(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
2020 年 1 月 2 日 (02.01.2020)

WIPO | PCT

(10) 国际公布号
WO 2020/000743 A1

(51) 国际专利分类号:
G06F 21/56 (2013.01)

(21) 国际申请号: PCT/CN2018/108472

(22) 国际申请日: 2018 年 9 月 28 日 (28.09.2018)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201810685031.3 2018年6月27日 (27.06.2018) CN

(71) 申请人: 平安科技（深圳）有限公司 (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) [CN/CN]; 中国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, Guangdong 518000 (CN)。

(72) 发明人: 刘立业 (LIU, Liye); 中国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, Guangdong 518000 (CN)。

(74) 代理人: 广州三环专利商标代理有限公司 (SCIHEAD IP LAW FIRM); 中国广东省广州市越秀区先烈中路80号汇华商贸大厦1508室, Guangdong 510070 (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, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国(除另有指明，要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT,

(54) **Title:** WEBSHELL DETECTION METHOD AND RELATED DEVICE

(54) 发明名称: 一种webshell检测方法及相关设备

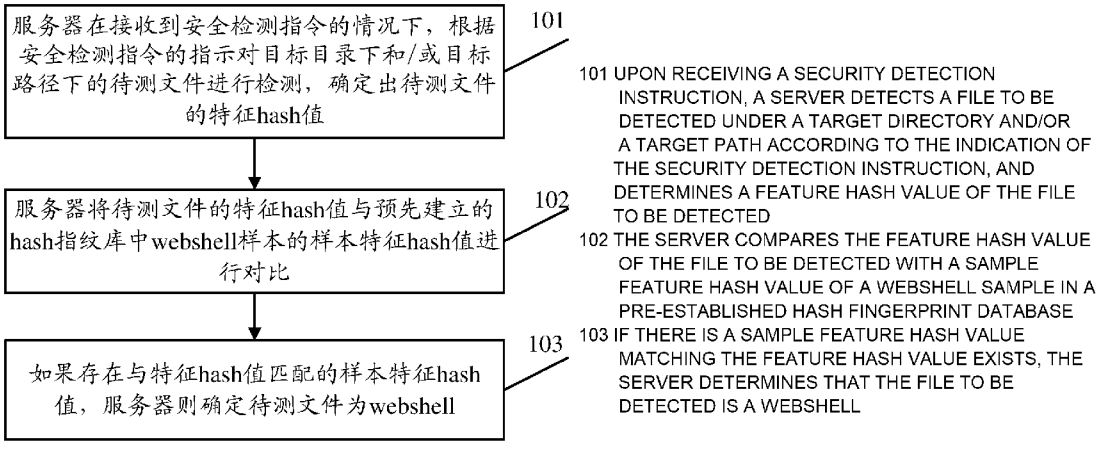


图 1

(57) **Abstract:** A malicious script back door webshell detection method and a related device, wherein the method comprises: Upon receiving a security detection instruction, a server detects a file to be detected under a target directory and/or a target path according to the indication of the security detection instruction, and determines a feature hash value of the file to be detected (101); The server compares the feature hash value of the file to be detected with a sample feature hash value of a webshell sample in a pre-established hash fingerprint database (102), if there is a sample feature hash value matching the feature hash value exists, the server determines that the file to be detected is a webshell (103). The method is beneficial to improve the detection efficiency of the webshell.

WO 2020/000743 A1

[见续页]

RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第21条(3))。

(57) 摘要: 一种恶意脚本后门webshell检测方法及相关设备，所述方法包括：服务器在接收到安全检测指令的情况下，根据安全检测指令的指示对目标目录下和/或目标路径下的待测文件进行检测，确定出所述待测文件的特征hash值(101)；服务器将待测文件的所述特征hash值与预先建立的hash指纹库中webshell样本的样本特征hash值进行对比(102)，如果存在与特征hash值匹配的样本特征hash值，服务器则确定所述待测文件为webshell(103)。该方法有利于提高webshell的检测效率。

一种 webshell 检测方法及相关设备

本申请要求于 2018 年 06 月 27 日提交中国专利局、申请号为 201810685031.3、申请名称为“一种 webshell 检测方法及相关设备”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

本申请涉及计算机技术领域，尤其涉及一种 webshell 检测方法及相关设备。

背景技术

恶意脚本后门 webshell 是一种黑客对服务器进行控制的后门程序。如果黑客发现 web 应用存在文件上传漏洞，可以上传 webshell 进行后续的攻击，并利用该 webshell 暗中对 web 服务器进行远程控制，上传、查看、修改、删除网站服务器上的文件，读取并修改网站数据库的数据，甚至可以直接在网站服务器上运行系统命令。

目前，webshell 检测主要通过恶意代码，字符串编码，危险函数等方法，需要将待测文件对应的代码或者函数与现有的恶意代码、危险函数进行对比，运算量大，耗时很长，检测效率较低。

发明内容

本申请实施例提供了一种 webshell 检测方法及相关设备，在进行 webshell 检测时，不用实时比对待测文件的字符串、代码、各种函数等等，而仅是比对 hash 值，有利于提高 webshell 检测效率。

第一方面，本申请实施例提供了一种 webshell 检测方法，该方法包括:

在接收到安全检测指令的情况下，根据所述安全检测指令的指示对目标目录下和/或目标路径下的待测文件进行检测，确定出所述待测文件的特征 hash 值;

将所述待测文件的所述特征 hash 值与预先建立的 hash 指纹库中 webshell 样本的样本特征 hash 值进行对比;

若存在与所述特征 hash 值匹配的样本特征 hash 值，则确定所述待测文件为 webshell。

第二方面，本申请实施例提供了一种 webshell 检测装置，该 webshell 检测装置包括用

于执行上述第一方面的方法的模块。

第三方面，本申请实施例提供了一种服务器，该服务器包括处理器、网络接口和存储器，所述处理器、网络接口和存储器相互连接，其中，所述网络接口受所述处理器的控制用于收发消息，所述存储器用于存储支持服务器执行上述方法的计算机程序，所述计算机程序包括程序指令，所述处理器被配置用于调用所述程序指令，执行上述第一方面的方法。

第四方面，本申请实施例提供了一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，所述计算机程序包括程序指令，所述程序指令当被处理器执行时使所述处理器执行上述第一方面的方法。

本申请实施例中，服务器可以将待测文件的特征 hash 值与预先建立的 hash 指纹库中恶意脚本后门 webshell 样本的样本特征 hash 值进行对比，如果存在与特征 hash 值匹配的样本特征 hash 值，则确定待测文件为 webshell。通过本申请实施例，有利于提高 webshell 的检测效率。

附图说明

图 1 是本申请实施例提供的一种 webshell 检测方法的流程示意图；

图 2 是本申请实施例提供的另一种 webshell 检测方法的流程示意图；

图 3 是本申请实施例提供的又一种 webshell 检测方法的流程示意图；

图 4 是本申请实施例提供的一种 webshell 检测装置的示意性框图；

图 5 是本申请实施例提供的一种服务器的示意性框图。

具体实施方式

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作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。

参见图 1，图 1 是本申请实施例提供的一种 webshell 检测方法的流程示意图，如图所示，该 webshell 检测方法可包括:

101、服务器在接收到安全检测指令的情况下，根据安全检测指令的指示对目标目录下和/或目标路径下的待测文件进行检测，确定出待测文件的特征 hash 值。

在一个实施例中，当运维人员发现有入侵事件或者发现业务系统有 webshell 存在时，可以针对服务器输入安全检测指令，用于指示服务器针对指定目录下或者指定路径下的待测文件进行扫描检测，例如该安全测试指令可以用于指示对 F 目录中的待测文件进行扫描检测。进一步地，服务器接收到用户输入的该安全检测指令后，则可以根据用户的指示对指定目录或者指定路径的任意一个待测文件进行检测，并通过文本相似度算法确定出待测文件的特征 hash 值。

