(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 109657459 A
(43)申请公布日 2019.04.19

(21)申请号 201811188296.9

(22)申请日 2018.10.11

(71)申请人 平安科技(深圳)有限公司
地址 518000 广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼

(72)发明人 李坤

(74)专利代理机构 深圳市世纪恒程知识产权代理事务所 44287
代理人 胡海国

(51)Int.Cl.
G06F 21/55(2013.01)
G06F 21/56(2013.01)
G06K 9/62(2006.01)

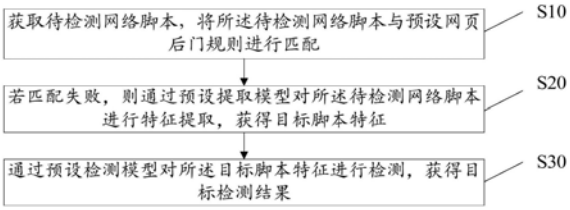
权利要求书2页 说明书11页 附图4页

(54)发明名称

网页后门检测方法、设备、存储介质及装置

(57)摘要

本发明公开了一种网页后门检测方法、设备、存储介质及装置，该方法包括：获取待检测网络脚本，将所述待检测网络脚本与预设网页后门规则进行匹配；若匹配失败，则通过预设提取模型对所述待检测网络脚本进行特征提取，获得目标脚本特征；通过预设检测模型对所述目标脚本特征进行检测，获得目标检测结果。本发明中，通过将规则检测和基于机器学习的模型检测相结合，提高系统检测所述待检测网络脚本是否为网页后门的准确性。



CN 109657459 A

1.一种网页后门检测方法,其特征在于,所述网页后门检测方法包括以下步骤:

获取待检测网络脚本,将所述待检测网络脚本与预设网页后门规则进行匹配;

若匹配失败,则通过预设提取模型对所述待检测网络脚本进行特征提取,获得目标脚本特征;

通过预设检测模型对所述目标脚本特征进行检测,获得目标检测结果。

2.如权利要求1所述的网页后门检测方法,其特征在于,所述通过预设检测模型对所述目标脚本特征进行检测,获得目标检测结果之后,所述网页后门检测方法还包括:

若所述目标检测结果为所述目标脚本特征是网页后门对应的特征,则根据所述目标脚本特征与对应的所述目标检测结果对所述预设检测模型进行训练。

3.如权利要求2所述的网页后门检测方法,其特征在于,所述若匹配失败,则通过预设提取模型对所述待检测网络脚本进行特征提取,获得目标脚本特征,包括:

若匹配失败,则对所述待检测网络脚本进行数据清洗,获得目标网络脚本;

通过所述预设提取模型对所述目标网络脚本进行特征提取,获得目标脚本特征。

4.如权利要求1-3中任一项所述的网页后门检测方法,其特征在于,所述获取待检测网络脚本,将所述待检测网络脚本与预设网页后门规则进行匹配,包括:

通过网关获取待检测网络脚本,对所述待检测网络脚本进行分析,提取出多个预设维度的特征;

将所述预设维度的特征与预设网页后门规则进行匹配。

5.如权利要求4所述的网页后门检测方法,其特征在于,所述通过预设检测模型对所述目标脚本特征进行检测,获得目标检测结果之前,所述网页后门检测方法还包括:

建立基础预测模型;

获取样本网络脚本及对应的样本检测结果;

将所述样本网络脚本通过所述预设提取模型进行特征提取,获得第一脚本特征;

根据所述第一脚本特征及对应的所述样本检测结果对所述基础预测模型进行训练,获得预设检测模型。

6.如权利要求5所述的网页后门检测方法,其特征在于,所述通过预设检测模型对所述目标脚本特征进行检测,获得目标检测结果之前,所述网页后门检测方法还包括:

获取第一数量的样本网页后门,将所述样本网页后门通过所述预设提取模型进行特征提取,获得第二脚本特征;

通过所述预设检测模型对所述第二脚本特征进行检测,获得评估检测结果,所述评估检测结果包括所述第二脚本特征是网页后门对应的特征的第一检测结果;

统计所述第一检测结果的第二数量,根据所述第一数量和所述第二数量计算所述预设检测模型的准确率;

在所述准确率超过预设阈值时,执行所述通过预设检测模型对所述目标脚本特征进行检测,获得目标检测结果的步骤。

7.如权利要求6所述的网页后门检测方法,其特征在于,所述评估检测结果包括所述第二脚本特征不是网页后门对应的特征的第二检测结果;

所述统计所述第一检测结果的第二数量,根据所述第一数量和所述第二数量计算所述预设检测模型的准确率之后,所述网页后门检测方法还包括:

获取所述第二检测结果对应的第二脚本特征作为误测网页后门特征；

设置所述误测网页后门特征的真实检测结果为所述误测网页后门特征是网页后门对应的特征，根据所述误测网页后门特征与对应的真实检测结果对所述预设检测模型进行训练。

8.一种网页后门检测设备，其特征在于，所述网页后门检测设备包括：存储器、处理器及存储在所述存储器上并可在所述处理器上运行的网页后门检测程序，所述网页后门检测程序被所述处理器执行时实现如权利要求1至7中任一项所述的网页后门检测方法的步骤。

9.一种存储介质，其特征在于，所述存储介质上存储有网页后门检测程序，所述网页后门检测程序被处理器执行时实现如权利要求1至7中任一项所述的网页后门检测方法的步骤。

10.一种网页后门检测装置，其特征在于，所述网页后门检测装置包括：

匹配模块，用于获取待检测网络脚本，将所述待检测网络脚本与预设网页后门规则进行匹配；

提取模块，用于若匹配失败，则通过预设提取模型对所述待检测网络脚本进行特征提取，获得目标脚本特征；

检测模块，用于通过预设检测模型对所述目标脚本特征进行检测，获得目标检测结果。

网页后门检测方法、设备、存储介质及装置

技术领域

[0001] 本发明涉及监控技术领域，尤其涉及一种网页后门检测方法、设备、存储介质及装置。

背景技术

[0002] 目前，网页后门(webshell)就是以动态服务器页面(Active Server Pages，ASP)、超文本预处理器(Hypertext Preprocessor，PHP)、java服务器页面(Java Server Pages，JSP)或者通用网关接口(Common Gateway Interface，CGI)等网页文件形式存在的一种命令执行环境。黑客在入侵了一个网站后，通常会将ASP或PHP后门文件与网站服务器WEB目录下正常的网页文件混在一起，然后就可以使用浏览器来访问ASP或PHP后门，得到一个命令执行环境，以达到控制网站服务器的目的。

