(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 115051845 A
(43)申请公布日 2022.09.13

(21)申请号 202210639313.6

(22)申请日 2022.06.08

(71)申请人 北京启明星辰信息安全技术有限公司
地址 100193 北京市海淀区东北旺西路8号21号楼启明星辰大厦102号
申请人 北京网御星云信息技术有限公司

(72)发明人 陈亘 刘敦辉

(74)专利代理机构 北京集佳知识产权代理有限公司 11227
专利代理师 高勇

(51)Int.Cl.
H04L 9/40(2022.01)

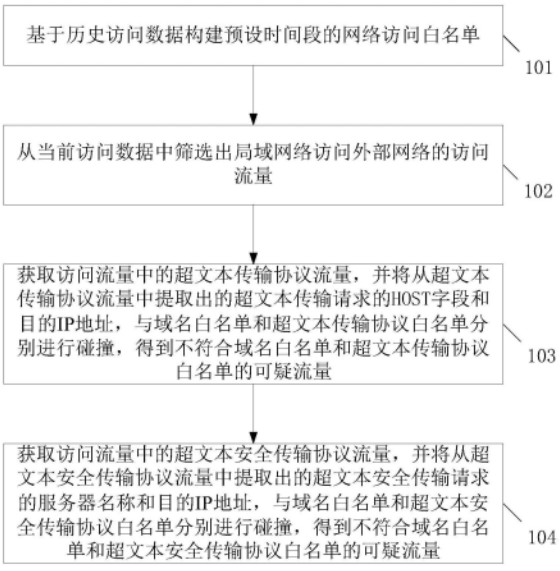
权利要求书2页 说明书8页 附图4页

(54)发明名称

一种可疑流量识别方法、装置、设备和存储介质

(57)摘要

本发明所提供的可疑流量识别方法，可以基于历史访问数据构建预设时间段的网络访问白名单，然后从当前访问数据中筛选出局域网络访问外部网络的访问流量。获取访问流量中的超文本传输协议流量，并将从超文本传输协议流量中提取出的超文本传输请求的HOST字段和目的IP地址，与域名白名单和超文本传输协议白名单分别进行碰撞，得到不符合域名白名单和超文本传输协议白名单的可疑流量。并将从超文本安全传输协议流量中提取出的超文本安全传输请求的服务器名称和目的IP地址，与域名白名单和所述超文本安全传输协议白名单分别进行碰撞，得到不符合域名白名单和超文本安全传输协议白名单的可疑流量。该流量过滤方法提高了对恶意流量过滤的精准度。



CN 115051845 A

1.一种可疑流量识别方法,其特征在于,包括:

基于历史访问数据构建预设时间段的网络访问白名单,所述网络访问白名单至少包括:域名白名单、超文本传输协议白名单和超文本安全传输协议白名单,所述域名白名单中存储有满足第一稀有度阈值的域名访问请求的域名,所述超文本传输协议白名单中存储有满足第二稀有度阈值的超文本传输请求的HOST字段和目的IP地址,所述超文本安全传输协议白名单中存储有满足第三稀有度阈值的超文本安全传输请求的服务器名称和目的IP地址;

从当前访问数据中筛选出局域网络访问外部网络的访问流量;

获取所述访问流量中的超文本传输协议流量,并将从所述超文本传输协议流量中提取出的超文本传输请求的HOST字段和目的IP地址,与所述域名白名单和所述超文本传输协议白名单分别进行碰撞,得到不符合所述域名白名单和所述超文本传输协议白名单的可疑流量;

获取所述访问流量中的超文本安全传输协议流量,并将从所述超文本安全传输协议流量中提取出的超文本安全传输请求的服务器名称和目的IP地址,与所述域名白名单和所述超文本安全传输协议白名单分别进行碰撞,得到不符合所述域名白名单和所述超文本安全传输协议白名单的可疑流量。

2.根据权利要求1所述的方法,其特征在于,所述基于历史访问数据构建预设时间段的网络访问白名单,包括:

获取所述历史访问数据中域名访问请求流量;

对所述预设时间段内包含相同域名的域名访问请求进行去重,得到去重后的域名访问请求中的源IP地址的个数;

将所述去重后的域名访问请求中的源IP地址的个数与所述第一稀有度阈值进行比较,得到满足所述第一稀有度阈值的域名访问请求的域名,并存入所述域名白名单中。

3.根据权利要求1所述的方法,其特征在于,所述基于历史访问数据构建预设时间段的网络访问白名单,包括:

获取所述历史访问数据中超文本传输请求流量;

对所述预设时间段内包含相同HOST字段和目的IP地址的超文本传输请求进行去重,得到去重后的超文本传输请求中的源IP地址的个数;

将所述去重后的超文本传输请求中的源IP地址的个数和所述第二稀有度阈值进行比较,得到满足所述第二稀有度阈值的超文本传输请求的HOST字段和目的IP地址,并存入所述超文本传输协议白名单中。

4.根据权利要求1所述的方法,其特征在于,所述基于历史访问数据构建预设时间段的网络访问白名单,包括:

获取所述历史访问数据中超文本安全传输请求流量;

将所述预设时间段内包含相同服务器名称和目的IP地址的超文本安全传输请求进行去重,得到去重后的超文本安全传输请求中的源IP地址的个数;

将所述去重后的超文本安全传输请求中的源IP地址的个数与所述第三稀有度阈值进行比较,得到满足所述第三稀有度阈值的超文本安全传输请求的服务器名称和目的IP地址,并存入所述超文本安全传输协议白名单中。

5.根据权利要求1所述的方法,其特征在于,所述从当前访问数据中筛选出局域网络访问外部网络的访问流量,包括:

获取局域网络内的地址列表,所述地址列表中存储有局域网络内设备的IP地址;

