一种基于哈希特征向量的恶意软件云检测方法

技术领域

本发明涉及计算终端的恶意软件检测领域，具体涉及云计算技术下如何在隐私和实际效率兼顾的情况下，利用哈希特征向量技术为终端提供高效的恶意软件扫描检测方法。

背景技术

随着移动智能设备和物联设备的迅速普及和云计算远程存储功能的发展，移动互联网的安全问题凸显。根据艾瑞《2013 年中国移动安全数据报告显示》，2013 年移动安全形势比较严峻，新增恶意软件 69 万个，是 2012 年的五倍多。大量经过重度混淆、加密的恶意软件涌现，且越来越多恶意软件或广告平台开始采用动态加载、延迟发作等方式试图规避安全软件的检测和查杀；另外，恶意软件的传播手段也在变化，内嵌子包或联网下载恶意软件等情况已十分常见，如何保证这样的恶意软件不会漏杀，成为安全厂商需要面对的一大挑战。

本发明旨在保护资源受限的中小型联网计算终端的安全，例如移动终端、物联设备、嵌入式设备以及追求效率的计算机终端，下文中将此类设备都简称为终端。目前针对终端的恶意软件扫描的研究越来越深入，主要的技术分为两种类型：第一种是与传统电脑安全软件类似，在终端上建立存储大量恶意软件特征码的特征码库，在终端上对文件进行特征匹配。这种技术原理简单，通过对已经发现了的恶意软件提取特定的字符串或者计算恶意软件 MD5（Message Digest Algorithm MD5）值等方法来创建恶意软件特征码，然后扫描文件，使用例如 BM（Boyer-Manber）和 AC（Aho-Corasick）算法等模式匹配算法，判断文件是否与某种恶意软件特征码相匹配，如果匹配成功则认为该文件是恶意软件。但是使用这种方法，终端需要不断更新恶意软件特征库，消耗大量网络和计算资源；另外扫描过程也会大量占用 CPU 和内存资源，严重影响资源受限设备的可用性和电池的持续性。

第二种是基于云计算的在线病毒扫描，在云端建立存储大量恶意软件特征的黑名单数据库和已经被证明是安全文件的白名单数据库。当终端需要进行病毒查杀时，会对设备所有文件计算 MD5 校验和，然后发送数据到云端，云端会对发送来的数据进行扫描，根据黑名单数据库和白名单数据库识别发送来的数据，判断原始文件是否是恶意软件。这种技术利用互联网，通过联网查询，把对终端里的文件扫描检测从终端转到云端，终端不需要保存恶意软件特征库，也不需要对特征库进行更新，提高了恶意软件查杀和防护的及时性、有效性。同时，90%以上的安全检测由云端服务器承担，从而降低了终端的 CPU 和内存等资源的占用，使设备运行变快。但是这种技术会将终端上所有文件的信息发送到云端，

从而用户的隐私会受到很大的威胁。另外该类方法大都没有考虑字符串类型的特征码检测，且终端的所有文件都需跟云端的所有特征码进行匹配，云端的计算任务极其繁重。

目前国内主流安全厂商生产的运行在终端上的安全软件大致采用上述两种技术模式。另外，目前相关研究也大致符合上述思路，如中国申请号为 201110265295.1、名称为“手机恶意软件查杀方法及系统”中提出了一种基于移动网络侧恶意软件监控分析系统的手机恶意软件查杀方法，能提高手机恶意软件查杀效率，但是在查杀过程中存在用户的一些重要身份标识、敏感信息以及服务端特征库泄露的风险，安全性难以得到保障。中国申请号为 201010292928.3、名称为“一种信息安全检测方法及移动终端”中提出通过动态虚拟机的方法预先分析恶意软件的行为特征，能有效减少对移动终端的威胁，但是动态虚拟机本身会造成终端资源的大量消耗，造成整体效率的下降。

综上所述，将安全检测过程放在终端，不会对用户隐私构成威胁，但是存在计算、存储、网络资源消耗大的问题，严重影响资源受限终端设备的可用性和电池的持续性；利用云计算的思想和架构，将安全检测过程转移到云端，在终端资源消耗和及时效率方面会得到提高，但是用户隐私却存在泄露的风险，另外云端的计算任务会急剧增加。现有技术未能很好地兼顾隐私和效率两方面的需求，因此提出能够同时兼顾效率和隐私的新型恶意软件扫描策略和架构，对于移动互联网和物联网的安全很有意义。

为了能够尽量减小恶意软件检测终端的开销和隐私泄露风险，本发明采用哈希映射的方式将恶意软件特征库压缩成为特征信息向量，在终端和云服务器之间进行检测信息交互。且改变了现有云安全技术需要将用户的所有文件跟所有特征码进行匹配的问题，只需将用户的特定嫌疑文件跟特定特征码进行匹配。作为本发明的同系列专利，专利“一种基于哈希特征向量矩阵的恶意软件云检测方法”在本专利的基础上拓展特征向量为特征向量矩阵，在控制终端开销和隐私泄露风险的前提下，能够很好地减轻云服务器的开销，但终端的计算量会有所增加。

