通过用变换密钥来对数据项标识符进行加密，来产生已加密索引；

加密变换密钥产生器，用于针对该行中的每一个数据项标识符，选择一种子，根据行标识符、第一密钥、第五密钥和该种子来产生中间密钥，并通过用该中间密钥来对变换密钥进行加密，来产生已加密变换密钥；以及

盲化索引产生器，用于针对该行中的每一个数据项标识符，根据该种子、第三密钥、行标识符和第一密钥来产生盲化索引，

其中扩展盲化倒排索引表中的对应项至少包括根据所述数据项标识符所产生的已加密索引、已加密变换密钥、盲化文件信息项和盲化索引。

4. 根据权利要求 3 所述的设备，其中

加密索引产生器还对已加密索引中的文件加密密钥进行加密。

5. 根据权利要求 3 所述的设备，还包括：

加密密钥产生器，用于根据行标识符和第四密钥来产生加密密钥，并且

其中加密索引产生器产生已加密索引还包括：用加密密钥对变换密钥进行加密。

6. 根据权利要求 1 ～ 5 之一所述的设备，其中

将每一行的关键字用作该行的行标识符。

7. 根据权利要求 1 ～ 5 之一所述的设备，其中

根据每一行的关键字和第四密钥来产生每一行的盲化关键字，并将盲化关键字用作该行的行标识符。

8. 根据权利要求 1 ～ 5 之一所述的设备，其中

数据项标识符具有加密的形式，被称为已加密数据项标识符，是根据文件加密密钥由明文形式产生的。

9. 一种产生扩展盲化倒排索引表的方法，包括步骤：

针对倒排索引表的每一行，根据行标识符和第一密钥，产生第一和第二键控行标识符 $KRID_1$ 和 $KRID_2$；

针对该行中的每一个数据项标识符，

根据数据项标识符和第二密钥，产生键控数据项标识符 KFN；

根据第三密钥、第一和第二键控行标识符 $KRID_1$ 和 $KRID_2$、键控数据项标识符 KFN 以及生成子，产生盲化文件信息项；以及通过排列行标识符和盲化文件信息项，形成扩展盲化倒排索引表的一行。

10. 根据权利要求 9 所述的方法，还包括步骤：

针对每一个数据项标识符，选择文件加密密钥，

其中利用所选的文件加密密钥来对与每一个数据项标识符相对应的文件进行加密，以获得已加密文件数据。

11. 根据权利要求 9 所述的方法，还包括步骤：

针对该行中的每一个数据项标识符，

选择变换密钥；

通过用变换密钥来对数据项标识符进行加密，来产生已加密索引；

选择一种子；

根据行标识符、第一密钥、第五密钥和该种子来产生中间密钥；

通过用该中间密钥来对变换密钥进行加密，来产生已加密变换密钥；以及

根据该种子、第三密钥、行标识符和第一密钥来产生盲化索引，

其中扩展盲化倒排索引表中的对应项至少包括根据所述数据项标识符所产生的已加密索引、已加密变换密钥、盲化文件信息项和盲化索引。

12. 根据权利要求 11 所述的方法，其中

文件加密密钥还被加密在已加密索引中。

13. 根据权利要求 11 所述的方法，还包括步骤：

根据行标识符和第四密钥来产生加密密钥，并且

其中产生已加密索引还包括：用加密密钥对变换密钥进行加密。

14. 根据权利要求 9 ～ 13 之一所述的方法，其中

将每一行的关键字用作该行的行标识符。

15. 根据权利要求 9 ～ 13 之一所述的方法，其中

根据每一行的关键字和第四密钥来产生每一行的盲化关键字，并将盲化关键字用作该行的行标识符。

16. 根据权利要求 9 ～ 13 之一所述的方法，其中

数据项标识符具有加密的形式，被称为已加密数据项标识符，是根据文件加密密钥由明文形式产生的。

17. 一种发布扩展盲化倒排索引表的联合关键字搜索权限的设备，包括：

键控行标识符产生器，用于针对一组 t 个查询关键字，其中 t 是大于等于 1 的整数，根据各个查询关键字和第一密钥，产生与各个查询关键字一一对应的第一和第二键控行标识符 $KRID_1$ 和 $KRID_2$；以及

联合关键字搜索权限产生器，用于针对该组 t 个查询关键字，选择一种子，并根据第三密钥、所有产生的第一和第二键控行标识符 $KRID_1$ 和 $KRID_2$、种子和生成子，来产生联合关键字搜索权限。

18. 根据权利要求 17 所述的设备，还包括：

盲化关键字产生器，用于在最开始，根据第四密钥和各个查询关键字，产生与各个查询关键字一一对应的 t 个盲化关键字，

其中在键控行标识符产生器和联合关键字搜索权限产生器中，以所述 t 个盲化关键字替换所述 t 个查询关键字。

19. 根据权利要求 17 所述的设备，其中

扩展盲化倒排索引表中的项至少包括根据同一个数据项标识符而产生的已加密索引、已加密变换密钥、盲化文件信息项和盲化索引。

20. 根据权利要求 17 所述的设备，还包括：

解密权限产生器，用于从该组 t 个查询关键字中选择一个查询关键字，根据第三密钥、所有产生的第一键控行标识符 $KRID_1$、所选的一个查询关键字、第一密钥以及第五密钥，来产生解密权限。

21. 根据权利要求 17 ～ 20 之一所述的设备，还包括：

查询表达式解析器，用于在最开始，将关键字的复杂查询表达式解析为由逻辑 OR 运算符连接的一系列关键字 AND 子查询表达式，

其中对于每个关键字 AND 子查询表达式，发布一个联合关键字搜索权限。

22. 一种发布扩展盲化倒排索引表的联合关键字搜索权限的方法，包括步骤：

针对一组 t 个查询关键字，其中 t 是大于等于 1 的整数，

根据各个查询关键字和第一密钥，产生与各个查询关键字一一对应的第一和第二键控行标识符 $KRID_1$ 和 $KRID_2$；

选择一种子；以及

根据第三密钥、所有产生的第一和第二键控行标识符 $KRID_1$ 和 $KRID_2$、种子和生成子，来产生联合关键字搜索权限。

23. 根据权利要求 22 所述的方法，其中

在最开始，根据第四密钥和各个查询关键字，产生与各个查询关键字一一对应的 t 个盲化关键字，然后在各个步骤中，以所述 t 个盲化关键字替换所述 t 个查询关键字。

24. 根据权利要求 22 所述的方法，其中

扩展盲化倒排索引表中的项至少包括根据同一个数据项标识符而产生的已加密索引、已加密变换密钥、盲化文件信息项和盲化索引。

25. 根据权利要求 22 所述的方法，还包括：

从该组 t 个查询关键字中选择一个查询关键字；以及

根据第三密钥、所有产生的第一键控行标识符 $KRID_1$、所选的一个查询关键字、第一密钥以及第五密钥，来产生解密权限。

26. 根据权利要求 22 ～ 25 之一所述的方法，还包括：

在最开始，将关键字的复杂查询表达式解析为由逻辑 OR 运算符连接的一系列关键字 AND 子查询表达式，以及

对于每个关键字 AND 子查询表达式，

分别执行后续各个步骤。

27. 一种在扩展盲化倒排索引表中执行联合搜索的设备，包括：

接收器，用于接收一组 t 个查询标识符和与之对应的联合搜索权限，其中 t 是大于等于

1 的整数 ;

行定位器,用于通过使用所述 t 个查询标识符作为行标识符,定位扩展盲化倒排索引表中分别与 t 个查询标识符相对应的 t 行 ;

匹配结果命中器,用于针对来自所定位 t 行的每一行的 t 个盲化文件信息项的每个组合,

如果从 t 个盲化文件信息项的第一子项和联合搜索权限的第一子搜索权限所获得的第一标准值与第二标准值相等,其中第二标准值是按照与第一标准值相同的方式、从 t 个盲化文件信息项的第二子项和联合搜索权限的第二子搜索权限所获得的,则确定匹配结果命中,以及

基于组合的 t 个盲化文件信息项的任何一个来记录该匹配结果。

28. 根据权利要求 27 所述的设备,其中

该组 t 个查询标识符是明文形式的一组 t 个查询关键字。

29. 根据权利要求 27 所述的设备,其中

该组 t 个查询标识符是与各个查询关键字一一对应的、根据第四密钥和各个查询关键字所产生的一组 t 个盲化关键字。

30. 根据权利要求 27 所述的设备,其中

扩展盲化倒排索引表中的项至少包括根据同一个数据项标识符而产生的已加密索引、已加密变换密钥、盲化文件信息项和盲化索引,以及

每一个记录的匹配结果包括来自组合的 t 个盲化文件信息项中的任何一个所属的项的已加密索引和已加密变换密钥、和来自所述组合的 t 个盲化文件信息项分别所属的各个项的 t 个盲化索引。

31. 根据权利要求 30 所述的设备,其中

接收器还接收解密权限,以及

所述设备还包括 :

匹配结果解密器,用于针对每一个记录的匹配结果,根据解密权限和 t 个盲化索引,产生中间密钥,利用中间密钥对已加密变换密钥进行解密以获得变换密钥,并用变换密钥对已加密索引进行解密以获得文件加密密钥和数据项标识符。

32. 根据权利要求 31 所述的设备,其中

数据项标识符具有加密形式,被称为已加密数据项标识符,所述设备还包括已加密数据项标识符解密器,用于利用文件加密密钥来对已加密数据项标识符进行解密以获得明文形式的数据项标识符。

33. 根据权利要求 31 所述的设备,还包括 :

已加密文件数据解密器,用于针对每一个记录的匹配结果,利用文件加密密钥来对与匹配结果的数据项标识符相对应的已加密文件数据进行解密。

34. 根据权利要求 27 ～ 33 之一所述的设备,其中

所述设备还包括 :查询表达式解析器,用于在最开始,将查询标识符的复杂查询表达式解析为由逻辑 OR 运算符连接的一系列查询标识符 AND 子查询表达式,

其中对于每个查询标识符 AND 子查询表达式,执行一次扩展盲化倒排索引表中的联合搜索 ;以及

所述设备还包括：合并器，用于在匹配结果命中器记录了每个 AND 子查询表达式的匹配结果之后，合并所述一系列 AND 子查询表达式的所有记录的匹配结果。

35. 根据权利要求 34 所述的设备，其中

在所述合并中，合并器消除所记录的匹配结果中的冗余。

36. 一种在扩展盲化倒排索引表中执行联合搜索的方法，包括步骤：

接收一组 t 个查询标识符和与之对应的联合搜索权限，其中 t 是大于等于 1 的整数；

通过使用所述 t 个查询标识符作为行标识符，定位扩展盲化倒排索引表中分别与 t 个查询标识符相对应的 t 行；

针对来自所定位 t 行的每一行的 t 个盲化文件信息项的每个组合，

如果从 t 个盲化文件信息项的第一子项和联合搜索权限的第一子搜索权限所获得的第一标准值与第二标准值相等，其中第二标准值是按照与第一标准值相同的方式、从 t 个盲化文件信息项的第二子项和联合搜索权限的第二子搜索权限所获得的，则确定匹配结果命中；以及

