(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
2016 年 9 月 29 日 (29.09.2016)

WIPO | PCT

(10) 国际公布号
WO 2016/150313 A1

(51) 国际专利分类号:
G06F 21/56 (2013.01)

(21) 国际申请号: PCT/CN2016/076228

(22) 国际申请日: 2016 年 3 月 14 日 (14.03.2016)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201510124614.5 2015 年 3 月 20 日 (20.03.2015) CN

(71) 申请人: 阿里巴巴集团控股有限公司 (ALIBABA GROUP HOLDING LIMITED) [—/CN]; 开曼群岛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: 陈艳军 (CHEN, Yanjun); 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: 北京三友知识产权代理有限公司 (BEIJING SANYOU INTELLECTUAL PROPERTY AGENCY LTD.); 中国北京市金融街 35 号国际企业大厦 A 座 16 层, Beijing 100033 (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, 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)。

[见续页]

(54) Title: METHOD AND APPARATUS FOR DETECTING SUSPICIOUS PROCESS

(54) 发明名称 : 一种可疑进程的探测方法及装置

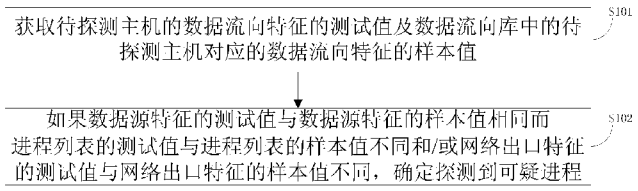

图 1

S101 Acquiring a test value of a data flow direction characteristic of a host to be detected and a sample value, in a data flow direction library, of the data flow direction characteristic corresponding to the host to be detected

S102 If a test value of a data source characteristic is identical to a sample value of the data source characteristic, a test value of a process list is different from a sample value of the process list and/or a test value of a network egress characteristic is different from a sample value of the network egress characteristic, determining that a suspicious process is detected

(57) Abstract: A method and apparatus for detecting a suspicious process. The method comprises: acquiring a test value of a data flow direction characteristic of a host to be detected and a sample value, in a data flow direction library, of the data flow direction characteristic corresponding to the host to be detected (S101), the data flow direction characteristic comprising at least one of a process list and a network egress characteristic, and a data source characteristic; and if a test value of the data source characteristic is identical to a sample value of the data source characteristic, a test value of the process list is different from a sample value of the process list and/or a test value of the network egress characteristic is different from a sample value of the network egress characteristic, determining that a suspicious process is detected (S102). It can be seen that the method and apparatus for detecting a suspicious process detect a suspicious process with reference to a data flow direction characteristic instead of an attack behaviour of an application program; and in the case of data theft, the data flow direction characteristic will change, such that the method and apparatus can accurately detect the suspicious process for data theft.

(57) 摘要:

[见续页]

WO 2016/150313 A1

本国际公布:

— 包括国际检索报告(条约第 21 条(3))。

一种可疑进程的探测方法及装置，获取待探测主机的数据流向特征的测试值及数据流向库中的待探测主机对应的数据流向特征的样本值(S101)，数据流向特征包括进程列表和网络出口特征中的至少一个、及数据源特征；如果在数据源特征的测试值与数据源特征的样本值相同的情况下，进程列表的测试值与进程列表的样本值不同和/或网络出口特征的测试值与网络出口特征的样本值不同，确定探测到可疑进程(S102)；可见，所述可疑进程探测方法及装置，以数据的流向特征为依据来探测可疑进程，而非依据应用程序的攻击行为，又因为一旦发生数据的盗用，数据的流向特征就会变化，所以所述方法及装置能够准确地检测到数据被盗取的可疑进程。

一种可疑进程的探测方法及装置

本申请要求2015年03月20日递交的申请号为201510124614.5、发明名称为“一种可疑进程的探测方法及装置”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及计算机领域，尤其涉及一种可疑进程的探测方法及装置。

背景技术

数据安全是云计算和数据开放面临的核心问题之一。以电商云为例，独立软件开发商（Independent Software Vendors，ISV）软件系统部署在电商云环境中，在获得天猫和淘宝商户的订购授权后，ISV能通过TOP访问商户在天猫和淘宝的敏感数据，如订单和客户关系等。ISV 任何软件或云资源管理上的漏洞，可能被人利用在云主机或应用中部署后门，非法读取、拷贝或外传这些敏感数据，造成大批量的数据泄漏。

而传统的病毒检测方法，通常针对病毒程序对系统的攻击行为设计防御策略，而在云主机或应用中盗取数据的后门程序，一般以获取数据为目的，并没有主动攻击系统的行为特征。

所以，传统的病毒检测技术，并不能准确地检测到数据被盗取的可疑进程。

发明内容

本申请提供了一种可疑进程的探测方法及装置，目的在于解决不能准确检测到数据被盗取的可疑进程的问题。

为了实现上述目的，本申请提供了以下技术方案：

一种可疑进程的探测方法，包括：

获取待探测主机的数据流向特征的测试值，以及，数据流向库中所述待探测主机对应的数据流向特征的样本值；所述数据流向特征包括进程列表和网络出口特征中的至少一个、以及数据源特征，所述数据源特征用于指示流入所述待探测主机的预设类型数据的数据源，所述进程列表中包括按照时间顺序排列的、调用所述数据源流出的数据的进程，所述网络出口特征用于指示所述数据源流出的数据在被所述进程列表中的进程调用后、流出所述待探测主机的出口；

如果在所述数据源特征的测试值与所述数据源特征的样本值相同的情况下，所述进程列表的测试值与所述进程列表的样本值不同，和/或，所述网络出口特征的测试值与所述网络出口特征的样本值不同，则确定探测到可疑进程。

可选地，所述数据流向库的建立过程包括：

按照以下方式，分别建立每一个数据源的数据流向特征：

从预先获取的所述待探测主机的网络事件表中确定与一个所述数据源特征相关的网络事件；

以所述网络事件的进程编号和时间戳为查找条件，通过关联所述网络事件表、所述待探测主机的进程事件表及所述待探测主机的文件读写事件表，依次获得调用所述数据源的数据的进程，以及，所述第二数据源的数据流出的第二网络出口。

可选地，还包括：

获取所述待探测主机中每一个应用程序的行为特征的测试值，以及，应用行为库中所述待探测主机的行为特征的样本值；所述行为特征至少包括以下一项：应用程序的级别、应用程序访问所述预设类型的数据的数据源的频率、应用程序对外连接的频率、应用程序对外连接的目的地址、应用程序对外连接的端口、运行应用程序的用户、应用程序的进程命令参数、应用程序的运行频率及应用程序的运行时长；