获取所述当前访问数据中的源IP地址和目的IP地址,若所述当前访问数据中的源IP地址存在于所述地址列表中且所述目的IP地址不存在所述地址列表中,则判定为局域网络访问外部网络的访问流量。

6.根据权利要求1所述的方法,其特征在于,所述第一稀有度阈值、所述第二稀有度阈值和所述第三稀有度阈值为相同的数值。

7.根据权利要求1至6任一项所述的方法,其特征在于,还包括:

在对所述网络访问白名单进行刷新后,基于更新后的网络访问白名单对正在使用中的网络访问白名单进行全部替换。

8.一种可疑流量识别装置,其特征在于,包括:

白名单模块,用于基于历史访问数据构建预设时间段的网络访问白名单,所述网络访问白名单至少包括:域名白名单、超文本传输协议白名单和超文本安全传输协议白名单,所述域名白名单中存储有满足第一稀有度阈值的域名访问请求的域名,所述超文本传输协议白名单中存储有满足第二稀有度阈值的超文本传输请求的HOST字段和目的IP地址,所述超文本安全传输协议白名单中存储有满足第三稀有度阈值的超文本安全传输请求的服务器名称和目的IP地址;

访问流量模块,用于从当前访问数据中筛选出局域网络访问外部网络的访问流量;

第一识别模块,用于获取所述访问流量中的超文本传输协议流量,并将从所述超文本传输协议流量中提取出的超文本传输请求的HOST字段和目的IP地址,与所述域名白名单和所述超文本传输协议白名单分别进行碰撞,得到不符合所述域名白名单和所述超文本传输协议白名单的可疑流量;以及

第二识别模块,用于获取所述访问流量中的超文本安全传输协议流量,并将从所述超文本安全传输协议流量中提取出的超文本安全传输请求的服务器名称和目的IP地址,与所述域名白名单和所述超文本安全传输协议白名单分别进行碰撞,得到不符合所述域名白名单和所述超文本安全传输协议白名单的可疑流量。

9.一种设备,其特征在于,包括:存储器和处理器;

所述存储器,用于存储程序;

所述处理器,用于执行所述程序,实现如权利要求1至7中任一项所述的可疑流量识别方法的各个步骤。

10.一种存储介质,其上存储有计算机程序,其特征在于,所述计算机程序被处理器执行时,实现如权利要求1至7中任一项所述的可疑流量识别方法的各个步骤。

一种可疑流量识别方法、装置、设备和存储介质

技术领域

[0001] 本发明涉及信息安全技术领域，具体涉及一种可疑流量识别方法、装置、设备和存储介质。

背景技术

[0002] 在网络的信号安全领域中，网络攻击的发起者为更好地隐藏自己，往往会使用HTTP(超文本传输协议)或HTTPS(超文本安全传输协议)的一些特性将自身伪装成大站流量以躲避检测。由于恶意流量和网络正常访问的流量具有非常高的相似性，非常容易绕过常规检测设备的流量预处理机制，造成大量的漏报产生。如何实现对可疑流量高效精准的识别，已经成为了亟需解决的技术问题。

发明内容

[0003] 为了解决现有技术存在的可疑流量漏报严重的问题，本发明提供了一种可疑流量识别方法、装置、设备和存储介质，其具有可疑流量识别更加精准等特点。

[0004] 根据本发明具体实施方式提供的一种可疑流量识别方法，包括：

[0005] 基于历史访问数据构建预设时间段的网络访问白名单，所述网络访问白名单至少包括：域名白名单、超文本传输协议白名单和超文本安全传输协议白名单，所述域名白名单中存储有满足第一稀有度阈值的域名访问请求的域名，所述超文本传输协议白名单中存储有满足第二稀有度阈值的超文本传输请求的HOST字段和目的IP地址，所述超文本安全传输协议白名单中存储有满足第三稀有度阈值的超文本安全传输请求的服务器名称和目的IP地址；

[0006] 从当前访问数据中筛选出局域网络访问外部网络的访问流量；

[0007] 获取所述访问流量中的超文本传输协议流量，并将从所述超文本传输协议流量中提取出的超文本传输请求的HOST字段和目的IP地址，与所述域名白名单和所述超文本传输协议白名单分别进行碰撞，得到不符合所述域名白名单和所述超文本传输协议白名单的可疑流量；

[0008] 获取所述访问流量中的超文本安全传输协议流量，并将从所述超文本安全传输协议流量中提取出的超文本安全传输请求的服务器名称和目的IP地址，与所述域名白名单和所述超文本安全传输协议白名单分别进行碰撞，得到不符合所述域名白名单和所述超文本安全传输协议白名单的可疑流量。

[0009] 进一步地，所述基于历史访问数据构建预设时间段的网络访问白名单，包括：

[0010] 获取所述历史访问数据中域名访问请求流量；

[0011] 对所述预设时间段内包含相同域名的域名访问请求进行去重，得到去重后的域名访问请求中的源IP地址的个数；

[0012] 将所述去重后的域名访问请求中的源IP地址的个数与所述第一稀有度阈值进行比较，得到满足所述第一稀有度阈值的域名访问请求的域名，并存入所述域名白名单中。

[0013] 进一步地,所述基于历史访问数据构建预设时间段的网络访问白名单,包括:

[0014] 获取所述历史访问数据中超文本传输请求流量;

[0015] 对所述预设时间段内包含相同HOST字段和目的IP地址的超文本传输请求进行去重,得到去重后的超文本传输请求中的源IP地址的个数;

[0016] 将所述去重后的超文本传输请求中的源IP地址的个数和所述第二稀有度阈值进行比较,得到满足所述第二稀有度阈值的超文本传输请求的HOST字段和目的IP地址,并存入所述超文本传输协议白名单中。

