

(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 11 月 16 日 (16.11.2017)



(10) 国际公布号
WO 2017/193271 A1

(51) 国际专利分类号:

H04L 29/06 (2006.01)

(21) 国际申请号:

PCT/CN2016/081503

(22) 国际申请日:

2016 年 5 月 10 日 (10.05.2016)

(25) 申请语言:

中文

(26) 公布语言:

中文

(71) 申请人: 华为技术有限公司 (HUAWEI TECHNOLOGIES CO., LTD.) [CN/CN]; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。

(72) 发明人: 付天福 (FU, Tianfu); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。周冲 (ZHOU, Chong); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。刘子宜 (LIU, Ziyi); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。

(74) 代理人: 北京三高永信知识产权代理有限公司 (BEIJING SAN GAO YONG XIN INTELLECTUAL PROPERTY AGENCY CO., LTD.); 中国北京市

海淀区学院路蓟门里和景园 A 座 1 单元 102 室, Beijing 100088 (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)。

(54) Title: METHOD AND DEVICE FOR DETECTING NETWORK ATTACK

(54) 发明名称: 检测网络攻击的方法及设备

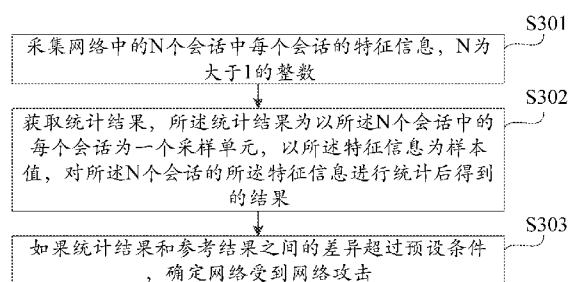


图 3

S301 COLLECTING FEATURE INFORMATION ABOUT EACH SESSION IN N SESSIONS IN A NETWORK, WHEREIN N IS AN INTEGER GREATER THAN 1
S302 ACQUIRING A STATISTICAL RESULT, WHEREIN THE STATISTICAL RESULT IS A RESULT OBTAINED AFTER COLLECTING STATISTICS OF THE FEATURE INFORMATION ABOUT THE N SESSIONS BY TAKING EACH SESSION IN THE N SESSIONS AS A SAMPLING UNIT, AND TAKING THE FEATURE INFORMATION AS A SAMPLE VALUE
S303 IF A DIFFERENCE BETWEEN THE STATISTICAL RESULT AND A REFERENCE RESULT EXCEEDS A PRE-SET CONDITION, DETERMINING THAT THE NETWORK IS UNDER A NETWORK ATTACK

(57) Abstract: Provided are a method and device for detecting a network attack, relating to the technical field of network security. The method comprises: collecting feature information about each session in N sessions in a network, wherein N is an integer greater than 1; acquiring a statistical result, wherein the statistical result is a result obtained after collecting statistics of the feature information about the N sessions by taking each session in the N sessions as a sampling unit, and taking the feature information as a sample value; and if a difference between the statistical result and a reference result exceeds a pre-set condition, determining that the network is under a network attack. The present application solves the technical problem that since it is difficult to collect enough network attack messages accounting for a relatively low proportion in a data flow, using a random message sampling technique makes it difficult to detect such a type of network attacks. In the embodiments of the present application, since a session is taken as a sampling unit rather than taking a message as a sampling unit, a session-type network attack can be detected effectively.

根据细则4.17的声明：

- 关于申请人有权申请并被授予专利(细则4.17(ii))

本国际公布：

- 包括国际检索报告(条约第21条(3))。

(57) 摘要：本申请实施例提供了一种检测网络攻击的方法及设备，涉及网络安全技术领域，所述方法包括：采集网络中的N个会话中每个会话的特征信息，N为大于1的整数；获取统计结果，该统计结果为以N个会话中的每个会话为一个采样单元，以特征信息为样本值，对N个会话的特征信息进行统计后得到的结果；如果统计结果和参考结果之间的差异超过预设条件，确定网络受到网络攻击。本申请解决了采用报文随机抽样技术，因难以采集到足够的在数据流量中所占比例较小的网络攻击报文，导致难以检测到此类网络攻击的技术问题。在本申请实施例中，由于以会话为抽样单位，而非以报文为抽样单位，故可有效检测出会话类网络攻击。

检测网络攻击的方法及设备

技术领域

5 本申请实施例涉及网络安全技术领域，特别涉及一种检测网络攻击的方法及设备。

背景技术

10 网络攻击是网络中重要安全问题。例如：拒绝服务(英文：denial of service；缩写：DoS) 攻击是一种常见的网络攻击。DoS 攻击通过向网络攻击针对的目标设备发送攻击报文，占用目标设备或者网络的资源，从而导致目标设备或网络无法正常的运行。

15 通常情况下，网络中的设备会采用报文随机抽样技术实现对网络攻击的检测。报文随机抽样检测网络攻击的方法是，采集目标时间段内，网络设备接收到的数据流量中一定比例的报文，获取每个报文的特征信息，然后对这些报文的特征进行统计分析，从而判断该网络设备所在的网络在该目标时段内是否受到网络攻击。

20 然而，存在一些网络攻击，这些网络攻击的攻击报文在网络整体的数据流量中所占比例较小，却对目标设备的资源消耗很大。通过报文随机抽样难以采集到足够的此类网络攻击报文，因此通过报文随机抽样的特征进行统计分析，难以检测到此类网络攻击。

发明内容

25 本申请实施例提供了一种检测网络攻击的方法及设备，用于检测一些报文随机抽样技术难以检测到的网络攻击。

30 第一方面，本发明实施例提供了一种检测网络攻击的方法。该方法包括：采集网络中的 N 个会话中每个会话的特征信息，所述 N 为大于 1 的整数；获取统计结果，所述统计结果为以所述 N 个会话中的每个会话为一个采样单元，以所述特征信息为样本值，对所述 N 个会话的所述特征信息进行统计后得到的结果；如果所述统计结果和参考结果之间的差异超过预设条件，确定所述网络

受到网络攻击。

通过会话随机抽样，并以会话为单位对会话的特征信息进行统计分析，可以有效检测出一些传统的报文随机抽样难以检测出的会话类网络攻击。会话类网络攻击存在如下特点：发起网络攻击的设备通过发起大量会话，占用目标设备的资源。由于发起一个会话只需要少量的报文，但是对设备的资源占用很多，因此在网络攻击报文的数量不大的情况下，就可以对目标设备造成较大的损害。若以报文为抽样单位，由于符合上述会话类网络攻击特性的网络攻击报文在数据流量中占比很小，导致此类网络攻击报文很难被采集。而本申请中以会话为抽样单位，无论是只有一个会话建立请求的网络攻击报文，还是包括数据流量较大的正常会话，在抽样统计中都作为一个会话来进行统计，因此有较大概率识别出所述会话类网络攻击。

在第一方面的第一种可能的实施方式中，所述特征信息是表示大小的数值，所述统计结果为基尼系数。所述获取统计结果，包括：按所述特征信息从小到大的顺序，对所述 N 个会话排序；从所述 N 个会话中选择 M 个会话，在坐标系中获取 M 个坐标，所述 M 个坐标与所述 M 个会话一一对应，所述 M 个会话中的第 a 个坐标对应于所述排序后的 N 个会话中的第 b 个会话，所述第 a 个坐标为 $(b/N, w/W)$ ，其中，所述 w 为所述 N 个会话中特征信息最小的 b 个会话的所述特征信息之和，所述 W 为所述 N 个会话的所述特征信息之和，所述 M 为大于 1 且小于等于所述 N 的整数，所述 a 为大于等于 1 且小于等于所述 M 的整数，所述 b 为大于等于 1 且小于等于所述 N 的整数；根据所述 M 个坐标，在所述坐标系中获取洛伦兹曲线；根据所述洛伦兹曲线计算所述基尼系数。

在第一方面的第二种可能的实施方式中，所述统计结果为信息熵。所述获取统计结果，包括：根据所述 N 个会话中每个会话的特征信息，计算 n 类会话中的每一类会话在所述 N 个会话中的出现概率，所述 N 个会话中的每一个会话属于所述 n 类会话中的一类会话，所述 n 为大于等于 1 且小于等于所述 N 的整数；根据所述 n 类会话中每一类会话的所述出现概率计算所述信息熵。

在第一方面的第三种可能的实施方式中，所述统计结果为卷积，所述 N 个会话中每个会话的特征信息包括第一类特征信息和第二类特征信息，所述第一类特征信息是表示大小的数值，所述第二类特征信息是表示大小的数值。所述获取统计结果，包括：根据所述 N 个会话中每个会话的所述第一类特征信息，构建第一离散序列；根据所述 N 个会话中每个会话的所述第二类特征信息，构

建第二离散序列；计算所述第一离散序列和所述第二离散序列的所述卷积。

在第一方面的第四种可能的实施方式中，所述统计结果为向量和，所述N个会话中每个会话的特征信息包括p类特征信息，所述p为大于1的整数。所述获取统计结果，包括：构建N个向量，所述N个向量与所述N个会话一一对应，所述N个向量中的每个向量具有p个维度，所述p个维度与所述p类特征信息一一对应；计算所述N个向量的向量和。

在第一方面的第五种可能的实施方式中，所述统计结果为向量积，所述N个会话中每个会话的特征信息包括p类特征信息，所述p为大于1的整数。所述获取统计结果，包括：构建N个向量，所述N个向量与所述N个会话一一对应，所述N个向量中的每个向量具有p个维度，所述p个维度与所述p类特征信息一一对应；计算所述N个向量的向量积。

本申请实施例提供了多种统计方法，使得方案的灵活性和适用性更佳。

结合第一方面或者第一方面的任意一种可能的实施方式，在第一方面的第六种可能的实施方式中，所述特征信息包括：五元组信息、报文数信息、字节数信息、时间信息、生存周期（英文：Time To Live；缩写：TTL）信息、报文长度信息、地理信息或报文平均长度中的一种或多种。

通过上述方式，采集多种类型的特征信息，有助于从多个不同维度检测网络攻击，有助于提高网络攻击的检测精度。

结合第一方面或者第一方面的任意一种可能的实施方式，在第一方面的第七种可能的实施方式中，所述确定网络受到网络攻击之后，还包括：根据所述统计结果，确定所述网络攻击的来源；或者，根据所述统计结果，确定所述网络攻击的目标。

通过上述方式，进一步确定网络攻击的来源或者目标，对网络攻击的发起方进行准确定位，有助于对其进行及时有效的遏制；另外，对网络攻击的遭受方进行准确定位，有助于其针对网络攻击采取及时有效的应对策略。

第二方面，本发明实施例提供了一种检测网络攻击的设备。该设备包括至少一个单元，该至少一个单元用于实现上述第一方面或者第一方面的任意一种可能的实施方式所提供的方法。

第三方面，本发明实施例提供了一种检测网络攻击的设备。所述检测网络攻击的设备可以实施上述第一方面或第一方面的任一种可能的实施方式所提供的方法。在一种可能的设计中，该设备包括处理器以及网络接口，处理器

用于执行以下操作：通过网络接口采集网络中的N个会话中每个会话的特征信息，N为大于1的整数；获取统计结果，该统计结果为以N个会话中的每个会话为一个采样单元，以特征信息为样本值，对N个会话的特征信息进行统计后得到的结果；如果统计结果和参考结果之间的差异超过预设条件，确定网络受到网络攻击。

附图说明

为了更清楚地说明本申请实施例中的技术方案，下面将对实施例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是本申请一个实施例提供的实施环境的示意图；

图2是本申请一个实施例提供的检测网络攻击的设备的框图；

图3是本申请一个实施例提供的检测网络攻击的方法的流程图；

图4A是本申请另一实施例提供的检测网络攻击的方法的流程图；

图4B是本申请另一实施例涉及的洛伦兹曲线的示意图；

图5是本申请另一实施例提供的检测网络攻击的方法的流程图；

图6是本申请另一实施例提供的检测网络攻击的方法的流程图；

图7是本申请另一实施例提供的检测网络攻击的方法的流程图；

图8是本申请另一实施例提供的检测网络攻击的方法的流程图；

图9是本申请另一实施例提供的检测网络攻击的设备的框图。

具体实施方式

为使本申请的目的、技术方案和优点更加清楚，下面将结合附图对本申请实施方式作进一步地详细描述。

在本文中提及的“模块”是可以由硬件实现，也可以由软硬件的结合实现，还可以由存储在存储器中的能够实现某些功能的程序或指令实现；在本文中提及的“单元”是指按照逻辑划分的功能性结构，该“单元”可以由纯硬件实现，或者由软硬件的结合实现。