[0003] 网页后门通常包含较为明显的静态特征，目前，根据所述静态特征对网页脚本进行检测，以检测所述网页脚本是否为网页后门，往往会产生很多的误报，因此，如何提高对网页后门检测的准确率是亟待解决的技术问题。

[0004] 上述内容仅用于辅助理解本发明的技术方案，并不代表承认上述内容是现有技术。

发明内容

[0005] 本发明的主要目的在于提供一种网页后门检测方法、设备、存储介质及装置，旨在解决现有技术中网页后门的检测误报率高的技术问题。

[0006] 为实现上述目的，本发明提供一种网页后门检测方法，所述网页后门检测方法包括以下步骤：

[0007] 获取待检测网络脚本，将所述待检测网络脚本与预设网页后门规则进行匹配；

[0008] 若匹配失败，则通过预设提取模型对所述待检测网络脚本进行特征提取，获得目标脚本特征；

[0009] 通过预设检测模型对所述目标脚本特征进行检测，获得目标检测结果。

[0010] 优选地，所述通过预设检测模型对所述目标脚本特征进行检测，获得目标检测结果之后，所述网页后门检测方法还包括：

[0011] 若所述目标检测结果为所述目标脚本特征是网页后门对应的特征，则根据所述目标脚本特征与对应的所述目标检测结果对所述预设检测模型进行训练。

[0012] 优选地，所述若匹配失败，则通过预设提取模型对所述待检测网络脚本进行特征提取，获得目标脚本特征，包括：

[0013] 若匹配失败，则对所述待检测网络脚本进行数据清洗，获得目标网络脚本；

[0014] 通过所述预设提取模型对所述目标网络脚本进行特征提取，获得目标脚本特征。

[0015] 优选地，所述获取待检测网络脚本，将所述待检测网络脚本与预设网页后门规则进行匹配，包括：

[0016] 通过网关获取待检测网络脚本，对所述待检测网络脚本进行分析，提取出多个预设维度的特征；

[0017] 将所述预设维度的特征与预设网页后门规则进行匹配。

[0018] 优选地，所述通过预设检测模型对所述目标脚本特征进行检测，获得目标检测结果之前，所述网页后门检测方法还包括：

[0019] 建立基础预测模型；

[0020] 获取样本网络脚本及对应的样本检测结果；

[0021] 将所述样本网络脚本通过所述预设提取模型进行特征提取，获得第一脚本特征；

[0022] 根据所述第一脚本特征及对应的所述样本检测结果对所述基础预测模型进行训练，获得预设检测模型。

[0023] 优选地，所述通过预设检测模型对所述目标脚本特征进行检测，获得目标检测结果之前，所述网页后门检测方法还包括：

[0024] 获取第一数量的样本网页后门，将所述样本网页后门通过所述预设提取模型进行特征提取，获得第二脚本特征；

[0025] 通过所述预设检测模型对所述第二脚本特征进行检测，获得评估检测结果，所述评估检测结果包括所述第二脚本特征是网页后门对应的特征的第一检测结果；

[0026] 统计所述第一检测结果的第二数量，根据所述第一数量和所述第二数量计算所述预设检测模型的准确率；

[0027] 在所述准确率超过预设阈值时，执行所述通过预设检测模型对所述目标脚本特征进行检测，获得目标检测结果的步骤。

[0028] 优选地，所述评估检测结果包括所述第二脚本特征不是网页后门对应的特征的第二检测结果；

[0029] 所述统计所述第一检测结果的第二数量，根据所述第一数量和所述第二数量计算所述预设检测模型的准确率之后，所述网页后门检测方法还包括：

[0030] 获取所述第二检测结果对应的第二脚本特征作为误测网页后门特征；

[0031] 设置所述误测网页后门特征的真实检测结果为所述误测网页后门特征是网页后门对应的特征，根据所述误测网页后门特征与对应的真实检测结果对所述预设检测模型进行训练。

[0032] 此外，为实现上述目的，本发明还提出一种网页后门检测设备，所述网页后门检测设备包括存储器、处理器及存储在所述存储器上并可在所述处理器上运行的网页后门检测程序，所述网页后门检测程序配置为实现如上文所述的网页后门检测方法的步骤。

[0033] 此外，为实现上述目的，本发明还提出一种存储介质，所述存储介质上存储有网页后门检测程序，所述网页后门检测程序被处理器执行时实现如上文所述的网页后门检测方法的步骤。

[0034] 此外，为实现上述目的，本发明还提出一种网页后门检测装置，所述网页后门检测装置包括：

[0035] 匹配模块，用于获取待检测网络脚本，将所述待检测网络脚本与预设网页后门规则进行匹配；

[0036] 提取模块，用于若匹配失败，则通过预设提取模型对所述待检测网络脚本进行特

征提取,获得目标脚本特征;

[0037] 检测模块,用于通过预设检测模型对所述目标脚本特征进行检测,获得目标检测结果。

[0038] 本发明中,获取待检测网络脚本,将所述待检测网络脚本与预设网页后门规则进行匹配,通过基于规则的匹配对所述待检测网络脚本进行检测,对于特征明显的网页后门能够被检测出来;若匹配失败,则通过预设提取模型对所述待检测网络脚本进行特征提取,获得目标脚本特征,通过预设检测模型对所述目标脚本特征进行检测,获得目标检测结果,通过将规则检测和基于机器学习的模型检测相结合,未能通过规则匹配检测出的网页后门,可通过基于机器学习的模型进一步进行检测,所述预设检测模型经过大量的样本学习和检测准确率的评估,具有较好的检测效果,从而提高系统检测网络脚本是否为网页后门的准确性。

附图说明

[0039] 图1是本发明实施例方案涉及的硬件运行环境的网页后门检测设备的结构示意图;

[0040] 图2为本发明网页后门检测方法第一实施例的流程示意图;

[0041] 图3为本发明网页后门检测方法第二实施例的流程示意图;

[0042] 图4为本发明网页后门检测方法第三实施例的流程示意图;