[0017] 进一步地,所述基于历史访问数据构建预设时间段的网络访问白名单,包括:

[0018] 获取所述历史访问数据中超文本安全传输请求流量;

[0019] 将所述预设时间段内包含相同服务器名称和目的IP地址的超文本安全传输请求进行去重,得到去重后的超文本安全传输请求中的源IP地址的个数;

[0020] 将所述去重后的超文本安全传输请求中的源IP地址的个数与所述第三稀有度阈值进行比较,得到满足所述第三稀有度阈值的超文本安全传输请求的服务器名称和目的IP地址,并存入所述超文本安全传输协议白名单中。

[0021] 进一步地,所述从当前访问数据中筛选出局域网络访问外部网络的访问流量,包括:

[0022] 获取局域网络内的地址列表,所述地址列表中存储有局域网络内设备的IP地址;

[0023] 获取所述当前访问数据中的源IP地址和目的IP地址,若所述当前访问数据中的源IP地址存在于所述地址列表中且所述目的IP地址不存在所述地址列表中,则判定为局域网络访问外部网络的访问流量。

[0024] 进一步地,所述第一稀有度阈值、所述第二稀有度阈值和所述第三稀有度阈值为相同的数值。

[0025] 进一步地,所述可疑流量识别方法还包括:

[0026] 在对所述网络访问白名单进行刷新后,基于更新后的网络访问白名单对正在使用中的网络访问白名单进行全部替换。

[0027] 根据本发明具体实施方式提供的一种可疑流量识别装置,包括:

[0028] 白名单模块,用于基于历史访问数据构建预设时间段的网络访问白名单,所述网络访问白名单至少包括:域名白名单、超文本传输协议白名单和超文本安全传输协议白名单,所述域名白名单中存储有满足第一稀有度阈值的域名访问请求的域名,所述超文本传输协议白名单中存储有满足第二稀有度阈值的超文本传输请求的HOST字段和目的IP地址,所述超文本安全传输协议白名单中存储有满足第三稀有度阈值的超文本安全传输请求的服务器名称和目的IP地址;

[0029] 访问流量模块,用于从当前访问数据中筛选出局域网络访问外部网络的访问流量;

[0030] 第一识别模块,用于获取所述访问流量中的超文本传输协议流量,并将从所述超文本传输协议流量中提取出的超文本传输请求的HOST字段和目的IP地址,与所述域名白名单和所述超文本传输协议白名单分别进行碰撞,得到不符合所述域名白名单和所述超文本传输协议白名单的可疑流量;以及

[0031] 第二识别模块,用于获取所述访问流量中的超文本安全传输协议流量,并将从所

述超文本安全传输协议流量中提取出的超文本安全传输请求的服务器名称和目的IP地址，与所述域名白名单和所述超文本安全传输协议白名单分别进行碰撞，得到不符合所述域名白名单和所述超文本安全传输协议白名单的可疑流量。

[0032] 根据本发明具体实施方式提供的一种设备，包括：存储器和处理器；

[0033] 所述存储器，用于存储程序；

[0034] 所述处理器，用于执行所述程序，实现如上所述的可疑流量识别方法的各个步骤。

[0035] 根据本发明具体实施方式提供的一种存储介质，其上存储有计算机程序，其特征在于，所述计算机程序被处理器执行时，实现如上所述的可疑流量识别方法的各个步骤。

[0036] 本发明所提供的可疑流量识别方法，可以基于历史访问数据构建预设时间段的网络访问白名单，其中网络访问白名单至少包括：域名白名单、超文本传输协议白名单和超文本安全传输协议白名单，域名白名单中存储有满足第一稀有度阈值的域名访问请求的域名，超文本传输协议白名单中存储有满足第二稀有度阈值的超文本传输请求的HOST字段和目的IP地址，超文本安全传输协议白名单中存储有满足第三稀有度阈值的超文本安全传输请求的服务器名称和目的IP地址。然后从当前访问数据中筛选出局域网络访问外部网络的访问流量。获取访问流量中的超文本传输协议流量，并将从超文本传输协议流量中提取出的超文本传输请求的HOST字段和目的IP地址，与域名白名单和超文本传输协议白名单分别进行碰撞，得到不符合域名白名单和超文本传输协议白名单的可疑流量。获取访问流量中的超文本安全传输协议流量，并将从超文本安全传输协议流量中提取出的超文本安全传输请求的服务器名称和目的IP地址，与域名白名单和所述超文本安全传输协议白名单分别进行碰撞，得到不符合域名白名单和超文本安全传输协议白名单的可疑流量。该流量过滤方法对超文本传输协议和超文本安全传输协议分别由独立的白名单进行过滤，并且基于白名单使用域名与对应的目的IP地址为过滤对象进行过滤，提高了对恶意流量过滤的精准度，减少漏报。

附图说明

[0037] 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，还可以根据提供的附图获得其他的附图。

[0038] 图1是根据一示例性实施例提供的可疑流量识别方法的流程图；

[0039] 图2是根据一示例性实施例提供的域名白名单的构建流程图；

[0040] 图3是根据一示例性实施例提供的超文本传输协议白名单的构建流程图；

[0041] 图4是根据一示例性实施例提供的超文本安全传输协议白名单的构建流程图；

[0042] 图5是根据一示例性实施例提供的访问流量的获取流程图；

[0043] 图6是根据一示例性实施例提供的可疑流量识别装置的结构图；

[0044] 图7是根据一示例性实施例提供的设备的结构图。

具体实施方式

[0045] 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、完

整地描述，显然，所描述的实施例仅仅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。

[0046] 参照图1所示，本发明的实施例提供了一种可疑流量识别方法，该方法可以包括以下步骤：

[0047] 101、基于历史访问数据构建预设时间段的网络访问白名单，网络访问白名单至少包括：域名白名单、超文本传输协议白名单和超文本安全传输协议白名单，域名白名单中存储有满足第一稀有度阈值的域名访问请求的域名，超文本传输协议白名单中存储有满足第二稀有度阈值的超文本传输请求的HOST字段(请求头字段)和目的IP地址，超文本安全传输协议白名单中存储有满足第三稀有度阈值的超文本安全传输请求的服务器名称和目的IP地址。

