

(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
2017 年 3 月 30 日 (30.03.2017)



(10) 国际公布号
WO 2017/050148 A1

- (51) 国际专利分类号:
G06F 11/36 (2006.01) *G06Q 10/06* (2012.01)
- (21) 国际申请号: PCT/CN2016/098816
- (22) 国际申请日: 2016 年 9 月 13 日 (13.09.2016)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201510609660.4 2015 年 9 月 22 日 (22.09.2015) CN
- (71) 申请人: 阿里巴巴集团控股有限公司 (ALIBABA GROUP HOLDING LIMITED) [—/CN]; 英属开曼群岛大开曼资本大厦一座四层 847 号邮箱, Grand Cayman (KY)。
- (72) 发明人: 及
- (71) 申请人 (仅对美国): 朱珠 (ZHU, Zhu) [CN/CN]; 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。 王晓哲 (WANG, Xiaozhe) [CN/CN]; 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。
- (74) 代理人: 北京国昊天诚知识产权代理有限公司 (CO-HORIZON INTELLECTUAL PROPERTY INC.);

中国北京市朝阳区小关北里甲 2 号渔阳置业大厦 B 座 605, Beijing 100029 (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, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, 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, GH, GM, KE, LR, LS, MW, MZ, NA, RW, 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, 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))。

(54) Title: DATA ANALYSIS METHOD AND DEVICE

(54) 发明名称: 一种数据分析方法和装置

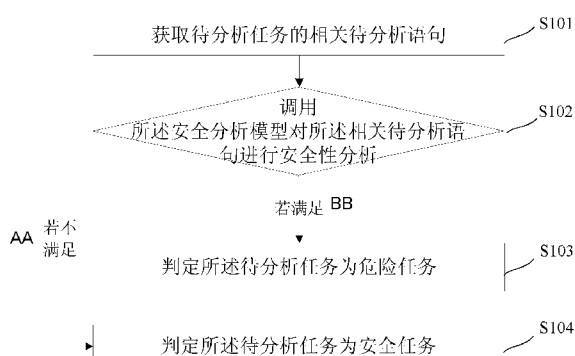


图 1

S101 Acquiring a related statement to be analysed of a task to be analysed

S102 Calling the security analysis model to analyse the security of the related statement to be analysed

S103 Determining that the task to be analysed is a dangerous task

S104 Determining that the task to be analysed is a safe task

AA If a security analysis condition is not met

BB If the security analysis condition is met

(57) Abstract: Disclosed are a data analysis method and device. The method comprises: according to a combinational relationship between statements, classifying the statements subjected to analysis, so that a sensitivity set and a security set are obtained through classification, establishing a security analysis model according to the sensitivity set and the security set, and acquiring a related statement to be analysed of a task to be analysed; calling the security analysis model to analyse the security of the related statement to be analysed, if the related statement to be analysed meets a security analysis condition, determining that the task to be analysed is a dangerous task, wherein the security analysis condition is set according to the sensitivity set and the security set; and if the related statement to be analysed does not meet the security analysis condition, determining that the task to be analysed is a safe task. Thus, automatic comparative analysis can be performed on a task to be analysed, so that an artificial influence in a conventional analysis method is removed, stable analysis accuracy is provided, and the security of a system is enhanced.

(57) 摘要:

[见续页]

WO 2017/050148 A1

一种数据分析方法和装置，根据语句的组合关系对用于分析的语句进行分类，分类得到敏感集合和安全集合，根据所述敏感集合和安全集合建立安全分析模型，获取待分析任务的相关待分析语句；调用所述安全分析模型对所述相关待分析语句进行安全性分析，若所述相关待分析语句满足安全分析条件，则判定所述待分析任务为危险任务，所述安全分析条件根据所述敏感集合和所述安全集合设置；若所述相关待分析语句不满足所述安全分析条件，则判定所述待分析任务为安全任务。由此可对待分析任务进行自动化比对分析，从而去除了传统分析方式中的人为影响，提供了稳定的分析准确性，增加了系统的安全性。

一种数据分析方法和装置

技术领域

本发明涉及数据处理领域，特别是涉及一种数据分析方法和装置。

背景技术

在数据研发的过程中，开发人员可以将研发完成的任务提交到系统，在该任务的分析合格后，该任务可以在该系统上线。上线后，该系统的用户可以查找到该任务，通过执行该任务可以实现该任务相应的功能。

针对任务的分析主要是分析任务所包含的语句是否具有风险，通过分析语句判断该任务是否会导致系统的敏感信息泄露，对系统造成不安全隐患。目前主要依靠专业的分析人员对任务包含的语句进行分析，分析的准确性和分析人员在数据分析方面的经验直接相关，经验匮乏的分析人员所得出的分析结果往往不够准确，例如将安全的任务分析成高危任务，或将具有安全漏洞的任务分析为安全任务并在系统上线。人为因素的影响导致了目前对数据任务的分析准确性的波动大，降低了系统的安全性。

发明内容

为了解决上述技术问题，本发明提供了一种数据分析方法和装置，以使得可对待分析任务进行自动化比对分析，从而去除了传统分析方式中的人为影响。

本发明实施例公开了如下技术方案：

一种数据分析方法，根据语句的组合关系对用于分析的语句进行分类，分类得到敏感集合和安全集合，所述敏感集合中包括能造成敏感数据泄露的语句组合，所述安全集合中包括安全的语句组合，根据所述敏感集合和安全集合建立安全分析模型，所述方法包括：