本发明是在 Bloom Filter 算法的基础上设计恶意软件扫描策略和架构的，下面对 Bloom Filter 算法进行简要的介绍。Bloom Filter（以下简称 BF）算法是由 B.H. Bloom 在 1970 年提出的二进制向量数据结构，它具有很好的空间和时间效率，它利用位数组很简洁地表示一个集合，并能判断一个元素是否属于这个集合。初始状态时，BF 是一个长度为 m 位的向量，每一位都置为 0。为了表达 $S=\{X_1, X_2,\ldots,Xn\}$这样的 n 个元素的集合，BF 使用 k 个相互独立的哈希函数 h_i（$1\leq i\leq k$）对每个元素进行 BF 映射处理，即首先将每个元素映射到{1,…,m}的范围中。对任意一个元素 X，第 i 个哈希函数映射的位置 $h_i(X)$就会

被置为 1（1≤i≤k）。若一个位置多次被置为 1，那么只有第一次会起作用，后面几次将没有任何效果。在判断 Y 是否属于这个集合时，我们对 Y 应用 k 次哈希函数，如果所有 $h_i(Y)$（1≤i≤k）的位置都是 1，那么我们就认为 Y 是集合中的元素，否则就认为 Y 不是集合中的元素。这种判断机制会因为哈希函数的碰撞而带来假阳率（False Positive），如已知 Y_1 在 X 中，若 $h_i(Y_1)$与 $h_i(Y_2)$值相同，那么 Y_2 则会被误判为也在 X 中，经计算可知这种假阳率的概率为$\left(1-e^{-kn/m}\right)^k$。

发明内容

本发明要解决的技术问题是：在云计算的背景下，如何提出一种终端与云服务端协同合作的恶意软件扫描方法，通过哈希特征向量技术实现特征库的压缩交互，在尽量减小恶意软件检测终端的网络、计算和存储开销的同时，使得终端尽量少的向云服务器提交文件信息，从而能够保护终端的隐私。

为此，本发明的技术策略为：云服务端负责维护和更新体量较大的恶意软件特征数据库，终端通过分段 BF 算法将特征数据库映射成体量很小的恶意软件特征向量。云服务端将恶意软件特征向量发送给终端，且每当恶意软件特征数据库发成更新时，向终端增量推送特征向量的更新。终端利用分段 BF 算法对本地待扫描文件进行映射处理后与恶意软件特征向量进行模糊扫描，并将匹配结果发送给云服务端。云服务端对匹配的结果进行进一步的精确扫描，之后将确认结果返回给终端。

为实现上述策略，本发明设计的技术包括两个主要方法：恶意软件特征向量处理方法和恶意软件云扫描确认方法。具体说明如下：

一、恶意软件特征向量处理方法包括以下三个步骤：

1）恶意软件特征向量生成。云服务端负责维护恶意软件特征数据库，该数据库主要由 MD5 特征和字符串特征构成。由于特征数据库体量较为庞大，若特征匹配时在云服务器端和终端直接交互将耗费大量的带宽资源，降低匹配的效率。因此本发明将特征数据库转化成为体量较小的特征向量。为了能够生成特征向量，云服务端分别对 MD5 特征和字符串特征进行不同的预处理，并通过 BF 映射得到恶意软件特征库的 MD5 特征向量 V_{md5} 和字符串特征向量 V_{str}。

本发明的分段 BF 映射机制是采用 256 个独立的 BF 向量 $V_i(0≤i≤255)$，每个 V_i 的长度为 m，每一位初始值为 0。根据特征首字节的内容将其通过唯一对应的 V_i 映射成为特征向量，且每个独立的 V_i 都只采用相同的公共哈希函数 H，即相当于把一个大长度的 BF

向量等分为256个小型BF向量，并按特征首字节进行分段映射，这样能够有效的降低误报率、特征映射和匹配的开销。例如，0号BF对应的特征首字节内容为“NULL”,65号BF对应的特征首字节内容为“A”。

针对MD5特征，若一个特征$X=\{x_1,x_2,\ldots,x_n\}$，映射过程包括如下两个步骤：

1．计算X的特征坐标。首先根据X的首字母x1的值，找到对应的BF向量V_{x_1}，再通过公共哈希函数计算特征在特征向量中的位置H(X)，我们把$L(X)=(x_1,H(X))$称为X的特征坐标。

2．将特征X映射到MD5特征向量V_{md5}中。即将X特征坐标的对应比特位置1，$V_{x_1,H(X)}=1$，若该位已经为1，则不操作。

对于字符串特征，每个特征的长度是不规则统一的，所以在映射成为字符串特征向量之前需要进行预处理。若$X=\{x_1,x_2,\ldots,x_n\}$是长度为n的字符串，特征映射过程包括如下三个步骤：

1．字符串特征切割。在本发明中为字符串特征设置一个长度为w的滑动窗口，将特征切割成为统一长度的特征片段。即按照w的滑动窗口切割后得出n-w+1个长度为w的片段，$X_1=\{x_1,x_2,\ldots,x_w\}$, $X_2=\{x_2,x_3,\ldots,x_{w+1}\}$,…, $X_{n-w+1}=\{x_{n-w+1},x_{n-w+2},\ldots,x_n\}$。