如果任意一个应用程序的行为特征中任意一项的测试值与此应用程序在所述行为特征库中此项行为特征的样本值的差别不在预设范围内，则确定探测到可疑进程。

可选地，在行为特征中的任意一项为多维数据的情况下，此行为特征的测试值与所述行为特征库中的此项行为特征的样本值的差别的确定方法包括：

计算该项的测试值与所述行为特征库中此项行为特征的样本值之间的距离值。

可选地，所述应用行为库的建立过程包括：

从预先获取的所述待探测主机的网络事件表及进程事件表中，获取每一个应用程序的行为特征；所述行为特征包括应用程序的级别、应用程序访问所述预设类型的数据的数据源的频率、应用程序对外连接的频率、应用程序对外连接的目的地址、应用程序对外连接的端口、运行应用程序的用户、应用程序的进程命令参数、应用程序的运行频率及应用程序的运行时长。

可选地，还包括：

按照预设的进程风险规则，从所述待探测主机的进程中确定可疑进程，所述进程风险规则包括：所述待探测主机发起对自身的网络连接及该连接的目标端口为远程登录端

口。

一种可疑进程的探测装置，包括：

第一获取模块，用于获取待探测主机的数据流向特征的测试值，以及，数据流向库中所述待探测主机对应的数据流向特征的样本值；所所述数据流向特征包括进程列表和网络出口特征中的至少一个、以及数据源特征，所述数据源特征用于指示流入所述待探测主机的预设类型数据的数据源，所述进程列表中包括按照时间顺序排列的、调用所述数据源流出的数据的进程，所述网络出口特征用于指示所述数据源流出的数据在被所述进程列表中的进程调用后、流出所述待探测主机的出口；

第一确定模块，用于如果在所述数据源特征的测试值与所述数据源特征的样本值相同的情况下，所述进程列表的测试值与所述进程列表的样本值不同，和/或，所述网络出口特征的测试值与所述网络出口特征的样本值不同，则确定探测到可疑进程。

可选地，还包括：

数据流向库建立模块，用于按照以下方式，分别建立每一个数据源的数据流向特征：从预先获取的所述待探测主机的网络事件表中确定与一个所述数据源特征相关的网络事件；以所述网络事件的进程编号和时间戳为查找条件，通过关联所述网络事件表、所述待探测主机的进程事件表及所述待探测主机的文件读写事件表，依次获得调用所述数据源的数据的进程，以及，所述第二数据源的数据流出的第二网络出口。

可选地，还包括：

第二获取模块，用于获取所述待探测主机中每一个应用程序的行为特征的测试值以及应用行为库中所述待探测主机的行为特征的样本值；所述行为特征至少包括以下一项：应用程序的级别、应用程序访问所述预设类型的数据的数据源的频率、应用程序对外连接的频率、应用程序对外连接的目的地址、应用程序对外连接的端口、运行应用程序的用户、应用程序的进程命令参数、应用程序的运行频率及应用程序的运行时长；

第二确定模块，用于如果任意一个应用程序的行为特征中任意一项的测试值与此应用程序在所述行为特征库中此项行为特征的样本值的差别不在预设范围内，则确定探测到可疑进程。

可选地，所述第二确定模块用于在行为特征中的任意一项为多维数据的情况下，确定此行为特征的测试值与所述行为特征库中的此项行为特征的样本值的差别的具体过程包括：

所述第二确定模块具体用于，计算该项的测试值与所述行为特征库中的此项行为特

征的样本值的距离值。

可选地，还包括：

应用行为库建立模块，用于从预先获取的所述待探测主机的网络事件表及进程事件表中获取每一个应用程序的行为特征；所述行为特征包括应用程序的级别、应用程序访问所述预设类型的数据的数据源的频率、应用程序对外连接的频率、应用程序对外连接的目的地址、应用程序对外连接的端口、运行应用程序的用户、应用程序的进程命令参数、应用程序的运行频率及应用程序的运行时长。

可选地，还包括：

第三确定模块，用于按照预设的进程风险规则，从所述待探测主机的进程中确定可疑进程，所述进程风险规则包括：所述待探测主机发起对自身的网络连接及该连接的目标端口为远程登录端口。

与现有技术相比，本申请实施例具有以下有益效果：

本申请所述的可疑进程的探测方法及装置，获取待探测主机的数据流向特征的测试值及数据流向库中的待探测主机对应的数据流向特征的样本值，其中，数据流向特征包括数据源特征、进程列表以及网络出口特征，如果在所述数据源特征的测试值与所述数据源特征的样本值相同的情况下，进程列表的测试值与进程列表的样本值不同或者网络出口特征的测试值与网络出口特征的样本值不同，则确定探测到可疑进程，可见，本申请所述的可疑进程探测方法及装置，以数据的流向特征为依据来探测可疑进程，而非依据应用程序的攻击行为，又因为一旦发生数据的盗用，数据的流向特征就会变化，所以本申请所述的可疑进程探测方法及装置能够准确地检测到数据被盗取的可疑进程。

当然，实施本申请的任一产品并不一定需要同时达到以上所述的所有优点。

附图说明

为了更清楚地说明本申请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申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

图 1 为申请实施例公开的一种可疑进程的探测方法的流程图；

图 2 为本申请实施例公开的又一种可疑进程的探测方法的流程图；

图 3 为本申请实施例公开的数据流向库及应用行为库的建立方法的流程图；

图 4 为本申请实施例公开的通过在待探测主机上部署的采集客户端，采集在预设时间段内的事件数据的方法的流程图；

图 5 为本申请实施例公开的一个数据源的数据流向特征的示意图；

图 6 为本申请实施例公开的一种可疑进程的探测装置的结构示意图；

图 7 为本申请实施例还公开的又一种可疑进程的探测装置的结构示意图；

图 8 为本申请实施例公开的可疑进程的探测装置与待探测主机的连接关系结构示意图。

具体实施方式

本申请实施例公开了一种可疑进程的探测方法及装置，可以应用在对于云主机的可疑进程的探测上，以便于能够准确发现盗取云主机中的数据的可疑进程。

下面将结合本申请实施例中的附图，对本申请实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，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，都属于本申请保护的范围。

本申请实施例公开的一种可疑进程的探测方法，如图 1 所示，包括以下步骤：

S101：获取待探测主机的数据流向特征的测试值及数据流向库中的待探测主机对应的数据流向特征的样本值。

本实施例中，数据流向特征可以包括：进程列表和网络出口特征中的至少一个、以及数据源特征。也就是说，数据流向特征中包括数据源特征，除此以外，可以包括进程列表和网络出口特征，也可以包括两者中的任意一个。在包括进程列表和网络出口特征的情况下，探测的准确性更高，本申请以下实施例中，均以数据流向特征中包括数据源特征、进程列表和网络出口特征为例进行描述。

其中，数据源特征用于指示流入待探测主机的预设类型数据的数据源，进程列表中包括按照时间顺序排列的、调用数据源流出的数据的进程，网络出口特征用于指示数据源流出的数据在被进程列表中的进程调用后、流出待探测主机的出口。