在本文中提及的“若干个”是指一个或多个，“多个”是指两个或两个以上。“和/或”，描述关联对象的关联关系，表示可以存在三种关系，例如，A和/或B，

可以表示：单独存在 A，同时存在 A 和 B，单独存在 B 这三种情况。字符“/”一般表示前后关联对象是一种“或”的关系。

在对本申请实施例进行介绍之前，首先对本申请实施例中涉及的会话进行说明。

5 本申请中，会话（英文：session）指的是在一个不中断的特定操作时间内，两个设备之间的通信交互。在一个会话期间，两个设备之间相互传输的所有报文都属于该会话。

 在“会话”的一种示例中，在第一设备和第二设备之间通信的报文是传输控制协议（英文：Transmission Control Protocol；缩写：TCP）报文的情况下，
10 或者在第一设备和第二设备之间通信的报文是用户数据报协议（英文：User Datagram Protocol；缩写：UDP）报文的情况下，该 TCP 报文或 UDP 报文中携带五元组信息。同一个会话的多个报文的五元组信息相匹配。即，第一设备给第二设备发送的报文所携带的五元组信息中，源 IP 地址是第一设备的 IP 地址，源端口号是第一设备的端口号，目的 IP 地址是第二设备的 IP 地址，目的
15 端口号是第二设备的端口号；第二设备给第一设备发送的报文所携带的五元组信息中，源 IP 地址是第二设备的 IP 地址，源端口号是第二设备的端口号，目的 IP 地址是第一设备的 IP 地址，目的端口号是第一设备的端口号；两个设备之间相互发送的报文所采用的传输层协议号都相同。这些报文都属于同一个 TCP/UDP 会话。

20 在“会话”的另一种示例中，在第一设备和第二设备之间通信的报文不是 TCP 报文或 UDP 报文的情况下，例如在第一设备和第二设备之间通信的报文是因特网控制消息协议（英文：Internet Control Message Protocol；缩写：ICMP）报文的情况下，也可以认为二元组信息相匹配的多个报文属于同一个会话的报文。即，第一设备给第二设备发送的报文所携带的二元组信息中，源 IP 地址
25 是第一设备的 IP 地址，目的 IP 地址是第二设备的 IP 地址；第二设备给第一设备发送的报文所携带的二元组信息中，源 IP 地址是第二设备的 IP 地址，目的 IP 地址是第一设备的 IP 地址；两个设备之间相互发送的报文所采用的传输层协议号都相同。这些报文都属于同一个 ICMP 会话。

 本申请中，对于会话的两个设备之间是否需要建立连接不作限定。一个会
30 话可以是有明确的连接建立过程和连接断开过程。例如，TCP 会话（英文：TCP session）过程中，两个设备的两个端口之间采用三次握手（英文：three-way

handshake) 建立 TCP 连接, 在会话完成时再通过四次挥手 (英文: four-way handshake) 断开 TCP 连接。一个会话也可以没有连接建立过程, 例如两个设备的两个端口之间在一个连续的时间段内使用 UDP 进行通信。

对于有明确的连接建立过程和连接断开过程的会话, 可将连接建立过程中的第一个报文作为会话的开始, 将连接断开过程中的最后一个报文作为会话的终止。例如 TCP 会话的三次握手中第一次握手的报文作为会话的开始, TCP 会话的四次挥手中第四次挥手的报文作为会话的终止。对于没有连接建立过程的会话, 在两个设备间传输的某个报文之前的第一预设时长内, 没有这两个设备间传输的报文, 则认为该报文是该会话的第一个报文; 在两个设备间传输某个报文之后, 如果在第二预设时长内有下一个报文传输, 则该下一个报文与之前传输的报文属于一个会话; 如果在所述第二预设时长内未有下一个报文传输, 则认为会话终止。

当然, 本领域技术人员可以理解的是, 如果报文是网络攻击报文, 那么即使在报文所使用的协议是有明确的连接建立过程和断开过程的协议, 例如 TCP 的情况下, 一个会话依然可能没有完整的建立过程和终止过程, 例如发起网络攻击的设备只进行三次握手之后不再发送任何报文, 或者发起网络攻击的设备在发起第一握手之后不进行后续的应答, 或者发起网络攻击的设备发送标识位与协议定义不符的畸形报文。这种情况下, 对会话开始和终止的判断, 可以采用与上述没有连接建立过程的会话的开始或终止相同的判断方法, 或者针对常见的网络攻击报文的类型对判断会话的开始或终止的条件进行设定。

在本申请实施例中, 采用会话随机抽样技术, 按照预设的抽样规则, 采集网络中的若干个会话中的报文, 通过从采集的报文中提取每个会话的特征信息, 并以会话为单位对这些会话的特征信息进行统计分析, 检测网络攻击。需要说明的是, 对于每一个会话, 在提取该会话的特征信息时, 所需采集的报文的数量依据所要提取的会话的特征信息而定。例如, 当所要提取的会话的特征信息为五元组信息时, 仅需采集每个会话的一个报文即可。再例如, 当所要提取的会话的特征信息为报文数信息时, 需要采集会话的所有报文。

下面, 通过几个实施例对本申请提供的技术方案进行详细介绍和说明。

请参考图 1, 其示出了本申请一个实施例提供的实施环境的示意图, 该实施环境可以是基于 TCP/IP 进行通信的网络系统。示例性地, 如图 1 所示, 该

实施环境包括：包括有若干台网络设备 110 的网络 10，以及检测网络攻击的设备（以下简称“检测设备”）120。

举例来说，网络 10 可以是运营商网络，也可以是局域网。网络设备 110 可以是交换机，路由器，波分复用设备，分组传送网设备，光传送网设备，防火墙，或数据中心等。在本实施例中，对网络 10 中的若干台网络设备 110 的组网结构并不限定。例如，该组网结构可以是集中式组网结构、分布式组网结构，等等。

网络设备 110 在转发报文的过程中，按照预设的抽样规则进行会话抽样。即，网络设备 110 抽取若干个会话中的报文，获取该若干个会话中每个会话的特征信息。

举例来说，网络设备 110 抽取若干个会话中的报文，可以是在预设的时间段内，网络设备 110 每接收到一个报文，就镜像该报文，保存该镜像的报文用于获取该报文所属的会话的特征信息，并将原报文按照传输路径转发。本申请中，在没有特殊说明的情况下，在从报文中采集会话的特征信息均指从被镜像并保存的报文中采集会话的特征信息。

举例来说，网络设备 110 判断报文所属的会话，可以采用如下的方法进行。网络设备 110 获取报文中的五元组信息作为第一五元组信息。网络设备 110 判断会话信息表中是否存储了所述第一五元组信息，如果存储了所述第一五元组信息，则说明该报文与网络设备 110 此前接收的报文属于同一个会话，则网络设备 110 标识该报文所属的会话；如果所述会话信息表中没有存储所述第一五元组信息，则将该第一五元组信息中的源 IP 地址与目的 IP 地址互换，将该第一五元组信息中的源端口地址与目的端口地址互换，获得第二五元组信息，判断所述会话信息表中是否存储了所述第二五元组信息，如果所述会话信息表中存储了所述第二五元组信息，则同样说明该报文与网络设备 110 此前接收的报文属于同一个会话，网络设备 110 标识该报文所属的会话；如果所述会话信息表中既不存在所述第一五元组信息也不存在所述第二五元组信息，则说明网络设备 110 此前没有接收过属于该会话的报文，网络设备 110 在所述会话信息表中增加所述第一五元组信息或所述第二五元组信息。当然，所述第一五元组信息和所述第二五元组信息，也可以分别是所述第一五元组信息和所述第二五元组信息通过同一个哈希函数计算得到的第一哈希值和第二哈希值。通过上述步骤，网络设备 110 采集到了预设时间段内接收的报文，并且标识出了每个报文

所属的会话。

网络设备 110 上的会话的所述抽样规则可以是网络管理员在该网络设备 110 上配置的，也可以是检测设备 120 发送的。举例来说，所述抽样规则可以是预设时间段内抽取预设比例的会话。抽取所述会话后，网络设备 110 提取被
5 抽取的会话中每个会话的特征信息，并发送给检测设备 120。

检测设备 120 与网络 10 中的若干台网络设备 110 建立通信连接。具体来说，检测设备 120 与网络 10 中用于进行会话抽样的网络设备 110 建立通信连接。用于进行会话抽样的网络设备 110，可以是网络 10 中的一个网络设备 110，也可以是多个网络设备 110。检测设备 120 接收网络设备 110 发送的会话特征
10 信息。检测设备 120 用于对会话的特征信息进行统计分析，检测网络 10 是否受到网络攻击。

在一种实现方式中，检测设备 120 可以是一台独立的服务器。当然，在其它可能的实现方式中，检测设备 120 也可由网络 10 中的某一台网络设备 110 的软件或硬件模块实现。该检测设备 200 也可以是多个网络设备组成的集群。
15

请参考图 2，其示出了本申请一个实施例提供的检测网络攻击的设备（以下简称“检测设备”）200 的框图。该检测设备 200 可以包括：处理器 210、存储器 220、网络接口 230 以及总线 240。存储器 220 和网络接口 230 通过总线 240 与处理器 210 相连。检测设备 200 用于检测网络是否受到网络攻击。结合
20 参考图 1 所示实施环境，该检测设备 200 可以是图 1 中所示的检测设备 120。该检测设备 200 可以是一台独立的服务器，也可以是网络中的一台网络设备上的软件或硬件模块，该检测设备 200 也可以是多个网络设备组成的集群。

处理器 210 包括一个或者一个以上处理核心。处理器 210 通过运行软件程序以及模块，从而执行各种功能应用以及数据处理。处理器 210 包括运算逻辑
25 部件、寄存器部件以及控制部件等，处理器 210 可以是独立的中央处理器，或者也可以是嵌入式处理器，比如微处理器（英文：Micro Processor Unit；缩写：MPU）、微控制器（英文：Microcontroller Unit；缩写：MCU）或者数字信号处理器（英文：Embedded Digital Signal Processor；缩写：EDSP）等。

存储器 220 可以由任何类型的易失性或非易失性存储设备或者它们的组合
30 实现，如静态随机存取存储器（英文：Static Random Access Memory；缩写：SRAM），电可擦除可编程只读存储器（英文：Electrically Erasable Programmable

Read-Only Memory; 缩写:EEPROM),可擦除可编程只读存储器(英文:Erasable Programmable Read Only Memory; 缩写:EPROM),可编程只读存储器(英文:Programmable Read-Only Memory; 缩写:PROM),只读存储器(英文:Read Only Memory; 缩写:ROM),磁存储器,快闪存储器,磁盘或光盘。存储器 220 5 可用于存储软件程序以及模块等可执行的指令。

网络接口 230 可以有有线接口,例如光纤分布式数据接口(英文:Fiber Distributed Data Interface; 简称:FDDI)、以太网(英文:Ethernet)接口。网络接口 230 也可以是无无线接口,例如无线局域网接口。

处理器 210 被配置为执行存储器 220 中存储的指令。处理器 210 通过执行 10 该指令来实现下述方法:采集网络中的 N 个会话中每个会话的特征信息, N 为大于 1 的整数;获取统计结果,该统计结果为以 N 个会话中的每个会话为一个采样单元,以特征信息为样本值,对 N 个会话的特征信息进行统计后得到的结果;如果统计结果和参考结果之间的差异超过预设条件,确定网络受到网络攻击。

15 网络接口 230 用于对外通信,网络接口 230 可以包括多种类型接口。例如,网络接口 230 用于向网络中的网络设备发送对会话进行随机抽样的指令,从网络设备接收会话的特征信息,等等。