获取待分析任务的相关待分析语句；

调用所述安全分析模型对所述相关待分析语句进行安全性分析，

若所述相关待分析语句满足安全分析条件，则判定所述待分析任务为危险任务，所述安全分析条件根据所述敏感集合和所述安全集合设置；

若所述相关待分析语句不满足所述安全分析条件，则判定所述待分析任务为安全任务。

可选的，在所述获取待分析任务的相关待分析语句之前，还包括：

获取所述待分析任务的任务标识；

根据所述任务标识从数据库保存的数据中调取与所述待分析任务的待分析语句；

根据所述任务标识查询所述待分析任务是否具有子任务，所述子任务为对所述待分析任务输出数据具有依赖关系的任务；

若具有，从数据库保存的数据库中调取所述子任务的语句；

且若具有，所述获取待分析任务的相关待分析语句，包括：

获取所述待分析任务的待分析语句、所述子任务的语句以及根据所述依赖关系得到的所述待分析任务的待分析语句和所述子任务的语句之间的语句组合关系。

可选的，所述根据所述敏感集合和安全集合建立安全分析模型，包括：

根据语句的组合关系将所述敏感集合中的语句组合建立至少一个敏感抽象语法树 AST，将所述安全集合中的语句组合建立至少一个安全 AST；

所述调用所述安全分析模型对所述相关待分析语句进行安全性分析，包括：

根据所述相关待分析语句中的语句组合关系，将所述相关待分析语句建立至少一个待分析 AST；

将所述至少一个待分析 AST 与所述安全分析模型中的所述至少一个敏感 AST 和安全 AST 进行比对分析。

可选的，所述安全分析条件包括下述一个或多个的组合：

所述至少一个待分析 AST 中用于数据产出的子 AST 与所述至少一个敏感 AST 中的相同，且所述子 AST 与所述至少一个安全 AST 不同；

所述相关待分析语句中一个应用程序编程接口 API 对应的 AST 的至少一个层级与所述至少一个敏感 AST 中的相同；

所述相关待分析语句中多个应用程序编程接口 API 对应的 AST 的至少一个层级与所述至少一个敏感 AST 中的相同。

可选的，若所述相关待分析语句满足安全分析条件，还包括：

将判定为危险任务的所述待分析任务发至分析终端，由所述分析终端对所述待分析任务进行二次安全性分析。

一种数据分析装置，根据语句的组合关系对用于分析的语句进行分类，分类得到敏感集合和安全集合，所述敏感集合中包括能造成敏感数据泄露的语句组合，所述安全集合中包括安全的语句组合，根据所述敏感集合和安全集合建立安全分析模型，所述装置包括：

第一获取单元，用于获取待分析任务的相关待分析语句；

分析单元，用于调用所述安全分析模型对所述相关待分析语句进行安全性分析；

若所述分析结果为所述相关待分析语句满足安全分析条件，触发判断单元判定所述待分析任务为危险任务，所述安全分析条件根据所述敏感集合和所述安全集合设置；

若所述分析结果为所述相关待分析语句不满足所述安全分析条件，触发所述判定单元判定所述待分析任务为安全任务。

可选的，还包括：

第二获取单元，用于在触发所述第一获取单元之前，获取所述待分析任务的任务标识；

调取单元，用于根据所述任务标识从数据库保存的数据中调取与所述待分析任务的待分析语句；

查询单元，用于根据所述任务标识查询所述待分析任务是否具有子任务，所述子任务为对所述待分析任务输出数据具有依赖关系的任务；

若具有，所述调取单元还用于从数据库保存的数据库中调取所述子任务的语句；

且若具有，所述第一获取单元还用于获取所述待分析任务的待分析语句、所述子任务的语句以及根据所述依赖关系得到的所述待分析任务的待分析语句和所述子任务的语句之间的语句组合关系。

可选的，所述根据所述敏感集合和安全集合建立安全分析模型，包括：

根据语句的组合关系将所述敏感集合中的语句组合建立至少一个敏感抽

象语法树 AST，将所述安全集合中的语句组合建立至少一个安全 AST；

所述分析单元具体用于根据所述相关待分析语句中的语句组合关系，将所述相关待分析语句建立至少一个待分析 AST；将所述至少一个待分析 AST 与所述安全分析模型中的所述至少一个敏感 AST 和安全 AST 进行比对分析。

可选的，所述安全分析条件包括下述一个或多个的组合：

所述至少一个待分析 AST 中用于数据产出的子 AST 与所述至少一个敏感 AST 中的相同，且所述子 AST 与所述至少一个安全 AST 不同；

所述相关待分析语句中一个应用程序编程接口 API 对应的 AST 的至少一个层级与所述至少一个敏感 AST 中的相同；

所述相关待分析语句中多个应用程序编程接口 API 对应的 AST 的至少一个层级与所述至少一个敏感 AST 中的相同。

可选的，还包括：

发送单元，用于若所述分析结果为所述相关待分析语句满足安全分析条件，将判定为危险任务的所述待分析任务发至分析终端，由所述分析终端对所述待分析任务进行二次安全性分析。

由上述技术方案可以看出，首先根据语句的组合关系对用于分析的语句进行分类，分类得到敏感集合和安全集合，所述敏感集合中包括能造成敏感数据泄露的语句组合，所述安全集合中包括安全的语句组合，根据所述敏感集合和安全集合建立安全分析模型，当获取与待分析任务的相关待分析语句时，可以调用所述安全分析模型对所述相关待分析语句进行安全性分析，将预先得到的敏感语句组合和安全语句组合与待分析任务的相关语句进行比对分析，根据是否满足根据所述敏感集合和所述安全集合设置的安全分析条件来判定是否为安全任务或是否为危险任务，由此可对待分析任务进行自动化比对分析，从而去除了传统分析方式中的人为影响，提供了稳定的分析准确性，增加了系统的安全性。

附图说明

为了更清楚地说明本发明实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本发明的一些实施例，对于本领域普通技术人员来讲，