进程列表的测试值与样本值均可以为进程列表中包含的文件的名称或者编号。进程列表的测试值与样本值不同，说明调用数据的进行发生了改变，改变可以包括增加了进程，或者，进程按照时间排序发生了变化。

数据源特征的测试值与样本值均可以为数据源的地址或端口号，网络出口特征的测

试值与样本值均可以为网络出口的地址或端口号。

S102：在数据源特征的测试值与数据源特征的样本值相同的情况下，如果满足预设条件，则确定探测到可疑进程。其中预设条件至少包括以下任意一项：

1、进程列表的测试值与进程列表的样本值不同；

2、网络出口特征的测试值与网络出口特征的样本值不同。

因为数据盗取方需要通过后门应用程序读取数据或者将数据引流的方式盗取数据，所以，本实施例中，从这两方面入手，设置以上条件，以便从根本上发现数据盗取行为。

可疑进程即为异常的特征对应的进程。例如，对于进程列表而言，可疑进程可以为进程列表的测试值与样本值相比，多出来的名称或编号对应的进程为可疑进程；对于网络出口特征而言，测试样与样本值不同，则通过此网络出口向得探测主机外传输数据的进程即为可疑进程。

例如，本实施例所述的方法应用在电商平台，对于电商平台而言，预设类型数据可以为敏感数据，例如客户的订单信息，为了防止云主机中的敏感数据泄露，探测电商平台的云主机中的数据流向特征的测试值，并从数据流向库中获取云主机的数据流向特征的样本值。如果敏感信息的数据源特征的测试值与样本值相同，而其进程列表中的测试值与样本值不同，例如，调用敏感信息的进程多了一个，多出的进程可能为盗取数据的进程。因此，可以确定存在数据被盗取的风险，则确定探测到可疑进程。网络运维人员可以进一步确定是否确实为风险进程，如确实为风险进程，则采用相应的处理措施。

可见，与现有的检测病毒的技术相比，本实施例中所述的方法，从数据盗取的特点为出发点，以待探测主机中的数据流向特征为依据进行可疑进程的探测，因此，能够准确发现盗取数据的可疑进程。

在上述实施例所述的方法的基础上，还可以进一步增加其它步骤以提高探测可疑进程的准确性，本申请实施例公开的又一种可疑进程的探测方法，与上一实施例相比，探测可疑进程的依据不仅仅限于数据流向特征，如图 2 所示，包括以下具体步骤：

S201：获取待探测主机的数据流向特征的测试值及数据流向库中的待探测主机对应的数据流向特征的样本值。

本实施例中，数据流向特征可以参见上一实施例所示，这里不再赘述。

S202：在数据源特征的测试值与数据源特征的样本值相同的情况下，如果进程列表的测试值与进程列表的样本值不同和/或网络出口特征的测试值与网络出口特征的样本

值不同，则确定探测到可疑进程。

S203：获取待探测主机中的每一个应用程序的行为特征的测试值以及应用行为库中的、待探测主机的行为特征的样本值。

本实施例中，行为特征用于表示各个应用程序在待探测主机中的行为，至少可以包括以下一项：应用程序的级别、应用程序访问所述预设类型的数据的数据源的频率、应用程序对外连接的频率、应用程序对外连接的目的地址、应用程序对外连接的端口、运行应用程序的用户、应用程序的进程命令参数、应用程序的运行频率及应用程序的运行时长。

其中，本实施例中，应用程序可以分为四个级别，分别为：直接访问数据源或中间文件并向外发起网络连接的 L1 级程序，访问数据源或中间文件但无其它网络连接的 L2 级程序，不访问数据源或中间文件但有主动外连行为的 L3 级程序，以及，不访问数据源或中间文件且无主动外连行为的 L4 级程序。

S204：如果任意一个应用程序的行为特征中的任意一项的测试值与此应用程序在所述行为库中的特征中的此项行为特征的样本值的差别不在预设范围内，则确定探测到可疑进程。

本实施例中，在行为特征中的任意一项为多维数据的情况下，此行为特征的测试值与行为特征库中的此项行为特征的样本值的差别的确定方法可以为：计算该项的测试值与行为特征库中的此项行为特征的样本值的距离值，所述距离值为该项的测试值与行为特征库中的此项行为特征的样本值的差别，其中，距离值可以为 K 近邻距离值；在行为特征中的任意一项为多维数据的情况下，可以直接计算测试值与样本值之间的差别。

本实施例中，预设范围可以预先设定，范围越大，容忍性越大，则探测可疑进程的条件越宽松，范围越小，容忍性越小，则探测可疑进程的条件越严格，实际应用中，可以依据实际需求设定。

例如，在云主机中，某个应用程序向外连接的频率的测试值明显大于其样本值，则说明此应用程序可能在进行数据的盗取，从而确定探测到可疑进程，可疑进程即为此应用程序对应的进程。

S205：按照预设的进程风险规则，从所述待探测主机的进程中确定可疑进程。

其中，进程风险规则可以包括：待探测主机发起对自身的网络连接及该连接的目标端口为远程登录端口。

本实施例中，进程风险规则可以通过由网络维护人员通过先验知识推导得到。

设置S205的目的在于，可以将一些可能导致重大安全事故的特征加入进程风险规则中，一旦待探测主机中出现这些特征，可以直接确定存在可疑进程，而不必通过数据流向特征及行为特征的探测。

从图2中可以看出，S201和S202，S203和S204，以及，S205分别为探测可疑进程的3个分支，需要说明的是，本实施例中为步骤标记的序号仅为便于说明，在实际应用中，这三个分支的执行顺序不做限定。

可选地，在确定探测到可疑进程后，本实施例中还可以包括：

发出预警信号，并将可疑进程的编号加入可疑进程列表中，以便于网络运营人员查看可疑进程列表，并对可疑进程进行审核，如确认无风险，在系统中标定，该标定记录将被加入该主机的应用行为库或数据流向库中。如有风险，能直观地定位有风险的进程，或数据异常流动的模式，启动应急措施控制风险。

本实施例所述的方法应用在电商平台的云主机中，可以分别从三个方面探测盗取数据的可疑进程，云主机上部署的应用程序，由于没有人机交互行为，所以，数据在主机内的流转、以及主机内的应用程序的行为比较固定，有明显的特征。本实施例所述的方法，分别从待探测主机中的数据流向特征、行为特征及进程风险规则三个方面探测待探测主机中盗取数据的可疑进程，因此，探测的角度更为多样化，能够更为及时准确地发现数据访问行为中的异常，定位到可疑的后门或程序。即使后门程序本身发生大的改变，只要盗取数据的行为还存在，就能被快速发现。

需要说明的是，上述实施例中，测试值可以从待探测主机第一时间段（例如某一天）的事件记录中采集，而样本值可以依据待探测主机在第二时间段内（例如一个月）的时间记录中采集的数据而生成。