在一个实施例中，服务器根据安全检测指令的指示对目标目录下和/或目标路径下的待测文件进行检测之前，还可以识别安全测试指令，确定安全测试指令所需测试的目标目录和/或目标路径，并获取目标目录和/或目标路径中所有文件的扩展名，进而将目标目录和/或目标路径中扩展名为预设扩展名的文件确定为待测文件。采用这样的方式，可以不用对指定目录中或者指定路径中的所有文件进行扫描，减少运算量，进一步提高 webshell 检测效率。其中，该预设扩展名可以为不包括运维人员指定的安全文件的扩展名，例如不包括.doc、.pdf 和.rar 等的扩展名。

示例性地，运维人员想要对目录 F 中且扩展名为预设扩展名的文件进行 webshell 检测，这种情况下，运维人员可以通过服务器输入安全测试指令，用于指示对目录 F（即目标目录）中且扩展名为预设扩展名的文件（即待测文件）进行 webshell 检测。进一步地，服务器接收到该安全测试指令后，可以将目录 F 确定为目标目录，并识别目录 F 中所有文件的扩展名，将目录 F 中扩展名为预设扩展名的文件确定为待测文件，并对待测文件进行 webshell 检测。

102、服务器将待测文件的特征 hash 值与预先建立的 hash 指纹库中 webshell 样本的样本特征 hash 值进行对比。

103、如果存在与特征 hash 值匹配的样本特征 hash 值，服务器则确定待测文件为 webshell。

在一个实施例中，服务器可以预先获取不同脚本语言的 webshell 样本，并通过文本相似度算法确定各个 webshell 样本的样本特征 hash 值，进而将不同脚本语言的 webshell 样本，以及各个 webshell 样本的样本特征 hash 值关联存储至 hash 指纹库中，也即建立 hash 指纹库。这种情况下，当服务器确定出待测文件的特征 hash 值之后，可以将待测文件的特征 hash 值与指纹库中的任一个或者多个 webshell 样本的样本特征 hash 值进行对比，如果对比时确定出特征 hash 值与任一个样本特征 hash 值匹配（如：特征 hash 值与任一个样本特征 hash

值之间的汉明距离等于汉明距离阈值，或者特征 hash 值与任一个样本特征 hash 值之间的匹配值等于匹配阈值等），则确定该待测文件为 webshell。采用这样的检测方式，不用实时比对待测文件的字符串、代码特征、各种函数等等，而仅是比对 hash 值，有利于提高 webshell 检测效率。

本申请实施例中，服务器可以在接收到安全检测指令的情况下，根据安全检测指令的指示对目标目录下和/或目标路径下的待测文件进行检测，确定出待测文件的特征 hash 值，并将待测文件的特征 hash 值与预先建立的 hash 指纹库中 webshell 样本的样本特征 hash 值进行对比，若存在与特征 hash 值匹配的样本特征 hash 值，则确定待测文件为 webshell。通过本申请实施例，有利于提高 webshell 的检测效率。

参见图 2，图 2 是本申请实施例提供的另一种 webshell 检测方法的流程示意图，如图所示，该 webshell 检测方法可包括:

201、服务器获取不同脚本语言的 webshell 样本。

202、服务器采用 N 种文本相似度算法确定出每个 webshell 样本在 N 种文本相似度算法下的 N 个样本特征 hash 值，其中，N 为正整数。

203、服务器根据不同脚本语言的 webshell 样本和 webshell 样本在 N 种文本相似度算法下的 N 个样本特征 hash 值，建立 hash 指纹库。

在一个实施例中，服务器可以获取不同脚本语言的 webshell 样本，并采用不同的文本相似度算法计算 webshell 样本的 hash 值，也即对于同一个 webshell 样本而言，N 种算法则对应 N 个 hash 值。获取不同脚本语言的 webshell 样本，可以克服大多数检测 webshell 是基于 ASP、ASP.NET、PHP 等语言的脚本后门进行检查，对 jsp 等其他脚本语言支持不足的问题，有利于多种脚本语言文件检测支持。其中，文本相似度算法可以包括 simhash 算法、ssdeep 算法等。

进一步地，服务器可以将获取到的所有 webshell 样本，以及与各个 webshell 样本对应的 N 个 hash 样本值关联存储至 hash 指纹库。例如，当文本相似度算法包括 simhash 和 ssdeep 这两种算法时，hash 指纹库可以反映如表 1-1 所示的对应关系:

表 1-1

webshell 样本	simhash 下的 hash	ssdeep 下的 hash
a	1000000001	101000000001

b	1000000010	101100000000

或者，服务器也可以针对不同的文本相似度算法，建立各个文本相似度算法各自对应的子 hash 指纹库，也即 hash 指纹库中包括一个或者多个该子 hash 指纹库。例如，当文本相似度算法包括 simhash 和 ssdeep 这两种算法时，hash 指纹库可以分别建立 simhash 算法和 ssdeep 算法的子 hash 指纹库，建立后的子 hash 指纹库反映的关系可以分别如表 1-2 和 1-3 所示。针对不同的文本相似度算法建立各自对应的子 hash 指纹库，这样便于当检测到采用目标算法（如 simhash 算法或者 ssdeep 算法等）进行文本相似度比对时，可以直接在目标算法对应的子 hash 指纹库获取到目标算法下的样本特征 hash 值，进一步提高 webshell 检测效率。

表 1-2

webshell 样本	simhash 下的 hash
a	1000000001
b	1000000010

表 1-3

webshell 样本	ssdeep 下的 hash
a	101000000001
b	101100000000

204、服务器在接收到安全检测指令的情况下，检测当前运行环境，从预设的文本相似度算法列表中确定出与当前运行环境相匹配的目标算法，该目标算法为第一文本相似度算法或者第二文本相似度算法。该预设的文本相似度算法列表中包括 N 种文本相似度算法。

其中，该文本相似度算法列表中包括的 N 种文本相似度算法和步骤 202 中建立 hash 指纹库时所采用的 N 种文本相似度算法是保持一致的。

在一个实施例中，该 N 种文本相似度算法可以包括第一文本相似度算法和第二文本相似度算法中的至少一个。其中，该第一文本相似度算法例如可以为 simhash 算法和第二文本相似度算法例如可以为 ssdeep 算法，这两种算法都是相似性检测算法，都可用于对疑似样本（即待测文件）生成特征 hash 值后与上述 hash 指纹库中任一 webshell 样本的样本特征 hash 值进行 hash 比对。但就这两种算法而言，simhash 算法使用容易，检测时间短，对生产环境的部署配置没有要求，适用于远程检测，安全应急响应中快速检测等场景。ssdeep

算法相比 simhash 算法检测准确度更高，但是对部署环境有一定的要求，依赖额外的库（如功能模块依赖算法库），也即 ssdeep 算法一般适用于本地场景的 webshell 检测。

在一个实施例中，文本相似度算法列表中包括上述第一文本相似度算法和第二文本相似度算法，该第一文本相似度算法对生产环境的部署配置没有要求，第二文本相似度算法，对部署环境有一定的要求，运行时需要依赖额外的库。这种情况下，服务器在接收到安全检测指令后，可以检测当前运行环境是否配置有第二文本相似度算法运行所需的功能模块依赖算法库，如果配置有，则将第二文本相似度算法确定为目标算法，并基于第二文本相似度算法对安全检测指令指示的目标目录和/或目标路径中的待测文件进行检测，确定出待测文件在第二文本相似度算法下的特征 hash 值。如果未配置有，则将第一文本相似度算法确定为目标算法，并基于第一文本相似度算法对安全检测指令指示的目标目录和/或目标路径中的待测文件进行检测，确定出待测文件在第一文本相似度算法下的特征 hash 值。采用多种算法结合的方式进行 webshell 检测，一方面使得本方案可以在远程和本地不同场景执行 webshell 检测，提高了通用性；另一方面，在本地场景执行的情况下，可以优先选择检测精度更高的算法（如 ssdeep 算法），使得 webshell 检测结果更加准确。

205、如果该目标算法为第一文本相似度算法，服务器基于第一文本相似度算法对安全检测指令指示的目标目录下和/或目标路径下的待测文件进行检测，确定出待测文件在第一文本相似度算法下的特征 hash 值。