在不付出创造性劳动性的前提下，还可以根据这些附图获得其他的附图。

图 1 为本发明实施例提供的一种数据分析方法的方法流程图；

图 2 为本发明实施例提供的一种获取相关待分析语句方法的方法流程图；

图 3 为本发明实施例提供的一种数据分析装置的装置结构图。

具体实施方式

在数据研发的过程中，开发人员可以将研发完成的任务提交到系统，在该任务的分析合格后，该任务可以在该系统上线。其中，针对任务的分析主要是分析任务所包含的语句是否具有风险，通过分析语句判断该任务是否会导致系统的敏感信息泄露，例如调用敏感数据，或者通过计算生成敏感数据等。若这种不安全的任务被系统上线，可能会导致用户在执行该任务时泄露系统的敏感数据，由此对系统造成不安全隐患。目前主要依靠专业的分析人员对任务包含的语句进行分析，分析的准确性和分析人员在数据分析方面的经验直接相关，经验匮乏的分析人员所得出的分析结果往往不够准确，例如将安全的任务分析成高危任务，或将具有安全漏洞的任务分析为安全任务并在系统上线。人为因素的影响导致了目前对数据任务的分析准确性的波动大，降低了系统的安全性。

为此，本发明实施例提供了一种数据分析方法和装置，首先根据语句的组合关系对用于分析的语句进行分类，分类得到敏感集合和安全集合，所述敏感集合中包括能造成敏感数据泄露的语句组合，所述安全集合中包括安全的语句组合，根据所述敏感集合和安全集合建立安全分析模型，当获取与待分析任务的相关待分析语句时，可以调用所述安全分析模型对所述相关待分析语句进行安全性分析，将预先得到的敏感语句组合和安全语句组合与待分析任务的相关语句进行比对分析，根据是否满足根据所述敏感集合和所述安全集合设置的安全分析条件来判定是否为安全任务或是否为危险任务，由此可对待分析任务进行自动化比对分析，从而去除了传统分析方式中的人为影响，提供了稳定的分析准确性，增加了系统的安全性。

为使本发明实施例的目的、技术方案和优点更加清楚，下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚地描述，显然，所描述的实施例是本发明一部分实施例，而不是全部的实施例。基于本发明

中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本发明保护的范围。

实施例一

在描述本发明实施例提供的数据分析方法的具体步骤之前，先介绍安全分析模型的组成和建立。

本发明实施例提供的安全分析模型（英文：Security Analysis Model）建立方式是，根据语句的组合关系对用于分析的语句进行分类，分类得到敏感集合和安全集合，所述敏感集合中包括能造成敏感数据泄露的语句组合，所述安全集合中包括安全的语句组合，根据所述敏感集合和安全集合建立安全分析模型。

举例说明，可以针对具体的应用场景得到用于分析的语句，语句可以为计算机语句，例如目前系统平台中较为常用的结构化查询语言（英文：Structured Query Language，缩写：SQL）语句等。

可以通过关系代数理论，对用于分析的语句中的组合关系等进行解析，从中确定出可能出现敏感数据泄漏的语句组合，以及安全的语句组合。确定语句属于敏感集合或安全集合的标准可以为安全约束原则，安全约束原则可以包括禁止从系统获取敏感数据等，例如禁止从系统平台的数据库中获取敏感数据。敏感数据可以为用户真实信息、用户的交易数据、公司数据等。

确定出的敏感集合可以具体为敏感指标集合（英文：Sensitive Indicator Set，缩写：SIS），其中可以包括从业务系统数据仓库表触发的计算过程。确定出的安全集合可以具体为非敏感、有实际业务意义的指标列表（英文：Non-sensitive Indicator Set，缩写：NIS）。

所述敏感集合和安全集合可以作为所述安全分析模型的组成部分。为了更为便于后续的安全性分析，可以将所述敏感集合和安全集合建立为抽象语法树（英文：abstract syntax tree，缩写：AST）。可选的，所述根据所述敏感集合和安全集合建立安全分析模型，包括：

根据语句的组合关系将所述敏感集合中的语句组合建立至少一个敏感抽象语法树 AST，将所述安全集合中的语句组合建立至少一个安全 AST。

其中，建立 AST 的过程可以为，将语句（例如 SQL 语句）作为 AST 的

中间节点，根据语句之间的组合关系连接各个中间节点。

图 1 为本发明实施例提供的一种数据分析方法的方法流程图，如图 1 所示，所述方法包括：

S101：获取待分析任务的相关待分析语句。

举例说明，所述待分析任务可以为经过开发人员或数据开发系统等开发完成的、等待被系统上线到网络上的任务，在上线之前，需要对所述待分析任务进行安全性分析，判断运行所述待分析任务是否会导致泄漏系统的敏感数据等。通过分析避免上线的任务对系统造成安全隐患。

所述相关待分析语句可以为与所述待分析任务具有关联关系的语句，不需局限于仅为所述待分析任务所包括的语句，还可以包括对所述待分析任务具有依赖关系的其他任务所包括的语句。

可选的，本发明实施例提供了一种获取待分析任务的相关待分析语句的方法，在图 1 所对应实施例的基础上，图 2 为本发明实施例提供的一种获取相关待分析语句方法的方法流程图，包括：

S201：获取所述待分析任务的标识。

一般情况下，并不会将所述待分析任务的所有内容直接发送到用于安全性分析的服务器上，而是开发人员或数据开发系统先将所述待分析任务的内容保存到用于存储的服务器上，并由此得到与所述待分析任务相对应的任务标识。当需要对所述待分析任务进行分析时，可以将任务标识发送到用于安全性分析的服务器，从而避免发送大量数据的情况发生。