下面将详细说明数据流向库及应用行为库的建立方法。

如图3所示，数据流向库及应用行为库的建立方法可以包括以下具体步骤：

S301：通过在待探测主机上部署的采集客户端，采集在预设时间段（例如一个月）内的事件数据。

其中，事件可以具体包括网络事件、进程事件以及文件读写事件。

具体地，网络事件数据可以包括发起或接收网络连接的进程的标号、发起时间、来源IP和端口以及目的IP和端口。进程事件数据可以包括进程的编号、事件类型（包括启动或停止）、活动时间、进程名以及命令行参数。文件读写事件数据可以包括文件读

写操作记录进程的编号、读写类型（包括读或写）以及活动时间。

如图 4 所示，S301 的具体实现过程可以包括以下步骤：

1、通过 Windows 事件收集器使用 ETW 框架，或者 Linux 事件收集器使用 Audit 框架，从操作系统底层抓取网络事件数据、进程事件数据和文件事件数据，为了减小事件数据量，抓取的粒度和事件类型由事件处理器通过配置来控制和过滤，排除已知的无风险进程、网络或文件活动。

2、事件处理器可以将数据整理成统一格式，如果当前数据不足以建立数据流向库，事件处理器可以调用系统函数补全数据后，实时上传到日志收集服务器。

3、日志收集服务器在数据缓存到一定容量或者超过某个时间值后，将已接收的数据同步到大数据处理平台存储、以及等待进一步处理。

S302：依据上述采集到的数据，在大数据处理平台上，事件数据被写入网络事件表、进程事件表及文件读写事件表。

上述三种事件表可以按照时间维度进行分片存储。

S303：按照以下方式，分别建立每一个数据源特征的数据流向特征：从所述网络事件表中确定与一个所述数据源特征相关的网络事件，以所述网络事件的进程编号和时间戳-为查找条件，通过关联所述网络事件表、所述进程事件表及所述文件读写事件表，依次获得调用所述数据源的数据的进程以及所述第二数据源的数据流出的第二网络出口。

其中，每一个数据源特征均可以通过人工提取获得，以电商云 TOP 数据源为例，TOP 服务器的 IP 地址为一固定列表，且服务端口为 80。

如图 5 所示，一个数据源的数据流向特征为：数据从数据源（如电商云中的 TOP）通过网络事件到达云主机第一个进程，再通过文件写事件存到本地文件，服务进程（一般为 Web 服务器）通过文件读事件读取到数据，随后通过网络事件发送给客户或者第三方系统（如电商云中的物流系统）。可见，一个数据源的流向特征表示的是从此数据源流出的数据在待探测主机中的流动路径。

需要说明的是，文件读事件仅为连接前后进程的中间过程，本申请的实施例中，不将文件读写事件看作参与探测可疑进程的数据流向特征。

本实施例中，依据以下步骤建立应用行为库：

S304：从所述网络事件表及所述进程事件表中获取每一个应用程序的行为特征，所述行为特征包括应用程序的级别、应用程序访问所述预设类型的数据的数据源的频率、应用程序对外连接的频率、应用程序对外连接的目的地址、应用程序对外连接的端口、

运行应用程序的用户、应用程序的进程命令参数、应用程序的运行频率及应用程序的运行时长。

需要说明的是，S303 和 S304 的执行顺序可以互换。

从上述过程可以看出，数据流向库以及应用行为特征库从待探测主机的日程运行数据得到，因此，具有时效性高的特点，并且，数据采集过程不会影响待探测的主机的正常运行。

与图 1 所示的方法实施例相对应地，本申请实施例还公开了一种可疑进程的探测装置，如图 6 所示，包括：

第一获取模块 601，用于获取待探测主机的数据流向特征的测试值以及数据流向库中的所述待探测主机对应的数据流向特征的样本值，所述数据流向特征包括进程列表和网络出口特征中的至少一个、以及数据源特征，所述数据源特征用于指示流入所述待探测主机的预设类型数据的数据源，所述进程列表中包括按照时间顺序排列的、调用所述数据源流出的数据的进程，所述网络出口特征用于指示所述数据源流出的数据在被所述进程列表中的进程调用后、流出所述待探测主机的出口；

第一确定模块 602，用于如果在所述数据源特征的测试值与所述数据源特征的样本值相同的情况下，所述进程列表的测试值与所述进程列表的样本值不同和/或所述网络出口特征的测试值与所述网络出口特征的样本值不同，则确定探测到可疑进程。

本实施例中所述的装置，以数据盗取的特点为出发点，以待探测主机中的数据流向特征为依据进行可疑进程的探测，因此，能够准确发现盗取数据的可疑进程。

与图 2 所示的方法实施例相对应地，本申请实施例还公开的又一种可疑进程的探测装置如图 7 所示，包括：第一获取模块 701、第一确定模块 702、第二获取模块 703、第二确定模块 704 以及第三确定模块 705。

其中，第一获取模块 701 和第一确定模块 702 的功能与上一实施例相同，这里不再赘述。

第二获取模块 703，用于获取所述待探测主机中的每一个应用程序的行为特征的测试值以及应用行为库中的、所述待探测主机的行为特征的样本值，所述行为特征至少包括以下一项：应用程序的级别、应用程序访问所述预设类型的数据的数据源的频率、应用程序对外连接的频率、应用程序对外连接的目的地址、应用程序对外连接的端口、运

行应用程序的用户、应用程序的进程命令参数、应用程序的运行频率及应用程序的运行时长。

第二确定模块 704，用于如果任意一个应用程序的行为特征中的任意一项的测试值与此应用程序在所述行为特征库中的此项行为特征的样本值的差别不在预设范围内，则确定探测到可疑进程。

具体地，第二确定模块在行为特征中的任意一项为多维数据的情况下，确定此行为特征的测试值与所述行为特征库中的此项行为特征的样本值的差别的具体过程可以为：计算该项的测试值与所述行为特征库中的此项行为特征的样本值的距离值，所述距离值为该项的测试值与所述行为特征库中的此项行为特征的样本值的差别。

第三确定模块 705，用于按照预设的进程风险规则，从所述待探测主机的进程中确定可疑进程，所述进程风险规则包括：所述待探测主机发起对自身的网络连接及该连接的目标端口为远程登录端口。

可选地，本实施例所述装置还可以包括：

数据流向库建立模块 706，用于按照以下方式，分别建立每一个数据源的数据流向特征：从预先获取的所述待探测主机的网络事件表中确定与一个所述数据源特征相关的网络事件；以所述网络事件的进程编号和时间戳为查找条件，通过关联所述网络事件表、所述待探测主机的进程事件表及所述待探测主机的文件读写事件表，依次获得调用所述数据源的数据的进程以及所述第二数据源的数据流出的第二网络出口。