可选地,存储器 220 可存储操作系统 222 以及至少一个功能所需的应用程序模块 224。操作系统 222 可以是实时操作系统(英文:Real Time eXecutive; 20 缩写:RTX)、LINUX、UNIX、WINDOWS 或 OS X 之类的操作系统。应用程序模块 224 可以包括:采集模块 224a、获取模块 224b 和确定模块 224c。其中,采集模块 224a,用于采集网络中的 N 个会话中每个会话的特征信息, N 为大于 1 的整数;获取模块 224b,用于获取统计结果,该统计结果为以 N 个会话中的每个会话为一个采样单元,以特征信息为样本值,对 N 个会话的特征信息 25 进行统计后得到的结果;确定模块 224c,用于如果统计结果和参考结果之间的差异超过预设条件,确定网络受到网络攻击。

可选地,检测设备 200 还可以包括输入/输出组件(图中未示出)。输入/输出组件包括有用于显示信息的显示器和用于用户输入信息的诸如鼠标、键盘之类的输入设备。其中,显示器和输入设备与处理器 210 通信。

30 请参考图 3,其示出了本申请一个实施例提供的一种检测网络攻击的方法

的流程图。例如，该方法可应用于图 1 所示实施环境中，图 3 的步骤执行主体可以是检测设备 120 或者图 2 所示的检测设备 200。该方法可包括如下步骤。

S301, 采集网络中的 N 个会话中每个会话的特征信息, N 为大于 1 的整数。

5 具体来说，某个网络中的会话是指属于该会话的报文在该网络的设备中传输。参与所述会话的两个设备可以都是所述网络中的设备；也可以只有其中一个设备是所述网络中的设备；还可以参与所述会话的两个设备都不是所述网络中的设备，所述会话的报文在该网络的转发设备中被转发。

举例来说，会话的特征信息可以包括从会话所包括的报文中直接采集获取的特征信息，也可以包括对所述直接采集获取的特征信息进行加工处理后得到的特征信息。对于加工处理得到的特征信息，所述加工处理步骤可以是进行会
10 话抽样的网络设备实施的，也可以是检测设备实施的。

其中，网络设备可以是图 1 所示实施环境中的网络中的网络设备 110。例如，网络设备 110 接收到检测设备 120 下发的对会话进行随机抽样的指令之后，按照预设的抽样规则，采集网络中的若干个会话的报文，提取会话的特征信息，
15 并发送给检测设备 120。

举例来说，会话的特征信息可以包括一种或多种用于描述会话的特征的信息。

举例来说，会话的特征信息包括但不限于：五元组信息、会话的报文数信息、字节数信息、时间信息、TTL 信息、报文长度信息、地理信息和报文平均
20 长度中的一种或多种。五元组信息包括：目的 IP 地址、源 IP 地址、目的端口号、源端口号和协议号。报文数信息包括但不限于：上行报文数、下行报文数、上行分片（英文：fragment）报文数和下行分片报文数中的一种或多种。字节数信息包括但不限于：上行字节数和下行字节数中的一种或多种。时间信息包括但不限于：上行报文间隔时间、下行报文间隔时间和会话持续时间中的一种
25 或多种。TTL 信息包括但不限于：上行最大 TTL、下行最大 TTL、上行最小 TTL 和下行最小 TTL 中的一种或多种。报文长度信息包括但不限于：上行最大报文长度、下行最大报文长度、上行最小报文长度和下行最小报文长度中的一种或多种。地理信息包括源地理信息和目的地理信息中的一种或多种。地理信息可由 IP 地址转换得到，源 IP 地址的归属地即为源地理信息，目的 IP 地址
30 的归属地即为目的地理信息。所述地理信息也可以根据 IP 地址从地理信息系统（英文：Geographic Information System；简称：GIS）查询获取。报文平均

长度可以包括上行报文平均长度和下行报文平均长度中的一种或多种。报文平均长度可根据采样得到的总字节数除以采样得到的总报文数得到。

举例来说，在所述会话为 TCP 会话的示例中，会话的特征信息包括但不限于：五元组信息、会话的报文数信息、字节数信息、时间信息、TTL 信息、
5 报文长度信息、地理信息、报文平均长度、TCP 会话的健康状态或乱序报文的统计结果中的一种或多种。其中，报文数信息可以包括上行报文数、下行报文数、上行分片报文数、下行分片报文数、TCP 会话的同步（英文：synchronize；简称：syn）报文数、TCP 会话的结束（英文：finish；简称：fin）报文数、TCP 会话的确认（英文：acknowledge；简称：ack）报文数、TCP 会话的复位（英文：reset；简称：rst）报文数、TCP 会话的紧急（英文：urgent；简称：urg）
10 报文数和 TCP 会话的推送（英文：push；简称：psh）报文数中的一种或多种。时间信息可以包括上行报文间隔时间、下行报文间隔时间、会话持续时间、TCP 会话的服务器响应时间和 TCP 会话的客户端响应时间中的一种或多种。TCP 会话的健康状态包括与 TCP 会话的服务质量（英文：Quality of Service；缩写：QoS）性能相关的信息，如 TCP 连接是否成功建立、丢包数、延迟，等等中的一种或多种。乱序报文的统计结果是指对一个会话中在后发送的报文比在先发送的报文更早到达目标设备的情况的统计结果。

需要说明的是，对于进行某个会话的一台设备而言，上行报文是指发送至该设备的报文，下行报文是指从该设备发出的报文。对于获取会话的特征信息的网络设备 110 而言，上行和下行可以是网络设备 110 定义的。例如，第一设备
20 和第二设备进行会话，上行报文可以是由所述第一设备发送给所述第二设备的，下行报文可以是由所述第二设备发送给所述第一设备的；当然，上行报文也可以是由所述第二设备发送给所述第一设备的，下行报文可以是由所述第一设备发送给所述第二设备的。

需要说明的是，由于会话中的报文是双向的，例如包括上行报文和下行报文，因此本申请中会话的特征信息中，五元组信息、源地址、目的地址、源地理信息或目的地理信息，均指会话中双向报文中某一个方向传输的报文的五元
25 组信息、源地址、目的地址、源地理信息或目的地理信息。

S302，获取统计结果，所述统计结果为以所述 N 个会话中的每个会话为一个采样单元（英文：sampling unit），以所述特征信息为样本（英文：sample）
30 值，对所述 N 个会话的所述特征信息进行统计后得到的结果。

具体来说,所述统计结果是根据特定的统计方法,对每个采样单元的样本值进行统计获得的结果。具体来说,以每个会话为一个采样单元是指,一个会话是统计的最小单位。以所述特征信息为样本值是指,每一个特征信息都是某个会话的特征信息,所述统计结果反映的是所述特征信息的相关特性。

5 例如,当特征信息为报文平均长度时,统计结果用于反映各个会话的报文平均长度的分布特征,如各个会话的报文平均长度是否接近。再例如,当特征信息为目的地理信息时,统计结果用于反映各个会话的目的地理信息的分布特征,如各个会话的目的地理信息所指向的地理位置的分布。举例来说,所述统计结果可以是基尼系数(英文: Gini coefficient)、信息熵、卷积、向量和或向量积中的任意一种。对于上述列举的各种统计结果,以及获取上述各种统计结果的统计方法,在下文将分别进行详细介绍和说明。

当然,本领域技术人员可以理解的是,所述统计结果可以是对每个会话的一类特征信息进行统计获得的结果,也可以是对每个会话的多类特征信息进行统计获得的结果。例如,在下文的统计结果是基尼系数、特征信息是会话的报文平均长度的示例中,所述基尼系数是基于会话的报文平均长度这一类特征信息进行统计获取的结果。又例如,在下文的统计结果是向量和、第一类特征信息是会话的上行报文数、第二类特征信息是会话的下行报文数的示例中,所述向量和是基于会话的上行报文数以及下行报文数这两类特征信息进行统计获取的结果。

20 S303,如果统计结果和参考结果之间的差异超过预设条件,确定网络受到网络攻击。

举例来说,如果统计结果和参考结果之间的差异超过预设条件,检测设备确定网络受到网络攻击。其中,统计结果和参考结果的类型相同。

25 在一种示例中,统计结果和参考结果均为一个表示大小的数值,则所述预设的条件可以是预设值。在另一种示例中,统计结果和参考结果不是表示大小的数值,则所述预设条件可以是其他反映两个结果差异的条件,例如统计结果和参考结果分别是一个多维向量,所述预设的条件是:统计结果的向量长度与参考结果的向量长度之差不超过第一预设值,并且统计结果的向量角度与参考结果的向量角度之差不超过第二预设值。

30 参考结果是在网络未受到网络攻击的情况下会话的特征信息的统计结果。可选地,通过如下方式获取参考结果:在多个不同的时间,采用 S301 和 S302

获取多个统计结果，如果所述多个统计结果彼此之间的差异均在一定的范围内，则将这些统计结果的平均值作为所述参考结果。

可选地，在确定网络受到网络攻击的情况下，检测设备还根据所述统计结果确定网络攻击的来源，或者根据所述统计结果确定网络攻击的目标。例如，
5 网络攻击的来源可包括：发起网络攻击的设备的 IP 地址、发起网络攻击的设备的端口号、发起网络攻击的设备所在的地理位置中的一个或多个；网络攻击的目标可包括：受到网络攻击的设备的 IP 地址、受到网络攻击的设备的端口号、受到网络攻击的设备的地理位置中的一个或多个。

举例来说，获取对所述统计结果与所述参考结果的差异起主要作用的会话，
10 分析参与这些会话的设备的信息，从而确定网络攻击的来源或目标。例如，在下文关于统计结果是基尼系数的示例中，假设会话的特征信息是报文平均长度，某类网络攻击的报文的长度均相同，例如均为第一长度，则该网络攻击的会话的报文平均长度也约等于所述第一长度。即此类网络攻击的会话的报文平均长度与
15 所述第一长度之差小于预设值。大量此类网络攻击的会话导致会话的报文平均长度比正常情况下分布更均衡，因此统计结果的基尼系数显著小于参考结果的基尼系数。检测设备获取报文平均长度与所述第一长度之差小于预设值的会话，分析参与这些会话的设备的地理信息。假设检测设备确定这些会话中，会话的发起方的 IP 地址在第一子网网段的概率较大，例如会话的发起方的 IP 地址在所述第一子网网段的会话数量占这些会话的总数的 80%，即可确定发起
20 网络攻击的设备是所述第一 IP 子网网段的设备。

综上所述，本实施例提供的方法，通过采集网络中的多个会话中每个会话的特征信息，获取统计结果，并在统计结果和参考结果之间的差异超过预设条件的情况下，确定网络受到网络攻击。通过会话随机抽样，并以会话为单位对会话的特征信息进行统计分析，可以有效检测出一些传统的报文随机抽样难以
25 检测出的会话类网络攻击。会话类网络攻击存在如下特点：发起网络攻击的设备通过发起大量会话，占用目标设备的资源。由于发起一个会话只需要少量的报文，但是对设备的资源占用很多，因此在网络攻击报文的数量不大的情况下，就可以对目标设备造成较大的损害。若以报文为抽样单位，由于符合上述会话类网络攻击特性的网络攻击报文在数据流量中占比很小，导致此类网络攻击报
30 文很难被采集；而本申请中以会话为抽样单位，无论是只有一个会话建立请求的网络攻击报文，还是包括数据流量较大的正常会话，在抽样统计中都作为一

个会话来进行统计，因此有较大概率识别出所述会话类网络攻击。

此外，由于会话随机抽样技术能够很好地还原会话的特征，如会话的 IP 地址分布特征、地域分布特征等，达到了有效检测 IP 分布异常和地域分布异常的技术效果。

5

参考如下表-1，其示例性地示出了几种统计方法，以及每一种统计方法适用的特征信息（打√表示适用）。

类别	属性	基尼系数	信息熵	卷积	向量和	向量积
基本信息 (即五元组信息)	目的 IP 地址 (DstIP)		√			
	源 IP 地址 (SrcIP)		√			
	目的端口号 (DstPort)		√			
	源端口号 (SrcPort)		√			
	协议号 (Protocol)					
地理信息	目的地理信息		√		√	
	源地理信息		√		√	
报文数信息	上行、下行报文数	√		√	√	√
	上行、下行分片报文数	√		√	√	√
	TCP 会话的 syn 报文数	√		√	√	√
	TCP 会话的 fin 报文数	√		√	√	√
	TCP 会话的 ack 报文数	√		√	√	√
	TCP 会话的 rst 报文数	√		√	√	√
	TCP 会话的 urg 报文数	√		√	√	√
	TCP 会话的 psh 报文数	√		√	√	√
字节数信息	上行、下行字节数	√		√	√	√
时间信息	会话持续时间	√		√	√	√
	上行、下行报文间隔时间	√		√		√
	TCP 会话的服务器、客户端响应时间	√		√	√	√