[0043] 图5为本发明网页后门检测装置第一实施例的结构框图。

[0044] 本发明目的的实现、功能特点及优点将结合实施例,参照附图做进一步说明。

具体实施方式

[0045] 应当理解,此处所描述的具体实施例仅仅用以解释本发明,并不用于限定本发明。

[0046] 参照图1,图1为本发明实施例方案涉及的硬件运行环境的网页后门检测设备结构示意图。

[0047] 如图1所示,该网页后门检测设备可以包括:处理器1001,例如中央处理器(Central Processing Unit,CPU),通信总线1002、用户接口1003,网络接口1004,存储器1005。其中,通信总线1002用于实现这些组件之间的连接通信。用户接口1003可以包括显示屏(Display),可选用户接口1003还可以包括标准的有线接口、无线接口,对于用户接口1003的有线接口在本发明中可为USB接口。网络接口1004可选的可以包括标准的有线接口、无线接口(如无线保真(WIreless-FIdelity,WI-FI)接口)。存储器1005可以是高速的随机存取存储器(Random Access Memory,RAM)存储器,也可以是稳定的存储器(Non-volatile Memory,NVM),例如磁盘存储器。存储器1005可选的还可以是独立于前述处理器1001的存储装置。

[0048] 本领域技术人员可以理解,图1中示出的结构并不构成对网页后门检测设备的限定,可以包括比图示更多或更少的部件,或者组合某些部件,或者不同的部件布置。

[0049] 如图1所示,认定为一种计算机存储介质的存储器1005中可以包括操作系统、网络通信模块、用户接口模块以及网页后门检测程序。

[0050] 在图1所示的网页后门检测设备中,网络接口1004主要用于连接后台服务器,与所

述后台服务器进行数据通信;用户接口1003主要用于连接用户设备;所述网页后门检测设备通过处理器1001调用存储器1005中存储的网页后门检测程序,并执行本发明实施例提供的网页后门检测方法。

[0051] 基于上述硬件结构,提出本发明网页后门检测方法的实施例。

[0052] 参照图2,图2为本发明网页后门检测方法第一实施例的流程示意图,提出本发明网页后门检测方法第一实施例。

[0053] 在第一实施例中,所述网页后门检测方法包括以下步骤:

[0054] 步骤S10:获取待检测网络脚本,将所述待检测网络脚本与预设网页后门规则进行匹配。

[0055] 应理解的是,本实施例的执行主体是所述网页后门检测设备,其中,所述网页后门检测设备可为个人电脑或服务器等电子设备。所述预设网页后门(webshell)规则可以是恶意字符串库,例如,包括:“组专用大马|提权|木马|PHP\s?反弹提权cmd执行”和“WScript.Shell、Shell.Application、Eval()、Excute()、Set Server、Run()、Exec()以及ShellExcute()”等恶意字符串。将所述待检测网络脚本进行特征提取,是指对所述待检测脚本中所使用的关键词、高危函数、文件修改的时间、文件权限、文件的所有者以及和其它文件的关联性等多个维度的特征进行提取,从而获得所述脚本特征,将获得的脚本特征与所述预设webshell规则库进行匹配,获得匹配结果。若所述匹配结果为匹配成功,则说明所述待检测网络脚本为webshell;若所述匹配结果为匹配失败,则说明所述待检测网络脚本不是webshell,可能为正常的网络脚本,或者是检测失误的webshell。

[0056] 在本实施例中,所述步骤S10,包括:

[0057] 通过网关获取待检测网络脚本,对所述待检测网络脚本进行分析,提取出多个预设维度的特征;

[0058] 将所述预设维度的特征与预设网页后门规则进行匹配。

[0059] 需要说明的是,所述网关(Gateway)从代理服务器(Agent)获取所述待检测网络脚本,所述待检测网络脚本通常为多个,也可以是一个。对所述待检测网络脚本进行分析,通常是将所述待检测网络脚本拆分成字符串,从所述待检测网络脚本对应的所有字符串中提取出多个预设维度的特征,所述多个预设维度包括:关键词、高危函数、文件修改的时间、文件权限、文件的所有者以及和其它文件的关联性等。正常的网络脚本不会包含所述预设网页后门规则中的特征,则将所述预设维度的特征与所述预设网页后门规则进行匹配,从而识别出所述待检测网络脚本是否为网页后门,或者是正常的网络脚本。

[0060] 步骤S20:若匹配失败,则通过预设提取模型对所述待检测网络脚本进行特征提取,获得目标脚本特征。

[0061] 可理解的是,若匹配失败,说明所述待检测网络脚本不是webshell,可能为正常的网络脚本,或者是检测失误的webshell。为了进一步识别所述待检测脚本是否为webshell,可通过所述预设提取模型进行特征提取,所述预设提取模型包括卷积神经网络模型等。可预先建立基础提取模型,获取样本网络脚本与对应的特征对所述基础提取模型进行训练,获得所述预设提取模型。通过所述预设提取模型进行特征提取,获得合适的所述目标脚本特征。

[0062] 步骤S30:通过预设检测模型对所述目标脚本特征进行检测,获得目标检测结果。

[0063] 在具体实现中，所述预设检测模型包括神经网络模型，经过丰富的大量的训练样本的训练，保证所述预设检测模型对所述目标脚本特征检测的准确性。所述目标检测结果可以是所述目标脚本特征为网页后门对应的特征，即所述目标脚本特征对应的所述待检测网络脚本为网页后门；所述目标检测结果还可以是所述目标脚本特征为正常网络脚本对应的特征，即所述目标脚本特征对应的所述待检测网络脚本为正常的网络脚本。

[0064] 应理解的是，对于所述预设检测模型的建立过程，首先建立基础预测模型，从数据库中获取大量的样本网络脚本及对应的样本检测结果，所述样本网络脚本包括大量的正常的网页脚本和大量的网页后门，可将所述样本网络脚本进行数据清洗，将数据清洗后的样本网络脚本通过所述预设提取模型进行特征提取，获得所述样本网络脚本对应的第一脚本特征，则可根据大量的所述第一脚本特征及对应的所述样本检测结果对所述基础预测模型进行训练，获得所述预设检测模型。所述数据清洗包括对所述样本网络脚本中的无关数据、重复数据和平滑噪声数据，处理所述样本网络脚本中缺失值和异常值。本实施例中，所述步骤S30之前，还包括：建立基础预测模型；获取样本网络脚本及对应的样本检测结果；将所述样本网络脚本通过所述预设提取模型进行特征提取，获得第一脚本特征；根据所述第一脚本特征及对应的所述样本检测结果对所述基础预测模型进行训练，获得预设检测模型。

[0065] 本实施例中，获取待检测网络脚本，将所述待检测网络脚本与预设网页后门规则进行匹配，通过基于规则的匹配对所述待检测网络脚本进行检测，对于特征明显的网页后门能够被检测出来；若匹配失败，则通过预设提取模型对所述待检测网络脚本进行特征提取，获得目标脚本特征，通过预设检测模型对所述目标脚本特征进行检测，获得目标检测结果，通过将规则检测和基于机器学习的模型检测相结合，未能通过规则匹配检测出的网页后门，可通过基于机器学习的模型进一步进行检测，所述预设检测模型经过大量的样本学习和检测准确率的评估，具有较好的检测效果，从而提高系统检测网络脚本是否为网页后门的准确性。