[0048] 白名单是与“黑名单”相对应。例如：在电脑系统里，有很多软件都应用到了黑白名单规则，操作系统、防火墙、杀毒软件、邮件系统、应用软件等，凡是涉及到控制方面几乎都应用了黑白名单规则。

[0049] 黑名单启用后，被列入到黑名单的用户(或IP地址、IP包、邮件、病毒等)不能通过。如果设立了白名单，则在白名单中的用户(或IP地址、IP包、邮件等)会优先通过，不会被当成垃圾邮件拒收，安全性和快捷性都大大提高。正是基于白名单的这种特性。基于网络上的历史访问数据建立预设时间段的网络访问白名单，如在一小时内的网络访问白名单。其中网络访问白名单可包括域名白名单、超文本传输协议白名单和超文本安全传输协议白名单。其中域名白名单是对网络中DNS协议流量中每个DNS请求即域名访问请求的域名进行的统计，其统计的条件是域名访问请求的域名在一小时的统计时间内稀有度阈值，不能低于第一稀有度阈值，这里设置稀有度阈值是基于一个有限的网络内DNS请求中访问越多的域名越安全的原理而设定的。基于同样的原理对超文本传输协议白名单中存储有满足第二稀有度阈值的超文本传输请求的HOST字段和目的IP地址，超文本安全传输协议白名单中存储有满足第三稀有度阈值的超文本安全传输请求的服务器名称和目的IP地址。其中超文本传输请求的HOST字段为要访问的域名或者目的IP地址。

[0050] 102、从当前访问数据中筛选出局域网络访问外部网络的访问流量。

[0051] 因为网络的安全主要针对内网即局域网内部的访问外部互联网络的安全性，对内部与内部之间的互相访问可以不进行过滤。因此需要从当前访问数据中筛选出局域网络访问外部网络的访问流量。访问流量的刷选，可基于内置内网地址库进行，例如比如IPV4为：10.0.0./8,172.16.0.0/12,192.168.0.0/16，IPV6为本地用IPV6单播地址(包括链路本地单播地址和站点本地单播地址)，和手工配置的内网地址信息。根据访问流量中目的IP地址和源IP地址的归属即可得到内网对外网的访问流量。

[0052] 103、获取访问流量中的超文本传输协议流量，并将从超文本传输协议流量中提取出的超文本传输请求的HOST字段和目的IP地址，与域名白名单和超文本传输协议白名单分别进行碰撞，得到不符合域名白名单和超文本传输协议白名单的可疑流量。

[0053] 在超文本传输协议中以HOST+目的IP组合去碰撞域名白名单，命中此白名单的流量直接丢弃。

[0054] 以HOST+目的IP去碰撞超文本传输协议白名单，命中此白名单的流量直接丢弃。

[0055] 104、获取访问流量中的超文本安全传输协议流量，并将从超文本安全传输协议流量中提取出的超文本安全传输请求的服务器名称和目的IP地址，与域名白名单和超文本安全传输协议白名单分别进行碰撞，得到不符合域名白名单和超文本安全传输协议白名单的可疑流量。

[0056] 在超文本安全传输协议的访问流量中以SNI(服务器名称)+目的IP去碰撞域名白名单，命中此白名单的流量直接丢弃。

[0057] 以SNI+目的IP去碰撞超文本安全传输协议白名单，命中此白名单的流量直接丢弃。

[0058] 在得到访问流量后，以目的IP地址和域名相结合的方式对访问流量进行碰撞，对不符合上述域名白名单和超文本传输协议白名单中的HTTP流量则进入到后续可疑流量监测模块中，以及对不符合域名白名单和超文本安全传输协议白名单的可疑流量进入到后续可疑流量监测模块中，进行进一步的隔离确认处理。使得HTTP协议和HTTPS协议分别有独立的白名单，互不干扰。由于大部分网站已经切换到HTTPS加密协议，目前HTTP、HTTPS协议下的正常流量已经有很大的区别，因此将HTTP协议和HTTPS协议白名单区分开，过滤更加精细化。并且在白名单的过滤机制中，不是单纯IP或域名过滤，而是使用域名以及目的IP地址为对象进行过滤，有效避免了伪装HTTP HOST或者Domain Borrowing等类型的恶意流量绕过流量预过滤系统，提高了可疑流量识别的精准性。

[0059] 作为上述实施例可行的实现方式，参照图2所示，域名白名单的构建过程可以包括以下步骤：

[0060] 201、获取历史访问数据中域名访问请求流量。

[0061] 202、对预设时间段内包含相同域名的域名访问请求进行去重，得到去重后的域名访问请求中的源IP地址的个数。

[0062] 203、将去重后的域名访问请求中的源IP地址的个数与第一稀有度阈值进行比较，得到满足第一稀有度阈值的域名访问请求的域名，并存入域名白名单中。

[0063] 具体的，记录网络中DNS协议流量的DNS请求域名、目的IP地址和发起DNS请求的源IP，形成DNS域名+目的IP+源IP记录并入库。统计一小时内每个DNS域名的去重后源IP的个数，以源IP的个数为对象，设置一定的稀有度阈值(一般可大于90)，计算出符合不稀有条件的DNS域名，并将其加入到域名白名单中。

[0064] 参照图3所示，超文本传输协议白名单的构建过程可以包括以下步骤：

[0065] 301、获取历史访问数据中超文本传输请求流量.