以及，应用行为库建立模块 707，用于从预先获取的所述待探测主机的网络事件表及进程事件表中获取每一个应用程序的行为特征，所述行为特征包括应用程序的级别、应用程序访问所述预设类型的数据的数据源的频率、应用程序对外连接的频率、应用程序对外连接的目的地址、应用程序对外连接的端口、运行应用程序的用户、应用程序的进程命令参数、应用程序的运行频率及应用程序的运行时长。

数据流向库建立模块和应用行为库建立模块的工作流程可以参见图 3 所示的方法实施例。

本实施例中所述的装置，可以设置在数据处理平台上，例如电商的大数据处理平台。所述数据处理平台与待探测主机相连，图 8 所示为本实施例所述的装置与待探测主机的连接关系。数据处理平台可以通过现有的数据采集模块及数据传输模块，将待探测主机中的数据传输到数据处理平台，数据处理平台的事件数据存储模块可以将这些数据进行存储，本实施例所述的装置按照以上所述的功能对数据进行分析整理，并依据分析整理

的结果探测待探测主机中的可疑进程。

需要说明的是，本申请实施例中所述的装置可以设置在电子设备中，所述电子设备除了可以为专业的监测设备外，也可以为移动终端设备。

本实施例所述的可疑进程的探测方法，从多个角度探测可疑进程，因此具有更高的准确性以及更小的时延。

本申请实施例方法所述的功能如果以软件功能单元的形式实现并作为独立的产品销售或使用时，可以存储在一个计算设备可读取存储介质中。基于这样的理解，本申请实施例对现有技术做出贡献的部分或者该技术方案的部分可以以软件产品的形式体现出来，该软件产品存储在一个存储介质中，包括若干指令用以使得一台计算设备（可以是个人计算机，服务器，移动计算设备或者网络设备等）执行本申请各个实施例所述方法的全部或部分步骤。而前述的存储介质包括：U 盘、移动硬盘、只读存储器（ROM，Read-Only Memory）、随机存取存储器（RAM，Random Access Memory）、磁碟或者光盘等各种可以存储程序代码的介质。

本说明书中各个实施例采用递进的方式描述，每个实施例重点说明的都是与其它实施例的不同之处，各个实施例之间相同或相似部分互相参见即可。

对所公开的实施例的上述说明，使本领域专业技术人员能够实现或使用本申请。对这些实施例的多种修改对本领域的专业技术人员来说将是显而易见的，本文中所定义的一般原理可以在不脱离本申请的精神或范围的情况下，在其它实施例中实现。因此，本申请将不会被限制于本文所示的这些实施例，而是要符合与本文所公开的原理和新颖特点相一致的最宽的范围。

权 利 要 求 书

1、一种可疑进程的探测方法，其特征在于，包括：

获取待探测主机的数据流向特征的测试值，以及，数据流向库中所述待探测主机对应的数据流向特征的样本值；所述数据流向特征包括进程列表和网络出口特征中的至少一个、以及数据源特征，所述数据源特征用于指示流入所述待探测主机的预设类型数据的数据源，所述进程列表中包括按照时间顺序排列的、调用所述数据源流出的数据的进程，所述网络出口特征用于指示所述数据源流出的数据在被所述进程列表中的进程调用后、流出所述待探测主机的出口；

如果在所述数据源特征的测试值与所述数据源特征的样本值相同的情况下，所述进程列表的测试值与所述进程列表的样本值不同，和/或，所述网络出口特征的测试值与所述网络出口特征的样本值不同，则确定探测到可疑进程。

2、根据权利要求 1 所述的方法，其特征在于，所述数据流向库的建立过程包括：

按照以下方式，分别建立每一个数据源的数据流向特征：

从预先获取的所述待探测主机的网络事件表中确定与一个所述数据源特征相关的网络事件；

以所述网络事件的进程编号和时间戳为查找条件，通过关联所述网络事件表、所述待探测主机的进程事件表及所述待探测主机的文件读写事件表，依次获得调用所述数据源的数据的进程，以及，所述第二数据源的数据流出的第二网络出口。

3、根据权利要求 1 或 2 所述的方法，其特征在于，还包括：

获取所述待探测主机中每一个应用程序的行为特征的测试值，以及，应用行为库中所述待探测主机的行为特征的样本值；所述行为特征至少包括以下一项：应用程序的级别、应用程序访问所述预设类型的数据的数据源的频率、应用程序对外连接的频率、应用程序对外连接的目的地址、应用程序对外连接的端口、运行应用程序的用户、应用程序的进程命令参数、应用程序的运行频率及应用程序的运行时长；

如果任意一个应用程序的行为特征中任意一项的测试值与此应用程序在所述行为特征库中此项行为特征的样本值的差别不在预设范围内，则确定探测到可疑进程。

4、根据权利要求 3 所述的方法，其特征在于，在行为特征中的任意一项为多维数据的情况下，此行为特征的测试值与所述行为特征库中的此项行为特征的样本值的差别的确定方法包括：

计算该项的测试值与所述行为特征库中此项行为特征的样本值之间的距离值。

5、根据权利要求 4 所述的方法，其特征在于，所述应用行为库的建立过程包括：

从预先获取的所述待探测主机的网络事件表及进程事件表中，获取每一个应用程序的行为特征；所述行为特征包括应用程序的级别、应用程序访问所述预设类型的数据的数据源的频率、应用程序对外连接的频率、应用程序对外连接的目的地址、应用程序对外连接的端口、运行应用程序的用户、应用程序的进程命令参数、应用程序的运行频率及应用程序的运行时长。

6、根据权利要求 1 或 2 所述的方法，其特征在于，还包括：

按照预设的进程风险规则，从所述待探测主机的进程中确定可疑进程，所述进程风险规则包括：所述待探测主机发起对自身的网络连接及该连接的目标端口为远程登录端口。

7、一种可疑进程的探测装置，其特征在于，包括：

第一获取模块，用于获取待探测主机的数据流向特征的测试值，以及，数据流向库中所述待探测主机对应的数据流向特征的样本值；所述数据流向特征包括进程列表和网络出口特征中的至少一个、以及数据源特征，所述数据源特征用于指示流入所述待探测主机的预设类型数据的数据源，所述进程列表中包括按照时间顺序排列的、调用所述数据源流出的数据的进程，所述网络出口特征用于指示所述数据源流出的数据在被所述进程列表中的进程调用后、流出所述待探测主机的出口；

第一确定模块，用于如果在所述数据源特征的测试值与所述数据源特征的样本值相同的情况下，所述进程列表的测试值与所述进程列表的样本值不同，和/或，所述网络出口特征的测试值与所述网络出口特征的样本值不同，则确定探测到可疑进程。

8、根据权利要求 7 所述的装置，其特征在于，还包括：

数据流向库建立模块，用于按照以下方式，分别建立每一个数据源的数据流向特征：从预先获取的所述待探测主机的网络事件表中确定与一个所述数据源特征相关的网络事件；以所述网络事件的进程编号和时间戳为查找条件，通过关联所述网络事件表、所述待探测主机的进程事件表及所述待探测主机的文件读写事件表，依次获得调用所述数据源的数据的进程，以及，所述第二数据源的数据流出的第二网络出口。