S202：根据所述任务标识从数据库保存的数据中调取与所述待分析任务的待分析语句。

S203：根据所述任务标识查询所述待分析任务是否具有子任务，所述子任务为对所述待分析任务输出数据具有依赖关系的任务。若具有，触发 S204。

当用于安全性分析的服务器获取所述任务标识，将从用于存储待分析任务的服务器上更加所述任务标识调取与所述任务标识对应的待分析语句。并且，所述用于安全性分析的服务器还可以根据所述任务标识查找依赖所述待分析任务的子任务。通过查找与所述待分析任务具有依赖关系的子任务，可以筛选出原本所述待分析任务产出的数据符合安全性，但是可能会在子任务

中作为生产出敏感数据的基础的情况。可见，通过同时分析子任务的语句，可以更加全面的确定所述待分析任务的安全性。

这里所述的依赖可以理解为子任务需要通过所述待分析任务才能实现自身功能，例如子任务所需要的数据由所述待分析任务计算得到或由所述待分析任务调取，或者例如子任务功能的实现所需要依赖于所述待分析任务的全部或部分功能的实现等。

S204：从数据库保存的数据库中调取所述子任务的语句。

S205：获取待分析任务的相关待分析语句，所述相关待分析语句在所述待分析任务具有子任务时包括所述待分析任务的待分析语句、所述子任务的语句以及根据所述依赖关系得到的所述待分析任务的待分析语句和所述子任务的语句之间的语句组合关系。

举例说明，所述相关待分析语句可以包括两种组成情况，和是否能够找到子任务相关。第一种情况下，没有找到所述待分析任务的子任务，这种情况下，所述相关待分析语句可以包括所述待分析任务的待分析语句。第二种情况下，找到了所述待分析任务的子任务，这种情况下，所述相关待分析语句在所述待分析任务具有子任务时包括所述待分析任务的待分析语句、所述子任务的语句以及根据所述依赖关系得到的所述待分析任务的待分析语句和所述子任务的语句之间的语句组合关系。

S102：调用所述安全分析模型对所述相关待分析语句进行安全性分析。

举例说明，所述安全性分析可以为将所述相关待分析语句中的语句组合、调用关系等与所述安全分析模型中敏感集合和安全集合中的语句组合和调用关系进行比对，从而确定出所述待分析任务的安全性。本发明不限定安全性分析的具体方式。在所述安全分析模型如上述例子中建立了AST的可选情况下，也可以对所述相关待分析语句进行相似处理，从而提高安全性分析中比对分析的效率。也就是说，可选的，所述调用所述安全分析模型对所述相关待分析语句进行安全性分析，包括：

根据所述相关待分析语句中的语句组合关系，将所述相关待分析语句建立至少一个待分析AST。

将所述至少一个待分析AST与所述安全分析模型中的所述至少一个敏感

AST 和安全 AST 进行比对分析。

举例说明，其中，建立 AST 的过程可以为，将语句（例如 SQL 语句）作为 AST 的中间节点，根据语句之间的组合关系连接各个中间节点。

由于比对的双方都是 AST 的结构，使得比对的过程更为便捷，可以单独的比对一个层级、一个实现数据产出功能的子 AST 等。

在 S102 中，若所述相关待分析语句满足安全分析条件，执行 S103，若所述相关待分析语句不满足所述安全分析条件，执行 S104。

举例说明，所述安全分析条件的内容可以与具体应用场景相关，所要求的安全性可以对应不同的安全分析条件。

可选的，在上述 AST 的情况下，所述安全分析条件包括下述条件一、条件二和条件三中一个或多个的组合。在所述安全分析条件包括多个条件的情况下，即使所述相关待分析语句在安全性分析中仅满足了其中一个条件，也可以认为所述相关待分析语句满足安全分析条件。

条件一：所述至少一个待分析 AST 中用于数据产出的子 AST 与所述至少一个敏感 AST 中的相同，且所述子 AST 与所述至少一个安全 AST 不同。

若满足所述条件一，则可以理解为，所述待分析任务中的 API 可能直接生产出了敏感指标且未经过脱敏处理就对外暴露。

条件二：所述相关待分析语句中一个应用程序编程接口（英文：Application Programming Interface，缩写：API）对应的 AST 的至少一个层级与所述至少一个敏感 AST 中的相同。

若满足所述条件二，则可以理解为，所述待分析任务中的 API 可能生产出了计算敏感指标所需的所有前提数据，并在单个 API 中将它们对外暴露。

条件三：所述相关待分析语句中多个应用程序编程接口 API 对应的 AST 的至少一个层级与所述至少一个敏感 AST 中的相同。

若满足所述条件三，则可以理解为，所述待分析任务中的独立软件开发商（英文：Independent Software Vendors，缩写：ISV）可能生产出了计算敏感指标所需的所有前提数据，并在多个 API 中分别将它们对外暴露。

S103：判定所述待分析任务为危险任务。

S104：判定所述待分析任务为安全任务。

举例说明，在所述待分析任务被判定为安全任务时，则可以通过系统进行上线操作，上线后所述待分析任务能够被系统中的其他用户（例如其他开发人员等）所查到，也可以被其他用户所执行、引用等。

针对所述待分析任务被判定为危险任务的情况，可以直接将所述待分析任务发回开发方重新修改，或通知开发方上线不成功等信息。或者，为了避免错误的安全性分析可能，还可以对判定为危险任务的所述待分析任务进行二次安全性分析，以减少错判率。可选的，若所述相关待分析语句满足安全分析条件，还包括：

将判定为危险任务的所述待分析任务发至分析终端，由所述分析终端对所述待分析任务进行二次安全性分析。

举例说明，所述分析终端可以为其他自动分析的服务器，对判定为危险任务的所述待分析任务再一次进行安全性分析，以核实第一次安全性分析结果的正确性。所述分析终端也可以是专业数据分析人员，通过人工的方式对判定为危险任务的所述待分析任务进行人工分析，以核实所述待分析任务是否确实出现安全性问题。