206、服务器在预先建立的 hash 指纹库中获取任一个或者多个 webshell 样本在第一文本相似度算法下各自对应的样本特征 hash 值。

207、服务器获取待测文件在第一文本相似度算法下的特征 hash 值与任一个或者多个 webshell 样本在第一文本相似度算法下各自对应的样本特征 hash 值之间的各个汉明距离。

208、如果各个汉明距离中的任一个汉明距离等于第一汉明阈值，服务器则确定存在与特征 hash 值匹配的样本特征 hash 值，并确定待测文件为 webshell。

在一个实施例中，当服务器确定出目标算法为第一文本相似度算法时，可以在 hash 指纹库中获取任一个或者多个 webshell 样本在第一文本相似度算法下各自对应的样本特征 hash 值（以下简称第一样本 hash 值），并将获取到任一个或者多个 webshell 样本各自对应的该第一样本特征 hash 值与待测文件在第一文本相似度算法下的待测文件的特征 hash 值（以下简称第一 hash 值）进行汉明距离对比，获取到待测文件的第一 hash 值与任一个或者多个 webshell 样本各自对应第一样本 hash 值之间的各个汉明距离。进一步地，服务器可以

将各个汉明距离与预先设置的第一汉明阈值、第二汉明阈值进行比较，该第二汉明阈值大于第一汉明阈值。如果各个汉明距离中的任一个汉明距离等于第一汉明阈值（例如0），则可以确定存在与该特征hash值匹配的样本特征hash值，进而可以确定待测文件为webshell；或者，如果该任一个汉明距离大于第一汉明阈值，且不大于第二汉明阈值（例如大于 0，且不大于3），则可以确定该待测文件为变种 webshell；又或者，如果该任一个汉明距离大于第二汉明阈值（如大于3），则可以确定待测文件为非webshell。

其中，当服务器确定出待测文件为webshell或者变种webshell时，还可以输出报警信息，用于提示检测出webshell或者变种webshell。在一个实施例中，

服务器在对待测文件进行检测的过程中，还可以记录当前待检测文件的特征hash值，以及它所在的路径信息（也即是步骤 205 中的目标路径和目录目录），当检测到待测文件是webshell的时候，还可以在hash指纹库中获取与这个待测文件进行比对，并确定待测文件是webshell的目标webshell样本的路径信息，以及这个目标webshell样本的样本特征hash值。在确定出待测文件是webshell或者变种webshell后，还可以将上述记录的信息（即当前待检测文件的特征hash值，当前待检测文件的路径信息，目标webshell样本的路径信息，以及目标webshell样本的样本特征hash值）生成扫描日志，输出给用户，以便于用户查看。

在一个实施例中，当服务器确定出待测文件为变种 webshell 时，还可以将该变种webshell以及该变种webshell对应的特征hash值关联存储至之前建立的hash指纹库中，实现对hash指纹库的更新。

本申请实施例中，服务器可以检测当前运行环境，从预设的文本相似度算法列表中确定出与当前运行环境相匹配的目标算法，如果该目标算法为第一文本相似度算法，则基于第一文本相似度算法对安全检测指令指示的目标目录下和/或目标路径下的待测文件进行检测，有利于提高webshell的检测效率。

参见图3，图3是本申请实施例提供的又一种webshell检测方法的流程示意图，如图所示，该webshell检测方法可包括:

301、服务器获取不同脚本语言的webshell样本。

302、服务器采用N种文本相似度算法确定出每个webshell样本在N种文本相似度算法下的N个样本特征hash值，其中，N为正整数。

303、服务器根据不同脚本语言的webshell样本和webshell样本在N种文本相似度算法下的N个样本特征hash值，建立hash指纹库。

304、服务器在接收到安全检测指令的情况下，检测当前运行环境，从预设的文本相似度算法列表中确定出与当前运行环境相匹配的目标算法，该目标算法为第一文本相似度算法或者第二文本相似度算法。

其中，步骤 301-步骤 304 的具体实现方式可以参见上述实施例中步骤 201-步骤 204 的相关描述，此处不再赘述。

305、如果该目标算法为第二文本相似度算法，服务器则基于第二文本相似度算法对安全检测指令指示的目标目录下和/或目标路径下的待测文件进行检测，确定出待测文件在第二文本相似度算法下的特征 hash 值。

306、服务器在预先建立的 hash 指纹库中获取任一个或者多个 webshell 样本在第二文本相似度算法下各自对应的样本特征 hash 值。

307、服务器获取待测文件在第二文本相似度算法下的特征 hash 值与任一个或者多个 webshell 样本在第二文本相似度算法下各自对应的样本特征 hash 值之间的各个加权编辑距离。

308、服务器根据各个加权编辑距离确定第二文本相似度算法下的特征 hash 值与任一个或者多个 webshell 样本在第二文本相似度算法下各自对应的样本特征 hash 值之间的各个匹配值。

309、如果该各个匹配值中任一个匹配值等于第一匹配阈值，服务器则确定存在与特征 hash 值匹配的样本特征 hash 值，并确定待测文件为 webshell。

在一个实施例中，当服务器检测到目标算法为第二文本相似度算法时，可以在 hash 指纹库中获取任一个或者多个 webshell 样本在第二文本相似度算法下各自对应的样本特征 hash 值（以下简称第二样本 hash 值），并将获取到的任一个或者多个 webshell 样本各自对应的该第二样本特征 hash 值与待测文件在第二文本相似度算法下的待测文件的特征 hash 值（以下简称第二 hash 值）进行相似度对比，获取到待测文件的第二 hash 值与任一个或者多个 webshell 样本各自对应的第二样本 hash 值之间的各个加权编辑距离，进而根据各个加权编辑距离确定出第二 hash 值与各个第二样本 hash 值之间的匹配值。在一个实施例中，服务器可以在得到待测文件的第二 hash 值与任一个或者多个 webshell 样本各自对应的第二样本 hash 值之间的各个加权编辑距离之后，可以将各个加权编辑距离除以第二 hash 值和第二样本 hash 值的长度和，再将相除后的结果映射到 0-100 的一个整数值上，进而得到待测文件的第二 hash 值与任一个或者多个 webshell 样本各自对应的第二样本 hash 值之间的各个

匹配值。

进一步地，这种情况下，如果该各个匹配值中的任一个匹配值等于第一匹配阈值（如100），服务器则可以确定存在与该特征hash值匹配的样本特征hash值，进而确定待测文件为webshell。

或者，如果该任一个匹配值大于第二匹配阈值（如50）且小于该第一匹配阈值（如100），则确定待测文件为变种webshell，也即可能存在黑客对该Webshell的代码进行了混淆的情况，该第一匹配阈值大于第二匹配阈值。

又或者，如果该任一个匹配值不大于第二匹配阈值（如50），且不小于0，则可以确定待测文件为非webshell。

在一个实例中，服务器在执行步骤303，根据不同脚本语言的webshell样本和webshell样本在N种文本相似度算法下的N个样本特征hash值，建立hash指纹库时。还可以进一步结合判定待测文件为webshell的条件（如匹配值等于第一匹配阈值）、变种webshell的条件（如该匹配值大于第二匹配阈值（如50）且小于该第一匹配阈值）以及非webshell的条件（如匹配值不大于第二匹配阈值（如50），且不小于0），来建立hash指纹库。

示例性地，假设建立hash指纹库采用了第一文本相似度算法和第二文本相似度算法，分别为simhash算法和ssdeep算法，在建立hash指纹库时，可以将hash指纹库细分为两个子hash指纹库，一个为simhash算法下的子hash指纹库（以下简称simhash指纹库），另一个为ssdeep算法下的子hash指纹库（以下简称ssdeep指纹库），其中，该simhash指纹库包括各个webshell样本以及各个webshell样本在simhash算法下的样本特征hash值（以下简称simhash样本特征hash值）；该ssdeep指纹库包括各个webshell样本以及各个webshell样本在ssdeep算法下的样本特征hash值（以下简称ssdeep样本特征hash值）。这种情况下，服务器可以将与各个simhash样本特征hash值对比满足汉明距离等于第一汉明阈值、汉明距离大于第一汉明阈值，且不大于第二汉明阈值、该汉明距离大于第二汉明阈值（如大于3）的各类hash值关联存储至simhash指纹库中，在存储时可以分别给每一类的hash值添加标记信息。例如，可以将与各个simhash样本特征hash值对比满足汉明距离等于第一汉明阈值的这类hash值，添加用于标记为webshell的标记信息；可以将与各个simhash样本特征hash值对比满足汉明距离大于第一汉明阈值，且不大于第二汉明阈值的这类hash值，添加用于标记为变种webshell的标记信息；可以将与各个simhash样本特征hash值对比满足大于第二汉明阈值的这类hash值，添加用于标记为非webshell的标记信息。