基于组合的 t 个盲化文件信息项的任何一个来记录该匹配结果。

37. 根据权利要求 36 所述的方法，其中

该组 t 个查询标识符是明文形式的一组 t 个查询关键字。

38. 根据权利要求 36 所述的方法，其中

该组 t 个查询标识符是与各个查询关键字一一对应的、根据第四密钥和各个查询关键字所产生的一组 t 个盲化关键字。

39. 根据权利要求 36 所述的方法，其中

扩展盲化倒排索引表中的项至少包括根据同一个数据项标识符而产生的已加密索引、已加密变换密钥、盲化文件信息项和盲化索引，以及

每一个记录的匹配结果包括来自组合的 t 个盲化文件信息项中的任何一个所属的项的已加密索引和已加密变换密钥、和来自所述组合的 t 个盲化文件信息项分别所属的各个项的 t 个盲化索引。

40. 根据权利要求 39 所述的方法，还包括：

接收解密权限；

针对每一个记录的匹配结果，

根据解密权限和 t 个盲化索引，产生中间密钥；

利用中间密钥对已加密变换密钥进行解密以获得变换密钥；

以及

用变换密钥对已加密索引进行解密以获得文件加密密钥和数据项标识符。

41. 根据权利要求 40 所述的方法，其中

数据项标识符具有加密形式，被称为已加密数据项标识符，

所述方法还包括步骤：

利用文件加密密钥来对已加密数据项标识符进行解密，以获得明文形式的数据项标识符。

42. 根据权利要求 40 所述的方法，还包括：

针对每一个记录的匹配结果，利用文件加密密钥来对与匹配结果的数据项标识符相对

应的已加密文件数据进行解密。

43. 根据权利要求 36 ～ 42 之一所述的方法，还包括：

在最开始，将查询标识符的复杂查询表达式解析为由逻辑 OR 运算符连接的一系列查询标识符 AND 子查询表达式；

对于每个查询标识符 AND 子查询表达式，

分别执行后续各个步骤；以及

在记录了每个 AND 子查询表达式的匹配结果之后，合并所述一系列 AND 子查询表达式的所有记录的匹配结果。

44. 根据权利要求 43 所述的方法，其中

在所述合并步骤中，消除所记录的匹配结果中的冗余。

扩展盲化倒排索引表产生方法和设备、联合关键字搜索方法和设备

技术领域

[0001] 本发明涉及计算机通信网络安全领域，更具体地，涉及一种包括扩展盲化倒排索引表 (EBIT) 的联合关键字搜索方法和设备。

背景技术

[0002] 数据存储外包是当前互联网上的一种趋势，即以全球存档服务来存储数据，而不是使用自身的本地存储器来存储数据。现在，基于互联网的在线存档服务为其终端用户提供大量的存储空间，其终端用户包括个人用户和企业。存在提供各种用户数据的存储的存档服务。例如，Amazon Simple Storage Service(Amazon S3)（参考文献 [1]）提供了一种网络服务接口，可用于存储和检索不限量的分类数据，以 GB/ 月和数据传输量来计费。网络上还存在提供特定数据类型的存储的其它存档服务，尤其是敏感数据类型，例如健康记录。例如，Google Health（参考文献 [2]）和 Microsoft HealthVault（参考文献 [3]）二者均提供个人健康信息综合服务，有助于其用户将分离的健康记录合并为一份综合的档案。

[0003] 尽管这些存档服务带来了便捷和易用的优点，但是它们也引起了对安全性的深度担忧。尽管所有这些服务提供商都提出了适当的书面安全和隐私策略，并且采取一些信息安全和系统安全措施来执行这些策略，但是，用户仅仅依赖于存档服务提供商来确保其数据安全和隐私是危险的。无疑服务提供商可能会无法适当地执行它们的书面安全和隐私策略。

[0004] 以存储客户的信用卡数据的企业为例。在 2008 年 6 月，BBC 新闻报道了服装厂 Cotton Traders 的多达 3.8 万名客户的信用卡细目被盗（参考文献 [4]）。这种情况并不少见，而且也不是最严重的事件。Securityfocus.com（参考文献 [5]）报道了在 2005 年 7 月至 2007 年 1 月间，未知攻击者入侵 TJX 公司的计算机交易处理系统，盗取了至少 4560 万张信用卡的数据。

[0005] 信用卡信息被认为至少与分类数据或健康记录同样敏感。因而，可以推断出，存储信用卡信息的公司具有适当的书面安全和隐私策略并且应该运用了表面上有力的安全措施来执行其策略。这些安全措施至少应该与用于保护分类数据或健康记录的安全措施一样有力。由于信用卡信息被披露的多次报道，同时注意到大量用户数据的高价值，因此，没有理由坚持认为存档服务提供商所存储的数据不会被盗和被暴露。

[0006] 无论如何，存在一种应对数据安全入侵的简单对策，即在输出敏感数据之前对其加密。结果，即使存档服务受到危害，所暴露的也只是大量密文，攻击者无法从中获利。然而，这种简单对策的代价是可用性。具体地说，难以搜索输出到外部的数据。例如，如果健康记录的所有方对健康记录进行了加密，则允许授权用户搜索健康有关信息的 Microsoft Live Search Health(Microsoft HealthVault 的搜索组件）无法工作。（当然，我们假定数据的所有方对其隐私充分关注，因此不会与 Microsoft 共享他们的解密密钥。）

[0007] 我们所关注的系统有三方，即数据所有方、服务器和搜索方。数据所有方对其数据

文件进行索引、对其数据文件进行加密并将索引和文件输出到服务器。服务器存储加密的文件及其索引（索引表），并提供对加密文件的搜索。搜索服务器的搜索方通常不是数据所有方自己，但是，当然，搜索方也可以是数据所有方自己。为了能够搜索加密数据，搜索方需要获得从数据所有方发出的搜索权限 (SC)，并且搜索方需要将 SC 提交给服务器。服务器可以通过将 SC 运用于索引来搜索加密数据。除了 SC 之外，搜索方还需要获得数据所有方发出的解密权限 (DC)。在从服务器接收到搜索结果时，搜索方将使用 DC 来对搜索结果进行解密，从而将数据文件恢复为明文。

[0008] 一些基本的安全要求包括：

[0009] 1) 服务器不知道搜索方查找什么，例如，如果搜索方正在搜索包含关键字“网络”的文献，服务器应该不知道。

[0010] 2) 搜索方无法根据经验伪造搜索权限，例如，如果搜索方曾经被发给了搜索包含关键字“网络”的文献的 SC，他应该不能够制造针对关键字“网”或“络”的 SC。这同样适用于服务器，即使搜索方与服务器串通一气。

[0011] 3) 解密权限与 SC 唯一关联，例如，如果 SC 允许搜索包含关键字“网络”的文献，DC 则仅能够对该特定 SC 的搜索结果进行解密。这同样适用于服务器，即使搜索方与服务器串通，即，服务器也许对其所存储的所有加密文件尝试使用 DC，但是除了 SC 的搜索结果之外，对于其他加密文件不会发生效力。

[0012] 除了上述安全要求之外，还有效率要求，例如 SC 的大小、索引的大小以及搜索所花费的时间等。

[0013] 自从 Song 等人（参考文献 [6]）首次提出了关于如何有效地对加密数据进行关键字搜索的问题以来，加密数据的搜索引起了广泛的关注。

[0014] 加密数据的搜索是不同领域的技术的融合，因而具有不同的分类标准。

[0015] 1) 从加密技术角度看，在秘密密钥设置中和公共密钥设置中考虑关键字搜索的加密，在秘密密钥设置中这被称为可搜索对称加密 (SSE)（参考文献 [6]），在公共密钥设置中这被称为公共密钥加密搜索 (PEKS)（参考文献 [7]）。然而，值得注意的是，通过使公共密钥保密，任何 PEKS 方案都简单地在 SSE 设置中也行得通。

[0016] 2) 从索引技术角度看，在正向索引设置中和倒排索引设置中考虑关键字索引的加密，在正向索引设置中这被称为盲化正向索引表 (BFT)，在倒排索引设置中这被称为盲化倒排索引表 (BIT)。

[0017] 3) 从搜索权限角度看，在单个关键字搜索 (SKS) 和联合关键作搜索 (CKS) 中考虑关键字搜索的加密。

[0018] 4) 从搜索关键字角度看，在域特定关键字 (DSK) 和非限定域关键字 (DFK) 中考虑关键字搜索的加密。

[0019] 就我们所知，现有技术大多数符合 SSE、BFT、SKS 和 DFK。也就是说，现有技术是秘密密钥设置的、基于盲化正向索引表、利用非限定域关键字、仅能够单个关键字搜索。然而，存在一些另外情况。

[0020] 参考文献 [8] 公开了秘密密钥设置 (SSE) 的方案，可以以非限定域关键字 (DFK) 实现盲化倒排索引表 (BIT) 的单关键字搜索 (SKS)。

[0021] 注意，明文倒排索引表在处理联合关键字搜索方面非常有效。使用明文倒排索引

表，仅仅需要针对每个关键字获取一组匹配文件，然后找到所有文件组的交集。例如，如果文件 F_1 和 F_2 符合关键字 KW_1 而文件 F_1 和 F_3 符合关键字 KW_2，则显然，“KW_1 和 KW_2”下的联合关键字搜索的结果得到文件 F1。

[0022] 具体地，图 1 和图 2 以两个阶段示出了参考文献 [8] 的详细过程，即索引阶段（图 1）和搜索阶段（图 2）。参考图 1 和图 2，数据所有方、搜索方和服务器的各个单元如下：

[0023] ◆密钥产生单元 1001 产生随机文件加密密钥。

[0024] ◆盲化单元 101 以主秘密密钥、文件加密密钥、加密文件的文件名 (EFN) 和（明文）倒排索引表作为输入，输出盲化倒排索引表 (BIT)。

[0025] ◆加密单元 102 以文件加密密钥和数据作为输入，输出加密数据 (EF)。

[0026] ◆权限发布单元 201 以秘密密钥和关键字作为输入，输出搜索权限 (SC) 和解密权限 (DC)。

[0027] ◆单关键字匹配单元 203 以 SC 和 BIT 作为输入，输出包括 BIT 中匹配行的相应已加密索引 ef_{ij} 的 BIT 匹配结果。

[0028] ◆ BIT 解密单元 202 以 DC 和 BIT 匹配结果作为输入，输出匹配 EFN 及其相应解密密钥。

[0029] ◆ EF 获取单元 204 以匹配 EFN 作为输入，并从服务器中检索出 EF。

[0030] ◆ EF 加密单元 205 以得到的 EF 和文件解密密钥作为输入，输出已解密数据 F。

[0031] 表 1 示例明文倒排索引表

[0032]

关键字			
KW_1	FN_1	FN_2	FN_3
KW_2	FN_1	FN_2	
KW_3	FN_2	FN_4	
.....	...	...	...