其它信息	TCP 会话的健康状态			√		√
	上行、下行 TCP 乱序统计			√	√	√
	上行、下行最大或最小 TTL	√	√	√		√
	上行、下行最大或最小报文长度	√	√	√		√

表-1

需要说明的是,在上述表-1 中,仅是示例性地列举了会话的几种特征信息,在实际应用中,会话的特征信息还可包括报文平均长度、源地理信息到目的地理信息的经纬度距离、源地理信息到目的地理信息的地表距离,等等。因此,每一种统计方法的适用范围包括但不限于上述表-1 所提供的特征信息。

下面,针对上文列举的几种统计方法,分别进行介绍和说明。

在基于图 3 所示实施例提供的一个可选实施例中,所述特征信息是表示大小的数值,统计方法为基尼系数统计方法,统计结果为基尼系数。

基尼系数是 1943 年美国经济学家阿尔伯特·赫希曼根据洛伦兹曲线(英文: Lorenz curve)所定义的判断收入分配公平程度的指标。基尼系数是一个比例值,其取值范围在 0 和 1 之间。在本实施例中,采用基尼系数来反映各个会话的特征信息的取值的分布平均程度。在本实施例中,将基尼系数作为网络攻击分析的特征参考,当基尼系数相较于参考结果发生较大幅度变化时,可判断网络存在异常,作为确认网络受到网络攻击的有效判据之一。在本实施例中,参考结果是在网络未受到网络攻击的情况下,采用下述步骤计算得到的基尼系数。

如图 4A 所示,图 3 所示的 S302 可以具体包括:

S401,按 N 个会话中每个会话的特征信息从小到大的顺序,对所述 N 个会话排序;

S402,从所述 N 个会话中选择 M 个会话,在坐标系中获取 M 个坐标,所述 M 个坐标与所述 M 个会话一一对应,所述 M 个会话中的第 a 个坐标对应于所述排序后的 N 个会话中的第 b 个会话,所述第 a 个坐标为 $(b/N, w/W)$,其中,所述 w 为所述 N 个会话中特征信息的值最小的 b 个会话的所述特征信息之和,所述 W 为所述 N 个会话的所述特征信息之和,所述 M 为大于 1 且小于等于所述 N 的整数,所述 a 为大于等于 1 且小于等于所述 M 的整数,所述 b 为大于等于 1 且小于等于所述 N 的整数;

S403,根据所述 M 个坐标,在所述坐标系中获取洛伦兹曲线;

S404, 根据所述洛伦兹曲线计算基尼系数。

结合参考图 4B, 横轴表示会话比例, 纵轴表示会话比例对应的特征信息比例的累积分布值。具体来说, 假设 N 个会话中特征信息的取值最小的 b 个会话的特征信息的取值之和为 w , 全部 N 个会话中所有会话的特征信息的取值之和为 W , 则步骤 401 排序后的 N 个会话中, 第 b 个会话对应的坐标为 $(\frac{b}{N}, \frac{w}{W})$, 也即上述 M 个坐标中的第 a 个坐标为 $(\frac{b}{N}, \frac{w}{W})$, 如图 4B 所示。

举例来说, 按照 S402 中的方法获取的所述 M 个坐标, 在坐标系中构建洛伦兹曲线。如图 4B 所示, 连接坐标 $(0, 0)$ 和坐标 $(1, 1)$ 的直线与洛伦兹曲线围合而成的非阴影部分区域的面积为 A , 所述洛伦兹曲线与坐标轴围合而成的阴影部分区域的面积为 B , 基尼系数 $= \frac{A}{A+B}$ 。用此定义, 在 N 个会话中每个会话的特征信息的取值均相等的情况下, 基尼系数为 0; 在 1 个会话的特征信息的取值为 W , 而其它各个会话的特征信息的取值均为 0 的情况下, 基尼系数为 1。由上述定义可以确定: 基尼系数越大, 表示 N 个会话中每个会话的特征信息的取值的分布越不均匀; 反之, 基尼系数越小, 表示 N 个会话中每个会话的特征信息的取值的分布越平均。

下面, 通过几个示意性例子对基尼系数统计方法进行举例说明。

1、会话的特征信息为上行报文平均长度。

以会话比例为横轴, 上行报文平均长度比例的累积分布值为纵轴, 构建洛伦兹曲线, 根据洛伦兹曲线计算基尼系数。示例性地, 假设 $N=10$, 该 10 个会话的上行报文平均长度从小到大依次为: 4、4、6、8、8、8、10、12、16、24。例如, 选择 $M=5$, 即从上述 10 个会话中选择 5 个会话。选择的 5 个会话分别是排序后的第 2 个会话、排序后的第 4 个会话、排序后的第 6 个会话、排序后的第 8 个会话以及排序后的第 10 个会话。则步骤 401 中获取的 M 个坐标分别是 $(20\%, 8\%)$ 、 $(40\%, 22\%)$ 、 $(60\%, 38\%)$ 、 $(80\%, 60\%)$ 和 $(100\%, 100\%)$, 依次连接上述各个坐标得到洛伦兹曲线, 而后便可根据洛伦兹曲线计算基尼系数。

在此示意性例子中, 基尼系数反映了各个会话的上行报文平均长度的分布平均程度。当网络在目标时段内受到报文长度固定的会话类网络攻击时, 基尼系数相较于参考结果会有较大差异, 例如基尼系数显著小于参考结果。因此, 根据基尼系数相较于参考结果的变化程度, 即可检测出是否存在网络攻击。

2、会话的特征信息为 TCP 会话中的 syn 报文数。

以会话比例为横轴，syn 报文数比例的累积分布值为纵轴，构建洛伦兹曲线，根据洛伦兹曲线计算基尼系数。在此示意性例子中，基尼系数反映了各个 TCP 会话中的 syn 报文数的分布平均程度。syn 报文是采用三次握手建立 TCP 连接时客户端向服务器发送的报文，syn 报文数通常较少且数量较为稳定。当网络在目标时段内受到 syn 报文数量多于正常值的会话类网络攻击时，则包括正常的会话和网络攻击的会话在内的 N 个会话中，syn 报文的数量分布与参考结果相比会变得更不平均，此时基尼系数大于参考结果，并且所述基尼系数与参考结果之差大于预设值。例如，正常会话在 TCP 三次握手的过程中会出现两个 syn 报文，而上述会话类网络攻击在一个会话中 syn 报文的数量超过两个。因此，根据基尼系数相较于参考结果的差异程度，即可检测出是否存在网络攻击。

3、会话的特征信息为源设备和目的设备之间的地表距离。

以会话比例为横轴，源设备和目的设备之间的地表距离比例的累积分布值为纵轴，构建洛伦兹曲线，根据洛伦兹曲线计算基尼系数。其中，源设备和目的设备之间的地表距离可根据源地理信息和目的地理信息确定。在此示意性例子中，基尼系数反映了各个会话对应的地表距离的分布平均程度。当网络在目标时段内受到从某一源设备到某一目的设备的会话类网络攻击时，N 个会话中地表距离相同的会话的数量与参考结果相比增多，地表距离的分布与参考结果相比会变得更加平均，基尼系数小于参考结果，且基尼系数与参考结果之差大于预设值。因此，根据基尼系数相较于参考结果的差异程度，即可检测出是否存在网络攻击。

当然，上述例子仅是示意性和解释性的，除上述例子中涉及的特征信息之外，基尼系数统计方法还可适用于统计分析其它多种特征信息。

在基于图 3 所示实施例提供的另一可选实施例中，统计方法为信息熵统计方法，统计结果为信息熵。相应地，如图 5 所示，S302 可以包括：

S501，根据 N 个会话中每个会话的特征信息，计算 n 类会话中的每一类会话在 N 个会话中的出现概率，所述 N 个会话中每一个会话属于所述 n 类会话中的一类会话，n 为大于 1 且小于等于 N 的整数；

S502，根据所述 n 类会话中每一类会话的出现概率计算信息熵。

信息熵是用于衡量一段信息所含的信息量的指标。信息熵在数学上是一个

较为抽象的概念，可将信息熵理解为某种特定信息的出现概率。信息熵的计算公式如下：

$$H=E[-\log p_i]=-\sum_{i=1}^n p_i \log p_i ; \dots\dots(1)$$

其中，H 表示信息熵。n 表示元素的取值种数，即本实施例中会话种类的数量。 p_i 表示第 i 种取值的元素的出现概率， $1 \leq i \leq n$ 且 i、n 均为整数，即本实施例中第 i 类会话出现的概率。信息熵越大，表示各元素的出现概率越均等；反之，信息熵越小，表示各元素的出现概率越不均等。

在本实施例中，属于同一类会话的多个会话的特征信息相同或者属于同一分类标准。所述 N 个会话中的每一个会话属于 n 类会话中的某一类会话，并且不属于其他 n-1 类会话。例如，会话的特征信息为该会话中上行报文的源 IP 地址，所述 N 个会话中上行报文的源 IP 地址有 n 种不同的取值，上行报文的源 IP 地址相同的会话属于一类会话。再例如，会话的特征信息为上行报文的源 IP 地址，所述 N 个会话中的上行报文的源 IP 地址所属的子网网段有 n 种不同的取值，上行报文的源 IP 地址属于同一子网网段的会话属于一类会话。

在本实施例中，将信息熵作为网络攻击，当信息熵相较于参考结果的差异较大时，可确定网络存在异常，作为确认网络受到网络攻击的有效判据之一。在本实施例中，参考结果是在网络未受到网络攻击的情况下，采用上述步骤计算得到的信息熵。

下面，通过几个示意性例子对信息熵统计方法进行举例说明。

1、会话的特征信息为上行报文的源 IP 地址，上行报文的源 IP 地址相同的会话属于同一类会话。

示例性地，假设 N=10，该 10 个会话的上行报文的源 IP 地址依次为：IP 地址 1、IP 地址 2、IP 地址 3、IP 地址 1、IP 地址 1、IP 地址 1、IP 地址 4、IP 地址 4、IP 地址 2、IP 地址 1。则 n=4，即上述 10 个会话可分为 4 类，上行报文的源 IP 地址为 IP 地址 1 的一类会话的出现概率为 0.5、上行报文的源 IP 地址为 IP 地址 2 的一类会话的出现概率为 0.2、上行报文的源 IP 地址为 IP 地址 3 的一类会话的出现概率为 0.1、上行报文的源 IP 地址为 IP 地址 4 的一类会话的出现概率为 0.2。将会话的分类数量和每一类会话的出现概率代入公式 (1) 计算信息熵。

在此示意性例子中，信息熵反映了上行报文的源 IP 地址相同的各类会话

的出现概率的均等程度。例如，信息熵越大，说明上述 4 类会话的出现概率越平均；信息熵越小，说明上述 4 类会话的出现概率越不平均。举例来说，当网络未受到攻击时，上行报文的源 IP 地址是上述 4 个源 IP 地址的会话出现的概率是均等的，当网络在目标时段内受到来自 IP 地址 1 的设备发起的网络攻击时，上行报文的源 IP 地址为 IP 地址 1 的会话比例高于其它几类会话，则此时信息熵小于参考结果。因此，根据信息熵相较于参考结果的差异，即可检测出是否存在网络攻击。例如，信息熵小于参考结果，并且参考结果与所述信息熵之差大于预设值时，确定存在网络攻击。在此示意性例子中，参考结果是在网络未受到网络攻击的情况下，根据采集的多个会话的上行报文的源 IP 地址计算得到的信息熵。