由上述实施例可以看出，首先根据语句的组合关系对用于分析的语句进行分类，分类得到敏感集合和安全集合，所述敏感集合中包括能造成敏感数据泄露的语句组合，所述安全集合中包括安全的语句组合，根据所述敏感集合和安全集合建立安全分析模型，当获取与待分析任务的相关待分析语句时，可以调用所述安全分析模型对所述相关待分析语句进行安全性分析，将预先得到的敏感语句组合和安全语句组合与待分析任务的相关语句进行比对分析，根据是否满足根据所述敏感集合和所述安全集合设置的安全分析条件来判定是否为安全任务或是否为危险任务，由此可对待分析任务进行自动化比对分析，从而去除了传统分析方式中的人为影响，提供了稳定的分析准确性，增加了系统的安全性。

实施例二

图 3 为本发明实施例提供的一种数据分析装置的装置结构图，根据语句

的组合关系对用于分析的语句进行分类，分类得到敏感集合和安全集合，所述敏感集合中包括能造成敏感数据泄露的语句组合，所述安全集合中包括安全的语句组合，根据所述敏感集合和安全集合建立安全分析模型，所述装置包括：

第一获取单元 301，用于获取待分析任务的相关待分析语句。

分析单元 302，用于调用所述安全分析模型对所述相关待分析语句进行安全性分析。

若所述分析结果为所述相关待分析语句满足安全分析条件，触发判断单元 303 判定所述待分析任务为危险任务，所述安全分析条件根据所述敏感集合和所述安全集合设置。

若所述分析结果为所述相关待分析语句不满足所述安全分析条件，触发所述判定单元 303 判定所述待分析任务为安全任务。

可选的，还包括：

第二获取单元，用于在触发所述第一获取单元 301 之前，获取所述待分析任务的任务标识；

调取单元，用于根据所述任务标识从数据库保存的数据中调取与所述待分析任务的待分析语句；

查询单元，用于根据所述任务标识查询所述待分析任务是否具有子任务，所述子任务为对所述待分析任务输出数据具有依赖关系的任务；

若具有，所述调取单元还用于从数据库保存的数据库中调取所述子任务的语句；

且若具有，所述第一获取单元 301 还用于获取所述待分析任务的待分析语句、所述子任务的语句以及根据所述依赖关系得到的所述待分析任务的待分析语句和所述子任务的语句之间的语句组合关系。

可选的，所述根据所述敏感集合和安全集合建立安全分析模型，包括：

根据语句的组合关系将所述敏感集合中的语句组合建立至少一个敏感抽象语法树 AST，将所述安全集合中的语句组合建立至少一个安全 AST；

所述分析单元 302 具体用于根据所述相关待分析语句中的语句组合关系，将所述相关待分析语句建立至少一个待分析 AST；将所述至少一个待分析

AST 与所述安全分析模型中的所述至少一个敏感 AST 和安全 AST 进行比对分析。

可选的，所述安全分析条件包括下述一个或多个的组合：

所述至少一个待分析 AST 中用于数据产出的子 AST 与所述至少一个敏感 AST 中的相同，且所述子 AST 与所述至少一个安全 AST 不同；

所述相关待分析语句中一个应用程序编程接口 API 对应的 AST 的至少一个层级与所述至少一个敏感 AST 中的相同；

所述相关待分析语句中多个应用程序编程接口 API 对应的 AST 的至少一个层级与所述至少一个敏感 AST 中的相同。

可选的，还包括：

发送单元，用于若所述分析结果为所述相关待分析语句满足安全分析条件，将判定为危险任务的所述待分析任务发至分析终端，由所述分析终端对所述待分析任务进行二次安全性分析。

由上述实施例可以看出，首先根据语句的组合关系对用于分析的语句进行分类，分类得到敏感集合和安全集合，所述敏感集合中包括能造成敏感数据泄露的语句组合，所述安全集合中包括安全的语句组合，根据所述敏感集合和安全集合建立安全分析模型，当获取与待分析任务的相关待分析语句时，可以调用所述安全分析模型对所述相关待分析语句进行安全性分析，将预先得到的敏感语句组合和安全语句组合与待分析任务的相关语句进行比对分析，根据是否满足根据所述敏感集合和所述安全集合设置的安全分析条件来判定是否为安全任务或是否为危险任务，由此可对待分析任务进行自动化比对分析，从而去除了传统分析方式中的人为影响，提供了稳定的分析准确性，增加了系统的安全性。

本领域普通技术人员可以理解：实现上述方法实施例的全部或部分步骤可以通过程序指令相关的硬件来完成，前述程序可以存储于一计算机可读取存储介质中，该程序在执行时，执行包括上述方法实施例的步骤；而前述的存储介质可以是下述介质中的至少一种：只读存储器（英文：read-only memory，缩写：ROM）、RAM、磁碟或者光盘等各种可以存储程序代码的介

质。

需要说明的是，本说明书中的各个实施例均采用递进的方式描述，各个实施例之间相同相似的部分互相参见即可，每个实施例重点说明的都是与其他实施例的不同之处。尤其，对于设备及系统实施例而言，由于其基本相似于方法实施例，所以描述得比较简单，相关之处参见方法实施例的部分说明即可。以上所描述的设备及系统实施例仅仅是示意性的，其中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实施例方案的目的。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。

以上所述，仅为本发明较佳的具体实施方式，但本发明的保护范围并不局限于此，任何熟悉本技术领域的技术人员在本发明揭露的技术范围内，可轻易想到的变化或替换，都应涵盖在本发明的保护范围之内。因此，本发明的保护范围应该以权利要求的保护范围为准。