[0066] 302、对预设时间段内包含相同HOST字段和目的IP地址的超文本传输请求进行去重，得到去重后的超文本传输请求中的源IP地址的个数。

[0067] 303、将去重后的超文本传输请求中的源IP地址的个数和第二稀有度阈值进行比较，得到满足第二稀有度阈值的超文本传输请求的HOST字段和目的IP地址，并存入超文本传输协议白名单中。

[0068] 具体的，记录网络中GET、POST方法的HTTP协议流量的HOST值、对应的目的IP地址和发起HTTP请求的源IP地址，形成HOST+目的IP+源IP记录并入库。统计一小时内基于HOST+目的IP对的去重后源IP的个数，以源IP地址的个数为对象，设置第二稀有度阈值(大于90)，计算出符合不稀有条件的HOST+目的IP，并将其加入到超文本传输协议白名单(其中以访问

IP个数为对象计算不稀有的原因是基于一个有限的网络内HTTP请求中访问越多的域名越安全,以HOST+目的IP为限制条件的原因是为了防止某些伪造HOST的情况)。

[0069] 参照图4所示,超文本安全传输协议白名单的构建过程可以包括以下步骤:

[0070] 401、获取历史访问数据中超文本安全传输请求流量。

[0071] 402、将预设时间段内包含相同服务器名称和目的IP地址的超文本安全传输请求进行去重,得到去重后的超文本安全传输请求中的源IP地址的个数。

[0072] 403、将去重后的超文本安全传输请求中的源IP地址的个数与第三稀有度阈值进行比较,得到满足第三稀有度阈值的超文本安全传输请求的服务器名称和目的IP地址,并存入超文本安全传输协议白名单中。

[0073] 通过记录网络中的HTTPS协议SNI(服务器名称)以及对应的目的IP地址和发起HTTPS请求的源IP地址,形成SNI+目的IP地址+源IP地址记录并入库。统计一小时内每个SNI对应的源IP地址个数,以源IP地址的个数为对象,设置一定的稀有度阈值(大于90),计算出符合不稀有条件的SNI+目的IP地址,并将其加入到超文本安全传输协议白名单中。

[0074] 可以理解的是,上述第一稀有度阈值、第二稀有度阈值和第三稀有度阈值为相同的数值,也可为不同的数值,本发明在此不做限制。

[0075] 在本发明的一些具体实施例中,参照图5所示从当前访问数据中筛选出局域网络访问外部网络的访问流量,可以包括以下步骤:

[0076] 501、获取局域网络内的地址列表,地址列表中存储有局域网络内设备的IP地址。

[0077] 502、获取当前访问数据中的源IP地址和目的IP地址,若当前访问数据中的源IP地址存在于地址列表中且目的IP地址不存在地址列表中,则判定为局域网络访问外部网络的访问流量。

[0078] 具体的,提取流量数据中的源IP地址和目的IP地址,判断如果源IP地址是在内网地址列表中,而目的IP地址是非内网地址列表中的地址,则判定为是内网主机访问公网的流量。然后根据接收到的所有流量,如果不满足上面的内网访问公网的规则,则判定为不是此检测模型需要检测的数据,则直接丢弃。而满足上面内网访问公网的流量,则保留。

[0079] 其中在对网络访问白名单进行刷新后,基于更新后的网络访问白名单对正在使用中的网络访问白名单进行全部替换。即白名单刷新机制采用全量替换的方案,新的白名单全量替换正在进行流量过滤的老白名单列表,解决网络链路发生变化之后,同一个域名解析后的地址发生变化,不再有流量匹配的情况。

[0080] 基于同样的设计思路参照图6所示,本发明的实施例还提供了一种可疑流量识别装置,该装置可以执行上述实施例所述的可疑流量识别方法的各个步骤,该装置可以包括:

[0081] 白名单模块601,用于基于历史访问数据构建预设时间段的网络访问白名单,网络访问白名单至少包括:域名白名单、超文本传输协议白名单和超文本安全传输协议白名单,域名白名单中存储有满足第一稀有度阈值的域名访问请求的域名,超文本传输协议白名单中存储有满足第二稀有度阈值的超文本传输请求的HOST字段和目的IP地址,超文本安全传输协议白名单中存储有满足第三稀有度阈值的超文本安全传输请求的服务器名称和目的IP地址。

[0082] 访问流量模块602,用于从当前访问数据中筛选出局域网络访问外部网络的访问流量。

[0083] 第一识别模块603，用于获取访问流量中的超文本传输协议流量，并将从超文本传输协议流量中提取出的超文本传输请求的HOST字段和目的IP地址，与域名白名单和超文本传输协议白名单分别进行碰撞，得到不符合域名白名单和超文本传输协议白名单的可疑流量。以及

[0084] 第二识别模块604，用于获取访问流量中的超文本安全传输协议流量，并将从超文本安全传输协议流量中提取出的超文本安全传输请求的服务器名称和目的IP地址，与域名白名单和超文本安全传输协议白名单分别进行碰撞，得到不符合域名白名单和超文本安全传输协议白名单的可疑流量。

[0085] 该装置具有和上述可疑流量识别方法相同的有益效果，本发明在此不再赘述。

[0086] 参照图7所示，本发明的实施例还提供了一种设备，该设备包括：存储器701和处理器702；

[0087] 存储器701，用于存储程序；

[0088] 处理器702，用于执行程序，实现如上所述的可疑流量识别方法的各个步骤。

[0089] 本发明的实施例还提供了一种存储介质，其上存储有计算机程序，该计算机程序被处理器执行时，实现如上所述的可疑流量识别方法的各个步骤。

[0090] 本发明上述实施例所提供的可疑流量识别方法、装置、设备和存储介质，在白名单过滤机制中，不是单纯IP或域名过滤，而是使用域名以及返回(对应的)IP地址为对象进行过滤，有效避免了伪装HTTP HOST或者Domain Borrowing等类型的恶意流量绕过流量预过滤系统。