可选地，在确定网络受到网络攻击的情况下，可提取出现概率相较于正常值显著增大的上行报文的源 IP 地址，进一步分析确定发起网络攻击的 IP 地址。

2、会话的特征信息为上行报文的源 IP 地址，上行报文的源 IP 地址属于同一子网网段的会话属于同一类会话。

示例性地，仍然以上述 10 个会话的上行报文的源 IP 地址为例，假设 IP 地址 1 和 IP 地址 2 属于同一子网网段，记为子网网段 1，IP 地址 3 和 IP 地址 4 属于同一子网网段，记为子网网段 2。则上述 10 个会话可分为 2 类，上行报文的源 IP 地址属于子网网段 1 的一类会话的出现概率为 0.7、上行报文的源 IP 地址属于子网网段 2 的一类会话的出现概率为 0.3。将会话的分类数量和每一类会话的出现概率代入公式 (1) 计算信息熵。

在此示意性例子中，信息熵反映了上行报文的源 IP 地址属于同一子网网段的各类会话的出现概率的均等程度。当网络在目标时段内受到来自某一子网网段的会话类网络攻击时，该子网网段所对应的一类会话的出现概率远远大于其它几类会话，此时信息熵显著小于参考结果。因此，根据信息熵相较于参考结果的差异，即可检测出是否存在网络攻击。在此示意性例子中，参考结果是在网络未受到网络攻击的情况下，根据采集的多个会话的上行报文的源 IP 地址计算得到的信息熵。可选地，在确定网络受到网络攻击的情况下，可提取出现概率相较于正常值显著增大的子网网段，进一步分析确定发起网络攻击的子网网段。

3、会话的特征信息为上行报文数，上行报文数属于同一取值区间的会话属于同一类会话。

示例性地, 所述 $n=3$, 即将 N 个会话依据上行报文数这一特征信息将 N 个会话分为 3 类: (a) 上行报文数大于等于 1 且小于等于 2, (b) 上行报文数大于等于 3 且小于等于 10, (c) 上行报文数大于 10。假设 $N=10$, 该 10 个会话的上行报文数依次为: 1、1、1、4、6、1、1、6、12、1。则上行报文数属于上述第 (a) 种取值区间的一类会话的出现概率为 0.6, 上行报文数属于上述第 (b) 种取值区间的一类会话的出现概率为 0.3, 上行报文数属于上述第 (c) 种取值区间的一类会话的出现概率为 0.1。将会话的分类数量和每一类会话的出现概率代入公式 (1) 计算信息熵。

在此示意性例子中, 信息熵反映了上行报文数属于各个不同取值区间的各类会话的出现概率的均等程度。当网络在目标时段内受到大量空会话的会话类网络攻击时, 例如网络攻击会话中仅包括用于请求建立会话的报文, 而不包括用于传输数据的报文, 则第 (a) 种取值区间的一类会话的出现概率远远大于其它几类会话, 此时信息熵显著小于参考结果。因此, 根据信息熵相较于参考结果的差异, 即可检测出是否存在网络攻击。

当然, 上述例子仅是示意性和解释性的, 除上述例子中涉及的特征信息之外, 信息熵统计方法还可适用于统计分析其它多种特征信息。

在基于图 3 所示实施例提供的另一可选实施例中, 统计方法为卷积统计方法, 统计结果为卷积。所示 N 个会话中每个会话的特征信息包括第一类特征信息和第二类特征信息, 所述第一类特征信息是表示大小的数值, 所述第二类特征信息是表示大小的数值。相应地, 如图 6 所示, S302 可以包括:

S601, 根据 N 个会话中每个会话的第一类特征信息, 构建第一离散序列;

S602, 根据 N 个会话中每个会话的第二类特征信息, 构建第二离散序列;

S603, 计算所述第一离散序列和所述第二离散序列的卷积。

此外, 在本实施例中, 对 S601 和 S602 的执行先后顺序不做限定, S601 可以在 S602 之前执行, 也可在 S602 之后执行, 或者和 S602 同时执行。在图 6 中, 仅以 S601 在 S602 之前执行为例, 但并不对本申请构成限定。

卷积是两个变量在某范围内相乘后求和的结果。在本实施例中, 两个变量均为离散变量, 相应地卷积为离散卷积。卷积的计算公式如下:

$$y(n) = x(n) * h(n) = \sum_{i=1}^N x(i)h(n-i); \dots\dots(2)$$

其中,*表示卷积运算, $x(n)$ 和 $h(n)$ 分别对应第一离散序列和第二离散序列, $y(n)$ 是经过卷积运算后得到的离散序列。其中,第一离散序列 $x(n)$ 是N个会话中每个会话的第一类特征信息构成的序列,第一离散序列 $x(n)$ 中的第*i*个元素表示第*i*个会话的第一类特征信息;第二离散序列 $h(n)$ 是N个会话中每个会话的第二类特征信息构成的序列,第二离散序列 $x(n)$ 中的第*i*个元素表示第*i*个会话的第二类特征信息。卷积运算结果 $y(n)$ 体现了第一类特征信息与第二类特征信息之间的关联关系。上述第一类特征信息和第二类特征信息均以指示大小的数值表示。

在本实施例中,将卷积作为网络攻击分析的特征参考,当卷积相较于参考结果发生较大幅度变化时,可判断网络存在异常,作为确认网络受到网络攻击的有效判据之一。在本实施例中,参考结果是在网络未受到网络攻击的情况下,采用上述步骤计算得到的信息熵。

下面,通过几个示意性例子对卷积统计方法进行举例说明。

1、第一类特征信息为会话的起始时刻,第二类特征信息为会话持续时间。

举例来说,可以通过将参考时刻设置为0,将会话的起始时刻表示为与参考时刻的时间间隔,将会话的起始时刻转换为指示大小的数值。例如参考时刻为0,参考时刻为11时20分,起始时刻在参考时刻之后15秒,则起始时刻为15。会话的起始时刻,可以是获取会话特征信息的网络设备接收到会话中第一个报文的时刻,也可以是会话的第一个报文中的时间戳信息,本申请不作限定。会话持续时间是指从会话的起始时刻至终止时刻所经过的时长。根据N个会话中每个会话的起始时刻构建第一离散序列 $x(n)$,第一离散序列 $x(n)$ 中的第*i*个元素表示第*i*个会话的起始时刻;根据N个会话中每个会话的会话持续时间构建第二离散序列 $h(n)$,第二离散序列 $h(n)$ 中的第*i*个元素表示第*i*个会话的会话持续时间。按照公式(2),计算第一离散序列 $x(n)$ 和第二离散序列 $h(n)$ 的卷积。

在此示意性例子中,卷积用于反映会话持续时间在时域上的概率密度分布,体现了N个会话的会话持续时间在时域上的分布特性。当网络在目标时段内受到大量的会话持续时间很短的会话类网络攻击时,会话持续时间很短的会话在时域上的分布密度与参考结果相比会显著增大。例如,持续时间很短的会话,可以是只包含一到两个报文的会话。例如网络攻击的方式是只发起会话建立请求,不再进行回应。那么该会话在受攻击设备发送回应报文之后就结束了。而正常的会话有完整的建立过程,以及数据的交互,因此会话持续时间长于此类

网络攻击的会话。因此，根据卷积计算结果和参考结果之间的差异，即可检测出是否存在网络攻击。

2、每个会话的第一类特征信息为会话持续时间，第二类特征信息为报文平均长度。

5 根据 N 个会话中每个会话的会话持续时间构建第一离散序列 $x(n)$ ，第一离散序列 $x(n)$ 中的第 i 个元素表示第 i 个会话的会话持续时间；根据 N 个会话中每个会话的报文平均长度构建第二离散序列 $h(n)$ ，第二离散序列 $h(n)$ 中的第 i 个元素表示第 i 个会话的报文平均长度。按照公式(2)，计算第一离散序列 $x(n)$ 和第二离散序列 $h(n)$ 的卷积。

10 当网络在目标时段内受到大量的会话持续时间很短且报文平均长度较小的会话类网络攻击时，如同步洪水攻击，卷积计算结果相较于参考结果有显著差异。因此，根据卷积计算结果和参考结果之间的差异，即可检测出是否存在网络攻击。

3、每个会话的第一类特征信息为会话的起始时刻，第二类特征信息为上
15 行最小报文长度。

根据 N 个会话中每个会话的起始时刻构建第一离散序列 $x(n)$ ，第一离散序列 $x(n)$ 中的第 i 个元素表示第 i 个会话的起始时刻；根据 N 个会话中每个会话的上行最小报文长度构建第二离散序列 $h(n)$ ，第二离散序列 $h(n)$ 中的第 i 个元素表示第 i 个会话的上行最小报文长度。按照公式(2)，计算第一离散序列 $x(n)$ 和
20 第二离散序列 $h(n)$ 的卷积。

在此示意性例子中，卷积用于反映上行最小报文长度在时域上的概率密度分布。当网络在目标时段内受到报文长度很短的大量会话的会话类网络攻击时，统计结果与参考结果相比会大，且统计结果与参考结果之差大于预设值。因此，根据卷积计算结果和参考结果之间的差异，即可检测出是否存在网络攻击。

25 当然，上述例子仅是示意性和解释性的，除上述例子中涉及的特征信息之外，卷积统计方法还可适用于统计分析其它多种特征信息。

在基于图 3 所示实施例提供的另一可选实施例中，统计方法为向量和统计方法，统计结果为向量和，所述 N 个会话中每个会话的特征信息包括 p 类特征
30 信息，所述 p 为大于 1 的整数。相应地，如图 7 所示，S302 包括：

S701，构建 N 个向量，所述 N 个向量与所述 N 个会话一一对应，所述 N

个向量中的每个向量具有 p 个维度，所述 p 个维度与所述 p 类特征信息一一对应；

S702，计算 N 个向量的向量和。

具体来说，向量的某个维度与会话的某个特征信息对应，可以是指所述某个维度的取值与所述某个特征信息的取值相等，也可以是指所述某个维度的取值与所述某个特征信息成固定的比例。

例如，向量的第一维度的取值与会话的上行报文数对应，向量的第二维度的取值与会话的下行报文数对应。再例如，向量的第一维度的取值与会话的上行报文的目的地地理信息的经度对应，向量的第二维度的取值与会话的上行报文的目的地地理信息的纬度对应。

在本实施例中，将向量和作为网络攻击分析的特征参考，当向量和相较于参考结果发生较大幅度变化时，可判断网络存在异常，作为确认网络受到网络攻击的有效判据之一。在本实施例中，参考结果是在网络未受到网络攻击的情况下，采用上述步骤计算得到的向量和。例如，当向量和的大小相较于参考结果的大小发生大幅变化，且向量和的方向与参考结果的方向之间的夹角 θ 大于一定值时，确认存在网络攻击。

下面，通过几个示意性例子对向量和统计方法进行举例说明。

1、每个会话的特征信息包括上行报文的目的地地理信息的经度和上行报文的目的地地理信息的纬度。

以每一个会话的上行报文的目的地地理信息的经纬度坐标，构建每一个会话对应的特征向量，并计算各个会话对应的特征向量的向量和。其中，每一个会话对应的特征向量为二维特征向量，特征向量的两个维度分别对应上行报文的目的地地理信息的经度和纬度。

当网络在目标时段内受到针对某一地域的会话类网络攻击时，向量和相较于参考结果差异大于预设条件，例如统计结果与参考结果的向量长度和/或上述夹角 θ 大于预设值。因此，根据向量和相较于参考结果的差异，即可检测出是否存在网络攻击。

可选地，在确认存在网络攻击的情况下，可提取引起向量和发生上述显著变化的上行报文的目的地地理信息，进一步分析确定受到网络攻击的地域。

2、每个会话的特征信息包括上行报文数和下行报文数。

以每一个会话的上行报文数和下行报文数构建每一个会话对应的特征向

量，并计算各个会话对应的特征向量的向量和。其中，每一个会话对应的特征向量为二维特征向量，特征向量的两个维度分别对应上行报文数和下行报文数。

当网络在目标时段内受到具有大量上行报文的会话类网络攻击时，由于上行报文数与正常情况相比显著增加，使得统计结果与参考结果的差异大于预设条件。因此，根据向量和相较于参考结果的差异，即可检测出是否存在网络攻击。