权 利 要 求 书

1、一种数据分析方法，其特征在于，根据语句的组合关系对用于分析的语句进行分类，分类得到敏感集合和安全集合，所述敏感集合中包括能造成敏感数据泄露的语句组合，所述安全集合中包括安全的语句组合，根据所述敏感集合和安全集合建立安全分析模型，所述方法包括：

获取待分析任务的相关待分析语句；

调用所述安全分析模型对所述相关待分析语句进行安全性分析，

若所述相关待分析语句满足安全分析条件，则判定所述待分析任务为危险任务，所述安全分析条件根据所述敏感集合和所述安全集合设置；

若所述相关待分析语句不满足所述安全分析条件，则判定所述待分析任务为安全任务。

2、根据权利要求1所述的方法，其特征在于，在所述获取待分析任务的相关待分析语句之前，还包括：

获取所述待分析任务的任务标识；

根据所述任务标识从数据库保存的数据中调取与所述待分析任务的待分析语句；

根据所述任务标识查询所述待分析任务是否具有子任务，所述子任务为对所述待分析任务输出数据具有依赖关系的任务；

若具有，从数据库保存的数据库中调取所述子任务的语句；

且若具有，所述获取待分析任务的相关待分析语句，包括：

获取所述待分析任务的待分析语句、所述子任务的语句以及根据所述依赖关系得到的所述待分析任务的待分析语句和所述子任务的语句之间的语句组合关系。

3、根据权利要求1所述的方法，其特征在于，所述根据所述敏感集合和安全集合建立安全分析模型，包括：

根据语句的组合关系将所述敏感集合中的语句组合建立至少一个敏感抽象语法树AST，将所述安全集合中的语句组合建立至少一个安全AST；

所述调用所述安全分析模型对所述相关待分析语句进行安全性分析，包括：

根据所述相关待分析语句中的语句组合关系，将所述相关待分析语句建立至少一个待分析 AST；

将所述至少一个待分析 AST 与所述安全分析模型中的所述至少一个敏感 AST 和安全 AST 进行比对分析。

4、根据权利要求 3 所述的方法，其特征在于，所述安全分析条件包括下述一个或多个的组合：

所述至少一个待分析 AST 中用于数据产出的子 AST 与所述至少一个敏感 AST 中的相同，且所述子 AST 与所述至少一个安全 AST 不同；

所述相关待分析语句中一个应用程序编程接口 API 对应的 AST 的至少一个层级与所述至少一个敏感 AST 中的相同；

所述相关待分析语句中多个应用程序编程接口 API 对应的 AST 的至少一个层级与所述至少一个敏感 AST 中的相同。

5、根据权利要求 1 至 4 任一项所述的方法，其特征在于，若所述相关待分析语句满足安全分析条件，还包括：

将判定为危险任务的所述待分析任务发至分析终端，由所述分析终端对所述待分析任务进行二次安全性分析。

6、一种数据分析装置，其特征在于，根据语句的组合关系对用于分析的语句进行分类，分类得到敏感集合和安全集合，所述敏感集合中包括能造成敏感数据泄露的语句组合，所述安全集合中包括安全的语句组合，根据所述敏感集合和安全集合建立安全分析模型，所述装置包括：

第一获取单元，用于获取待分析任务的相关待分析语句；

分析单元，用于调用所述安全分析模型对所述相关待分析语句进行安全性分析；

若所述分析结果为所述相关待分析语句满足安全分析条件，触发判断单元判定所述待分析任务为危险任务，所述安全分析条件根据所述敏感集合和所述安全集合设置；

若所述分析结果为所述相关待分析语句不满足所述安全分析条件，触发所述判定单元判定所述待分析任务为安全任务。

7、根据权利要求 6 所述的装置，其特征在于，还包括：

第二获取单元，用于在触发所述第一获取单元之前，获取所述待分析任务的标识；

调取单元，用于根据所述任务标识从数据库保存的数据中调取与所述待分析任务的待分析语句；

查询单元，用于根据所述任务标识查询所述待分析任务是否具有子任务，所述子任务为对所述待分析任务输出数据具有依赖关系的任务；

若具有，所述调取单元还用于从数据库保存的数据库中调取所述子任务的语句；

且若具有，所述第一获取单元还用于获取所述待分析任务的待分析语句、所述子任务的语句以及根据所述依赖关系得到的所述待分析任务的待分析语句和所述子任务的语句之间的语句组合关系。

8、根据权利要求6所述的装置，其特征在于，所述根据所述敏感集合和安全集合建立安全分析模型，包括：

根据语句的组合关系将所述敏感集合中的语句组合建立至少一个敏感抽象语法树AST，将所述安全集合中的语句组合建立至少一个安全AST；

所述分析单元具体用于根据所述相关待分析语句中的语句组合关系，将所述相关待分析语句建立至少一个待分析AST；将所述至少一个待分析AST与所述安全分析模型中的所述至少一个敏感AST和安全AST进行比对分析。

9、根据权利要求8所述的装置，其特征在于，所述安全分析条件包括下述一个或多个的组合：

所述至少一个待分析AST中用于数据产出的子AST与所述至少一个敏感AST中的相同，且所述子AST与所述至少一个安全AST不同；

所述相关待分析语句中一个应用程序编程接口API对应的AST的至少一个层级与所述至少一个敏感AST中的相同；

所述相关待分析语句中多个应用程序编程接口API对应的AST的至少一个层级与所述至少一个敏感AST中的相同。

10、根据权利要求6至9任一项所述的装置，其特征在于，还包括：
发送单元，用于若所述分析结果为所述相关待分析语句满足安全分析条件，将判定为危险任务的所述待分析任务发至分析终端，由所述分析终端对所述待分析任务进行二次安全性分析。