[0066] 参照图3，图3为本发明网页后门检测方法第二实施例的流程示意图，基于上述图2所示的第一实施例，提出本发明网页后门检测方法的第二实施例。

[0067] 在第二实施例中，所述步骤S30之后，还包括：

[0068] 步骤S40：若所述目标检测结果为所述目标脚本特征是网页后门对应的特征，则根据所述目标脚本特征与对应的所述目标检测结果对所述预设检测模型进行训练。

[0069] 可理解的是，若所述目标检测结果为所述目标脚本特征是网页后门对应的特征，说明所述目标脚本特征对应的待检测网络脚本是网页后门，可将所述待检测网络脚本与对应的所述目标检测结果存入数据库中，后续可作为在线训练或离线训练的样本数据，还可根据所述目标脚本特征与对应的所述目标检测结果对所述预设检测模型进行训练，以提高所述预设检测模型的训练量，从而提高所述预设检测模型进行检测的准确率。

[0070] 在本实施例中，所述步骤S20，包括：

[0071] 步骤S201：若匹配失败，则对所述待检测网络脚本进行数据清洗，获得目标网络脚本。

[0072] 应理解的是，若将所述待检测脚本与预设网页后门规则进行匹配的结果为匹配失败，则说明所述待检测网络脚本不是webshell。需要将所述待检测网络脚本发送至所述预设高吞吐量的分布式发布订阅消息系统(Kafka)，所述预设高吞吐量的分布式发布订阅消

息系统作为消息队列，可以缓存数据，也可以对数据分流，通常将所述待检测网络脚本进行复制分发。所述预设高吞吐量的分布式发布订阅消息系统首先将所述待检测脚本复制一份存入Hadoop，Hadoop是一个由Apache基金会所开发的分布式系统基础架构，Hadoop将所述待检测网络脚本用于离线学习以及回溯事件的时候使用。所述预设高吞吐量的分布式发布订阅消息系统还复制一份所述待检测脚本进行在线学习，发送至所述网页后门检测设备。在线学习和离线学习都需要将所述待检测网络脚本经过数据清洗和特征提取两个流程，再对所述预设检测模型进行训练。

[0073] 需要说明的是，所述网页后门检测设备接收所述预设高吞吐量的分布式发布订阅消息系统发送的待检测网络脚本，进行在线学习，通常需先将所述待检测网络脚本进行数据清洗，数据清洗主要是负责将不符合规则的数据过滤掉，对敏感的数据脱敏和对数据进行格式化，以方便特征提取。例如，删除所述待检测网络脚本中的无关数据、重复数据和平滑噪声数据，处理所述待检测网络脚本中缺失值和异常值。可通过设置哪些字段是允许的，哪些字段是不允许的，构成清洗规则，通过所述清洗规则来过滤掉所述待检测脚本中不符合格式的数据。对所述待检测网络脚本进行数据清洗之后，获得目标网络脚本，将所述目标网络脚本再进行特征提取，获得所述目标脚本特征。

[0074] 步骤S202：通过所述预设提取模型对所述目标网络脚本进行特征提取，获得目标脚本特征。

[0075] 在具体实现中，若匹配失败，说明所述待检测网络脚本不是webshell，可能为正常的网络脚本，或者是检测失误的webshell。为了进一步识别所述待检测脚本是否为webshell，可通过所述预设提取模型对所述目标网络脚本进行特征提取，所述目标网络脚本经过数据清洗，避免处理过多重复数据和不相关的数据，提取出更合适的所述目标脚本特征，提高特征提取的效率和质量。

[0076] 本实施例中，若匹配失败，则对所述待检测网络脚本进行数据清洗，获得目标网络脚本，通过所述预设提取模型对所述目标网络脚本进行特征提取，获得目标脚本特征，经过数据清洗，过滤掉所述待检测网络脚本中的重复数据和无关数据，从而提高特征提取的效率和质量。

[0077] 参照图4，图4为本发明网页后门检测方法第三实施例的流程示意图，基于上述图3所示的第二实施例，提出本发明网页后门检测方法的第三实施例。

[0078] 在第三实施例中，所述步骤S30之前，还包括：

[0079] 步骤S203：获取第一数量的样本网页后门，将所述样本网页后门通过所述预设提取模型进行特征提取，获得第二脚本特征。

[0080] 应理解的是，为了确保所述待检测网络脚本通过所述预设检测模型进行检测的准确率，在通过所述预设检测模型进行检测之前，需要对所述预设检测模型进行准确率或召回率的计算，在所述准确率超过预设阈值(比如80％)时，使用所述预设检测模型进行检测，或者，在所述召回率小于预设召回阈值(比如20％)时，使用所述预设检测模型进行检测。可通过从数据库中获取所述第一数量的样本网页后门，根据所述预设检测模型对所述第一数量的样本网页后门的检测结果来计算其准确率或召回率。由于所述预设检测模型进行检测的是网络脚本对应的特征，则需将所述样本网页后门通过所述预设提取模型进行特征提取，获得所述第二脚本特征。

[0081] 步骤S204:通过所述预设检测模型对所述第二脚本特征进行检测,获得评估检测结果,所述评估检测结果包括所述第二脚本特征是网页后门对应的特征的第一检测结果。

[0082] 需要说明的是,所述第二脚本特征为所述样本网页后门对应的特征,将所述第二脚本特征通过所述预设检测模型,从而检测出所述第二脚本特征是否为网页后门的特征,若所有的所述第二脚本特征都能够被所述预设检测模型成功检测出来,说明所述预设检测模型的准确率为100%。所述评估检测结果中包括第一检测结果和第二检测结果,所述第一检测结果为所述第二脚本特征是网页后门对应的特征,所述第二检测结果为所述第二脚本特征不是网页后门对应的特征。

[0083] 步骤S205:统计所述第一检测结果的第二数量,根据所述第一数量和所述第二数量计算所述预设检测模型的准确率。

[0084] 在具体实现中,通常所述预设检测模型不能完全将所有的第一数量的样本网页后门成功检测出来,则可对所述评估检测结果进行分析,统计出所述评估检测结果为所述第二脚本特征是网页后门对应的特征的所述第一检测结果的第二数量,所述第二数量即所述预设检测模型能够正确检测出的所述第二脚本特征的数量,即能够正确检测出对应的样本网页后门的数量,将所述第二数量除以所述第一数量,即获得所述预设检测模型的准确率。将所述第一数量减去所述第二数量,获得差值数量,将所述差值数量除以所述第一数量,即获得所述预设检测模型的召回率。