3、每个会话的特征信息包括上行字节数和下行字节数。

以每一个会话的上行字节数和下行字节数构建每一个会话对应的特征向量，并计算各个会话对应的特征向量的向量和。其中，每一个会话对应的特征向量为二维特征向量，特征向量的两个维度分别对应上行字节数和下行字节数。

当网络在目标时段内受到具有大量字节的会话类网络攻击时，由于上行字节数显著增加，使得统计结果与参考结果的差异大于预设条件。因此，根据向量和相较于参考结果的差异，即可检测出是否存在网络攻击。

当然，上述例子仅是示意性和解释性的，除上述例子中涉及的特征信息之外，向量和统计方法还可适用于统计分析其它多种特征信息。

在基于图 3 所示实施例提供的另一可选实施例中，统计方法为向量积统计方法，统计结果为向量积。所述 N 个会话中每个会话的特征信息包括 p 类特征信息，所述 p 为大于 1 的整数。相应地，如图 7 所示，S302 包括：

S801，构建 N 个向量，所述 N 个向量与所述 N 个会话一一对应，所述 N 个向量中的每个向量具有 p 个维度，所述 p 个维度与所述 p 类特征信息一一对应；

S802，计算 N 个向量的向量积。

在本实施例中，将向量积作为网络攻击分析的特征参考，当向量积相较于参考结果发生较大幅度变化时，可判断网络存在异常，作为确认网络受到网络攻击的有效判据之一。在本实施例中，参考结果是在网络未受到网络攻击的情况下，采用上述步骤计算得到的向量积。例如，统计结果与参考结果大于预设条件时，确认存在网络攻击。此外，采用向量积统计方法的示意性例子可参考上述采用向量和统计方法的示意性例子，此处不再赘述。

在本申请实施例中，提供了多种统计方法。在实际应用中，可采用上述任意一种或多种的组合，对会话类网络攻击进行分析检测。

下述为本申请设备实施例，可以用于执行本申请方法实施例。对于本申请设备实施例中未披露的细节，请参照本申请方法实施例。

请参考图 9，其示出了本申请另一实施例提供的检测网络攻击的设备的框图。该设备可以包括：采集单元 910、获取单元 920 和确定单元 930。

采集单元 910，用于采集网络中的 N 个会话中每个会话的特征信息，N 为大于 1 的整数。

获取单元 920，用于获取统计结果，并将统计结果发送给确定单元 930；所述统计结果为以所述采集单元 910 采集的 N 个会话中的每个会话为一个采样单元，以所述特征信息为样本值，对所述 N 个会话的所述特征信息进行统计后得到的结果。

确定单元 930，用于在从获取单元 920 接收的统计结果和参考结果之间的差异超过预设条件的情况下，确定网络受到网络攻击。

综上所述，本实施例提供的设备，通过采集网络中的多个会话中每个会话的特征信息，获取统计结果，并在统计结果和参考结果之间的差异超过预设条件的情况下，确定网络受到网络攻击。通过会话随机抽样，并以会话为单位对会话的特征信息进行统计分析，可以有效检测出一些传统的报文随机抽样难以检测出的会话类网络攻击。会话类网络攻击存在如下特点：发起网络攻击的设备通过发起大量会话，占用目标设备的资源。由于发起一个会话只需要少量的报文，但是对设备的资源占用很多，因此在网络攻击报文的数量不大的情况下，就可以对目标设备造成较大的损害。若以报文为抽样单位，由于符合上述会话类网络攻击特性的网络攻击报文在数据流量中占比很小，导致此类网络攻击报文很难被采集；而本申请中以会话为抽样单位，无论是只有一个会话建立请求的网络攻击报文，还是包括数据流量较大的正常会话，在抽样统计中都作为一个会话来进行统计，因此有较大概率识别出所述会话类网络攻击。

在基于图 9 所示实施例提供的一个可选实施例中，所述特征信息是表示大小的数值，所述统计结果为基尼系数。所述获取单元 920，具体用于：按采集单元 910 采集的所述 N 个会话中每个会话的特征信息从小到大的顺序，对所述 N 个会话排序；从所述 N 个会话中选择 M 个会话，在坐标系中获取 M 个坐标，所述 M 个坐标与所述 M 个会话一一对应，所述 M 个会话中的第 a 个坐标对应于所述 N 个会话中的第 b 个会话，所述第 a 个坐标为 $(b/N, w/W)$ ，其中，所述

w 为所述 N 个会话中特征信息最小的 b 个会话的所述特征信息之和, 所述 W 为所述 N 个会话的所述特征信息之和, 所述 M 为大于 1 且小于等于所述 N 的整数, 所述 a 为大于等于 1 且小于等于所述 M 的整数, 所述 b 为大于等于 1 且小于等于所述 N 的整数; 根据所述 M 个坐标, 在所述坐标系中获取洛伦兹曲线; 根据所述洛伦兹曲线计算所述基尼系数。

在基于图 9 所示实施例提供的另一可选实施例中, 所述统计结果为信息熵。所述获取单元 920, 具体用于: 根据采集单元 910 采集的 N 个会话中每个会话的特征信息, 计算 n 类会话中的每一类会话在 N 个会话中的出现概率, 所述 N 个会话中的每一个会话属于所述 n 类会话中的一类会话, n 为大于 1 且小于等于 N 的整数; 根据所述 n 类会话中每一类会话的出现概率计算所述信息熵。

在基于图 9 所示实施例提供的另一可选实施例中, 所述统计结果为卷积, 所述 N 个会话中每个会话的特征信息包括第一类特征信息和第二类特征信息, 所述第一类特征信息是表示大小的数值, 所述第二类特征信息是表示大小的数值。所述获取单元 920, 具体用于: 根据采集单元 910 采集的 N 个会话中每个会话的第一类特征信息, 构建第一离散序列; 根据采集单元 910 采集的 N 个会话中每个会话的第二类特征信息, 构建第二离散序列; 计算第一离散序列和第二离散序列的卷积。

在基于图 9 所示实施例提供的另一可选实施例中, 所述统计结果为向量和, 所述 N 个会话中每个会话的特征信息包括 p 类特征信息, 所述 p 为大于 1 的整数。所述获取单元 920, 具体用于: 构建 N 个向量, 所述 N 个向量与所述 N 个会话一一对应, 所述 N 个向量中的每个向量具有 p 个维度, 所述 p 个维度与所述 p 类特征信息一一对应; 计算所述 N 个向量的向量和。

在基于图 9 所示实施例提供的另一可选实施例中, 所述统计结果为向量积, 所述 N 个会话中每个会话的特征信息包括 p 类特征信息, 所述 p 为大于 1 的整数。所述获取单元 920, 具体用于: 构建 N 个向量, 所述 N 个向量与所述 N 个会话一一对应, 所述 N 个向量中的每个向量具有 p 个维度, 所述 p 个维度与所述 p 类特征信息一一对应; 计算所述 N 个向量的向量积。

可选地, 所述特征信息包括: 五元组信息、报文数信息、字节数信息、时间信息、TTL 信息、报文长度信息、地理信息或报文平均长度中的一种或多种。

可选地, 在基于图 9 所示实施例提供的另一可选实施例中, 所述确定单元 930, 还用于: 根据从获取单元 920 接收的统计结果, 确定网络攻击的来源;

或者，根据从获取单元 920 接收的统计结果，确定网络攻击的目标。

需要说明的是：上述实施例提供的设备在实现其功能时，仅以上述各功能模块的划分进行举例说明，实际应用中，可以根据需要而将上述功能分配由不同的功能模块完成，即将设备的内部结构划分成不同的功能模块，以完成以上描述的全部或者部分功能。另外，上述实施例提供的设备与方法实施例属于同一构思，其具体实现过程详见方法实施例，这里不再赘述。

本领域普通技术人员可以理解实现上述实施例的全部或部分步骤可以通过硬件来完成，也可以通过程序来指令相关的硬件完成，所述的程序可以存储于一种计算机可读存储介质中，上述提到的存储介质可以是只读存储器，磁盘或光盘等。

以上所述仅为本申请的较佳实施例，并不用以限制本申请，凡在本申请的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本申请的保护范围之内。

权 利 要 求 书

1、一种检测网络攻击的方法，其特征在于，所述方法包括：

采集网络中的 N 个会话中每个会话的特征信息，所述 N 为大于 1 的整数；

5 获取统计结果，所述统计结果为以所述 N 个会话中的每个会话为一个采样单元，以所述特征信息为样本值，对所述 N 个会话的所述特征信息进行统计后得到的结果；

如果所述统计结果和参考结果之间的差异超过预设条件，确定所述网络受到网络攻击。

10

2、根据权利要求 1 所述的方法，其特征在于，所述特征信息是表示大小的数值，所述统计结果为基尼系数；所述获取统计结果，包括：

按所述特征信息从小到大的顺序，对所述 N 个会话排序；

15 从所述 N 个会话中选择 M 个会话，在坐标系中获取 M 个坐标，所述 M 个坐标与所述 M 个会话一一对应，所述 M 个会话中的第 a 个坐标对应于所述排序后 N 个会话中的第 b 个会话，所述第 a 个坐标为 $(b/N, w/W)$ ，其中，所述 w 为所述 N 个会话中特征信息最小的 b 个会话的所述特征信息之和，所述 W 为所述 N 个会话的所述特征信息之和，所述 M 为大于 1 且小于等于所述 N 的整数，所述 a 为大于等于 1 且小于等于所述 M 的整数，所述 b 为大于等于 1 且小于等于
20 所述 N 的整数；

根据所述 M 个坐标，在所述坐标系中获取洛伦兹曲线；

根据所述洛伦兹曲线计算所述基尼系数。

25 3、根据权利要求 1 所述的方法，其特征在于，所述统计结果为信息熵，所述获取统计结果，包括：

根据所述 N 个会话中每个会话的特征信息，计算 n 类会话中的每一类会话在所述 N 个会话中的出现概率，所述 N 个会话中的每一个会话属于所述 n 类会话中的一类会话，所述 n 为大于等于 1 且小于等于所述 N 的整数；

根据所述 n 类会话中每一类会话的所述出现概率计算所述信息熵。

30

4、根据权利要求 1 所述的方法，其特征在于，所述统计结果为卷积，所述

N 个会话中每个会话的特征信息包括第一类特征信息和第二类特征信息, 所述第一类特征信息是表示大小的数值, 所述第二类特征信息是表示大小的数值; 所述获取统计结果, 包括:

根据所述 N 个会话中每个会话的所述第一类特征信息, 构建第一离散序列;

5 根据所述 N 个会话中每个会话的所述第二类特征信息, 构建第二离散序列;
计算所述第一离散序列和所述第二离散序列的所述卷积。

5、根据权利要求 1 所述的方法, 其特征在于, 所述统计结果为向量和, 所述 N 个会话中每个会话的特征信息包括 p 类特征信息, 所述 p 为大于 1 的整数;
10 所述获取统计结果, 包括:

构建 N 个向量, 所述 N 个向量与所述 N 个会话一一对应, 所述 N 个向量中的每个向量具有 p 个维度, 所述 p 个维度与所述 p 类特征信息一一对应;

计算所述 N 个向量的向量和。

15 6、根据权利要求 1 所述的方法, 其特征在于, 所述统计结果为向量积, 所述 N 个会话中每个会话的特征信息包括 p 类特征信息, 所述 p 为大于 1 的整数;
所述获取统计结果, 包括:

构建 N 个向量, 所述 N 个向量与所述 N 个会话一一对应, 所述 N 个向量中的每个向量具有 p 个维度, 所述 p 个维度与所述 p 类特征信息一一对应;

20 计算所述 N 个向量的向量积。

7、根据权利要求 1 至 6 任一项所述的方法, 其特征在于, 所述确定所述网络受到网络攻击之后, 还包括:

根据所述统计结果, 确定所述网络攻击的来源; 或者,

25 根据所述统计结果, 确定所述网络攻击的目标。

8、一种检测网络攻击的设备, 其特征在于, 所述设备包括:

采集单元, 用于采集网络中的 N 个会话中每个会话的特征信息, 所述 N 为大于 1 的整数;