[0033] 在表 1 中，每一行可由不同的关键字 KW_i 标识。跟随 KW_i 的是包含 KW_i 的所有文件 FN_u。很容易看出，明文倒排索引表的联合关键字搜索是容易的。例如，“KW_2 和 KW_3”的联合关键字搜索的唯一结果是 FN_2。

[0034] 根据参考文献 [8]，盲化单元 101 输出的 BFT 如下表 2 所示。

[0035] 表 2 示例盲化倒排索引表

[0036]

关键字			
EK_1	ef_{11}	ef_{12}	ef_{13}
EK_2	ef_{21}	ef_{22}	
EK_3	ef_{31}	ef_{32}	
……	…	…	…

[0037] 表 1 中的每个 KW_i 由 EK_i 代替，EK_i 是使用数据所有方的主秘密密钥作为密钥的 KW_i 的加密散列或键控散列 (keyed hash)。

[0038] 此外，表 1 中的每个 FN_u 由 ef_{ij}（已加密索引）代替。为了计算 ef_{ij}，数据所有方首先针对表的每一行，使用主秘密密钥和 KW_i 来产生不同的解密密钥 ek_i。ef_{ij} 实际上是使用 ek_i 的相应 EN_u 的（对称）加密。显然，BIT 的直接联合关键字搜索是困难的。即使从明文倒排索引表看出 FN_2 是关键字“KW_2 和 KW_3”的搜索的结果，由于 ef_{22} 和 ef_{32} 是经过不可分辨性的基本安全要求（也被称为“语义安全 semanticsecurity”）的（对称）加密的输出，所以也不可能（通过计算）知道 ef_{22} 和 ef_{32} 是否是相同文件的加密。

[0039] 简而言之，参考文献 [8] 公开了以下方法。

[0040] 密钥产生：

[0041] 选择对称加密算法 $Enc_{key}(msg)$，以 key 和明文消息 msg 作为输入，输出密文。在 msg 是密文的情况下输出明文；

[0042] 选择安全单向键控散列函数 $H_{key}(msg)$，以密钥 key 和明文消息 msg 作为输入，输出摘要；

[0043] 选择主秘密密钥 msk。

[0044] BIT 产生：

[0045] 以明文倒排索引表作为输入。不失一般性地，假设表 1 的明文倒排索引表是输入。

[0046] 选择文件加密密钥 fk_u，并按照 $EF_u = Enc_{fk_u}(F_u)$ 加密 FN_u 的内容，即 F_u。此外，按照 $CFN_u = Enc_{fk_u}(FN_u)$ 加密 FN_u 以获得加密文件名。

[0047] 计算 $EK_i = Enc_{msk}(KW_i)$ 和 $ef_{ij} = Enc_{H_{msk}(KW_i)}(fk_u, CFN_u)$。

[0048] 如表 2 所示，以 EK_i 和 ef_{ij} 填充 BIT。

[0049] 注意，处于安全考虑，对称加密必须以起始向量作为附加输入。例如，计算 $EK_i = Enc_{msk}(KW_i)$ 的起始向量可以计算为 $IV_i = H_{msk}(KW_i \| \text{“IV”})$，其中“|”表示级联。此外，存在以上简述的其它可选方案。例如，可以按照 $EK_i = H_{msk}(KW_i \| \text{“0”})$ 计算 EK_i，可以按照 $ef_{ij} = Enc_{H_{msk}(KW_i \| \text{“1”})}(fk_u, CFN_u)$ 计算 ef_{ij}。关于更详细的说明，请查阅参考文献 [8]。

[0050] 产生 SC 和 DC：

[0051] 在搜索方想要搜索关键字 KW_i 时，按照 $SC = Enc_{msk}(KW_i)$ 计算搜索权限 (SC)，并按照 $DC = H_{msk}(KW_i)$ 计算解密权限 (DC)。

[0052] 搜索：

[0053] 在接收到SC时，服务器定位BIT中以SC作为起始的行，并将行中 ef_{ij} 的返回给搜索方。

[0054] 解密搜索结果：

[0055] 搜索方使用DC来解密 $ef_{ij} = Enc_{H_{msk}(KW_i)}(fk_u, CFN_u)$，并获得 fk_u 和 CFN_u。

[0056] 搜索方向服务器请求 CFN_u 的内容，即 EF_u。

[0057] 最后，搜索方使用 fk_u 来解密 CFN_u 和 EF_u，获得 FN_u 及其内容 F_u。

[0058] 搜索方（例如正在查找包含关键字“KW_1”的文件的人）将进行以下动作：

[0059] 1）向数据搜索方请求域“KW_1”相对应的搜索权限和解密权限；

[0060] 2）以搜索权限查询服务器，从服务器获得匹配结果；

[0061] 3）对匹配结果进行解密，获得加密的文件名和文件加密密钥；

[0062] 4）将加密的文件名提交给服务器，并获得加密的文件内容；

[0063] 5）利用文件加密密钥来解密加密文件名和加密文件内容，最终产生明文的文件名和文件内容。

[0064] 然而，明文倒排索引表的联合关键字搜索的简单性并不直接适用于盲化倒排索引表。为了保密，BIT的每一项都被加密以使得它们不可（通过计算）彼此区分开。加密的该特性被称为语义安全或不可分辨性。例如，如果使用对称加密机制来对BIT的表项进行加密，则必须使用不同的加密密钥或不同的初始向量来加密每一表项。因此，即使在明文倒排索引表中，一个文件在被转换为加密倒排索引表之后，在两行中出现（即与两个关键字匹配），因此难以通过计算分辨出。因此，即使简单地模拟明文倒排索引表的联合关键字搜索方法，也不可能进行基于BIT的联合关键字搜索。

[0065] 因此，参考文献[8]不能够处理联合关键字搜索。

发明内容

[0066] 鉴于现有技术的上述缺点，本发明提出了一种扩展盲化倒排索引表产生方法和设备、以及一种秘密密钥设置下的联合关键字搜索方法和设备。

[0067] 根据本发明的第一方案，提出一种产生扩展盲化倒排索引表的设备，包括：键控行标识符产生器，用于针对倒排索引表的每一行，根据行标识符和第一密钥，产生第一和第二键控行标识符 $KRID_1$ 和 $KRID_2$；键控数据项标识符产生器，用于针对该行中的每一个数据项标识符，根据数据项标识符和第二密钥，产生键控数据项标识符KFN；盲化文件信息项产生器，用于针对该行中的每一个数据项标识符，根据第三密钥、第一和第二键控行标识符 $KRID_1$ 和 $KRID_2$、键控数据项标识符KFN以及生成子，产生盲化文件信息项；以及排列器，用于通过排列行标识符和盲化文件信息项，形成扩展盲化倒排索引表的一行。

[0068] 优选地，所述设备还可包括：文件加密器，用于针对每一个数据项标识符，选择文件加密密钥，并利用所选的文件加密密钥来对与每一个数据项标识符相对应的文件进行加密，以获得已加密文件数据。

[0069] 优选地，所述设备还可包括加密索引产生器，用于针对该行中的每一个数据项标识符，选择变换密钥，并通过用变换密钥来对数据项标识符进行加密，来产生已加密索引；已加密变换密钥产生器，用于针对该行中的每一个数据项标识符，选择一种子，根据行标识

符、第一密钥、第五密钥和该种子来产生中间密钥，并通过用该中间密钥来对变换密钥进行加密，来产生已加密变换密钥；以及盲化索引产生器，用于针对该行中的每一个数据项标识符，根据该种子、第三密钥、行标识符和第一密钥来产生盲化索引，其中扩展盲化倒排索引表中的对应项至少包括根据所述数据项标识符所产生的已加密索引、已加密变换密钥、盲化文件信息项和盲化索引。

[0070] 更优选地，加密索引产生器还可对已加密索引中的文件加密密钥进行加密。

[0071] 更优选地，所述设备还可包括加密密钥产生器，用于根据行标识符和第四密钥来产生加密密钥，其中加密索引产生器产生已加密索引还包括：用加密密钥对变换密钥进行加密。

[0072] 优选地，将每一行的关键字用作该行的行标识符。或者可选地，根据每一行的关键字和第四密钥来产生每一行的盲化关键字，并将盲化关键字用作该行的行标识符。

[0073] 优选地，数据项标识符具有加密的形式，被称为已加密数据项标识符，是根据文件加密密钥由明文形式产生的。

[0074] 根据本发明的第二方案，提出一种产生扩展盲化倒排索引表的方法，包括步骤：针对倒排索引表的每一行，根据行标识符和第一密钥，产生第一和第二键控行标识符 $KRID_1$ 和 $KRID_2$；针对该行中的每一个数据项标识符，根据数据项标识符和第二密钥，产生键控数据项标识符 KFN；根据第三密钥、第一和第二键控行标识符 $KRID_1$ 和 $KRID_2$、键控数据项标识符 KFN 以及生成子，产生盲化文件信息项；以及通过排列行标识符和盲化文件信息项，形成扩展盲化倒排索引表的一行。

[0075] 优选地，所述方法还可包括步骤：针对每一个数据项标识符，选择文件加密密钥，其中利用所选的文件加密密钥来对与每一个数据项标识符相对应的文件进行加密，以获得已加密文件数据。

[0076] 优选地，所述方法还可包括步骤：针对该行中的每一个数据项标识符，选择变换密钥；通过用变换密钥来对数据项标识符进行加密，来产生已加密索引；选择一种子；根据行标识符、第一密钥、第五密钥和该种子来产生中间密钥；通过用该中间密钥来对变换密钥进行加密，来产生已加密变换密钥；以及根据该种子、第三密钥、行标识符和第一密钥来产生盲化索引，其中扩展盲化倒排索引表中的对应项至少包括根据所述数据项标识符所产生的已加密索引、已加密变换密钥、盲化文件信息项和盲化索引。

[0077] 更优选地，文件加密密钥还被加密在已加密索引中。

[0078] 更优选地，所述方法还可包括步骤：根据行标识符和第四密钥来产生加密密钥，并且其中产生已加密索引还包括步骤：用加密密钥对变换密钥进行加密。

[0079] 优选地，将每一行的关键字用作该行的行标识符。或者可选地，根据每一行的关键字和第四密钥来产生每一行的盲化关键字，并将盲化关键字用作该行的行标识符。

[0080] 优选地，数据项标识符具有加密的形式，被称为已加密数据项标识符，是根据文件加密密钥由明文形式产生的。

[0081] 根据本发明的第三方案，提出一种发布扩展盲化倒排索引表的联合关键字搜索权限的设备，包括：键控行标识符产生器，用于针对一组 t 个查询关键字，根据各个查询关键字和第一密钥，产生与各个查询关键字一一对应的第一和第二键控行标识符 $KRID_1$ 和 $KRID_2$；以及联合关键字搜索权限产生器，用于针对该组 t 个查询关键字，选择一种子，并根

据第三密钥、所有产生的第一和第二键控行标识符 $KRID_1$ 和 $KRID_2$、种子和生成子，来产生联合关键字搜索权限。

[0082] 优选地，所述设备还可包括：盲化关键字产生器，用于在最开始，根据第四密钥和各个查询关键字，产生与各个查询关键字一一对应的 t 个盲化关键字，其中在键控行标识符产生器和联合关键字搜索权限产生器中，以所述 t 个盲化关键字替换所述 t 个查询关键字。