[0085] 步骤S206:在所述准确率超过预设阈值时,执行所述步骤S30。

[0086] 可理解的是,在所述准确率超过所述预设阈值时,或者,所述召回率小于预设召回阈值时,说明所述预设检测模型进行检测的准确率较高,值得信赖,则可通过所述预设检测模型对所述待检测网络脚本对应的目标脚本特征进行检测,以检测出所述目标脚本特征是否为网页后门对应的特征。

[0087] 在本实施例中,所述评估检测结果包括所述第二脚本特征不是网页后门对应的特征的第二检测结果;所述步骤S205之后,还包括:

[0088] 获取所述第二检测结果对应的第二脚本特征作为误测网页后门特征;

[0089] 设置所述误测网页后门特征的真实检测结果为所述误测网页后门特征是网页后门对应的特征,根据所述误测网页后门特征与对应的真实检测结果对所述预设检测模型进行训练。

[0090] 应理解的是,所述评估检测结果包括所述第二检测结果,所述第二检测结果为所述第二脚本特征不是网页后门对应的特征,获取所述第二检测结果对应的第二脚本特征,即所述预设检测模型未能成功检测出的第二脚本特征,可将其作为误测网页后门特征,并设置所述误测网页后门特征的真实检测结果为所述误测网页后门特征是网页后门对应的特征。根据所述误测网页后门特征与对应的真实检测结果对所述预设检测模型进行训练,以使所述预设检测模型在后续检测时能够识别出所述误测网页后门特征为网页后门对应的特征,以提高所述预设检测模型进行检测的准确率。再对所述样本网页后门进行特征提取之前,可先对所述样本网页后门进行数据清洗,再将数据清洗后的样本网页后门通过所述预设提取模型进行特征提取,获得第二脚本特征。数据清洗包括删除所述样本网页后门中的无关数据、重复数据和平滑噪声数据,处理所述样本网页后门中缺失值和异常值。可通过设置哪些字段是允许的,哪些字段是不允许的,构成清洗规则,通过所述清洗规则来过滤

掉所述样本网页后门中不符合格式的数据。

[0091] 本实施例中，获取第一数量的样本网页后门，将所述样本网页后门通过所述预设提取模型进行特征提取，获得第二脚本特征，通过所述预设检测模型对所述第二脚本特征进行检测，获得评估检测结果，所述评估检测结果包括所述第二脚本特征是网页后门对应的特征的第一检测结果，统计所述第一检测结果的第二数量，根据所述第一数量和所述第二数量计算所述预设检测模型的准确率，在所述准确率超过预设阈值时，执行所述通过预设检测模型对所述目标脚本特征进行检测，获得目标检测结果的步骤，在所述准确率超过所述预设阈值时，说明所述预设检测模型进行检测的准确率较高，值得信赖，从而确保所述预设检测模型对所述待检测脚本的目标脚本特征进行检测的准确性。

[0092] 此外，本发明实施例还提出一种存储介质，所述存储介质上存储有网页后门检测程序，所述网页后门检测程序被处理器执行时实现如上文所述的网页后门检测方法的步骤。

[0093] 此外，参照图5，本发明实施例还提出一种网页后门检测装置，所述网页后门检测装置包括：

[0094] 匹配模块10，用于获取待检测网络脚本，将所述待检测网络脚本与预设网页后门规则进行匹配；

[0095] 提取模块20，用于若匹配失败，则通过预设提取模型对所述待检测网络脚本进行特征提取，获得目标脚本特征；

[0096] 检测模块30，用于通过预设检测模型对所述目标脚本特征进行检测，获得目标检测结果。

[0097] 应理解的是，本实施例的执行主体是所述网页后门检测设备，其中，所述网页后门检测设备可为个人电脑或服务器等电子设备。所述预设网页后门(webshell)规则可以是恶意字符串库，例如，包括："组专用大马|提权|木马|PHP\s?反弹提权cmd执行"和"WScript.Shell、Shell.Application、Eval()、Excute()、Set Server、Run()、Exec()、ShellExcute()"等恶意字符串。将所述待检测网络脚本进行特征提取，是指对所述待检测脚本中所使用的关键词、高危函数、文件修改的时间、文件权限、文件的所有者以及和其它文件的关联性等多个维度的特征进行提取，从而获得所述脚本特征，将获得的脚本特征与所述预设webshell规则库进行匹配，获得匹配结果。若所述匹配结果为匹配成功，则说明所述待检测网络脚本为webshell；若所述匹配结果为匹配失败，则说明所述待检测网络脚本不是webshell，可能为正常的网络脚本，或者是检测失误的webshell。

[0098] 在本实施例中，所述提取模块20，还用于通过网关获取待检测网络脚本，对所述待检测网络脚本进行分析，提取出多个预设维度的特征；

[0099] 所述匹配模块10，还用于将所述预设维度的特征与预设网页后门规则进行匹配。

[0100] 需要说明的是，所述网关(Gateway)从代理服务器(Agent)获取所述待检测网络脚本，所述待检测网络脚本通常为多个，也可以是一个。对所述待检测网络脚本进行分析，通常是将所述待检测网络脚本拆分成字符串，从所述待检测网络脚本对应的所有字符串中提取出多个预设维度的特征，所述多个预设维度包括：关键词、高危函数、文件修改的时间、文件权限、文件的所有者以及和其它文件的关联性等。正常的网络脚本不会包含所述预设网页后门规则中的特征，则将所述预设维度的特征与所述预设网页后门规则进行匹配，从而

识别出所述待检测网络脚本是否为网页后门,或者是正常的网络脚本。

[0101] 可理解的是,若匹配失败,说明所述待检测网络脚本不是webshell,可能为正常的网络脚本,或者是检测失误的webshell。为了进一步识别所述待检测脚本是否为webshell,可通过所述预设提取模型进行特征提取,所述预设提取模型包括卷积神经网络模型等。可预先建立基础提取模型,获取样本网络脚本与对应的特征对所述基础提取模型进行训练,获得所述预设提取模型。通过所述预设提取模型进行特征提取,获得合适的所述目标脚本特征。