30 获取单元, 用于获取统计结果, 所述统计结果为以所述 N 个会话中的每个会话为一个采样单元, 以所述特征信息为样本值, 对所述 N 个会话的所述特征

信息进行统计后得到的结果；

确定单元，用于在所述统计结果和参考结果之间的差异超过预设条件的情况下，确定所述网络受到网络攻击。

- 5 9、根据权利要求8所述的设备，其特征在于，所述特征信息是表示大小的数值，所述统计结果为基尼系数；

所述获取单元，具体用于：

按所述特征信息从小到大的顺序，对所述N个会话排序；

- 10 从所述N个会话中选择M个会话，在坐标系中获取M个坐标，所述M个坐标与所述M个会话一一对应，所述M个会话中的第a个坐标对应于所述排序后的N个会话中的第b个会话，所述第a个坐标为 $(b/N, w/W)$ ，其中，所述w为所述N个会话中特征信息最小的b个会话的所述特征信息之和，所述W为所述N个会话的所述特征信息之和，所述M为大于1且小于等于所述N的整数，所述a为大于等于1且小于等于所述M的整数，所述b为大于等于1且小于等于
15 于所述N的整数；

根据所述M个坐标，在所述坐标系中获取洛伦兹曲线；

根据所述洛伦兹曲线计算所述基尼系数。

- 20 10、根据权利要求8所述的设备，其特征在于，所述统计结果为信息熵；
所述获取单元，具体用于：

根据所述N个会话中每个会话的特征信息，计算n类会话中的每一类会话在所述N个会话中的出现概率，所述N个会话中的每一个会话属于所述n类会话中的一类会话，所述n为大于等于1且小于等于所述N的整数；

根据所述n类会话中每一类会话的所述出现概率计算所述信息熵。

- 25 11、根据权利要求8所述的设备，其特征在于，所述统计结果为卷积，所述N个会话中每个会话的特征信息包括第一类特征信息和第二类特征信息，所述
所述第一类特征信息是表示大小的数值，所述第二类特征信息是表示大小的数值；

所述获取单元，具体用于：

- 30 根据所述N个会话中每个会话的所述第一类特征信息，构建第一离散序列；
根据所述N个会话中每个会话的所述第二类特征信息，构建第二离散序列；

计算所述第一离散序列和所述第二离散序列的所述卷积。

12、根据权利要求 8 所述的设备，其特征在于，所述统计结果为向量和，
所述 N 个会话中每个会话的特征信息包括 p 类特征信息，所述 p 为大于 1 的整
5 数；

所述获取单元，具体用于：

构建 N 个向量，所述 N 个向量与所述 N 个会话一一对应，所述 N 个向量中的
每个向量具有 p 个维度，所述 p 个维度与所述 p 类特征信息一一对应；

计算所述 N 个向量的向量和。

10

13、根据权利要求 8 所述的设备，其特征在于，所述统计结果为向量积，
所述 N 个会话中每个会话的特征信息包括 p 类特征信息，所述 p 为大于 1 的整
数；

所述获取单元，具体用于：

15 构建 N 个向量，所述 N 个向量与所述 N 个会话一一对应，所述 N 个向量中的
每个向量具有 p 个维度，所述 p 个维度与所述 p 类特征信息一一对应；

计算所述 N 个向量的向量积。

14、根据权利要求 8 至 13 任一项所述的设备，其特征在于，所述确定单元，
20 还用于：

根据所述统计结果，确定所述网络攻击的来源；或者，

根据所述统计结果，确定所述网络攻击的目标。

15、一种检测网络攻击的设备，其特征在于，所述设备包括：处理器以及
25 网络接口，所述处理器用于执行以下操作：

通过所述网络接口采集网络中的 N 个会话中每个会话的特征信息，所述 N
为大于 1 的整数；

获取统计结果，所述统计结果为以所述 N 个会话中的每个会话为一个采样
单元，以所述特征信息为样本值，对所述 N 个会话的所述特征信息进行统计后
30 得到的结果；

如果所述统计结果和参考结果之间的差异超过预设条件，确定所述网络受

到网络攻击。

16、根据权利要求 15 所述的设备，其特征在于，所述特征信息是表示大小的数值，所述统计结果为基尼系数；所述获取统计结果，包括：

5 按所述特征信息从小到大的顺序，对所述 N 个会话排序；

从所述 N 个会话中选择 M 个会话，在坐标系中获取 M 个坐标，所述 M 个坐标与所述 M 个会话一一对应，所述 M 个会话中的第 a 个坐标对应于所述排序后的 N 个会话中的第 b 个会话，所述第 a 个坐标为 $(b/N, w/W)$ ，其中，所述 w 为所述 N 个会话中特征信息最小的 b 个会话的所述特征信息之和，所述 W 为所述 N 个会话的所述特征信息之和，所述 M 为大于 1 且小于等于所述 N 的整数，
10 所述 a 为大于等于 1 且小于等于所述 M 的整数，所述 b 为大于等于 1 且小于等于所述 N 的整数；

根据所述 M 个坐标，在所述坐标系中获取洛伦兹曲线；

根据所述洛伦兹曲线计算所述基尼系数。

15

17、根据权利要求 15 所述的设备，其特征在于，所述统计结果为信息熵，所述获取统计结果，包括：

根据所述 N 个会话中每个会话的特征信息，计算 n 类会话中的每一类会话在所述 N 个会话中的出现概率，所述 N 个会话中的每一个会话属于所述 n 类会话中的一类会话，所述 n 为大于等于 1 且小于等于所述 N 的整数；
20

根据所述 n 类会话中每一类会话的所述出现概率计算所述信息熵。

18、根据权利要求 15 所述的设备，其特征在于，所述统计结果为卷积，所述 N 个会话中每个会话的特征信息包括第一类特征信息和第二类特征信息，所述
25 所述第一类特征信息是表示大小的数值，所述第二类特征信息是表示大小的数值；所述获取统计结果，包括：

根据所述 N 个会话中每个会话的所述第一类特征信息，构建第一离散序列；
根据所述 N 个会话中每个会话的所述第二类特征信息，构建第二离散序列；
计算所述第一离散序列和所述第二离散序列的所述卷积。

30

19、根据权利要求 15 所述的设备，其特征在于，所述统计结果为向量和，

所述 N 个会话中每个会话的特征信息包括 p 类特征信息，所述 p 为大于 1 的整数；所述获取统计结果，包括：

构建 N 个向量，所述 N 个向量与所述 N 个会话一一对应，所述 N 个向量中的每个向量具有 p 个维度，所述 p 个维度与所述 p 类特征信息一一对应；

5 计算所述 N 个向量的向量和。

20、根据权利要求 15 所述的设备，其特征在于，所述统计结果为向量积，所述 N 个会话中每个会话的特征信息包括 p 类特征信息，所述 p 为大于 1 的整数；所述获取统计结果，包括：

10 构建 N 个向量，所述 N 个向量与所述 N 个会话一一对应，所述 N 个向量中的每个向量具有 p 个维度，所述 p 个维度与所述 p 类特征信息一一对应；