[0083] 优选地，扩展盲化倒排索引表中的项至少包括根据同一个数据项标识符而产生的已加密索引、已加密变换密钥、盲化文件信息项和盲化索引。

[0084] 优选地，所述设备还可包括：解密权限产生器，用于从该组 t 个查询关键字中选择一个查询关键字，根据第三密钥、所有产生的第一键控行标识符 $KRID_1$、所选的一个查询关键字、第一密钥以及第五密钥，来产生解密权限。

[0085] 优选地，所述设备还可包括查询表达式解析器，用于在最开始，将关键字的复杂查询表达式解析为由逻辑 OR 运算符连接的一系列关键字 AND 子查询表达式，其中对于每个关键字 AND 子查询表达式，发布一个联合关键字搜索权限。

[0086] 根据本发明的第四方案，提出一种发布扩展盲化倒排索引表的联合关键字搜索权限的方法，包括步骤：针对一组 t 个查询关键字，根据各个查询关键字和第一密钥，产生与各个查询关键字一一对应的第一和第二键控行标识符 $KRID_1$ 和 $KRID_2$；选择一种子；以及根据第三密钥、所有产生的第一和第二键控行标识符 $KRID_1$ 和 $KRID_2$、种子和生成子，来产生联合关键字搜索权限。

[0087] 优选地，在最开始，根据第四密钥和各个查询关键字，产生与各个查询关键字一一对应的 t 个盲化关键字，然后在各个步骤中，以所述 t 个盲化关键字被用于替换所述 t 个查询关键字。

[0088] 优选地，扩展盲化倒排索引表中的项至少包括根据同一个数据项标识符而产生的已加密索引、已加密变换密钥、盲化文件信息项和盲化索引。

[0089] 优选地，所述方法还可包括：从该组 t 个查询关键字中选择一个查询关键字；根据第三密钥、所有产生的第一键控行标识符 $KRID_1$、所选的一个查询关键字、第一密钥以及第五密钥，来产生解密权限。

[0090] 优选地，所述方法还可包括：在最开始，将关键字的复杂查询表达式解析为由逻辑 OR 运算符连接的一系列关键字 AND 子查询表达式，以及对于每个关键字 AND 子查询表达式，分别执行后续各个步骤。

[0091] 根据本发明的第五方案，提出一种在扩展盲化倒排索引表中执行联合搜索的设备，包括：接收器，用于接收一组 t 个查询标识符和与之对应的联合搜索权限；行定位器，用于通过使用所述 t 个查询标识符作为行标识符，定位扩展盲化倒排索引表中分别与 t 个查询标识符相对应的 t 行；匹配结果命中器，用于针对来自所定位 t 行的每一行的 t 个盲化文件信息项的每个组合，如果从 t 个盲化文件信息项的第一子项和联合搜索权限的第一子搜索权限所获得的第一标准值与第二标准值相等，其中第二标准值是按照与第一标准值相同的方式、从 t 个盲化文件信息项的第二子项和联合搜索权限的第二子搜索权限所获得的，则确定匹配结果命中，并基于组合的 t 个盲化文件信息项的任何一个来记录该匹配结果。

[0092] 优选地，该组 t 个查询标识符是明文形式的一组 t 个查询关键字。或者可选地，该

组 t 个查询标识符是与各个查询关键字一一对应的、根据第四密钥和各个查询关键字所产生的一组 t 个盲化关键字。

[0093] 优选地，扩展盲化倒排索引表中的项至少包括根据同一个数据项标识符而产生的已加密索引、已加密变换密钥、盲化文件信息项和盲化索引，每一个记录的匹配结果包括来自组合的 t 个盲化文件信息项中的任何一个所属的项的已加密索引和已加密变换密钥、和来自组合的 t 个盲化文件信息项分别所属的各个项的 t 个盲化索引。

[0094] 更优选地，接收器还可接收解密权限，所述设备还包括：匹配结果解密器，用于针对每一个记录的匹配结果，根据解密权限和 t 个盲化索引，产生中间密钥，利用中间密钥对已加密变换密钥进行解密以获得变换密钥，并用变换密钥对已加密索引进行解密以获得文件加密密钥和数据项标识符。

[0095] 更优选地，数据项标识符具有加密形式，被称为已加密数据项标识符，所述设备还包括已加密数据项标识符解密器，用于利用文件加密密钥来对已加密数据项标识符进行解密以获得明文形式的数据项标识符。

[0096] 优选地，所述设备还包括：已加密文件数据解密器，用于针对每一个记录的匹配结果，利用文件加密密钥来对与匹配结果的数据项标识符相对应的已加密文件数据进行解密。

[0097] 优选地，所述设备还可包括：查询表达式解析器，用于在最开始，将查询标识符的复杂查询表达式解析为由逻辑 OR 运算符连接的一系列查询标识符 AND 子查询表达式，其中对于每个查询标识符 AND 子查询表达式，执行一次扩展盲化倒排索引表中的联合搜索；并且所述设备还可包括：合并器，用于在匹配结果命中器记录了每个 AND 子查询表达式的匹配结果之后，合并器合并所述一系列 AND 子查询表达式的所有记录的匹配结果。

[0098] 更优选地，在所述合并中，合并器可消除记录的匹配结果中的冗余。

[0099] 根据本发明的第六方案，提出一种在扩展盲化倒排索引表中执行联合搜索的方法，包括步骤：接收一组 t 个查询标识符和与之对应的联合搜索权限；通过使用所述 t 个查询标识符作为行标识符，定位扩展盲化倒排索引表中分别与 t 个查询标识符相对应的 t 行；针对来自所定位 t 行的每一行的 t 个盲化文件信息项的每个组合，如果从 t 个盲化文件信息项的第一子项和联合搜索权限的第一子搜索权限所获得的第一标准值与第二标准值相等，其中第二标准值是按照与第一标准值相同的方式、从 t 个盲化文件信息项的第二子项和联合搜索权限的第二子搜索权限所获得的，则确定匹配结果命中；以及基于组合的 t 个盲化文件信息项的任何一个来记录该匹配结果。

[0100] 优选地，该组 t 个查询标识符是明文形式的一组 t 个查询关键字。或者可选地，该组 t 个查询标识符是与各个查询关键字一对应的、根据第四密钥和各个查询关键字所产生的一组 t 个盲化关键字。

[0101] 优选地，扩展盲化倒排索引表中的项至少包括根据同一个数据项标识符而产生的已加密索引、已加密变换密钥、盲化文件信息项和盲化索引，每一个记录的匹配结果包括来自组合的 t 个盲化文件信息项中的任何一个所属的项的已加密索引和已加密变换密钥、和来自组合的 t 个盲化文件信息项分别所属的各个项的 t 个盲化索引。

[0102] 更优选地，所述方法还可包括：接收解密权限；针对每一个记录的匹配结果，根据解密权限和 t 个盲化索引，产生中间密钥；利用中间密钥对已加密变换密钥进行解密以获

得变换密钥；以及用变换密钥对已加密索引进行解密以获得文件加密密钥和数据项标识符。

[0103] 更优选地，数据项标识符具有加密形式，被称为已加密数据项标识符，所述方法还包括步骤：利用文件加密密钥来对已加密数据项标识符进行解密，以获得明文形式的数据项标识符。

[0104] 优选地，所述方法还包括：针对每一个记录的匹配结果，利用文件加密密钥来对与匹配结果的数据项标识符相对应的已加密文件数据进行解密。

[0105] 优选地，所述方法还可包括：在最开始，将查询标识符的复杂查询表达式解析为由逻辑 OR 运算符连接的一系列查询标识符 AND 子查询表达式；对于每个查询标识符 AND 子查询表达式，分别执行后续各个步骤；以及在记录了每个 AND 子查询表达式的匹配结果之后，合并所述一系列 AND 子查询表达式的所有记录的匹配结果。

[0106] 更优选地，在所述合并步骤中，消除所记录的匹配结果中的冗余。

[0107] 与最接近的现有技术相比，本发明的积极效果包括：

[0108] (I) 能够对盲化倒排索引表进行联合关键字搜索；以及

[0109] (II) 能够以合理的计算和存储开销来进行联合关键字搜索。

附图说明

[0110] 结合附图，根据下面对本发明的非限制性实施例的详细描述，本发明的上述及其他目的、特征和优点将变得更加清楚，附图中：

[0111] 图 1 是示出了根据参考文献 [8]、在索引阶段工作的各个单元的方框图；

[0112] 图 2 是示出了根据参考文献 [8]、在搜索阶段工作的各个单元的方框图；

[0113] 图 3 是示出了根据本发明第一实施例、在索引阶段工作的各个单元的方框图；

[0114] 图 4 是示出了根据本发明第一实施例、在搜索阶段工作的各个单元的方框图；以及

[0115] 图 5 是示出了根据本发明第二实施例、在搜索阶段工作的各个单元的方框图。

具体实施方式

[0116] 下面，根据附图描述本发明。在以下描述中，一些具体实施例仅用于描述目的，而不应该理解为对本发明有任何限制，而只是本发明的示例。省略了常规结构或构造，以免导致对本发明的理解不清楚。

[0117] [第一实施例]

[0118] 根据本发明的第一实施例，图 3 和图 4 以两个阶段示出了所提出的联合关键字搜索方案的详细过程，即索引阶段 (图 3) 和搜索阶段 (图 4)。在所提出的联合关键字搜索方案中，涉及扩展盲化倒排索引表 (EBIT)。参考图 3 和图 4，数据所有方、搜索方和服务器的各个单元如下：

[0119] ◆密钥产生单元 3001 产生随机文件加密密钥和变换密钥。

[0120] ◆扩展盲化单元 301 以主秘密密钥、文件加密密钥、加密文件的文件名 (EFN)、变换密钥和 (明文) 倒排索引表作为输入，输出扩展盲化倒排索引表 (EBIT)。除了已加密索引之外，EBIT 中的每一个表项还包含盲化文件信息和盲化变换索引。

[0121] ◆扩展权限发布单元 401 以主秘密密钥和关键字作为输入，输出扩展搜索权限 (SC)，扩展搜索权限 (SC) 允许对每个关键字进行单关键字搜索和联合关键字搜索。此外，还输出扩展解密权限 (DC)，扩展解密权限 (DC) 仅能够对联合搜索结果进行解密。

[0122] ◆联合关键字匹配单元 407 以单关键字匹配单元 403 的输出以及扩展 SC 作为输入，输出第二级 EBIT 匹配结果，第二级 EBIT 匹配结果仅包含联合关键字搜索结果的已加密索引和盲化变换密钥。

[0123] ◆变换密钥恢复单元 406 以盲化变换密钥和扩展 DC 作为输入，输出变换密钥。

[0124] ◆ EBIT 解密单元 402 以变换密钥和已加密索引作为输入，输出匹配的 EFN 及其对应解密密钥。

[0125] ◆所有其它单元 (302、402、404 和 405) 与在本说明书“背景技术”部分所描述的单元 (102、202、204 和 205) 执行相同的操作，因此为了简明起见，在此省略了对这些单元的详细说明。