[0091] 对于前述的各方法实施例，为了简单描述，故将其都表述为一系列的动作组合，但是本领域技术人员应该知悉，本发明并不受所描述的动作顺序的限制，因为依据本发明，某些步骤可以采用其他顺序或者同时进行。其次，本领域技术人员也应该知悉，说明书中所描述的实施例均属于优选实施例，所涉及的动作和模块并不一定是本发明所必须的。

[0092] 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，每个实施例重点说明的都是与其他实施例的不同之处，各个实施例之间相同相似的部分互相参见即可。对于装置类实施例而言，由于其与方法实施例基本相似，所以描述的比较简单，相关之处参见方法实施例的部分说明即可。

[0093] 本发明各实施例方法中的步骤可以根据实际需要进行顺序调整、合并和删减，各实施例中记载的技术特征可以进行替换或者组合。

[0094] 本发明各实施例种装置及终端中的模块和子模块可以根据实际需要进行合并、划分和删减。

[0095] 本发明所提供的几个实施例中，应该理解到，所揭露的终端，装置和方法，可以通过其它的方式实现。例如，以上所描述的终端实施例仅仅是示意性的，例如，模块或子模块的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式，例如多个子模块或模块可以结合或者可以集成到另一个模块，或一些特征可以忽略，或不执行。另一点，所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口，装置或模块的间接耦合或通信连接，可以是电性，机械或其它的形式。

[0096] 作为分离部件说明的模块或子模块可以是或者也可以不是物理上分开的，作为模块或子模块的部件可以是或者也可以不是物理模块或子模块，即可以位于一个地方，或者

也可以分布到多个网络模块或子模块上。可以根据实际的需要选择其中的部分或者全部模块或子模块来实现本实施例方案的目的。

[0097] 另外，在本发明各个实施例中的各功能模块或子模块可以集成在一个处理模块中，也可以是各个模块或子模块单独物理存在，也可以两个或两个以上模块或子模块集成在一个模块中。上述集成的模块或子模块既可以采用硬件的形式实现，也可以采用软件功能模块或子模块的形式实现。

[0098] 专业人员还可以进一步意识到，结合本文中所公开的实施例描述的各示例的单元及算法步骤，能够以电子硬件、计算机软件或者二者的结合来实现，为了清楚地说明硬件和软件的可互换性，在上述说明中已经按照功能一般性地描述了各示例的组成及步骤。这些功能究竟以硬件还是软件方式来执行，取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能，但是这种实现不应认为超出本发明的范围。

[0099] 结合本文中所公开的实施例描述的方法或算法的步骤可以直接用硬件、处理器执行的软件单元，或者二者的结合来实施。软件单元可以置于随机存储器(RAM)、内存、只读存储器(ROM)、电可编程ROM、电可擦除可编程ROM、寄存器、硬盘、可移动磁盘、CD-ROM、或技术领域内所公知的任意其它形式的存储介质中。

[0100] 最后，还需要说明的是，在本文中，诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来，而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、物品或者设备所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括所述要素的过程、方法、物品或者设备中还存在另外的相同要素。

[0101] 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发明。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发明的精神或范围的情况下，在其它实施例中实现。因此，本发明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。