计算所述 N 个向量的向量积。

21、根据权利要求 15 至 20 任一项所述的设备，其特征在于，所述处理器，
15 还用于执行以下操作：

根据所述统计结果，确定所述网络攻击的来源；或者，

根据所述统计结果，确定所述网络攻击的目标。

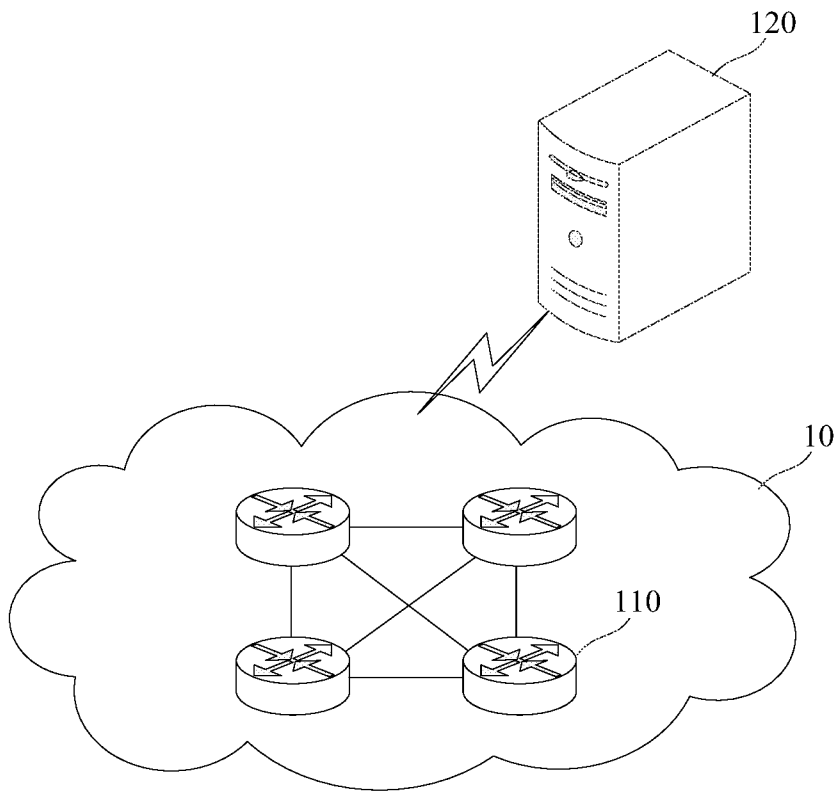


图 1

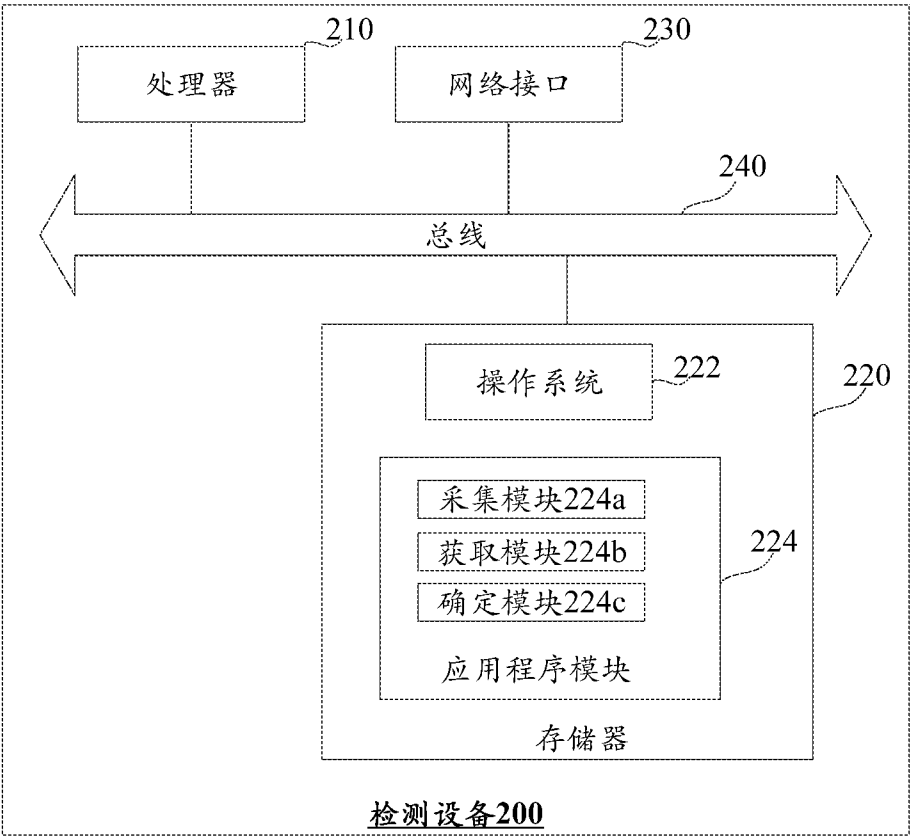


图 2

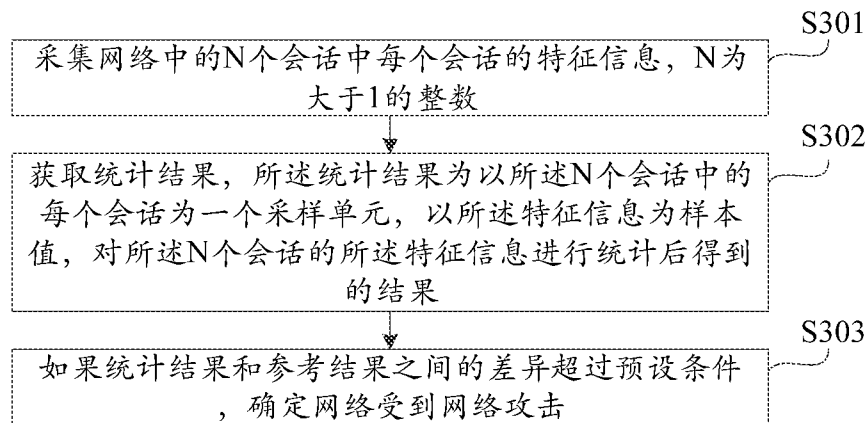


图 3

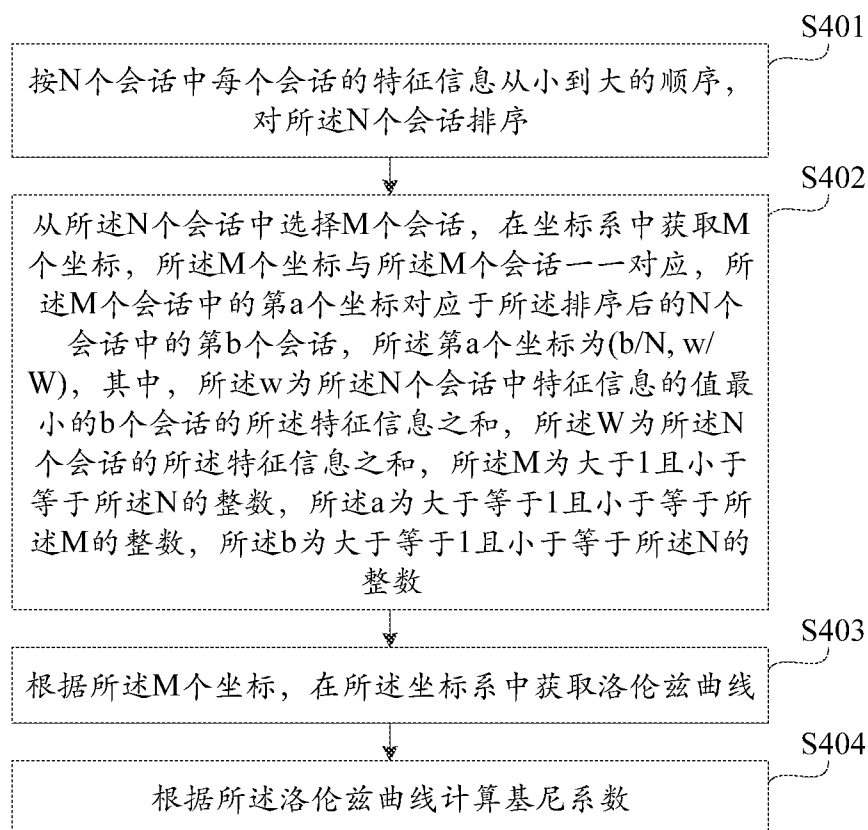


图 4A

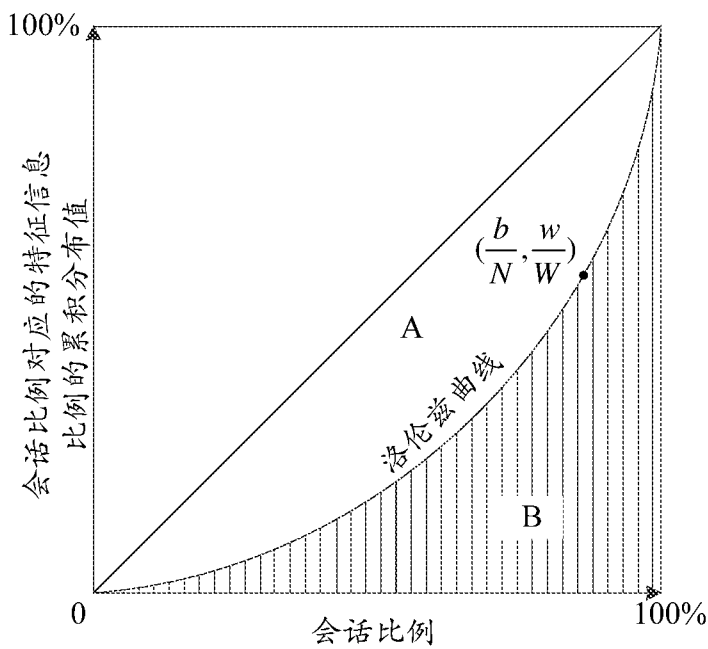


图 4B

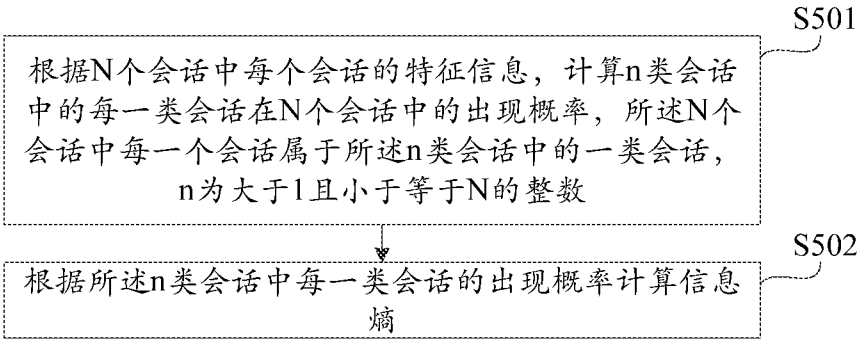


图 5

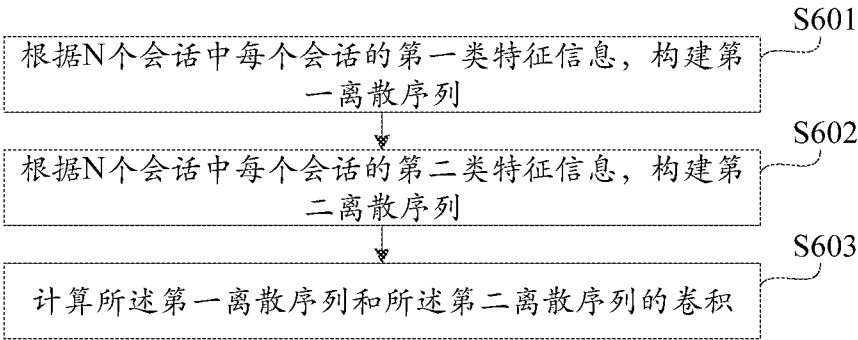


图 6

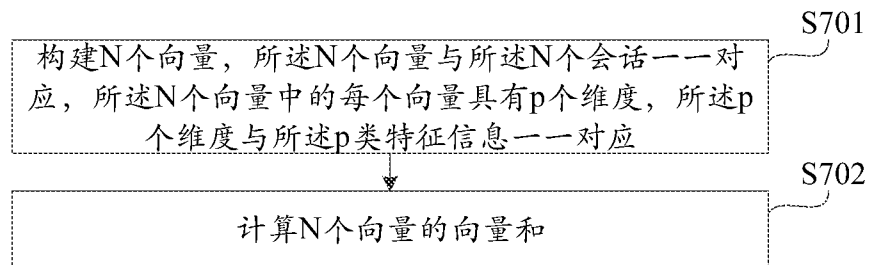


图 7

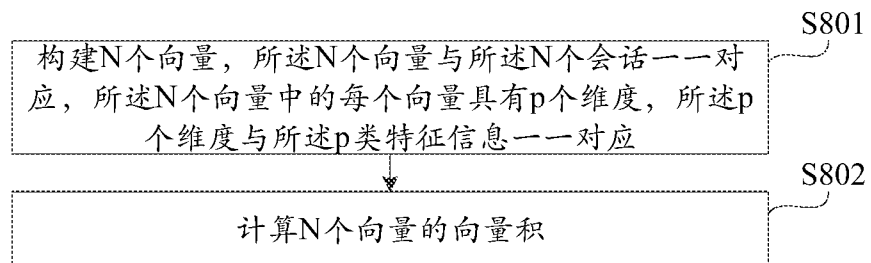


图 8

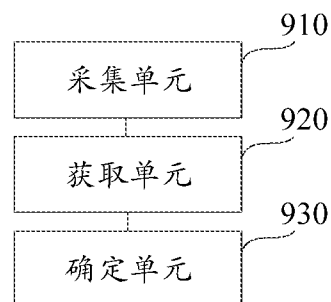


图 9

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2016/081503

A. CLASSIFICATION OF SUBJECT MATTER

H04L 29/06 (2006.01) 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)

H04L

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, CNKI, WPI, EPODOC, GOOGLE: information entropy, network, attack, statistics, session, feature, threshold, length, address, size, entropy, Gini coefficient, vector, convolution

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 102882894 A (HANGZHOU DPTECH TECHNOLOGIES CO., LTD.), 16 January 2013 (16.01.2013), description, paragraphs 0009 and 0024-0055, and figures 1-3	1-21
A	CN 104796405 A (SANGFOR INC.), 22 July 2015 (22.07.2015), the whole document	1-21
A	JP 2007060379 A (NIPPON TELEGRAPH & TELEPHONE CORP.), 08 March 2007 (08.03.2007), the whole document	1-21
A	CN 105530243 A (INFORMATION CENTER OF CHINA SOUTHERN POWER GRID CO., LTD. et al.), 27 April 2016 (27.04.2016), the whole document	1-21
A	US 2013042322 A1 (ELECTRONICS AND TELECOMMUNICATIONS RESEARCH INSTITUTE), 14 February 2013 (14.02.2013), the whole document	1-21
A	CN 102202064 A (LIU, Shengli), 28 September 2011 (28.09.2011), the whole document	1-21

☐ 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	
"E" earlier application or patent but published on or after the international filing date	"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
"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)	"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
"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	"&" document member of the same patent family

Date of the actual completion of the international search
12 January 2017 (12.01.2017)

Date of mailing of the international search report
26 January 2017 (26.01.2017)

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
WANG, Xia
Telephone No.: (86-10) **62413245**

International application No.
PCT/CN2016/081503

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

国际检索报告

国际申请号

PCT/CN2016/081503

A. 主题的分类

H04L 29/06 (2006.01) 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)

H04L

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, CNKI, WPI, EPDOC, GOOGLE: 网络, 攻击, 统计, 会话, 特征, 门限, 阈值, 长度, 地址, 大小, 信息熵, 基尼系数, 向量, 卷积, network, attack, statistics, session, feature, threshold, length, address, size, entropy, Gini coefficient, vector, convolution

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 102882894 A (杭州迪普科技有限公司) 2013年 1月 16日 (2013 - 01 - 16) 说明书第0009、0024-0055段, 图1-3	1-21
A	CN 104796405 A (深信服网络科技深圳有限公司) 2015年 7月 22日 (2015 - 07 - 22) 全文	1-21
A	JP 2007060379 A (NIPPON TELEGRAPH & TELEPHONE CORP.) 2007年 3月 8日 (2007 - 03 - 08) 全文	1-21
A	CN 105530243 A (中国南方电网有限责任公司信息中心等) 2016年 4月 27日 (2016 - 04 - 27) 全文	1-21
A	US 2013042322 A1 (ELECTRONICS AND TELECOMMUNICATIONS RESEARCH INSTITUTE) 2013年 2月 14日 (2013 - 02 - 14) 全文	1-21
A	CN 102202064 A (刘胜利) 2011年 9月 28日 (2011 - 09 - 28) 全文	1-21

☐ 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

2017年 1月 12日

国际检索报告邮寄日期

2017年 1月 26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

王侠

电话号码 (86-10) 62413245

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2016/081503

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	102882894	A	2013年 1月 16日	无			
CN	104796405	A	2015年 7月 22日	无			
JP	2007060379	A	2007年 3月 8日	无			
CN	105530243	A	2016年 4月 27日	无			
US	2013042322	A1	2013年 2月 14日	KR	20130017333	A	2013年 2月 20日
CN	102202064	A	2011年 9月 28日	无			

表 PCT/ISA/210 (同族专利附件) (2009年7月)