[0126] 场景描述

[0127] 首先给出所提出的包括扩展盲化倒排索引表 (EBIT) 的联合关键字搜索方案的概述。

[0128] 不失一般性地，以示例的明文倒排索引表开始。示例的明文倒排索引表如表 1，与在本说明书“背景技术”部分中所给出的表 1 相同。

[0129] 表 1 示例明文倒排索引表

[0130]

关键字			
KW_1	FN_1	FN_2	FN_3
KW_2	FN_1	FN_2	
KW_3	FN_2	FN_4	
……	…	…	…

[0131] 在表 1 中，每一行可由不同的关键字 KW_i 标识。跟随 KW_i 的是包含 KW_i 的所有文件 FN_u。很容易看出，明文倒排索引表的联合关键字搜索是容易的。例如，“KW_2 和 KW_3”的联合关键字搜索的唯一结果是 FN_2。

[0132] 下面描述使用新颖 EBIT 方案的针对 BIT 的联合关键字搜索，其中上表 1 被扩展盲化单元 301 转换为下表 3。

[0133] 表 3 示例的扩展盲化倒排索引表

[0134]

关键字			
EK_1	$ef_{11}, EF_{11}, W_{11}, V_{11}$	$ef_{12}, EF_{12}, W_{12}, V_{12}$	$ef_{13}, EF_{13}, W_{13}, V_{13}$
EK_2	$ef_{21}, EF_{21}, W_{21}, V_{21}$	$ef_{22}, EF_{22}, W_{22}, V_{22}$	
EK_3	$ef_{31}, EF_{31}, W_{31}, V_{31}$	$ef_{32}, EF_{32}, W_{32}, V_{32}$	
.....	...	...	...

[0135] 表 3 中的 EK_i 与表 2 中的 EK_i 完全相同。

[0136] 表 3 中的 ef_{ij} 与表 2 中的 ef_{ij} 稍有不同。为了计算 ef_{ij}，针对表的每一行，数据所有方首先使用主秘密密钥和 KW_i 来产生不同的加密密钥 eki。然后，数据所有方针对每一个 FN_u 而选择变换密钥 tk_{ij}。最后，ef_{ij} 具有两个部分，即 $ef_{ij.A}$ 和 $ef_{ij.B}$。$ef_{ij.A}$ 是使用 ek_i 的 tk_{ij} 的（对称）加密，$ef_{ij.B}$ 是使用 tk_{ij} 的对应 fk_u 和 CFN_u 的（对称）加密。

[0137] 与表 2 相比，表 3 中的 W_{ij}（盲化文件信息）、EF_{ij}（已加密变换密钥）和 V_{ij}（盲化索引）是新引入的。稍后，将对这三个数据项进行详细说明。接下来，概述搜索方和服务器如何基于所产生的 EBIT 来实现联合关键字搜索。

[0138] 搜索方（例如，正在查找包含关键字“KW_2 和 KW_3”的文件的人）从数据所有方接收到所需的 SC 和 DC。在该特定示例中，除了其它数据之外，SC 还包含 EK_2 和 EK_3。因此，服务器在从搜索方接收到 SC 时可以快速地定位以 EK_2 和 EK_3 作为起始的两行。接下来，使用 SC 中的其它数据，服务器可比较盲化文件信息 W_{2u} 和 W_{3v} 是否是相同文件信息的加密。在该特定示例中，服务器最后获得只有 W_{22} 和 W_{31} 是相同文件信息的加密的搜索结果，这表示：(EF_{22}, V_{22}) 和 (EF_{31}, V_{31}) 是关键字 EK_2 和 EK_3 的联合搜索的结果。

[0139] 由于 (EF_{22}, V_{22}) 和 (EF_{31}, V_{31}) 的解密是相同的文件，所以仅将已加密变换密钥中的一个返回给搜索方。在该特定示例中，不失一般性地，假设搜索方最后从服务器接收到 (EF_{22}, V_{22}, V_{31}) 作为搜索结果。搜索结果还包括 $ef_{22.B}$。

[0140] 注意，作为搜索结果，V_{22} 和 V_{31} 都需要。这是为了满足 DC 仅能够对该特定 SC 的搜索结果进行解密的安全要求。DC 对于 SC 的搜索结果之外的搜索结果均不起作用。

[0141] 在接收到 (EF_{22}, V_{22}, V_{31}) 时，搜索方首先使用 DC、V_{22} 和 V_{31} 恢复中间密钥 ink_{22}。然后，搜索方可以使用 ink_{22} 对 EF_{22} 进行解密，得到变换密钥 tk_{22}。由于 $ef_{22.B}$ 是使用 tk_{22} 的 fk_2 和 CFN_2 的加密，所以搜索方最终获得明文的内容 FN_2。

[0142] [第二实施例]

[0143] 上述第一实施例仅考虑 AND 查询，并未考虑组合了 AND 和 OR 的复杂查询表达式，例如“(KW_1OR KW_2)AND(KW_3OR KW_4)”。第二实施例可以处理这种复杂查询表达式。

[0144] 根据本发明的第二实施例，图 5 示出了所提出的联合关键字搜索方案在搜索阶段（图 5）的详细过程。在所提出的联合关键字搜索方案中，涉及扩展盲化倒排索引表 (EBIT)，并且索引阶段与本发明的第一实施例（图 3）类似。参考图 3 和图 5，数据所有方、搜索方和服务器的各个单元如下：

[0145] ◆除了查询解析单元608和后处理单元609之外的所有单元与第一实施例中的单元执行相同的操作，因此为了简明起见，在此省略了对这些单元的详细说明。

[0146] ◆查询解析单元608以（可能复杂的）联合关键字查询作为输入，将其解析为一系列AND子查询表达式。

[0147] ◆后处理单元609以匹配的EFN作为输入，消除其中冗余的EFN。

[0148] 场景描述

[0149] 在第二实施例中，搜索方和/或数据所有方还附加地配备有查询表达式解析单元608（具体地，在图5中，搜索方配备有查询表达式解析单元），并且搜索方还附加地配备有后处理单元609。

[0150] 在第二实施例中，查询表达式解析单元608将复杂查询表达式解析为一系列AND子查询表达式。例如，“$(KW_1OR\ KW_2)AND(KW_3ORKW_4)$”可被解析为“$(KW_1AND\ KW_3)OR(KW_1AND\ KW_4)OR(KW_2AND\ KW_3)OR(KW_2AND\ KW_4)$”。很容易可以看出，可以通过逐一进行AND子查询表达式，实现复杂查询表达式。由于AND子查询表达式的搜索结果具有冗余文件，所以后处理单元609合并AND子查询表达式的搜索结果，并消除冗余。

[0151] [详细原理说明]

[0152] 使用传统的乘法群标记，代替通常在椭圆曲线设置中使用的加法标记。

[0153] 假设$G_1 = \langle g_1 \rangle$和$G_2 = \langle g_2 \rangle$是两个有限循环群，具有附加的群$\mathcal{G} = \langle g \rangle$，使得$|G_1| = |G_2| = |\mathcal{G}| = p$，其中p是某个大的素数。双线性映射$e : \mathbb{G}_1 \times \mathbb{G}_2 \rightarrow \mathcal{G}$是具有如下效果的函数：

[0154] ■双线性的：对于所有$h_1 \in \mathbb{G}_1$，$h_2 \in \mathbb{G}_2$，对于所有$a, b \in \mathbb{Z}_p$，$e(h_1^a, h_2^b) = e(h_1, h_2)^{ab}$；

[0155] ■非退化的：$\exists h_1 \in \mathbb{G}_1$，$\exists h_2 \in \mathbb{G}_2$，使得$e(h_1, h_2) \neq I$，其中I是$\mathcal{G}$的单位元素；以及

[0156] ■可计算的：存在计算e的有效算法。

[0157] 假设存在针对输入安全参数1^k的设置算法Setup(·)，输出双线性映射的上述设置。该过程被表示为$(p, G_1, G_2, \mathcal{G}, g_1, g_2, e) \leftarrow Setup(1^k)$。

[0158] 由于$\mathbb{G}_1$、$\mathbb{G}_2$和$\mathcal{G}$都具有相同的素数阶p，因此根据双线性特性以及非退化特性，很容易可以看出$e(g_1, g_2) = g$。

[0159] 现在，详细描述涉及EBIT的联合关键字搜索方案。假设明文倒排索引表如下：对于以关键字KW_i起始的每一行，存在一组匹配明文文件F_u，其明文文件名由FN_u表示。表1示出了这种明文倒排索引表。

[0160] 密钥产生：

[0161] a) $(p, \mathbb{G}_1, \mathbb{G}_2, \mathcal{G}, g_1, g_2, e) \leftarrow Setup(1^\kappa)$。

[0162] b) 选择$(x, y, z) \in_R \mathbb{Z}_p^{*3}$。

[0163] c) 选择安全的单向散列函数$H^\kappa : \mathcal{G} \rightarrow \mathbb{K}$。

[0164] d) 选择键控散列函数$H_K : K \in \mathbb{Z}_p^* \times \{0,1\}^* \rightarrow \mathbb{Z}_p^*$，其中K是密钥。

[0165] 公共密钥是$(p, \mathbb{G}_1, \mathbb{G}_2, \mathcal{G}, g_1, g_2, e)$、$H^\kappa$和$H_K$。秘密密钥是(x, y, z)。

[0166] EBIT产生：

[0167] 针对每一行，基于明文倒排索引表（如表1所示），计算KW_i的加密或键控散列：$EK_i = Enc_{msk}(KW_i)$，加密密钥 $ek_i = H_{msk}(KW_i)$。

[0168] 对于每个FN_u，选择文件加密密钥$fk_u \in_R \mathbb{Z}_p^*$，并将其用于对$FN_u$及其内容$F_u$进行加密，产生密文文件名$CFN_u$和密文的$F_u$。

[0169] 假设FN_u出现在KW_i以作为起始的第 i 行的第 j 列，则选择变换密钥$tk_{ij} \in_R \mathbb{Z}_p^*$，并计算已加密索引：

[0170] $ef_{ij} = \left\{ef_{ij.A} = Enc_{ek_i}(tk_{ij}), ef_{ij.B} = Enc_{tk_{ij}}(fk_u, CFN_u)\right\}$，将其填入表项。

[0171] 通过以下过程来计算表 3 所示的附加数据项：

[0172] a) 对于每一个文件FN_u，随机地选择种子$h_u \in \mathbb{G}_1$。例如，首先选择$s_u \in_R \mathbb{Z}_p^*$，然后计算$h_u = g_1^{s_u} \in \mathbb{G}_1$。

[0173] b) 对于每个$ef_{ij} = \left\{ef_{ij.A} = Enc_{ek_i}(tk_{ij}), ef_{ij.B} = Enc_{tk_{ij}}(fk_u, CFN_u)\right\}$，计算中间密钥$K_{ui} = H_{\mathbb{K}}\left(e(h_u^{\ y}, g_2^{\ H_x(KW_i,0)})\right)$和 + 已加密变换密钥$EF_{ij} = Enc_{K_{ui}}(tk_{ij})$。