9、根据权利要求 7 所述的装置，其特征在于，还包括：

第二获取模块，用于获取所述待探测主机中每一个应用程序的行为特征的测试值以及应用行为库中所述待探测主机的行为特征的样本值；所述行为特征至少包括以下一

项：应用程序的级别、应用程序访问所述预设类型的数据的数据源的频率、应用程序对外连接的频率、应用程序对外连接的目的地址、应用程序对外连接的端口、运行应用程序的用户、应用程序的进程命令参数、应用程序的运行频率及应用程序的运行时长；

第二确定模块，用于如果任意一个应用程序的行为特征中任意一项的测试值与此应用程序在所述行为特征库中此项行为特征的样本值的差别不在预设范围内，则确定探测到可疑进程。

10、根据权利要求 9 所述的装置，其特征在于，所述第二确定模块用于在行为特征中的任意一项为多维数据的情况下，确定此行为特征的测试值与所述行为特征库中的此项行为特征的样本值的差别的具体过程包括：

所述第二确定模块具体用于，计算该项的测试值与所述行为特征库中的此项行为特征的样本值的距离值。

11、根据权利要求 10 所述的装置，其特征在于，还包括：

应用行为库建立模块，用于从预先获取的所述待探测主机的网络事件表及进程事件表中获取每一个应用程序的行为特征；所述行为特征包括应用程序的级别、应用程序访问所述预设类型的数据的数据源的频率、应用程序对外连接的频率、应用程序对外连接的目的地址、应用程序对外连接的端口、运行应用程序的用户、应用程序的进程命令参数、应用程序的运行频率及应用程序的运行时长。

12、根据权利要求 7 或 8 所述的装置，其特征在于，还包括：

第三确定模块，用于按照预设的进程风险规则，从所述待探测主机的进程中确定可疑进程，所述进程风险规则包括：所述待探测主机发起对自身的网络连接及该连接的目标端口为远程登录端口。

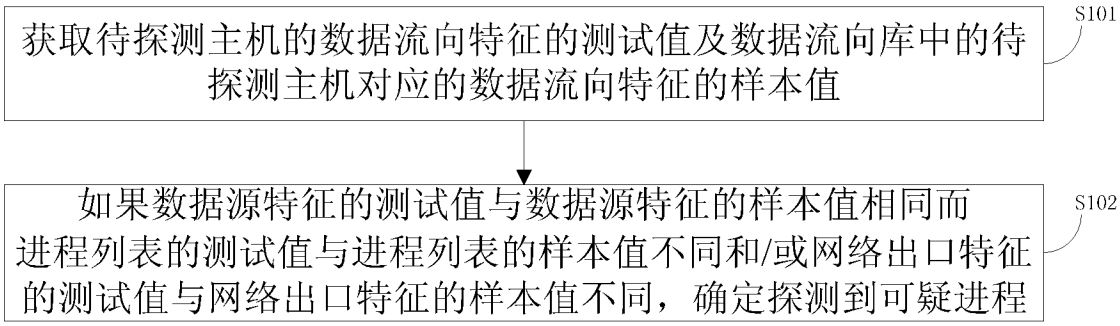
获取待探测主机的数据流向特征的测试值及数据流向库中的待探测主机对应的数据流向特征的样本值
S101
如果数据源特征的测试值与数据源特征的样本值相同而进程列表的测试值与进程列表的样本值不同和/或网络出口特征的测试值与网络出口特征的样本值不同，确定探测到可疑进程
S102

图 1

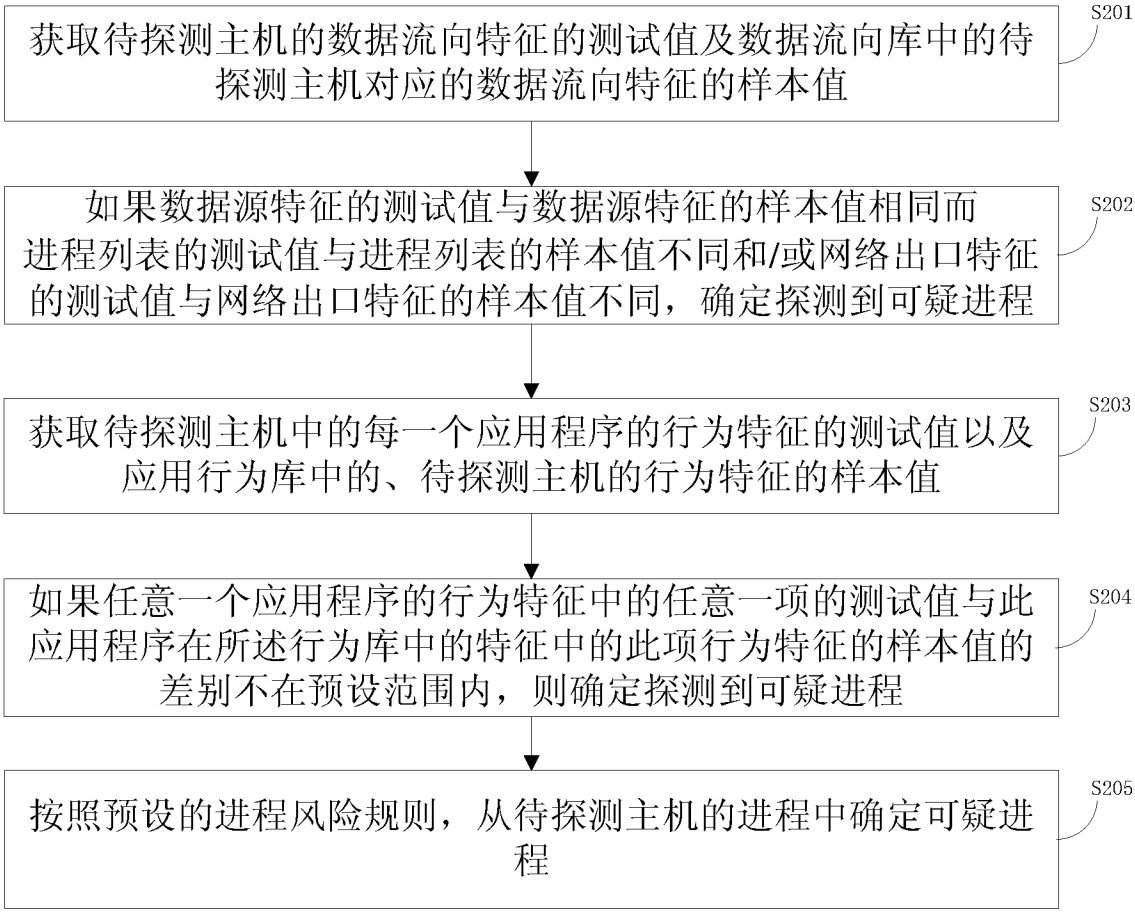
获取待探测主机的数据流向特征的测试值及数据流向库中的待探测主机对应的数据流向特征的样本值
S201
如果数据源特征的测试值与数据源特征的样本值相同而进程列表的测试值与进程列表的样本值不同和/或网络出口特征的测试值与网络出口特征的样本值不同，确定探测到可疑进程
S202
获取待探测主机中的每一个应用程序的行为特征的测试值以及应用行为库中的、待探测主机的行为特征的样本值
S203
如果任意一个应用程序的行为特征中的任意一项的测试值与此应用程序在所述行为库中的特征中的此项行为特征的样本值的差别不在预设范围内，则确定探测到可疑进程
S204
按照预设的进程风险规则，从待探测主机的进程中确定可疑进程
S205