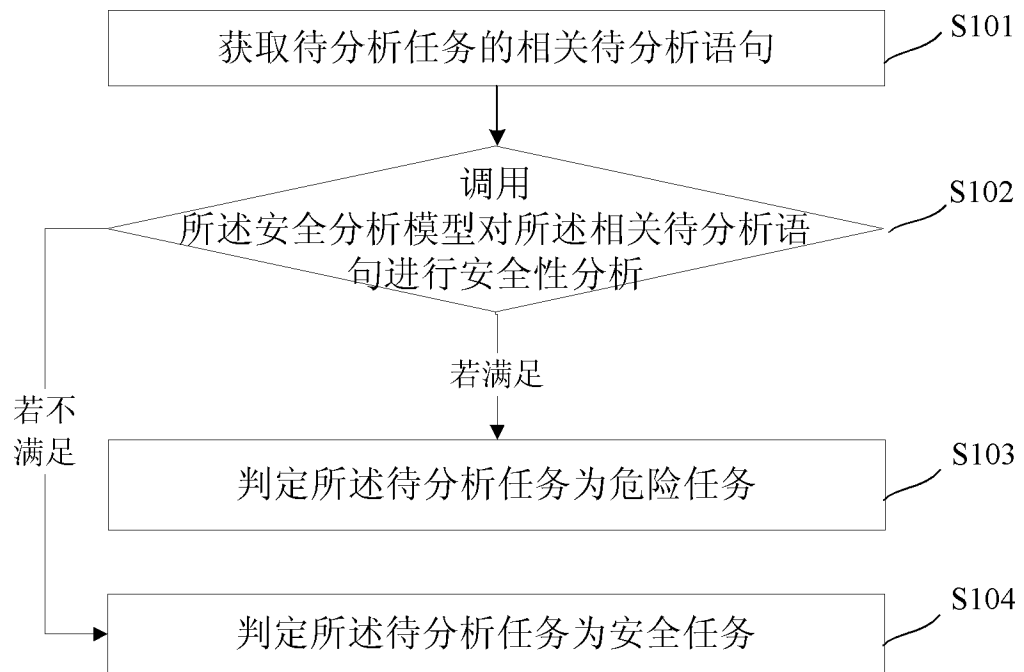


图 1

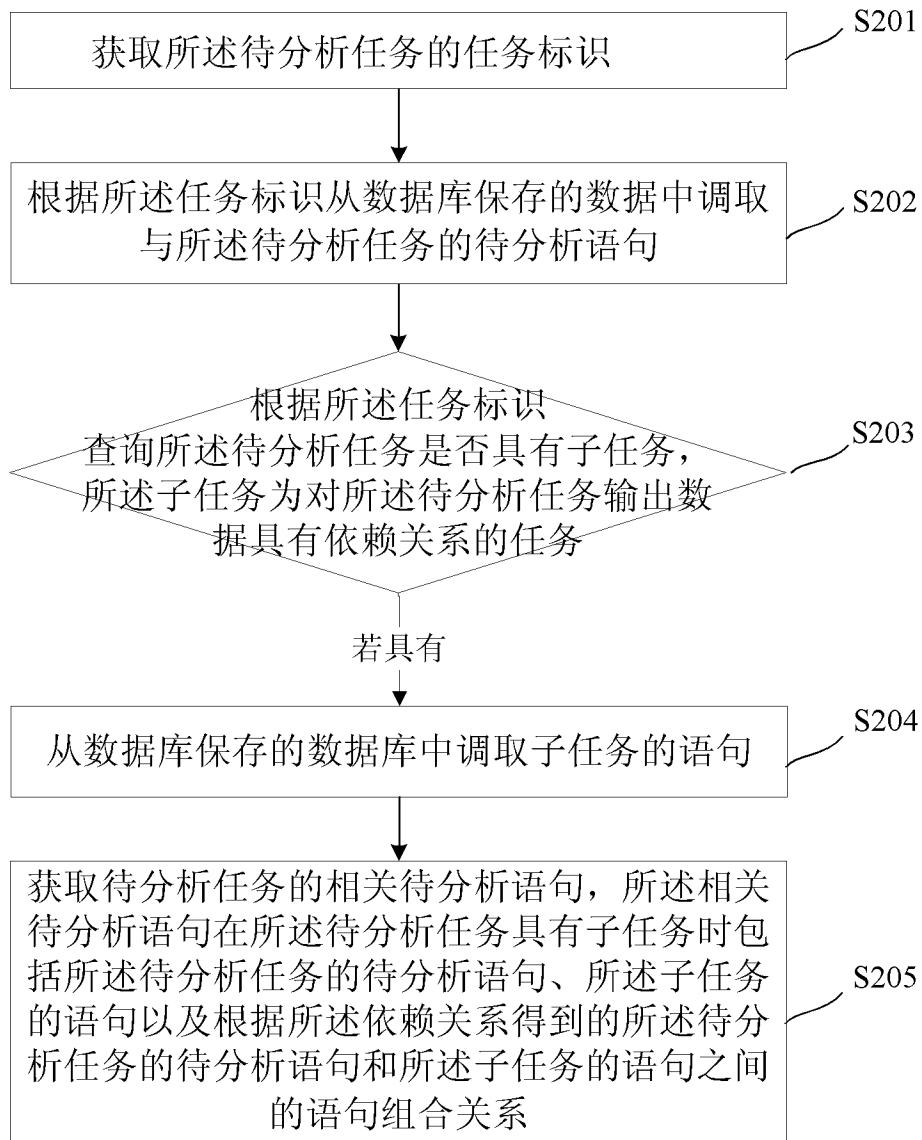


图 2

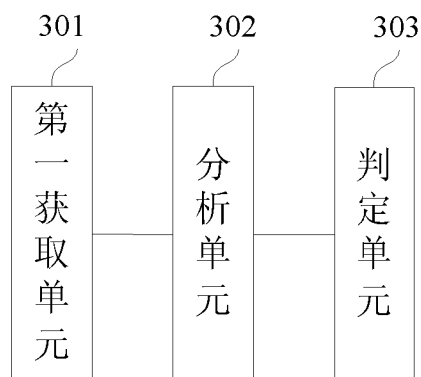


图 3

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2016/098816

A. CLASSIFICATION OF SUBJECT MATTER

G06F 11/36 (2006.01) i; G06Q 10/06 (2012.01) 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)