[0174] C) 对于每个$ef_{ij} = \left\{ef_{ij.A} = Enc_{ek_i}(tk_{ij}), ef_{ij.B} = Enc_{tk_{ij}}(fk_u, CFN_u)\right\}$，计算盲化文件信息项$W_{ij} = \begin{Bmatrix} W_{ij.A} = g_1^{(z+H_x(KW_i,1))\cdot H_\rho(CFN_u)} \\ W_{ij.B} = g_1^{(z+H_x(KW_i,2))\cdot H_\rho(CFN_u)} \end{Bmatrix}$和盲化索引$V_{ij} = h_u^{\ z+H_x(KW_i,1)}$。

[0175] 权限发布：

[0176] 执行以下过程来计算目标关键字$\{KW'_v\}$的搜索权限，其中$\{KW'_v\}$属于域R_v并且$1 \leqslant |\{KW'_v\}| = t$。

[0177] a) 按照$\{EK_v = H_{msk}(KW'_v\}$计算单关键字搜索（SKS）权限。

[0178] b) 如下计算根联合关键字搜索权限

[0179] $SC' = \left\{SC'_A = g_2^{\ 1/(tz+\sum_v H_x(KW'_v,1))}, SC'_B = g_2^{\ 1/(tz+\sum_v H_x(KW'_v,2))}\right\}$。

[0180] c) 计算根解密权限：$DC' = SC'^{\ y}_A$。

[0181] d) 选择一个目标关键字$KW'_\mu \in \{KW'_v\}$并计算其对应解密权限：

[0182] $DC = DC'^{\ H_x(KW'_\mu,0)}$。

[0183] e) 选择种子$r \in_R \mathbb{Z}_p^*$，计算联合关键字搜索权限：

[0184] $SC = \left\{\left\{EK_v\right\}, EK_\mu, SC_A = SC'^{\ r}_A, SC_B = SC'^{\ r}_B\right\}$。

[0185] 最后，权限是 SC 和 DC。

[0186] 搜索：

[0187] a) 首先使用 SKS 搜索权限$\{EK_v\}$来定位 EBIT 中的 t 行和对应的

[0188] $W_{vj_v} = \begin{Bmatrix} W_{vj_v.A} = g_1^{(z+H_x(KW'_v,1))\cdot H_\rho(CFN_u)} \\ W_{vj_v.B} = g_1^{(z+H_x(KW'_v,2))\cdot H_\rho(CFN_u)} \end{Bmatrix}$。

[0189] b) 对于 t 行中 t 个W_{vj_v}的每一种组合，分别计算

[0190]

$$\Upsilon A_{j_1,j_2,\ldots,j_t} = e(\prod_v W_{vj_v.A}, SC_A)、\ \Upsilon B_{j_1,j_2,\ldots,j_t} = e(\prod_v W_{vj_v.B}, SC_B)和\prod_v V_{vj_v}。$$

[0191] c) 如果$\Upsilon A_{j_1,j_2,\ldots,j_t} = \Upsilon B_{j_1,j_2,\ldots,j_t}$，则找到了联合关键字搜索的匹配。匹配项被记录为$(ef_{\mu j_\mu.B}, EF_{\mu j_\mu}, \prod_v V_{vj_v})$。

[0192] 最后，联合关键字搜索输出是匹配$\left\{ef_{\mu j_\mu.B}, EF_{\mu j_\mu}, \prod_v V_{vj_v}\right\}$。

[0193] 这里，$ef_{\mu j_\mu.B}$和$EF_{\mu j_\mu}$是由 EK_μ 标识的行中的表项。换言之，除了$\prod_v V_{vj_v}$，仅将属于由 EK_μ 标识的行的匹配项返回给搜索方。对于匹配项，必须在等式$\Upsilon A_{j_1,j_2,\ldots,j_t} = \Upsilon B_{j_1,j_2,\ldots,j_t}$中涉及$W_{\mu j_\mu}$。

[0194] 注意，当且仅当用相同的 $H_\rho(CFN_u)$ 产生了所有的W_{vj_v}时，$\Upsilon A_{j_1,j_2,\ldots,j_t} = \Upsilon B_{j_1,j_2,\ldots,j_t}$。

[0195] 解密：

[0196] a) 对于每一个搜索结果$\left\{ef_{\mu j_\mu.B}, EF_{\mu j_\mu}, \prod_v V_{vj_v}\right\}$，计算中间密钥$K_{u\mu} = H_{\mathbb{K}}\left(e(\prod_v V_{vj_v}, DC)\right)$。

[0197] b) 使用对应密钥 $K_{u\mu}$ 来解密每个$EF_{\mu j_\mu}$以获得变换密钥$tk_{\mu j_\mu}$。

[0198] c) 使用变换密钥$tk_{\mu j_\mu}$来解密$ef_{\mu j_\mu.B}$，以获得文件加密密钥 fk_u 和密文文件名 CFN_u。

[0199] d) 最后，检索由 CFN_u 标识的加密文件，并使用 fk_u 来对 CFN_u 和加密文件进行解密，得到明文文件名 FN_u 和明文文件内容。

[0200] 在前提条件$\Upsilon A_{j_1,j_2,\ldots,j_t} = \Upsilon B_{j_1,j_2,\ldots,j_t}$下，很容易验证

[0201]

$$H_{\mathbb{K}}\left(e(\prod_v V_{vj_v}, DC)\right) = H_{\mathbb{K}}\left(e(h_u{}^y, g_2{}^{H_x(KW'_\mu,0)})\right) = K_{u\mu}。$$

[0202] 应该注意，本领域技术人员所公知的，可以以多种显而易见的方式对上述方案进行改进。例如，不必使用相同的文件加密密钥 fk_u 来加密 FN_u 及其内容 F_u。此外，取决于所希望的加密等级，完全可以不必加密 FN_u。

[0203] 对于其它示例，可以按照$V_{ij} = h_u{}^{z+H_x(KW_i,2)}$或$V_{ij} = h_u{}^{z+H_x(KW_i,3)}$计算盲化索引$V_{ij} = h_u{}^{z+H_x(KW_i,1)}$，并相应地，按照$DC' = \left(g_2{}^{1/(tz+\sum_v H_x(KW'_v,2))}\right)^y$或$DC' = \left(g_2{}^{1/(tz+\sum_v H_x(KW'_v,3))}\right)^y$计算 DC′。此外，可以使用不同的 z_1、z_2 和 z_3 来计算 W_{ij} 和 V_{ij}，使得$W_{ij} = \left\{\begin{matrix} W_{ij.A} = g_1{}^{(z_1+H_x(KW_i,1))\cdot H_\rho(CFN_u)} \\ W_{ij.B} = g_1{}^{(z_2+H_x(KW_i,2))\cdot H_\rho(CFN_u)} \end{matrix}\right\}$和$V_{ij} = h_u{}^{z_3+H_x(KW_i,1)}$。结果，DC′ 和 SC′ 被计算为

$DC' = \left(g_2^{1/(tz_3 + \sum_v H_x(KW'_v,1))} \right)^y$ 和 $SC' = \left\{ SC'_A = g_2^{1/(tz_1 + \sum_v H_x(KW'_v,1))}, SC'_B = g_2^{1/(tz_2 + \sum_v H_x(KW'_v,2))} \right\}$。

此外，很容易想到如下变体：根据 $V_{ij} = h_u^{z_3 + H_{x_2}(KW_i,1)}$，使用不同的密钥 x_2 来计算 x_2。另一方面，取决于所希望的安全等级，可以减少密钥数目。例如，可以将相同的密钥用于 (msk，x，y，z，ρ)，即 msk = x = y = z = ρ。

[0204] 作为另一示例，存在实现 $H_x(KW_i,1)$ 以及 $H_x(KW_i,2)$ 的多种方法。一般地，可以将两个不同的比特串用作键控散列函数的输入，以分别代替“1”和“2”。

[0205] 最后但并非最不重要的，可以从 ef_{ij} 中省略 $ef_{ij.A} = Enc_{ek_i}(tk_{ij})$。由于本发明的方案允许 $1 \leqslant |\{KW'_v\}| = t$，所以 $|\{KW'_v\}| = 1$ 的特殊情况可以很好地给搜索方提供解密权限，而不需要 $ef_{ij.A}$。然而，在 ef_{ij} 中包括 $ef_{ij.A}$ 可以加速单关键字搜索情况下的关键字搜索速度。实际上，这只是参考文献 [8] 的教导。

[0206] [改进]

[0207] 可选方案 1：

[0208] 尽管上述基本方案仅涉及诸如 $KW_1 \wedge KW_2$ 之类的 AND 查询，但是很容易使用该基本方案作为构建模块来处理组合有 AND 和 OR 查询的复杂查询表达式。例如，可以通过执行 AND 子查询表达式 (ASQ)$(KW_1 \wedge KW_2)$ 和 $(KW_3 \wedge KW_4)$，并合并两个 AND 子查询表达式的搜索结果，来进行复杂查询表达式 $(KW_1 \wedge KW_2) \vee (KW_3 \wedge KW_4)$。再如，可以首先将复杂查询表达式 $(KW_1 \vee KW_2) \wedge (KW_3 \vee KW_4)$ 解析为如下一系列 AND 子查询表达式：

[0209] $(KW_1 \wedge KW_3) \vee (KW_1 \wedge KW_4) \vee (KW_2 \wedge KW_3) \vee (KW_2 \wedge KW_4)$。

[0210] 然后，分别执行 ASQ，并合并不同 ASQ 的搜索结果，得到复杂查询表达式 $(KW_1 \vee KW_2) \wedge (KW_3 \vee KW_4)$ 的搜索结果。

[0211] 以上关于如何处理合并有 AND 和 OR 查询的复杂查询表达式的说明表明还需要额外的附加单元，具体地，还需要查询表达式解析单元和后处理单元。

[0212] 搜索方和 / 或数据所有方可能需要查询表达式解析单元，将输入查询表达式（可能是复杂查询表达式）解析为一系列 ASQ。如上所述，ASQ 由逻辑 OR 运算符连接。

[0213] 此外，搜索方可能需要后处理器单元。后处理单元检查所有 ASQ 的输出，通过消除冗余来合并这些输出。冗余是由于一个密文文件名可能是多个 ASQ 的搜索结果而引起的。在后处理单元消除了冗余之后，在最终的（复杂查询表达式的）联合搜索结果中，不存在出现两次的密文文件名。

[0214] 以上描述仅给出了本发明的优选实施例，而并不是要以任何方式限制本发明。因此，在本发明精神和原理内进行的任何修改、替换、改进等应该由本发明范围所涵盖。

[0215] 参考文献列表

[0216] [1]Amazon Simple Storage Service(Amazon S3)，http://aws.amazon.com/s3；

[0217] [2]Google Health，https://www.google.com/health；

[0218] [3]Microsoft HealthVault，http://www.healthvault.com；

[0219] [4]Card details stolen in web hack，BBC news，http://news.bbc.co.uk/2/hi/technology/7446871.stm；