例如，simhash 算法下 webshell 样本 a 的样本特征 hash 值为 1001001，那么可以将与 1001001 的汉明距离=0 的字符串（即 hash 值）与 1001001 关联存储至 simhash 指纹库，并标记这类字符串为 webshell 所属的字符串；将与 1001001 的汉明距离大于 3 的字符串（即 hash 值）与 1001001 关联存储至 simhash 指纹库，并标记这类字符串为非 webshell 所属的字符串；将与 1001001 的汉明距离大于 0 且不小于 3 的字符串（即 hash 值）与 1001001 关联存储至 simhash 指纹库，并标记这类字符串为变种 webshell 所属的字符串。

这种情况下，当服务器确定出待测文件的特征 hash 值之后，可以在预先建立的 hash 指纹库中查找与待测文件的特征hash值相同的字符串，进而确定该字符串所属的标记信息，若标记信息指示该字符串为 webshell 所属的字符串，则确定待测文件为 webshell；若标记信息指示该字符串为非 webshell 所属的字符串，则确定待测文件为非 webshell；若标记信息指示该字符串为变种 webshell 所属的字符串，则确定待测文件为变种 webshell。采用这样的方式，可以减少将待测文件的特征hash值与预先建立的hash指纹库中的样本特征hash值进行对比时的计算量，进一步提高webshell检测效率。

本申请实施例中，服务器可以检测当前运行环境，从预设的文本相似度算法列表中确定出与当前运行环境相匹配的目标算法，如果该目标算法为第二文本相似度算法，则基于第二文本相似度算法对安全检测指令指示的目标目录下和/或目标路径下的待测文件进行检测。一方面，有利于提高 webshell 的检测效率，另一方面，采用与当前运行环境匹配的文本相似度算法对待测文件进行检测，有利于提高 webshell 检测的准确度。

本申请实施例还提供了一种 webshell 检测装置，该装置包括用于执行前述图 1、图 2 或者图 3 所述的方法的模块。具体地，参见图 4，是本申请实施例提供的一种 webshell 检测装置的示意框图。本实施例的 webshell 检测装置包括:

检测模块 40，用于在接收到安全检测指令的情况下，根据所述安全检测指令的指示对目标目录下和/或目标路径下的待测文件进行检测，确定出所述待测文件的特征 hash 值;

对比模块 41，用于将所述待测文件的所述特征 hash 值与预先建立的 hash 指纹库中 webshell 样本的样本特征 hash 值进行对比;

确定模块 42，用于若存在与所述特征 hash 值匹配的样本特征 hash 值，则确定所述待测文件为 webshell。

在一个实施例中，所述装置还包括：获取模块 43，建立模块 44，其中：

获取模块 43，用于获取不同脚本语言的 webshell 样本;

确定模块 42，还用于采用 N 种文本相似度算法确定出每个 webshell 样本在所述 N 种文本相似度算法下的 N 个样本特征 hash 值，其中，N 为正整数;

建立模块 44，用于根据所述不同脚本语言的 webshell 样本和所述 webshell 样本在所述 N 种文本相似度算法下的 N 个样本特征 hash 值，建立 hash 指纹库。

在一个实施例中，所述检测模块 40，具体用于检测当前运行环境，从预设的文本相似度算法列表中确定出与当前运行环境相匹配的目标算法，所述预设的文本相似度算法列表中包括所述 N 种文本相似度算法；基于所述目标算法对所述安全检测指令指示的目标目录下和/或目标路径下的待测文件进行检测，确定出所述待测文件在所述目标算法下的特征 hash 值。

在一个实施例中，对比模块 41，具体用于在预先建立的 hash 指纹库中获取任一个或者多个 webshell 样本在所述目标算法下各自对应的样本特征 hash 值若存在与所述特征 hash 值匹配的样本特征 hash 值; 将所述待测文件在所述目标算法下的所述特征 hash 值与所述任一个或者多个 webshell 样本在所述目标算法下各自对应的所述样本特征 hash 值进行对比。

在一个实施例中，所述目标算法为第一文本相似度算法，对比模块 41，具体还用于获取所述待测文件在所述第一文本相似度算法下的特征 hash 值与所述任一个或者多个 webshell 样本在所述第一文本相似度算法下各自对应的样本特征 hash 值之间的各个汉明距离；确定模块 42，具体用于如果所述各个汉明距离中的任一个汉明距离等于第一汉明阈值，则确定存在与所述特征 hash 值匹配的样本特征 hash 值。

在一个实施例中，所述目标算法为第二文本相似度算法，对比模块 41，具体用于获取所述待测文件在所述第二文本相似度算法下的特征 hash 值与所述任一个或者多个 webshell 样本在所述第二文本相似度算法下各自对应的样本特征 hash 值之间的各个加权编辑距离；根据所述各个加权编辑距离确定所述第二文本相似度算法下的特征 hash 值与所述任一个或者多个 webshell 样本在所述第二文本相似度算法下各自对应的样本特征 hash 值之间的各个匹配值;

其中，确定模块 42，具体还用于如果所述各个匹配值中任一个匹配值等于第一匹配阈值，则确定存在与所述特征 hash 值匹配的样本特征 hash 值。

在一个实施例中，如果确定模块 42 确定出所述任一个汉明距离大于所述第一汉明阈值且不大于第二汉明阈值，则确定所述待测文件为变种 webshell，所述第二汉明阈值大于所述第一汉明阈值;

在一个实施例中，如果确定模块 42 确定出所述任一个匹配值大于第二匹配阈值且小于所述第一匹配阈值，则确定所述待测文件为变种 webshell，所述第一匹配阈值大于所述第二匹配阈值。

在一个实施例中，所述确定模块 42，还用于识别所述安全检测指令，以确定所述安全检测指令所需测试的目标目录和/或目标路径；所述获取模块 43，还用于获取所述目标目录和/或所述目标路径中所有文件的扩展名，并将目标目录和/或目标路径中扩展名为预设扩展名的文件确定为待测文件。

需要说明的是，本申请实施例所描述的 webshell 检测装置的各功能模块的功能可根据图 1、图 2 或者图 3 所述的方法实施例中的方法具体实现，其具体实现过程可以参照图 1、图 2 或者图 3 的方法实施例的相关描述，此处不再赘述。

请参见图 5，图 5 是本申请实施例提供的一种服务器的示意性框图。如图 5 所示，该服务器包括，处理器 501、存储器 502 和网络接口 503。上述处理器 501、存储器 502 和网络接口 503 可通过总线或其他方式连接，在本申请实施例所示图 5 中以通过总线连接为例。其中，网络接口 503 受所述处理器的控制用于收发消息，存储器 502 用于存储计算机程序，所述计算机程序包括程序指令，处理器 501 用于执行存储器 502 存储的程序指令。其中，处理器 501 被配置用于调用所述程序指令执行：在接收到安全检测指令的情况下，根据所述安全检测指令的指示对目标目录下和/或目标路径下的待测文件进行检测，确定出所述待测文件的特征 hash 值；将所述待测文件的所述特征 hash 值与预先建立的 hash 指纹库中 webshell 样本的样本特征 hash 值进行对比；若存在与所述特征 hash 值匹配的样本特征 hash 值，则确定所述待测文件为 webshell。