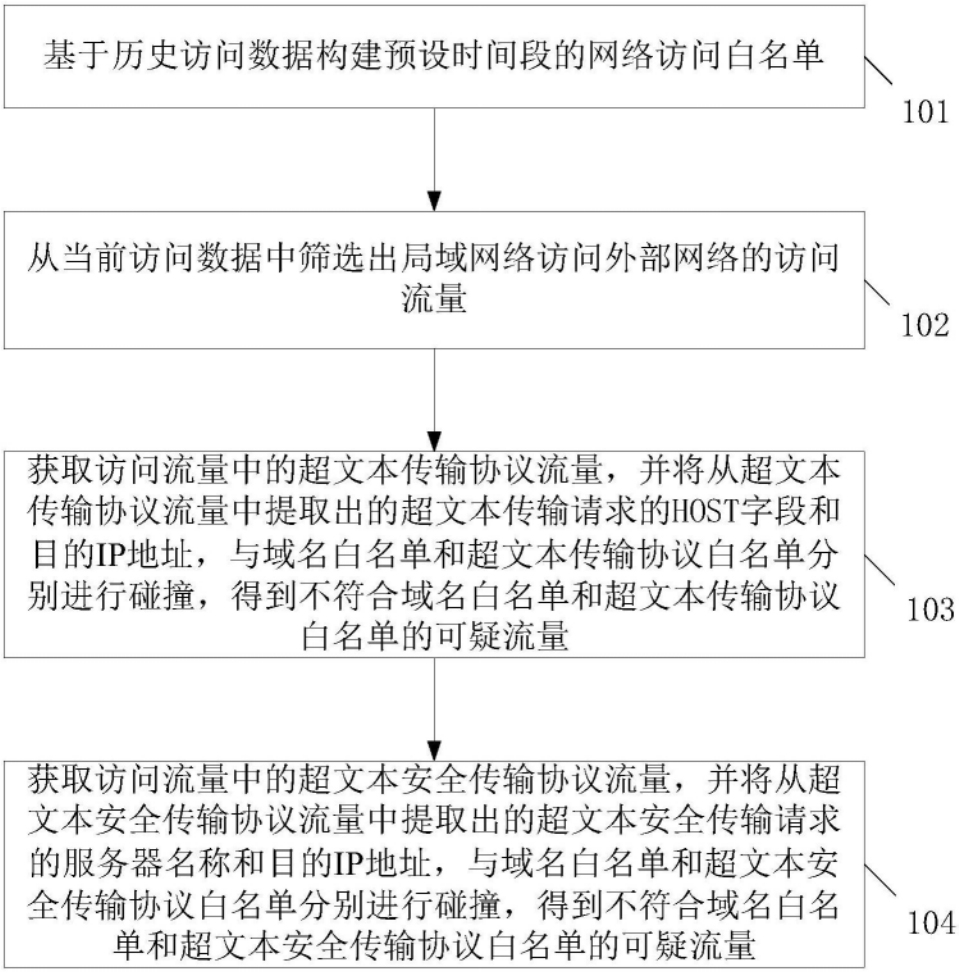

基于历史访问数据构建预设时间段的网络访问白名单
101
从当前访问数据中筛选出局域网络访问外部网络的访问流量
102
获取访问流量中的超文本传输协议流量，并将从超文本传输协议流量中提取出的超文本传输请求的HOST字段和目的IP地址，与域名白名单和超文本传输协议白名单分别进行碰撞，得到不符合域名白名单和超文本传输协议白名单的可疑流量
103
获取访问流量中的超文本安全传输协议流量，并将从超文本安全传输协议流量中提取出的超文本安全传输请求的服务器名称和目的IP地址，与域名白名单和超文本安全传输协议白名单分别进行碰撞，得到不符合域名白名单和超文本安全传输协议白名单的可疑流量
104

图1

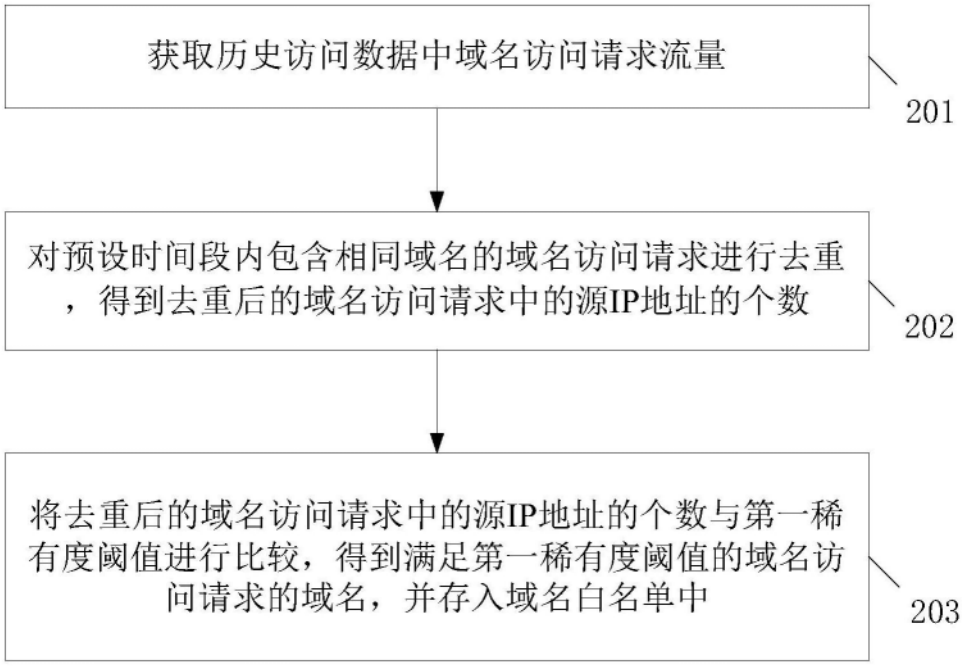

获取历史访问数据中域名访问请求流量
201
对预设时间段内包含相同域名的域名访问请求进行去重，得到去重后的域名访问请求中的源IP地址的个数
202
将去重后的域名访问请求中的源IP地址的个数与第一稀有度阈值进行比较，得到满足第一稀有度阈值的域名访问请求的域名，并存入域名白名单中
203

图2

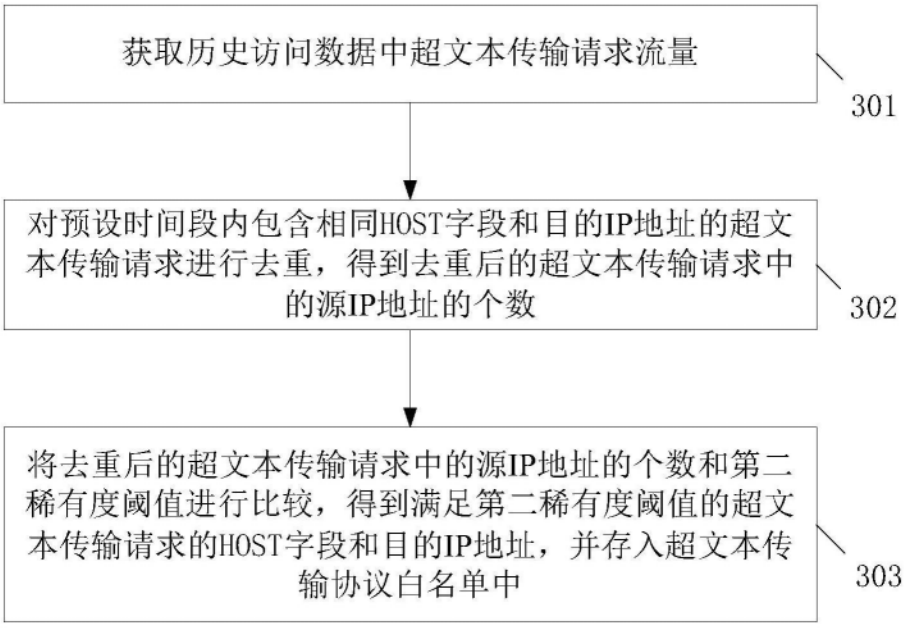

获取历史访问数据中超文本传输请求流量
301
对预设时间段内包含相同HOST字段和目的IP地址的超文本传输请求进行去重，得到去重后的超文本传输请求中的源IP地址的个数
302
将去重后的超文本传输请求中的源IP地址的个数和第二稀有度阈值进行比较，得到满足第二稀有度阈值的超文本传输请求的HOST字段和目的IP地址，并存入超文本传输协议白名单中
303