[0220] [5]TJX theft tops 45.6million card numbers，reported bySecurityFocus.

com，http://www.securityfocus.com/news/11455；

[0221] [6]D.Song，D.Wagner，A.Perrig，Practical techniques forsearches on encrypted data，in Proceedings of IEEESymposium on Security and Privacy'00，pp.44-55，2000；

[0222] [7]D.Boneh，G.D.Crescenzo，R.Ostrovsky，G.Persiano.Public Key Encryption with Keyword Search.InProceeding of EuroCrypt'04，LNCS 3027，pp.506-522，2004；

[0223] [8]中国发明专利申请，申请号：CN 200810145083.8，发明名称："用于快速密文检索的方法、装置和系统"。

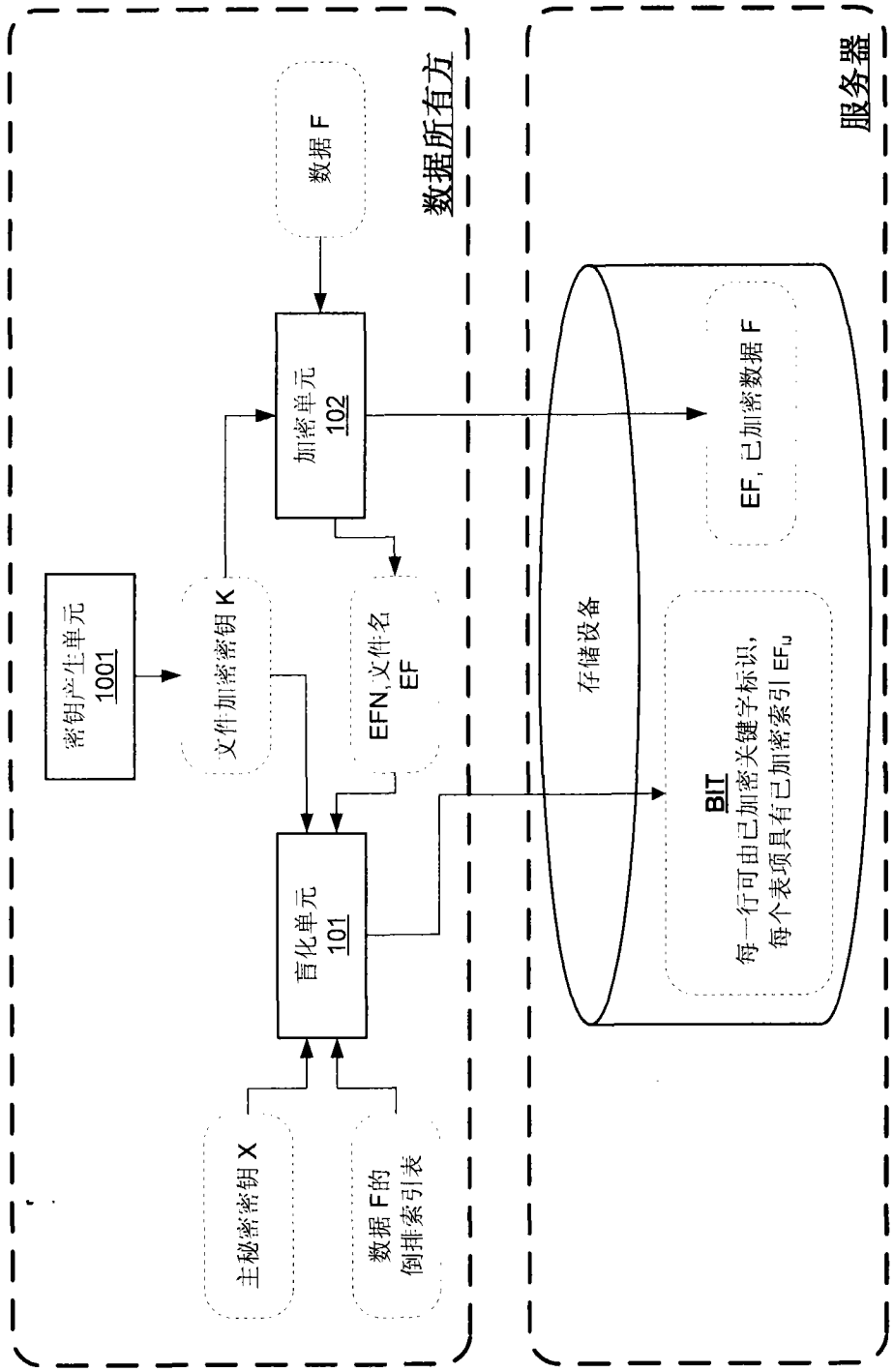

数据所有方
服务器
数据 F
加密单元
102
密钥产生单元
1001
文件加密密钥 K
EFN,文件名
EF
存储设备
EF, 已加密数据 F
BIT
每一行可由已加密关键字标识,
每个表项具有已加密索引 EFID
盲化单元
101
主秘密密钥 X
数据 F的
倒排索引表

图 1

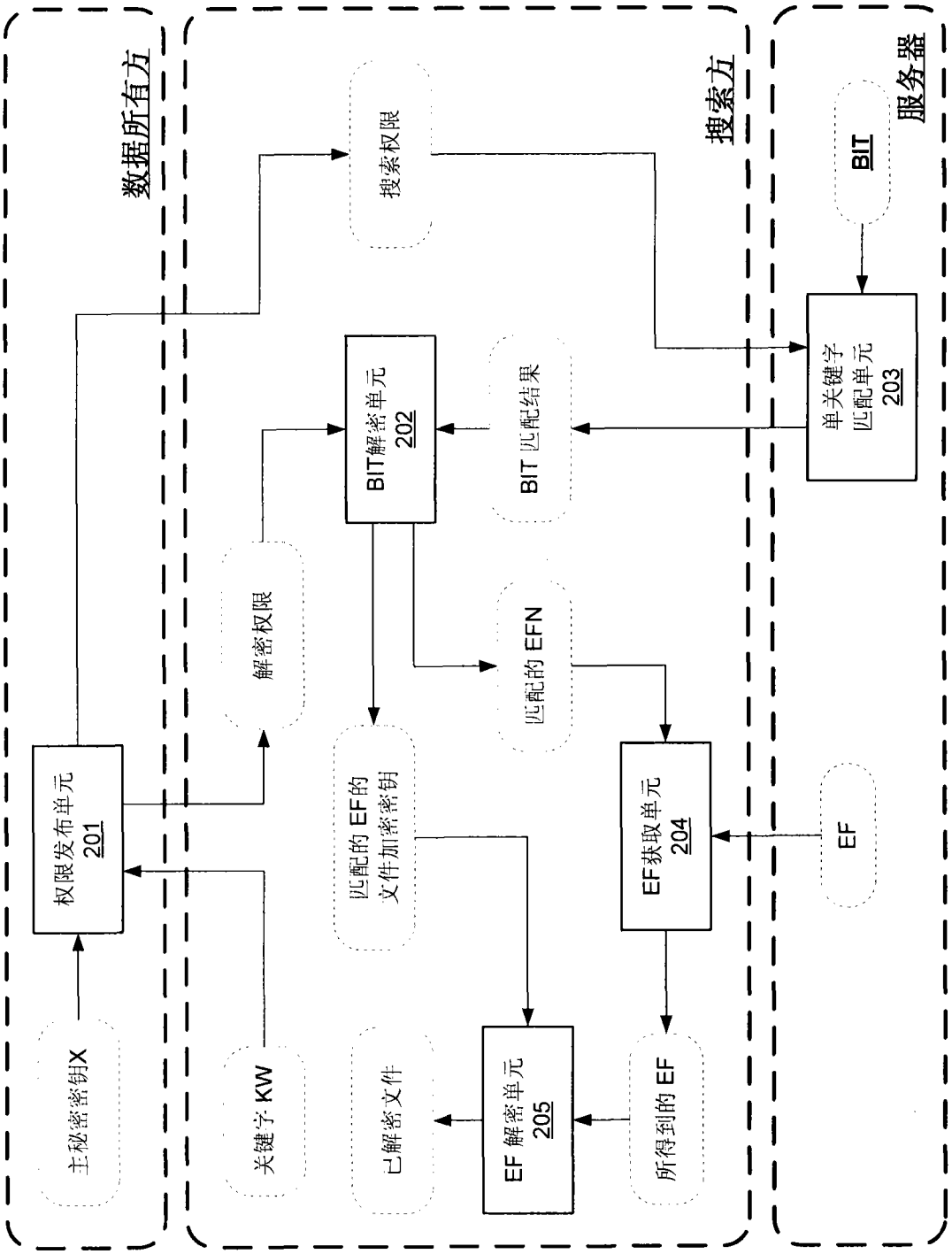

数据所有方
搜索方
服务器
主秘密密钥X
权限发布单元 201
关键字 KW
解密权限
搜索权限
BIT解密单元 202
匹配的 EF的 文件加密密钥
已解密文件
EF 解密单元 205
BIT 匹配结果
匹配的 EFN
所得到的 EF
EF获取单元 204
EF
BIT
单关键字 匹配单元 203