[0102] 在具体实现中,所述预设检测模型包括神经网络模型,经过丰富的大量的训练样本的训练,保证所述预设检测模型对所述目标脚本特征检测的准确性。所述目标检测结果可以是所述目标脚本特征为网页后门对应的特征,即所述目标脚本特征对应的所述待检测网络脚本为网页后门;所述目标检测结果还可以是所述目标脚本特征为正常网络脚本对应的特征,即所述目标脚本特征对应的所述待检测网络脚本为正常的网络脚本。

[0103] 应理解的是,对于所述预设检测模型的建立过程,首先建立基础预测模型,从数据库中获取大量的样本网络脚本及对应的样本检测结果,所述样本网络脚本包括大量的正常的网页脚本和大量的网页后门,可将所述样本网络脚本进行数据清洗,将数据清洗后的样本网络脚本通过所述预设提取模型进行特征提取,获得所述样本网络脚本对应的第一脚本特征,则可根据大量的所述第一脚本特征及对应的所述样本检测结果对所述基础预测模型进行训练,获得所述预设检测模型。所述数据清洗包括对所述样本网络脚本中的无关数据、重复数据和平滑噪声数据,处理所述样本网络脚本中缺失值和异常值。本实施例中,还包括:建立模块,用于建立基础预测模型;获取模块,用于获取样本网络脚本及对应的样本检测结果;所述提取模块20,还用于将所述样本网络脚本通过所述预设提取模型进行特征提取,获得第一脚本特征;训练模块,用于根据所述第一脚本特征及对应的所述样本检测结果对所述基础预测模型进行训练,获得预设检测模型。

[0104] 本实施例中,获取待检测网络脚本,将所述待检测网络脚本与预设网页后门规则进行匹配,通过基于规则的匹配对所述待检测网络脚本进行检测,对于特征明显的网页后门能够被检测出来;若匹配失败,则通过预设提取模型对所述待检测网络脚本进行特征提取,获得目标脚本特征,通过预设检测模型对所述目标脚本特征进行检测,获得目标检测结果,通过将规则检测和基于机器学习的模型检测相结合,未能通过规则匹配检测出的网页后门,可通过基于机器学习的模型进一步进行检测,所述预设检测模型经过大量的样本学习和检测准确率的评估,具有较好的检测效果,从而提高系统检测网络脚本是否为网页后门的准确性。

[0105] 在一实施例中,所述网页后门检测装置还包括:训练模块,用于若所述目标检测结果为所述目标脚本特征是网页后门对应的特征,则根据所述目标脚本特征与对应的所述目标检测结果对所述预设检测模型进行训练。

[0106] 在一实施例中,所述网页后门检测装置还包括:数据清洗模块,用于若匹配失败,则对所述待检测网络脚本进行数据清洗,获得目标网络脚本;

[0107] 所述提取模块20,还用于通过所述预设提取模型对所述目标网络脚本进行特征提取,获得目标脚本特征。

[0108] 在一实施例中,所述提取模块20,还用于通过网关获取待检测网络脚本,对所述待

检测网络脚本进行分析，提取出多个预设维度的特征；

[0109] 所述匹配模块10，还用于将所述预设维度的特征与预设网页后门规则进行匹配。

[0110] 在一实施例中，所述网页后门检测装置还包括：建立模块，用于建立基础预测模型；

[0111] 获取模块，用于获取样本网络脚本及对应的样本检测结果；

[0112] 所述提取模块20，还用于将所述样本网络脚本通过所述预设提取模型进行特征提取，获得第一脚本特征；

[0113] 训练模块，用于根据所述第一脚本特征及对应的所述样本检测结果对所述基础预测模型进行训练，获得预设检测模型。

[0114] 在一实施例中，所述获取模块，还用于获取第一数量的样本网页后门，将所述样本网页后门通过所述预设提取模型进行特征提取，获得第二脚本特征；

[0115] 所述检测模块30，还用于通过所述预设检测模型对所述第二脚本特征进行检测，获得评估检测结果，所述评估检测结果包括所述第二脚本特征是网页后门对应的特征的第一检测结果；

[0116] 所述网页后门检测装置还包括：计算模块，用于统计所述第一检测结果的第二数量，根据所述第一数量和所述第二数量计算所述预设检测模型的准确率；

[0117] 所述检测模块30，还用于在所述准确率超过预设阈值时，执行所述通过预设检测模型对所述目标脚本特征进行检测，获得目标检测结果的步骤。

[0118] 在一实施例中，所述评估检测结果包括所述第二脚本特征不是网页后门对应的特征的第二检测结果；

[0119] 所述获取模块，还用于获取所述第二检测结果对应的第二脚本特征作为误测网页后门特征；

[0120] 所述训练模块，还用于设置所述误测网页后门特征的真实检测结果为所述误测网页后门特征是网页后门对应的特征，根据所述误测网页后门特征与对应的真实检测结果对所述预设检测模型进行训练。

[0121] 本发明所述网页后门检测装置的其他实施例或具体实现方式可参照上述各方法实施例，此处不再赘述。

[0122] 需要说明的是，在本文中，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、物品或者系统不仅包括那些要素，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、物品或者系统所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括该要素的过程、方法、物品或者系统中还存在另外的相同要素。

[0123] 上述本发明实施例序号仅仅为了描述，不代表实施例的优劣。在列举了若干装置的单元权利要求中，这些装置中的若干个可以是通过同一个硬件项来具体体现。词语第一、第二、以及第三等的使用不表示任何顺序，可将这些词语解释为名称。

[0124] 通过以上的实施方式的描述，本领域的技术人员可以清楚地了解到上述实施例方法可借助软件加必需的通用硬件平台的方式来实现，当然也可以通过硬件，但很多情况下前者是更佳的实施方式。基于这样的理解，本发明的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来，该计算机软件产品存储在一个存储介质

(如只读存储器镜像(ReadOnly Memory image,ROM)/随机存取存储器(Random Access Memory,RAM)、磁碟、光盘)中,包括若干指令用以使得一台终端设备(可以是手机,计算机,服务器,空调器,或者网络设备等)执行本发明各个实施例所述的方法。

[0125] 以上仅为本发明的优选实施例,并非因此限制本发明的专利范围,凡是利用本发明说明书及附图内容所作的等效结构或等效流程变换,或直接或间接运用在其他相关的技术领域,均同理包括在本发明的专利保护范围内。