应当理解，在本申请实施例中，所称处理器 501 可以是中央处理单元（Central Processing Unit，CPU），该处理器 501 还可以是其他通用处理器、数字信号处理器（Digital Signal Processor，DSP）、专用集成电路（Application Specific Integrated Circuit，ASIC）、现成可编程门阵列（Field-Programmable Gate Array，FPGA）或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件等。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。

该存储器 502 可以包括只读存储器和随机存取存储器，并向处理器 501 提供指令和数据。存储器 502 的一部分还可以包括非易失性随机存取存储器。例如，存储器 502 还可以存储设备类型的信息。

具体实现中，本申请实施例中所描述的处理器 501、存储器 502 和网络接口 503 可执行本申请实施例提供的图 1、图 2 或者图 3 所述的方法实施例所描述的实现方式，也可执行本申请实施例所描述的 webshell 检测装置的实现方式，在此不再赘述。

在本申请的另一实施例中提供一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，所述计算机程序包括程序指令，所述程序指令被处理器执行时实现：在接收到安全检测指令的情况下，根据所述安全检测指令的指示对目标目录下和/或目标路径下的待测文件进行检测，确定出所述待测文件的特征 hash 值；将所述待测文件的所述特征 hash 值与预先建立的 hash 指纹库中 webshell 样本的样本特征 hash 值进行对比；若存在与所述特征 hash 值匹配的样本特征 hash 值，则确定所述待测文件为 webshell。

所述计算机可读存储介质可以是前述任一实施例所述的服务器的内部存储单元，例如服务器的硬盘或内存。所述计算机可读存储介质也可以是所述服务器的外部存储设备，例如所述服务器上配备的插接式硬盘，智能存储卡（Smart Media Card，SMC），安全数字（Secure Digital，SD）卡，闪存卡（Flash Card）等。进一步地，所述计算机可读存储介质还可以既包括所述服务器的内部存储单元也包括外部存储设备。所述计算机可读存储介质用于存储所述计算机程序以及所述服务器所需的其他程序和数据。所述计算机可读存储介质还可以用于暂时地存储已经输出或者将要输出的数据。

本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程，是可以通过计算机程序来指令相关的硬件来完成，所述的程序可存储于一计算机可读取存储介质中，该程序在执行时，可包括如上述各方法的实施例的流程。其中，所述的存储介质可为磁碟、光盘、只读存储记忆体（Read-Only Memory，ROM）或随机存储记忆体（Random Access Memory，RAM）等。

以上所揭露的仅为本申请的部分实施例而已，当然不能以此来限定本申请之权利范围，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流程，并依本申请权利要求所作的等同变化，仍属于发明所涵盖的范围。

权 利 要 求

1、一种恶意脚本后门 webshell 检测方法，其特征在于，包括:

在接收到安全检测指令的情况下，根据所述安全检测指令的指示对目标目录下和/或目标路径下的待测文件进行检测，确定出所述待测文件的特征 hash 值;

将所述待测文件的所述特征 hash 值与预先建立的 hash 指纹库中 webshell 样本的样本特征 hash 值进行对比;

若存在与所述特征 hash 值匹配的样本特征 hash 值，则确定所述待测文件为 webshell。

2、根据权利要求 1 所述的方法，其特征在于，所述根据所述安全检测指令的指示对目标目录下和/或目标路径下的待测文件进行检测，确定出所述待测文件的特征 hash 值，包括:

检测当前运行环境，从预设的文本相似度算法列表中确定出与当前运行环境相匹配的目标算法，所述预设的文本相似度算法列表中包括所述 N 种文本相似度算法;

基于所述目标算法对所述安全检测指令指示的目标目录下和/或目标路径下的待测文件进行检测，确定出所述待测文件在所述目标算法下的特征 hash 值;

其中，所述将所述待测文件的所述特征 hash 值与预先建立的 hash 指纹库中 webshell 样本的样本特征 hash 值进行对比，包括:

在预先建立的 hash 指纹库中获取任一个或者多个 webshell 样本在所述目标算法下各自对应的样本特征 hash 值;

将所述待测文件在所述目标算法下的所述特征 hash 值与所述任一个或者多个 webshell 样本在所述目标算法下各自对应的所述样本特征 hash 值进行对比。

3、根据权利要求 2 所述的方法，其特征在于，所述目标算法为第一文本相似度算法，所述将所述待测文件在所述目标算法下的特征 hash 值与所述任一个或者多个 webshell 样本在所述目标算法下各自对应的所述样本特征 hash 值进行对比，包括:

获取所述待测文件在所述第一文本相似度算法下的特征 hash 值与所述任一个或者多个 webshell 样本在所述第一文本相似度算法下各自对应的样本特征 hash 值之间的各个汉明距离;

如果所述各个汉明距离中的任一个汉明距离等于第一汉明阈值，则确定存在与所述特征 hash 值匹配的样本特征 hash 值。

4、根据权利要求 3 所述的方法，其特征在于，所述方法还包括:

如果所述任一个汉明距离大于所述第一汉明阈值且不大于第二汉明阈值，则确定所述

待测文件为变种 webshell，所述第二汉明阈值大于所述第一汉明阈值。

5、根据权利要求 2 所述的方法，其特征在于，所述目标算法为第二文本相似度算法，所述将所述待测文件在所述目标算法下的特征 hash 值与所述任一个或者多个 webshell 样本在所述目标算法下各自对应的样本特征 hash 值进行对比，包括:

获取所述待测文件在所述第二文本相似度算法下的特征 hash 值与所述任一个或者多个 webshell 样本在所述第二文本相似度算法下各自对应的样本特征 hash 值之间的各个加权编辑距离;

根据所述各个加权编辑距离确定所述第二文本相似度算法下的特征 hash 值与所述任一个或者多个 webshell 样本在所述第二文本相似度算法下各自对应的样本特征 hash 值之间的各个匹配值;

如果所述各个匹配值中的任一个匹配值等于第一匹配阈值，则确定存在与所述特征 hash 值匹配的样本特征 hash 值。

6、根据权利要求 5 所述的方法，其特征在于，所述方法还包括:

如果所述任一个匹配值大于第二匹配阈值且小于所述第一匹配阈值，则确定所述待测文件为变种 webshell，所述第一匹配阈值大于所述第二匹配阈值。

7、根据权利要求 1-6 任一项所述的方法，其特征在于，所述将所述待测文件的所述特征 hash 值与预先建立的 hash 指纹库中 webshell 样本的样本特征 hash 值进行对比之前，所述方法还包括:

获取不同脚本语言的 webshell 样本;

采用 N 种文本相似度算法确定出每个 webshell 样本在所述 N 种文本相似度算法下的 N 个样本特征 hash 值，其中，N 为正整数;

根据所述不同脚本语言的 webshell 样本和所述 webshell 样本在所述 N 种文本相似度算法下的 N 个样本特征 hash 值，建立 hash 指纹库。

8、根据权利要求 1-7 任一项所述的方法，其特征在于，所述根据所述安全检测指令的指示对目标目录下和/或目标路径下的待测文件进行检测之前，所述方法还包括:

识别所述安全检测指令，以确定所述安全检测指令所需测试的目标目录和/或目标路径;

获取所述目标目录和/或所述目标路径中所有文件的扩展名，并将目标目录和/或目标路径中扩展名为预设扩展名的文件确定为待测文件。

9、一种 webshell 检测装置，其特征在于，包括:

检测模块，用于在接收到安全检测指令的情况下，根据所述安全检测指令的指示对目标目录下和/或目标路径下的待测文件进行检测，确定出所述待测文件的特征 hash 值;

对比模块，用于将所述待测文件的所述特征 hash 值与预先建立的 hash 指纹库中 webshell 样本的样本特征 hash 值进行对比;

确定模块，用于若存在与所述特征 hash 值匹配的样本特征 hash 值，则确定所述待测文件为 webshell。

10、根据权利要求 9 所述的装置，其特征在于，所述检测模块，具体用于检测当前运行环境，从预设的文本相似度算法列表中确定出与当前运行环境相匹配的目标算法，所述预设的文本相似度算法列表中包括所述 N 种文本相似度算法；基于所述目标算法对所述安全检测指令指示的目标目录下和/或目标路径下的待测文件进行检测，确定出所述待测文件在所述目标算法下的特征 hash 值；其中，所述对比模块，具体用于在预先建立的 hash 指纹库中获取任一个或者多个 webshell 样本在所述目标算法下各自对应的样本特征 hash 值；将所述待测文件在所述目标算法下的所述特征 hash 值与所述任一个或者多个 webshell 样本在所述目标算法下各自对应的所述样本特征 hash 值进行对比。