图 2

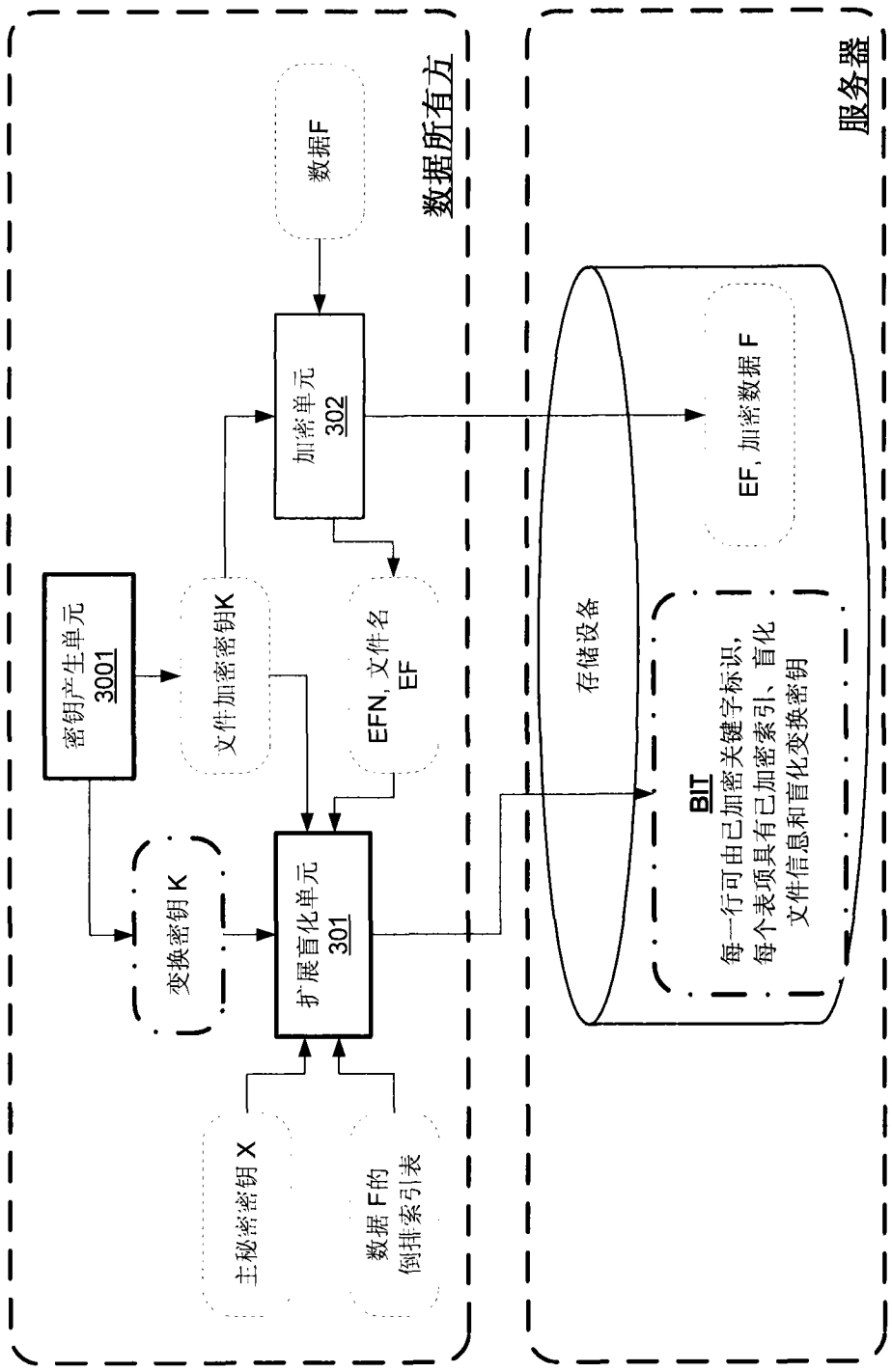

数据所有方
服务器
数据F
加密单元
302
密钥产生单元
3001
文件加密密钥K
EFN, 文件名
EF
变换密钥 K
扩展盲化单元
301
主秘密密钥 X
数据 F的
倒排索引表
存储设备
EF, 加密数据 F
BIT
每一行可由已加密关键字标识，
每个表项具有已加密索引、盲化
文件信息和盲化变换密钥

图 3

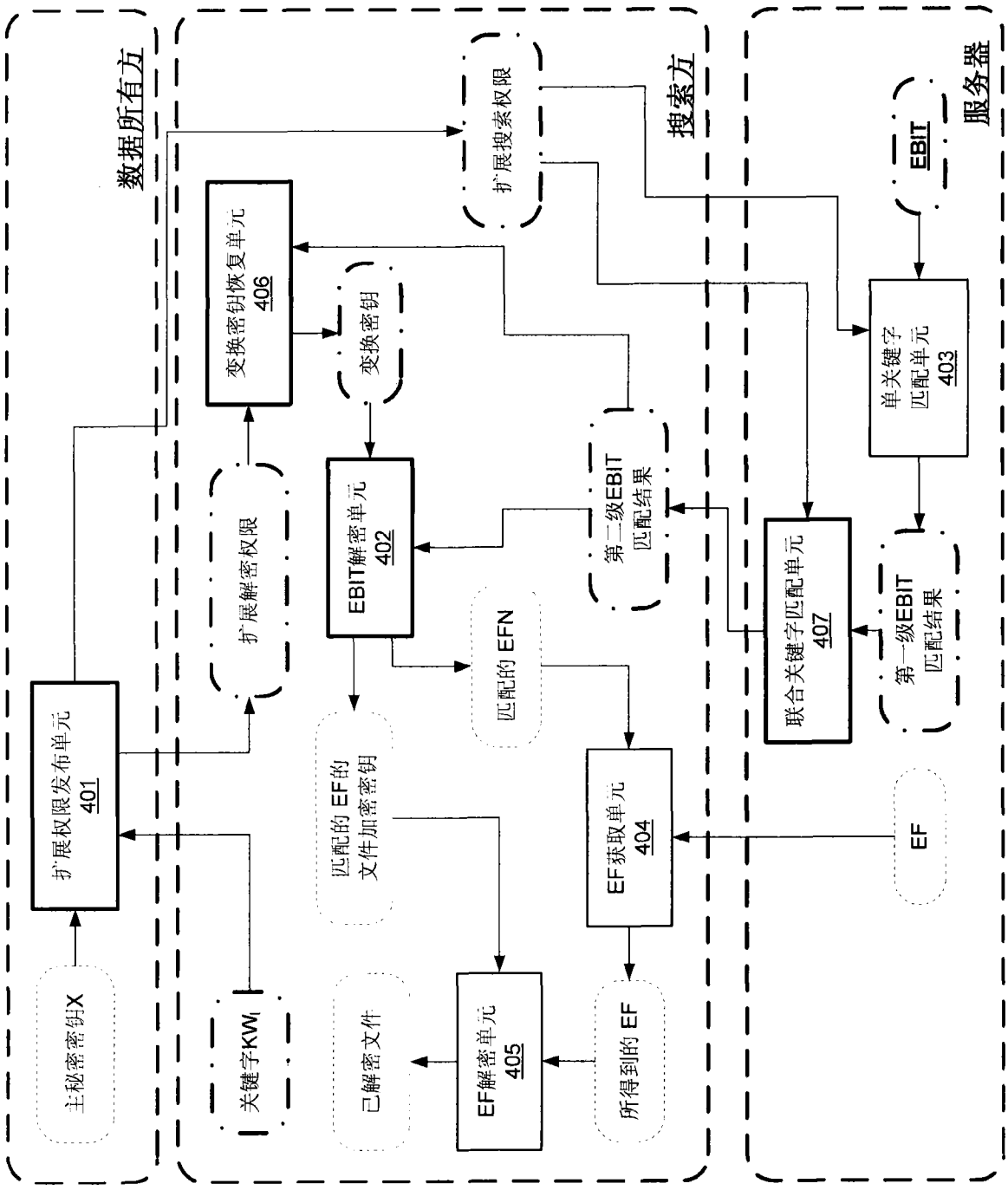

数据所有方
搜索方
服务器
扩展权限发布单元 401
主秘密密钥X
关键字KWi
扩展解密权限
变换密钥恢复单元 406
变换密钥
扩展搜索权限
EBIT解密单元 402
匹配的EF的文件加密密钥
匹配的 EFN
第二级EBIT匹配结果
EF获取单元 404
所得到的 EF
已解密文件
EF解密单元 405
联合关键字匹配单元 407
第一级EBIT匹配结果
单关键字匹配单元 403
EBIT
EF

图 4

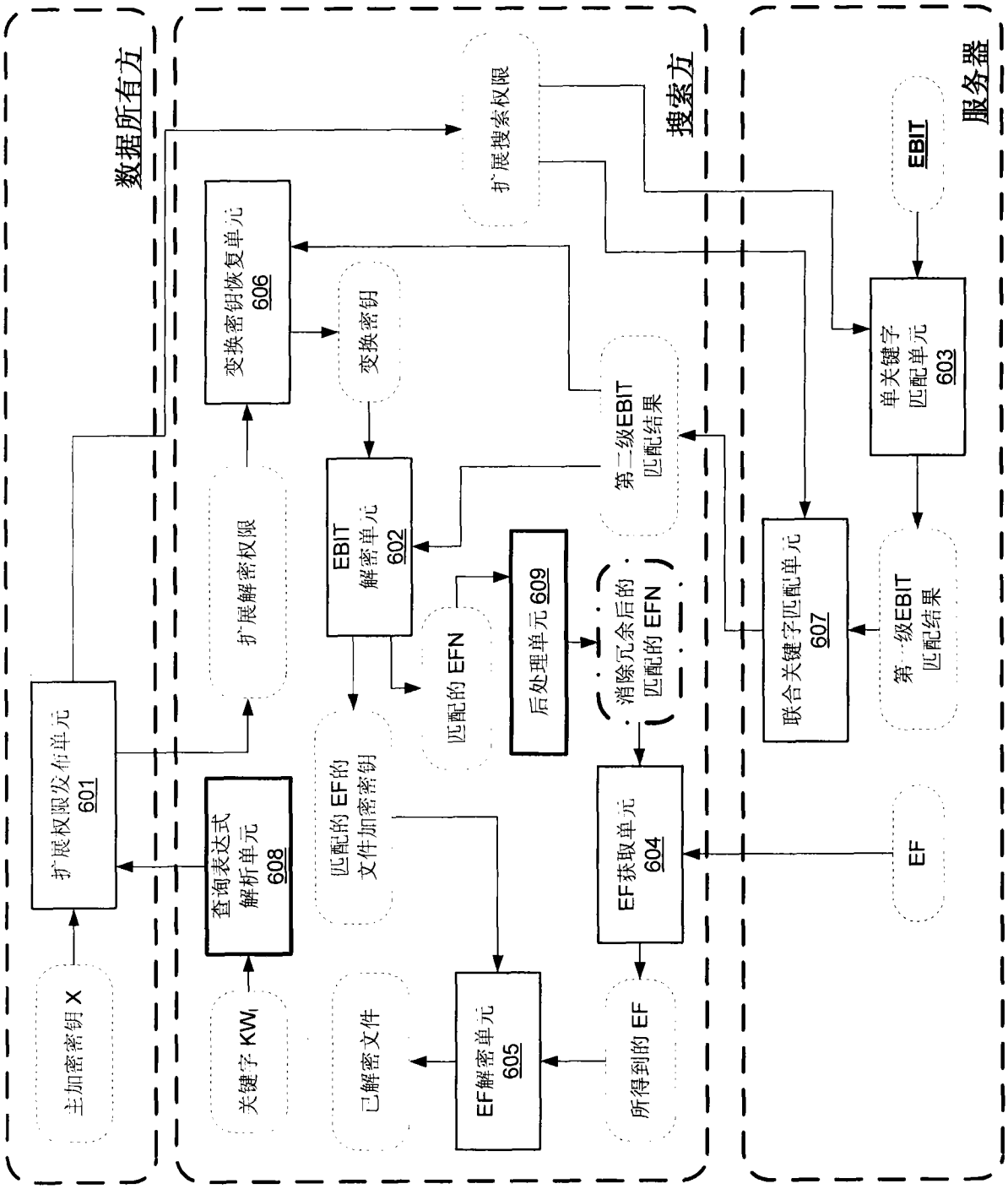

数据所有方
搜索方
服务器
主加密密钥 X
扩展权限发布单元 601
关键字 KWi
查询表达式解析单元 608
扩展解密权限
变换密钥恢复单元 606
扩展搜索权限
变换密钥
EBIT 解密单元 602
匹配的 EF 的文件加密密钥
匹配的 EFN
后处理单元 609
消除冗余后的匹配的 EFN
第二级EBIT匹配结果
EF获取单元 604
已解密文件
EF解密单元 605
所得到的 EF
联合关键字匹配单元 607
单关键字匹配单元 603
EBIT
第一级EBIT匹配结果
EF

图 5