图3

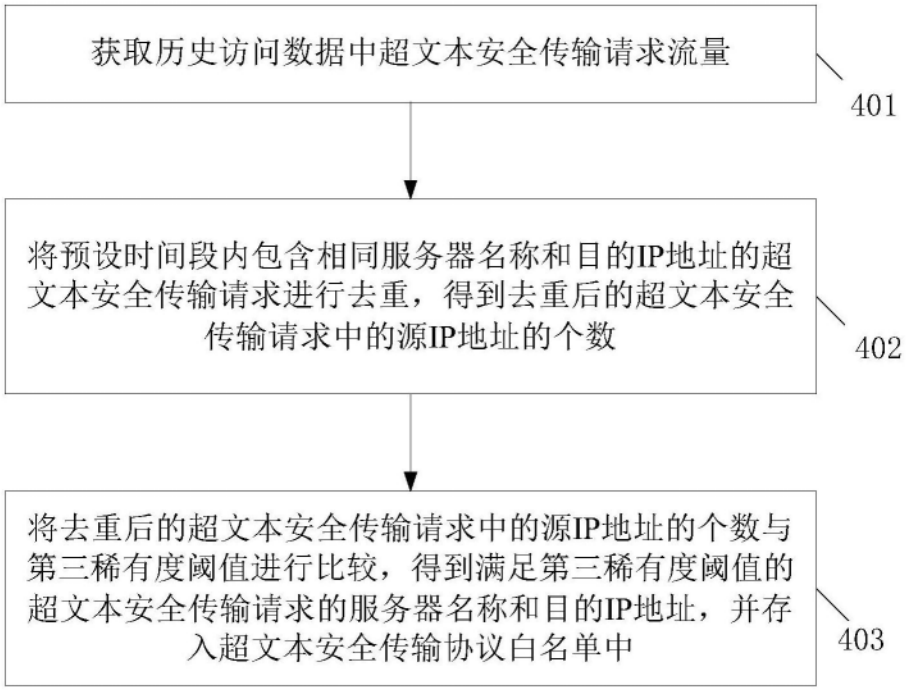

获取历史访问数据中超文本安全传输请求流量
401
将预设时间段内包含相同服务器名称和目的IP地址的超文本安全传输请求进行去重，得到去重后的超文本安全传输请求中的源IP地址的个数
402
将去重后的超文本安全传输请求中的源IP地址的个数与第三稀有度阈值进行比较，得到满足第三稀有度阈值的超文本安全传输请求的服务器名称和目的IP地址，并存入超文本安全传输协议白名单中
403

图4

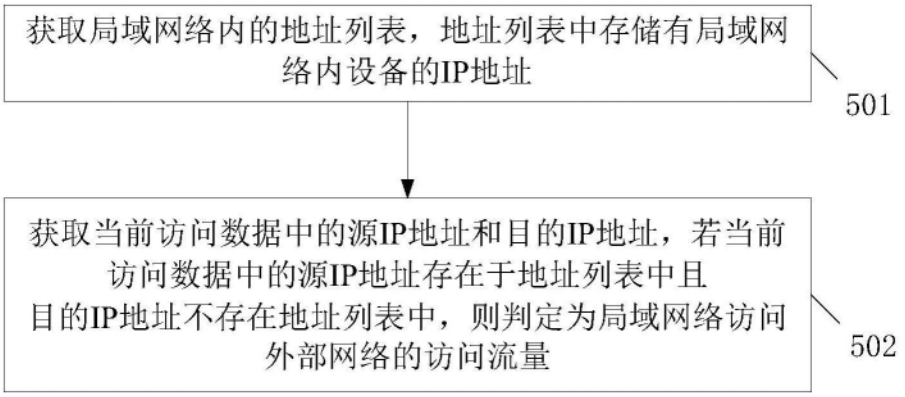

获取局域网络内的地址列表，地址列表中存储有局域网络内设备的IP地址
501
获取当前访问数据中的源IP地址和目的IP地址，若当前访问数据中的源IP地址存在于地址列表中且
目的IP地址不存在地址列表中，则判定为局域网络访问外部网络的访问流量
502

图5

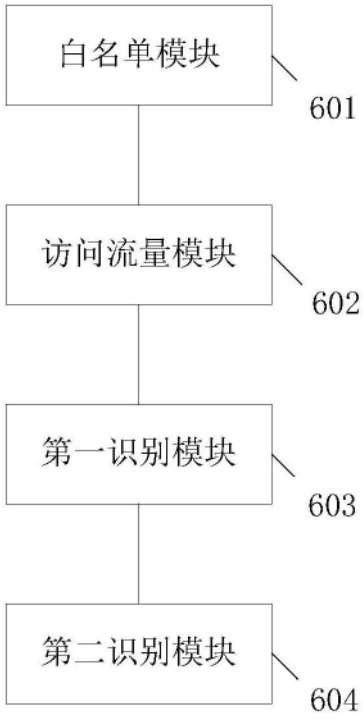


图6

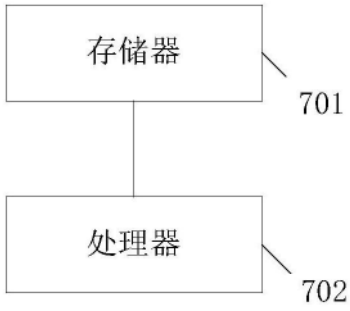


图7