11、根据权利要求 10 所述的装置，其特征在于，所述目标算法为第一文本相似度算法，所述对比模块，具体还用于获取所述待测文件在所述第一文本相似度算法下的特征 hash 值与所述任一个或者多个 webshell 样本在所述第一文本相似度算法下各自对应的样本特征 hash 值之间的各个汉明距离；所述确定模块，具体用于如果所述各个汉明距离中的任一个汉明距离等于第一汉明阈值，则确定存在与所述特征 hash 值匹配的样本特征 hash 值。

12、根据权利要求 11 所述的装置，其特征在于，所述确定模块，还用于如果确定出所述任一个汉明距离大于所述第一汉明阈值且不大于第二汉明阈值，则确定所述待测文件为变种 webshell，所述第二汉明阈值大于所述第一汉明阈值。

13、根据权利要求 10 所述的装置，其特征在于，所述目标算法为第二文本相似度算法，所述对比模块，具体用于获取所述待测文件在所述第二文本相似度算法下的特征 hash 值与所述任一个或者多个 webshell 样本在所述第二文本相似度算法下各自对应的样本特征 hash 值之间的各个加权编辑距离；根据所述各个加权编辑距离确定所述第二文本相似度算法下的特征 hash 值与所述任一个或者多个 webshell 样本在所述第二文本相似度算法下各自对应的样本特征 hash 值之间的各个匹配值；其中，所述确定模块，具体还用于如果所述各个匹

配值中任一个匹配值等于第一匹配阈值，则确定存在与所述特征 hash 值匹配的样本特征 hash 值。

14、根据权利要求 13 所述的装置，其特征在于，所述确定模块，还用于如果确定出所述任一个匹配值大于第二匹配阈值且小于所述第一匹配阈值，则确定所述待测文件为变种 webshell，所述第一匹配阈值大于所述第二匹配阈值。

15、根据权利要求 9-14 所述的装置，其特征在于，所述装置还包括：获取模块、建立模块，其中：

所述获取模块，用于获取不同脚本语言的 webshell 样本；

所述确定模块，还用于采用 N 种文本相似度算法确定出每个 webshell 样本在所述 N 种文本相似度算法下的 N 个样本特征 hash 值，其中，N 为正整数；

所述建立模块，用于根据所述不同脚本语言的 webshell 样本和所述 webshell 样本在所述 N 种文本相似度算法下的 N 个样本特征 hash 值，建立 hash 指纹库。

16、根据权利要求 9-15 任一项所述的装置，其特征在于，所述确定模块，还用于识别所述安全检测指令，以确定所述安全检测指令所需测试的目标目录和/或目标路径；所述获取模块，还用于获取所述目标目录和/或所述目标路径中所有文件的扩展名，并将目标目录和/或目标路径中扩展名为预设扩展名的文件确定为待测文件。

17、一种服务器，其特征在于，包括处理器和存储器，所述处理器和所述存储器相互连接，其中，所述存储器用于存储计算机程序，所述计算机程序包括程序指令，所述处理器被配置用于调用所述程序指令执行：在接收到安全检测指令的情况下，根据所述安全检测指令的指示对目标目录下和/或目标路径下的待测文件进行检测，确定出所述待测文件的特征 hash 值；将所述待测文件的所述特征 hash 值与预先建立的 hash 指纹库中 webshell 样本的样本特征 hash 值进行对比；若存在与所述特征 hash 值匹配的样本特征 hash 值，则确定所述待测文件为 webshell。

18、根据权利要求 17 所述的服务器，其特征在于，所述处理器，还用于；

检测当前运行环境，从预设的文本相似度算法列表中确定出与当前运行环境相匹配的目标算法，所述预设的文本相似度算法列表中包括所述 N 种文本相似度算法；

基于所述目标算法对所述安全检测指令指示的目标目录下和/或目标路径下的待测文件进行检测，确定出所述待测文件在所述目标算法下的特征 hash 值；

在预先建立的 hash 指纹库中获取任一个或者多个 webshell 样本在所述目标算法下各自

对应的样本特征 hash 值;

将所述待测文件在所述目标算法下的所述特征 hash 值与所述任一个或者多个 webshell 样本在所述目标算法下各自对应的所述样本特征 hash 值进行对比。

19、根据权利要求 18 所述的服务器，其特征在于，所述目标算法为第一文本相似度算法，所述处理器，还用于:

获取所述待测文件在所述第一文本相似度算法下的特征 hash 值与所述任一个或者多个 webshell 样本在所述第一文本相似度算法下各自对应的样本特征 hash 值之间的各个汉明距离;

如果所述各个汉明距离中的任一个汉明距离等于第一汉明阈值，则确定存在与所述特征 hash 值匹配的样本特征 hash 值。

20、一种计算机可读存储介质，其特征在于，所述计算机可读存储介质存储有计算机程序，所述计算机程序包括程序指令，所述程序指令当被处理器执行时使所述处理器执行如权利要求 1-8 任一项所述的方法。

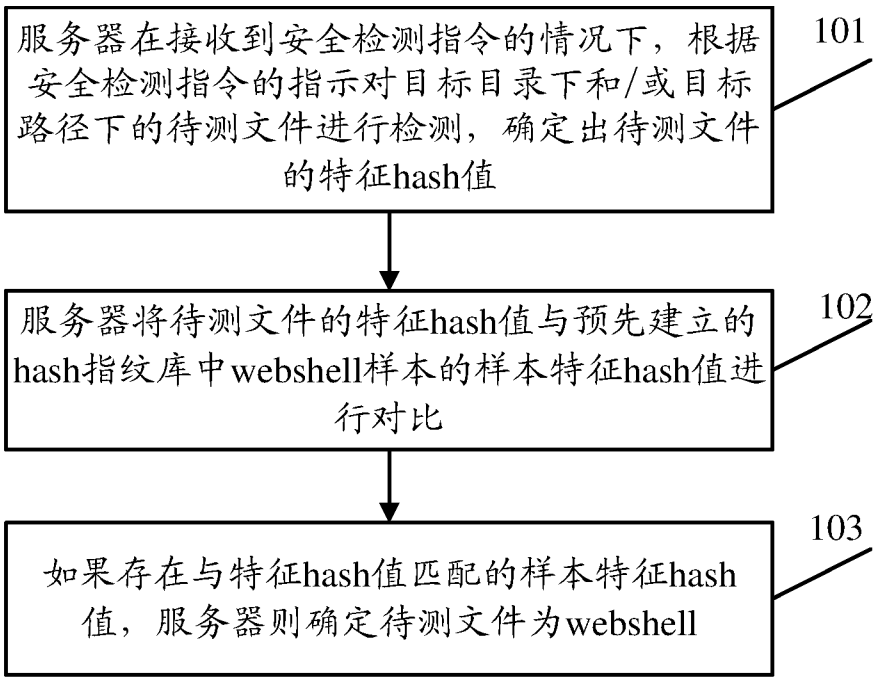

服务器在接收到安全检测指令的情况下，根据安全检测指令的指示对目标目录下和/或目标路径下的待测文件进行检测，确定出待测文件的特征hash值
101
服务器将待测文件的特征hash值与预先建立的hash指纹库中webshell样本的样本特征hash值进行对比
102
如果存在与特征hash值匹配的样本特征hash值，服务器则确定待测文件为webshell
103