图 2

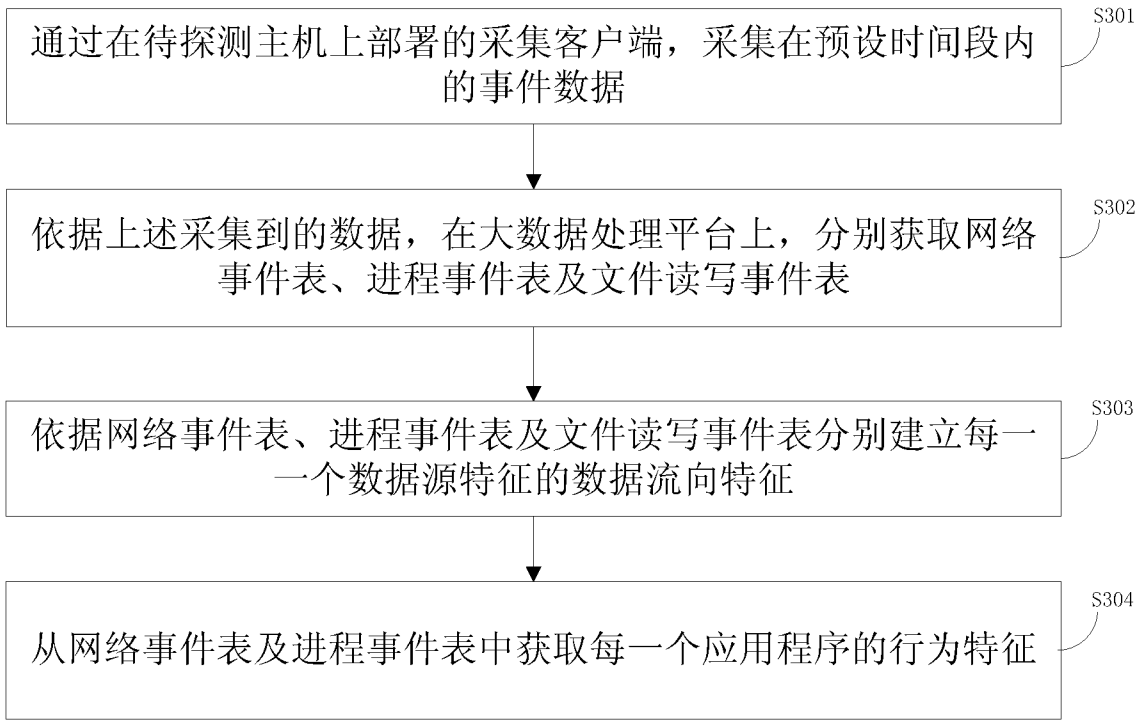
通过在待探测主机上部署的采集客户端，采集在预设时间段内的事件数据
S301
依据上述采集到的数据，在大数据处理平台上，分别获取网络事件表、进程事件表及文件读写事件表
S302
依据网络事件表、进程事件表及文件读写事件表分别建立每一一个数据源特征的数据流向特征
S303
从网络事件表及进程事件表中获取每一个应用程序的行为特征
S304

图 3

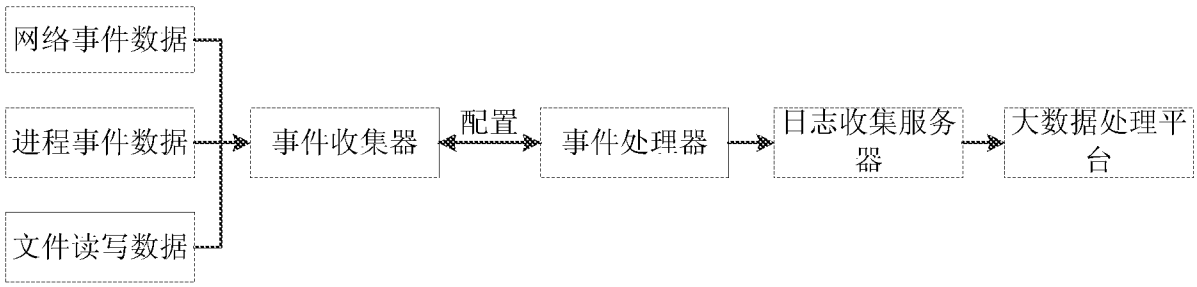
网络事件数据
进程事件数据
文件读写数据
事件收集器
配置
事件处理器
日志收集服务器
大数据处理平台

图 4

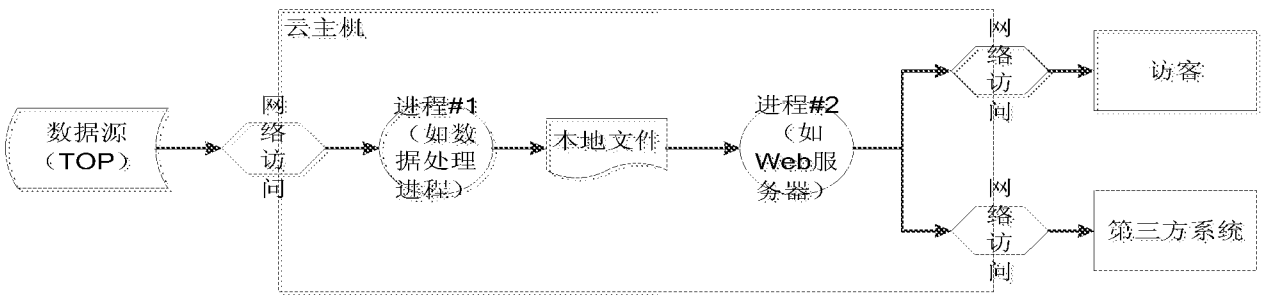
云主机
数据源（TOP）
网络访问
进程#1（如数据处理进程）
本地文件
进程#2（如Web服务器）
网络访问
访客
网络访问
第三方系统