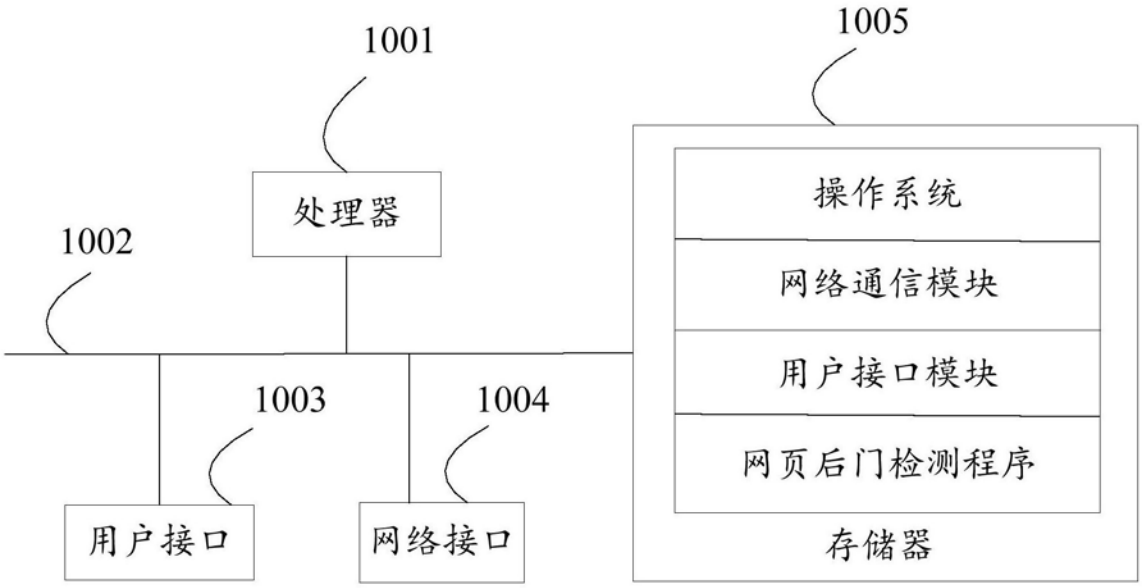

1001
1005
处理器
操作系统
1002
网络通信模块
用户接口模块
1003
1004
网页后门检测程序
用户接口
网络接口
存储器

图1

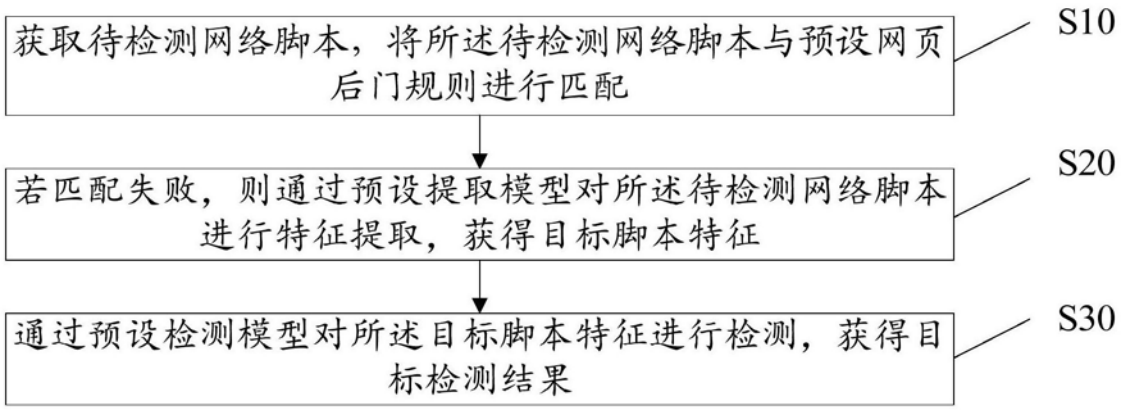

获取待检测网络脚本，将所述待检测网络脚本与预设网页后门规则进行匹配
S10
若匹配失败，则通过预设提取模型对所述待检测网络脚本进行特征提取，获得目标脚本特征
S20
通过预设检测模型对所述目标脚本特征进行检测，获得目标检测结果
S30

图2

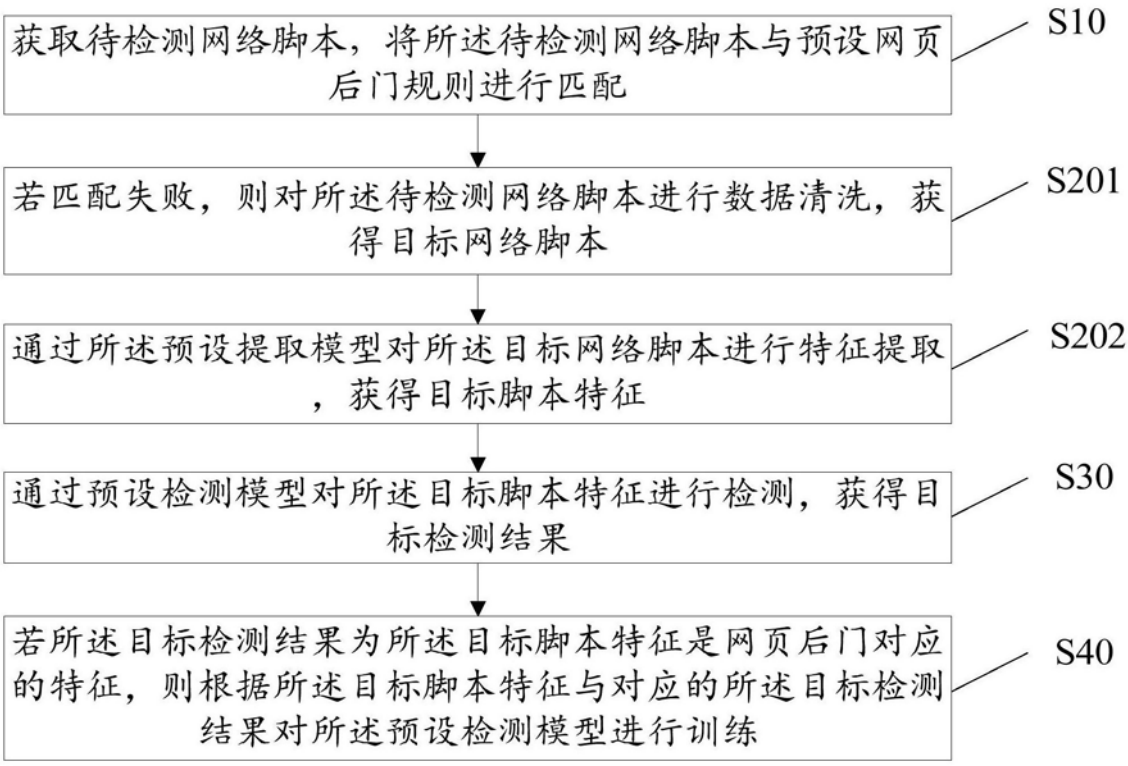

获取待检测网络脚本，将所述待检测网络脚本与预设网页后门规则进行匹配
S10
若匹配失败，则对所述待检测网络脚本进行数据清洗，获得目标网络脚本
S201
通过所述预设提取模型对所述目标网络脚本进行特征提取，获得目标脚本特征
S202
通过预设检测模型对所述目标脚本特征进行检测，获得目标检测结果
S30
若所述目标检测结果为所述目标脚本特征是网页后门对应的特征，则根据所述目标脚本特征与对应的所述目标检测结果对所述预设检测模型进行训练
S40