图 1

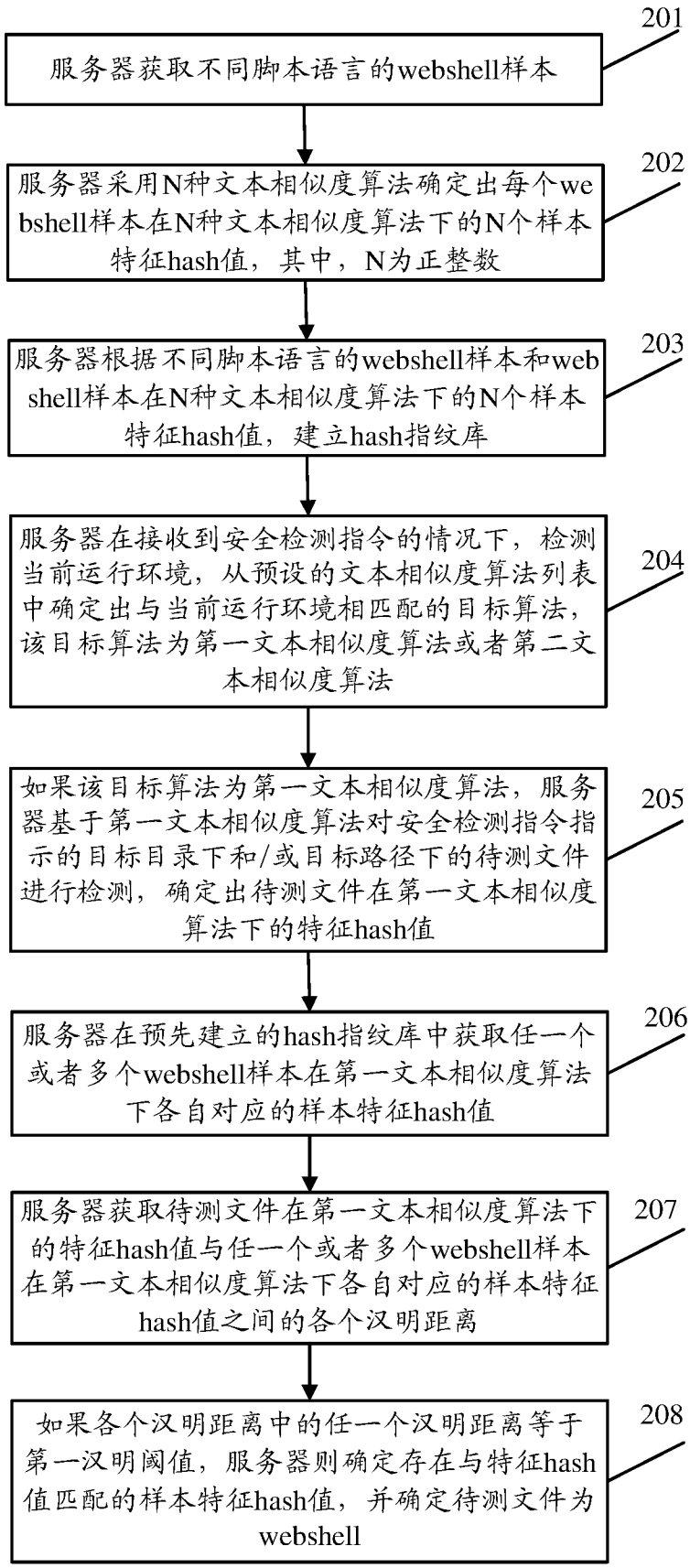

201
服务器获取不同脚本语言的webshell样本
202
服务器采用N种文本相似度算法确定出每个webshell样本在N种文本相似度算法下的N个样本特征hash值，其中，N为正整数
203
服务器根据不同脚本语言的webshell样本和webshell样本在N种文本相似度算法下的N个样本特征hash值，建立hash指纹库
204
服务器在接收到安全检测指令的情况下，检测当前运行环境，从预设的文本相似度算法列表中确定出与当前运行环境相匹配的目标算法，该目标算法为第一文本相似度算法或者第二文本相似度算法
205
如果该目标算法为第一文本相似度算法，服务器基于第一文本相似度算法对安全检测指令指示的目标目录下和/或目标路径下的待测文件进行检测，确定出待测文件在第一文本相似度算法下的特征hash值
206
服务器在预先建立的hash指纹库中获取任一个或者多个webshell样本在第一文本相似度算法下各自对应的样本特征hash值
207
服务器获取待测文件在第一文本相似度算法下的特征hash值与任一个或者多个webshell样本在第一文本相似度算法下各自对应的样本特征hash值之间的各个汉明距离
208
如果各个汉明距离中的任一个汉明距离等于第一汉明阈值，服务器则确定存在与特征hash值匹配的样本特征hash值，并确定待测文件为webshell

图 2

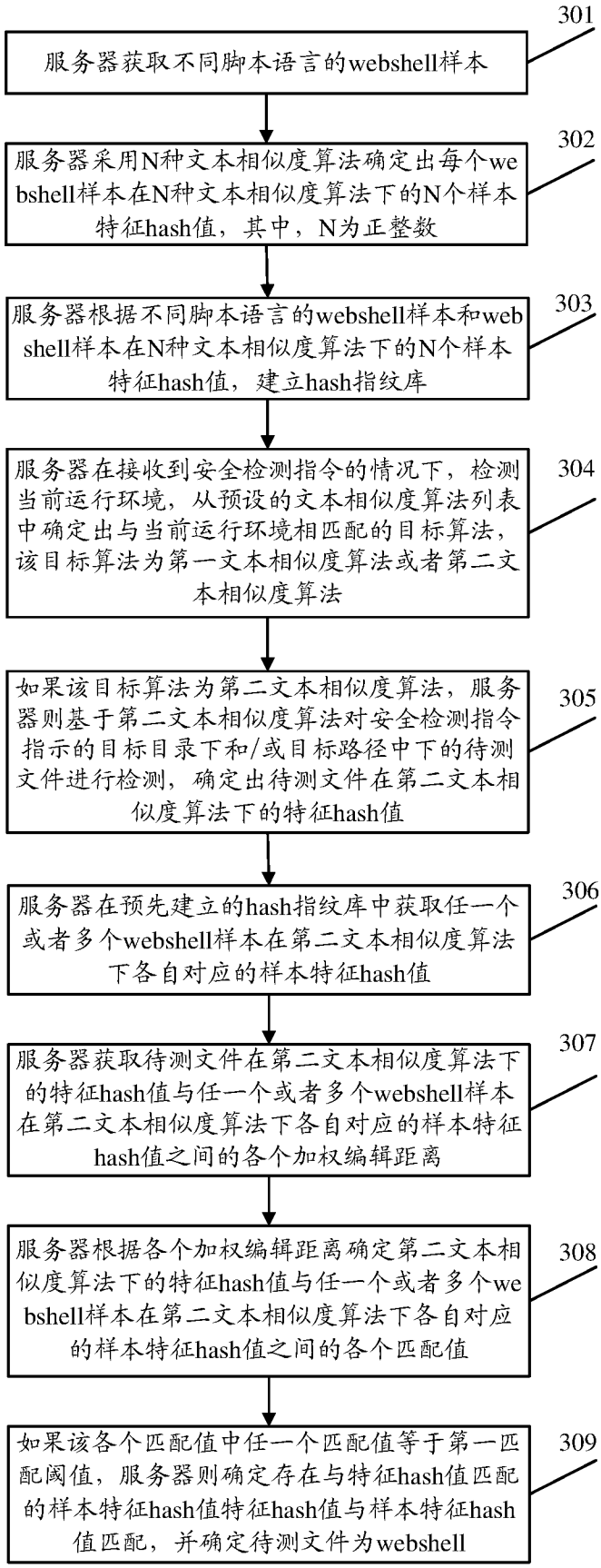

301
服务器获取不同脚本语言的webshell样本
302
服务器采用N种文本相似度算法确定出每个webshell样本在N种文本相似度算法下的N个样本特征hash值，其中，N为正整数
303
服务器根据不同脚本语言的webshell样本和webshell样本在N种文本相似度算法下的N个样本特征hash值，建立hash指纹库
304
服务器在接收到安全检测指令的情况下，检测当前运行环境，从预设的文本相似度算法列表中确定出与当前运行环境相匹配的目标算法，该目标算法为第一文本相似度算法或者第二文本相似度算法
305
如果该目标算法为第二文本相似度算法，服务器则基于第二文本相似度算法对安全检测指令指示的目标目录下和/或目标路径中下的待测文件进行检测，确定出待测文件在第二文本相似度算法下的特征hash值
306
服务器在预先建立的hash指纹库中获取任一个或者多个webshell样本在第二文本相似度算法下各自对应的样本特征hash值
307
服务器获取待测文件在第二文本相似度算法下的特征hash值与任一个或者多个webshell样本在第二文本相似度算法下各自对应的样本特征hash值之间的各个加权编辑距离
308
服务器根据各个加权编辑距离确定第二文本相似度算法下的特征hash值与任一个或者多个webshell样本在第二文本相似度算法下各自对应的样本特征hash值之间的各个匹配值
309
如果该各个匹配值中任一个匹配值等于第一匹配阈值，服务器则确定存在与特征hash值匹配的样本特征hash值特征hash值与样本特征hash值匹配，并确定待测文件为webshell