图 5

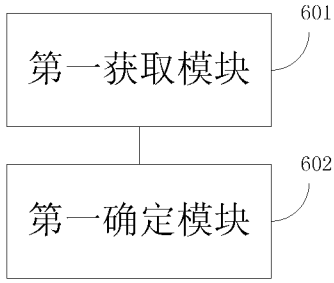
601
第一获取模块
602
第一确定模块

图 6

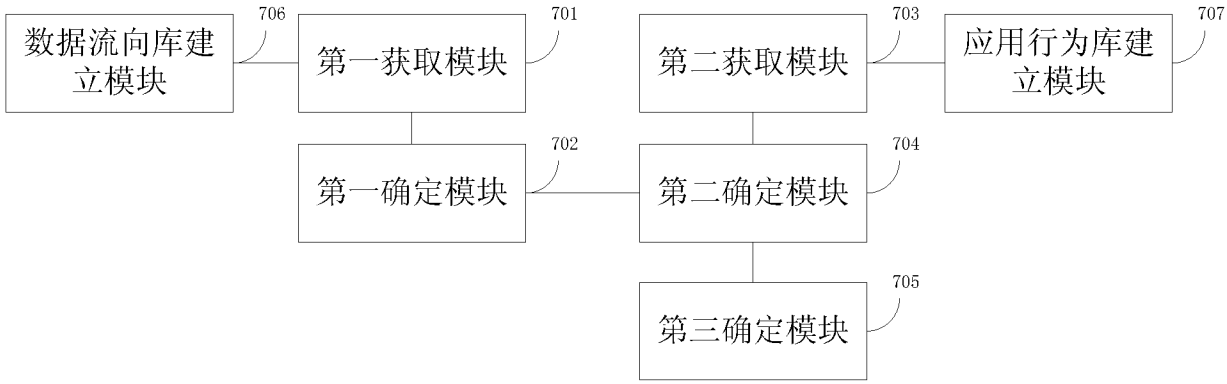
706
数据流向库建立模块
701
第一获取模块
703
第二获取模块
707
应用行为库建立模块
702
第一确定模块
704
第二确定模块
705
第三确定模块

图 7

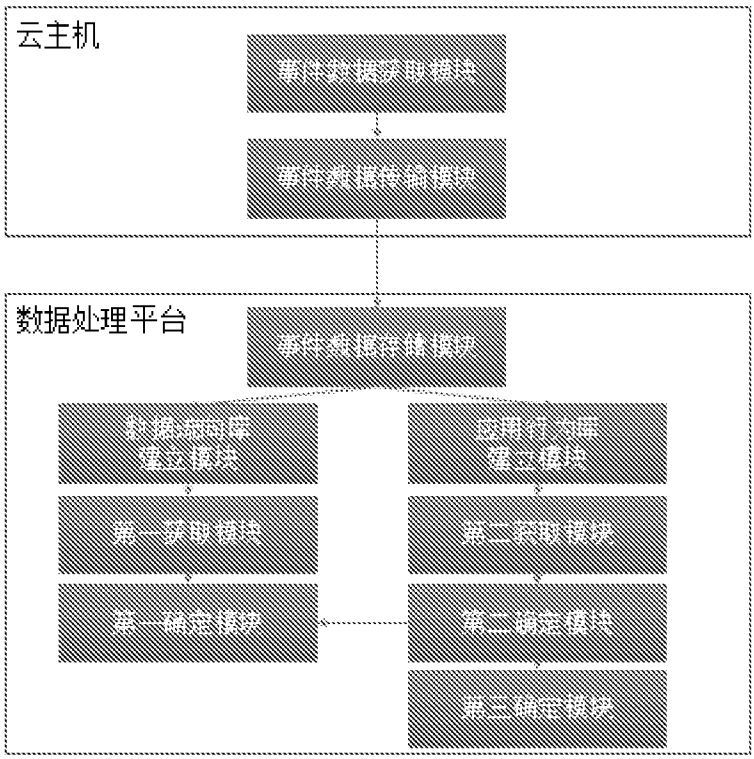
云主机
事件数据获取模块
事件数据传输模块
数据处理平台
事件数据存储模块
数据流向库建立模块
应用行为库建立模块
第一获取模块
第二获取模块
第一确定模块
第二确定模块
第三确定模块

图 8

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2016/076228

A. CLASSIFICATION OF SUBJECT MATTER

G06F 21/56 (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)

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)

CNPAT, WPI, EPODOC, CNKI, GOOGLE: sample, doubt, suspicious, process, detect, examine, swatch, exit, differ, contrast, data, source

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 103957205 A (STATE GRID CORPORATION OF CHINA et al.) 30 July 2014 (30.07.2014) description, paragraphs [0032]-[0053]	1-12
A	CN 102629308 A (QIZHI SOFTWARE (BEIJING) CO., LTD.) 08 August 2012 (08.08.2012) the whole document	1-12
A	CN 101895521 A (GRADUATE UNIVERSITY OF CHINESE ACADEMY OF SCIENCES (GUCAS)) 24 November 2010 (24.11.2010) the whole document	1-12
A	US 2007272744 A1 (HONEYWELL INTERNATIONAL INC.) 29 November 2007 (29.11.2007) the whole document	1-12

☐ 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

"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	Date of mailing of the international search report
26 May 2016	27 June 2016
Name and mailing address of the ISA 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 ZHAO, Tianqi Telephone No. (86-10) 61648112

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2016/076228

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN 103957205 A	30 July 2014	None	
CN 102629308 A	08 August 2012	None	
CN 101895521 A	24 November 2010	None	
US 2007272744 A1	29 November 2007	None	

国际检索报告

国际申请号 PCT/CN2016/076228

A. 主题的分类

G06F 21/56(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)

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）)

CNPAT, WPI, EPODOC, CNKI, GOOGLE: 可疑, 进程, 探测, 检测, 检查, 样本, 出口, 不同, 对比, 数据源, doubt, suspicious, process, detect, examine, swatch, exit, differ, contrast, data, source

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 103957205 A (国家电网公司等) 2014年 7月 30日 (2014-07-30) 说明书第32-53段	1-12
A	CN 102629308 A (奇智软件北京有限公司) 2012年 8月 8日 (2012-08-08) 全文	1-12
A	CN 101895521 A (中国科学院研究生院) 2010年 11月 24日 (2010-11-24) 全文	1-12
A	US 2007272744 A1 (HONEYWELL INTERNATIONAL INC.) 2007年 11月 29日 (2007-11-29) 全文	1-12

☐ 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	国际检索报告邮寄日期
2016年 5月 26日	2016年 6月 27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 赵天奇 电话号码 (86-10)61648112

表 PCT/ISA/210 (第2页) (2009年7月)

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2016/076228

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	103957205	A	2014年 7月 30日	无	
CN	102629308	A	2012年 8月 8日	无	
CN	101895521	A	2010年 11月 24日	无	
US	2007272744	A1	2007年 11月 29日	无	