G06F, G06Q

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)

CNPAT, WPI, EPODOC, CNKI, GOOGLE: AST, API, data, analy+, sentence?, statement?, relation+, classf+, sensitive, safe+, secur+, leak+, model+, condition?, determin+, judg+, danger+, mark+, identif+, database?, subtask?, depend+, syntax w tree?, compar+, application, program+, interface

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 101833505 A (TIANJIN UNIVERSITY), 15 September 2010 (15.09.2010), description, paragraphs [0034]-[0066]	1, 3-6, 8-10
Y	CN 101833505 A (TIANJIN UNIVERSITY), 15 September 2010 (15.09.2010), description, paragraphs [0034]-[0066]	2, 7
Y	CN 104850493 A (BAIDU ONLINE NETWORK TECHNOLOGY (BEIJING) CO., LTD.), 19 August 2015 (19.08.2015), description, paragraphs [0030]-[0059]	2, 7
X	CN 103793652 A (GUANGDONG POWER GRID CORPORATION INFORMATION CENTRE), 14 May 2014 (14.05.2014), description, paragraphs [0029]-[0047]	1, 3-6, 8-10
Y	CN 103793652 A (GUANGDONG POWER GRID CORPORATION INFORMATION CENTRE), 14 May 2014 (14.05.2014), description, paragraphs [0029]-[0047]	2, 7
X	CN 102945203 A (INFORMATION CENTER OF SHENZHEN ENTRY-EXIT INSPECTION AND QUARANTINE BUREAU et al.), 27 February 2013 (27.02.2013), description, paragraphs [0024]-[0027] and [0039]-[0051]	1, 3-6, 8-10

☒ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	“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
“A” document defining the general state of the art which is not considered to be of particular relevance	“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
“E” earlier application or patent but published on or after the international filing date	“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
“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)	“&” document member of the same patent family
“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	

Date of the actual completion of the international search 08 December 2016 (08.12.2016)	Date of mailing of the international search report 19 December 2016 (19.12.2016)
Name and mailing address of the ISA/CN: State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No.: (86-10) 62019451	Authorized officer WEI, Jingyao Telephone No.: (86-10) 010-62414031

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2016/098816

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 102945203 A (INFORMATION CENTER OF SHENZHEN ENTRY-EXIT INSPECTION AND QUARANTINE BUREAU et al.), 27 February 2013 (27.02.2013), description, paragraphs [0024]-[0027] and [0039]-[0051]	2, 7
A	CN 104899043 A (BEIJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS), 09 September 2015 (09.09.2015), the whole document	1-10

International application No.
PCT/CN2016/098816

Form PCT/ISA/210 (patent family annex) (July 2009)

国际检索报告

国际申请号

PCT/CN2016/098816

A. 主题的分类

G06F 11/36(2006.01)i; G06Q 10/06(2012.01)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)

G06F, ; G06Q

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))

CNPAT, WPI, EPDOC, CNKI, GOOGLE: 数据, 分析, 语句, 关系, 分类, 敏感, 安全, 泄露, 模型, 条件, 判定, 判断, 危险, 标识, 子任务, 依赖, 语法树, AST, 比对, API, data, analy+, sentence?, statement?, relation+, classf+, sensitive, safe+, secur+, leak+, model+, condition?, determin+, judg+, danger+, mark+, identif+, database?, subtask?, depend+, syntax w tree?, compar+, application, program+, interface

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
X	CN 101833505 A (天津大学) 2010年 9月 15日 (2010 - 09 - 15) 说明书第[0034]-[0066]段	1、3-6、8-10
Y	CN 101833505 A (天津大学) 2010年 9月 15日 (2010 - 09 - 15) 说明书第[0034]-[0066]段	2、7
Y	CN 104850493 A (百度在线网络技术北京有限公司) 2015年 8月 19日 (2015 - 08 - 19) 说明书第[0030]-[0059]段	2、7
X	CN 103793652 A (广东电网公司信息中心) 2014年 5月 14日 (2014 - 05 - 14) 说明书第[0029]-[0047]段	1、3-6、8-10
Y	CN 103793652 A (广东电网公司信息中心) 2014年 5月 14日 (2014 - 05 - 14) 说明书第[0029]-[0047]段	2、7
X	CN 102945203 A (深圳出入境检验检疫局信息中心 等) 2013年 2月 27日 (2013 - 02 - 27) 说明书第[0024]-[0027]、[0039]-[0051]段	1、3-6、8-10

☒ 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

“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

2016年 12月 8日

国际检索报告邮寄日期

2016年 12月 19日

ISA/CN的名称和邮寄地址

中华人民共和国国家知识产权局(ISA/CN)
中国北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10)62019451

受权官员

魏晶瑶

电话号码 (86-10)010-62414031

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
Y	CN 102945203 A (深圳出入境检验检疫局信息中心 等) 2013年 2月 27日 (2013 - 02 - 27) 说明书第[0024]-[0027]、[0039]-[0051]段	2、7
A	CN 104899043 A (北京航空航天大学) 2015年 9月 9日 (2015 - 09 - 09) 全文	1-10

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2016/098816

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	101833505	A	2010年 9月 15日	无	
CN	104850493	A	2015年 8月 19日	无	
CN	103793652	A	2014年 5月 14日	无	
CN	102945203	A	2013年 2月 27日	无	
CN	104899043	A	2015年 9月 9日	无	