图 3

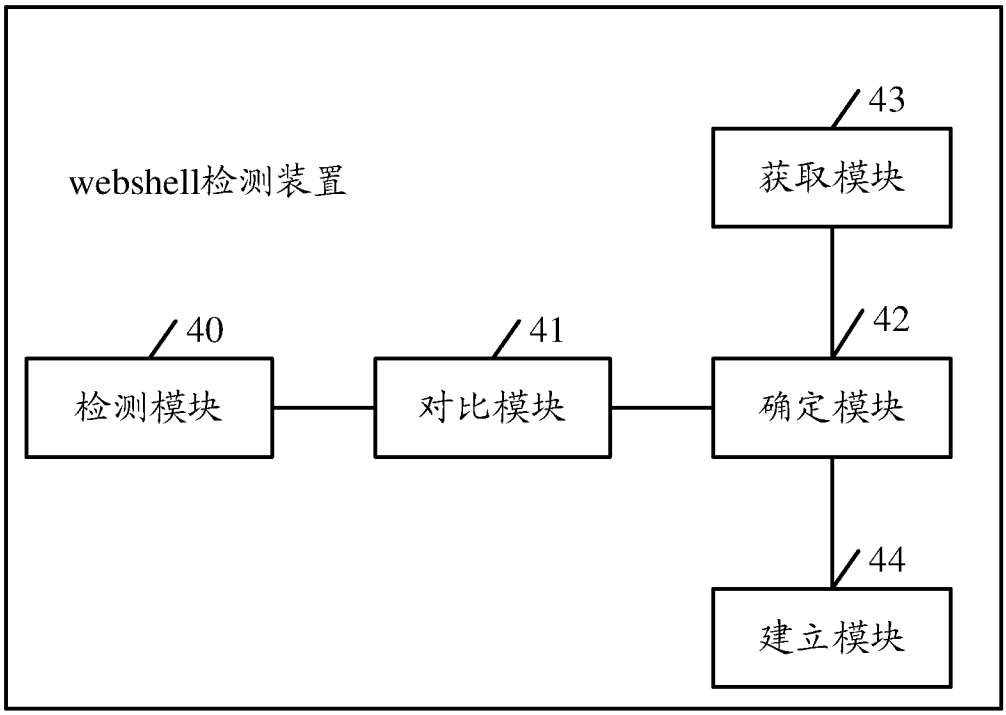

webshell检测装置
43
获取模块
42
40
41
检测模块
对比模块
确定模块
44
建立模块

图 4

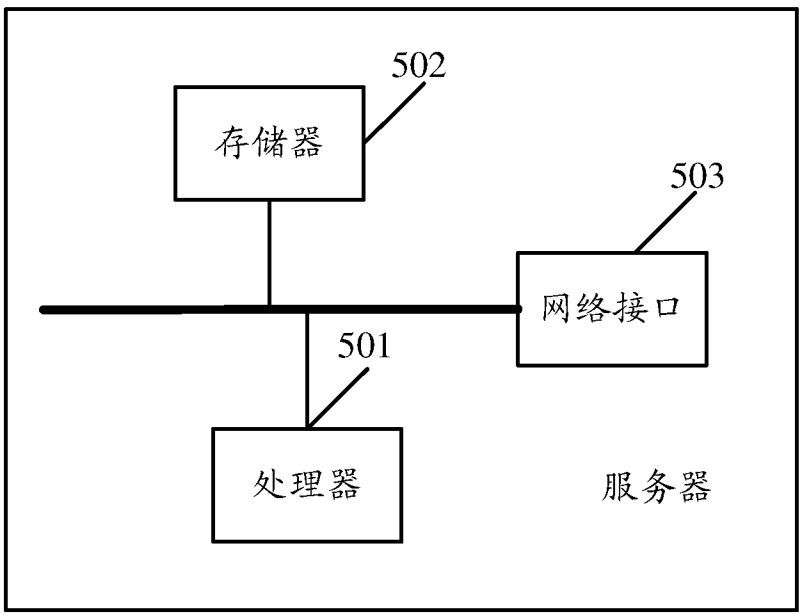

502
存储器
503
网络接口
501
处理器
服务器

图 5

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2018/108472

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, IEEE, CNKI, GOOGLE: 后门, 脚本, 检测, 安全, 哈希, 匹配, 对比, backdoor, script, webshell, detect +, security, hash, match, contrast

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 107103237 A (ALIBABA GROUP HOLDING LIMITED) 29 August 2017 (2017-08-29) description, paragraphs [0069], [0089]-[0096] and [0115]-[0118]	1-20
A	CN 108156131 A (SHANGHAI INFORMATION & DATA SECURITY SOLUTIONS CO., LTD.) 12 June 2018 (2018-06-12) entire document	1-20
A	CN 104811447 A (SANGFOR NETWORK TECHNOLOGY (SHENZHEN) CO., LTD.) 29 July 2015 (2015-07-29) entire document	1-20
A	CN 105933268 A (CHINA UNIONPAY CO., LTD.) 07 September 2016 (2016-09-07) entire document	1-20
A	CN 103561012 A (INSTITUTE OF INFORMATION ENGINEERING, CHINESE ACADEMY OF SCIENCES) 05 February 2014 (2014-02-05) entire document	1-20

☐ 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
10 March 2019	**29 March 2019**

Name and mailing address of the ISA/CN	Authorized officer
National Intellectual Property Administration, PRC (ISA/CN) **No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088** **China**	
Facsimile No. **(86-10)62019451**	Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2018/108472

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN	107103237	A	29 August 2017	None	
CN	108156131	A	12 June 2018	None	
CN	104811447	A	29 July 2015	None	
CN	105933268	A	07 September 2016	None	
CN	103561012	A	05 February 2014	None	

国际检索报告

国际申请号
PCT/CN2018/108472

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, IEEE, CNKI, GOOGLE: 后门，脚本，检测，安全，哈希，匹配，对比，backdoor, script, webshell, detect+, security, hash, match, contrast

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 107103237 A (阿里巴巴集团控股有限公司) 2017年 8月 29日 (2017-08-29) 说明书第[0069]、[0089]-[0096]、[0115]-[0118]段	1-20
A	CN 108156131 A (上海观安信息技术股份有限公司) 2018年 6月 12日 (2018-06-12) 全文	1-20
A	CN 104811447 A (深信服网络科技深圳有限公司) 2015年 7月 29日 (2015-07-29) 全文	1-20
A	CN 105933268 A (中国银联股份有限公司) 2016年 9月 7日 (2016-09-07) 全文	1-20
A	CN 103561012 A (中国科学院信息工程研究所) 2014年 2月 5日 (2014-02-05) 全文	1-20

☐ 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
2019年 3月 10日	2019年 3月 29日

ISA/CN的名称和邮寄地址	受权官员
中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088	高丹丹
传真号 (86-10)62019451	电话号码 86-(10)-53961301

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2018/108472

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	107103237	A	2017年 8月 29日	无	
CN	108156131	A	2018年 6月 12日	无	
CN	104811447	A	2015年 7月 29日	无	
CN	105933268	A	2016年 9月 7日	无	
CN	103561012	A	2014年 2月 5日	无	