图3

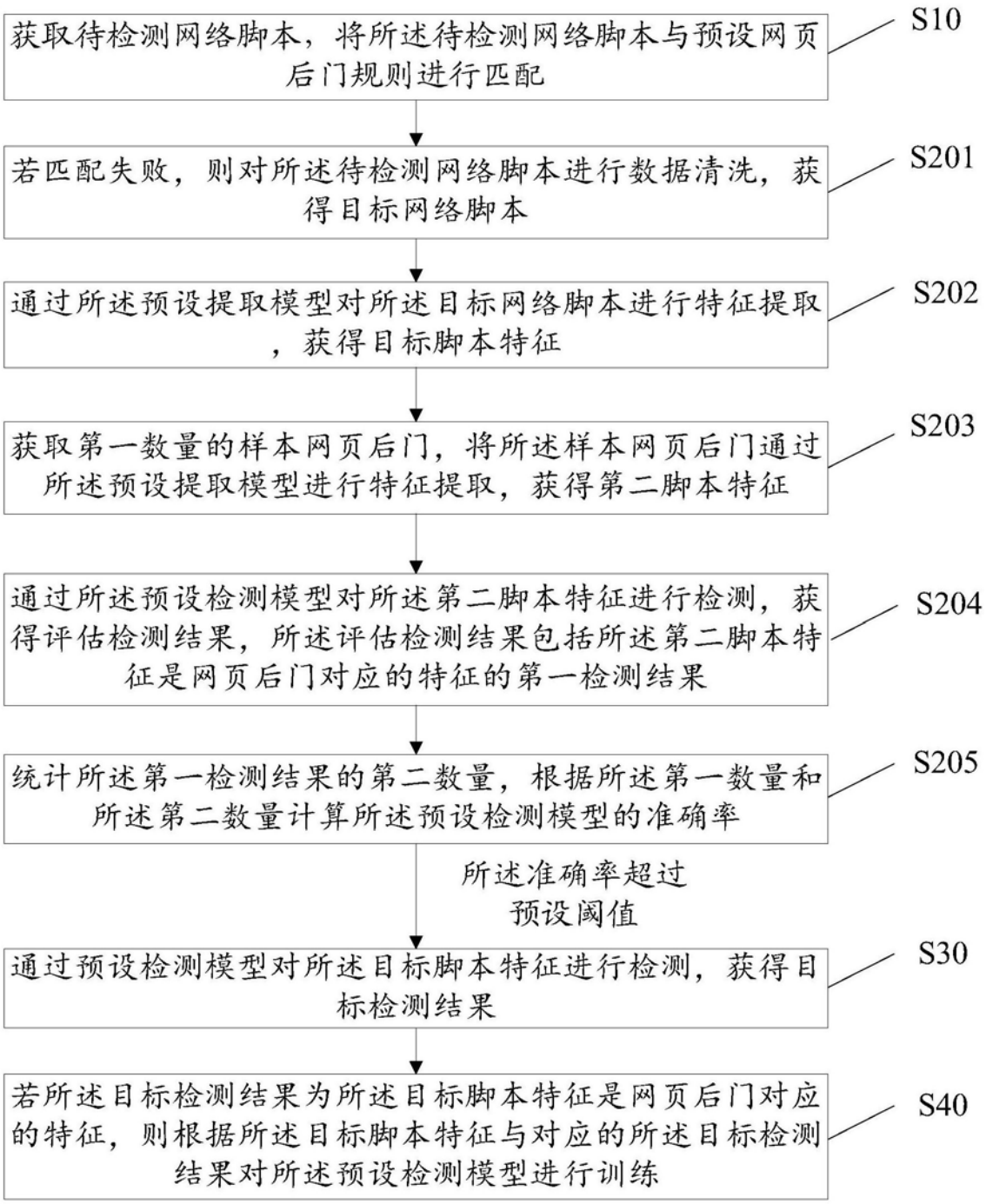

获取待检测网络脚本，将所述待检测网络脚本与预设网页后门规则进行匹配
S10
若匹配失败，则对所述待检测网络脚本进行数据清洗，获得目标网络脚本
S201
通过所述预设提取模型对所述目标网络脚本进行特征提取，获得目标脚本特征
S202
获取第一数量的样本网页后门，将所述样本网页后门通过所述预设提取模型进行特征提取，获得第二脚本特征
S203
通过所述预设检测模型对所述第二脚本特征进行检测，获得评估检测结果，所述评估检测结果包括所述第二脚本特征是网页后门对应的特征的第一检测结果
S204
统计所述第一检测结果的第二数量，根据所述第一数量和所述第二数量计算所述预设检测模型的准确率
S205
所述准确率超过预设阈值
通过预设检测模型对所述目标脚本特征进行检测，获得目标检测结果
S30
若所述目标检测结果为所述目标脚本特征是网页后门对应的特征，则根据所述目标脚本特征与对应的所述目标检测结果对所述预设检测模型进行训练
S40

图4

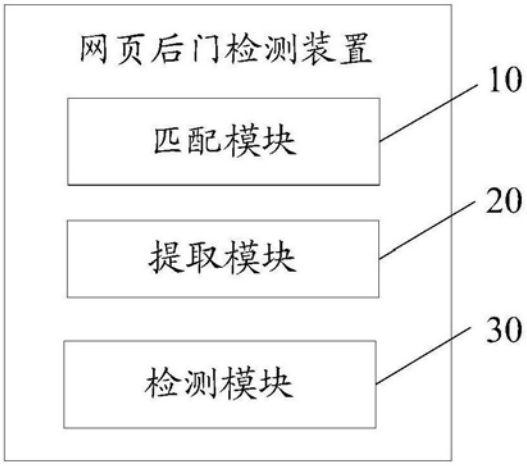


图5