2．计算特征片段$X_p(1\le p\le n-w)$的特征坐标。考虑到切分后可能片段数目较多，通过特征映射所带来的计算量会较大，所以在字符串特征的映射过程中的公共函数H采用递归哈希函数$R(x_1,x_2,\ldots,x_w)$。递归哈希函数的计算是根据输入字符串X的内容决定的，而切割后得到的相邻的两个片段之中会有w-1长度的重叠部分，所以在递归式的哈希函数中，上一文件片段的哈希结果$R_p=R(x_p,\ldots,x_{p+w-1})$可用于下一片段的哈希结果$R_{p+1}=R(x_{p+1},\ldots,x_{p+w})$计算中，从而能够有效的减少计算带来的开销。比较常用的递归式哈希函数有Rabin指纹函数等。得到哈希结果后，再根据每个片段的首字母x_p的值找到对应的BF向量V_{x_p}，从而得出X_p的特征坐标$L(X_p)=(x_p, R_p)$。

3．将特征片段$X_p(1\le p\le n-w)$映射到字符串特征向量V_{str}中，即将X_p特征坐标的对应比特位置1，$V_{x_p,R_p}=1$，若该位已经为1，则不操作。

对于长度小于w的特征(称为短字符特征),其在总的字符串特征中所占的比例较小，所以都将这些特征在后续的特征模糊扫描中进行单独的扫描,对整个系统的性能影响较小。

云服务端将特征通过分段BF映射的过程中，将每个特征或特征片段与其特征坐标建

立映射关系，即给每个特征或特征片段 X 加上一个标签 T，$T_X=L(X)$，我们称 T_X 为映射记录。映射记录可以为之后的精确匹配提供准确快速的定位。

2）恶意软件特征向量推送。终端初始时需从云服务器端获得恶意软件特征数据库进行恶意软件的模糊扫描，考虑到终端数据带宽和资费的限制，云服务端只向终端推送特征向量，即云服务端在接收到终端推送请求后，特征向量 V_{md5} 和 V_{str} 压缩存储后推送给终端。由于特征向量属于大型稀疏型向量，采用常用的一些压缩方法，如 gzip、xz 等，就能达到较高的压缩率，从而减少交互的信息量。对于短字符特征，云服务器端同时将该特征集合压缩加密发送至终端。

3）恶意软件特征向量更新。云服务端负责更新特征数据库（包括短字符特征集），当恶意软件特征数据库有更新时，云服务端启动终端的特征向量增量更新。在现有恶意软件特征向量的基础上，服务器对新增的恶意软件特征进行步骤 1）处理，从而得到新的恶意软件特征向量，然后通过对新旧恶意软件特征向量进行异或运算得到恶意软件特征向量更新。云服务端将更新经压缩后推送给所有终端，终端接收到更新向量后再与本地的特征向量进行异或运算即可得出新的恶意软件特征向量，从而完成系统的特征向量更新。对于短字符特征更新，则单独的将需要更新的短字符特征集合压缩加密发送至终端。

二、恶意软件云扫描确认方法包括以下两个步骤：

1）终端模糊扫描。该部分在终端进行，目的是高效快速地筛选出嫌疑文件集 $S=\{S_{md5},S_{str}\}$ 以及其对应的嫌疑特征哈希坐标 $\prod=\{\prod_{md5}, \prod_{str}\}$，其中，$S_{md5}$ 为嫌疑文件的 MD5 值集合，$\prod_{md5}$ 为嫌疑 MD5 值对应的特征坐标集合，S_{str} 为嫌疑文件的字符碎片集合，$\prod_{str}$ 为嫌疑字符碎片对应的特征坐标集合。

在扫描时，对于 MD5 特征，将带扫描的文件通过 MD5 映射算法处理成为 MD5 值 Y，然后通过分段 BF 的映射机制得到 MD5 值对应的特征坐标 $L(Y)=(y_1,H(Y))$。若特征坐标 L(Y)在 MD5 特征向量 V_{md5} 中的对应的位值为 1，则表示该文件为嫌疑文件，将其 MD5 值 Y 加入到 S_{md5} 中，特征坐标 L(Y)插入到 $\prod_{md5}$。

对于字符串文件特征，首先对待扫描文件进行切分，同样设置一个长度为 w 的滑动窗口，从文件的第一个字节开始向后滑动。切分后得到若干个规整的文件碎片 F，将这些碎片采用模式匹配的方法通过短字符特征集的预扫描，如果扫描匹配，则直接可以将该字符串碎片列为恶意碎片；否则碎片通过同样的递归哈希函数进行映射处理，得到字符串文件特征坐标 $L(F)=(f_p, R_p)$。若 L(F)在 MD5 特征向量 V_{str} 中的对应的位值为 1，那么该碎片是嫌疑碎片，将碎片的值 F 插入到 S_{str} 中，特征坐标 L(F)插入到 $\prod_{str}$ 中。

由于特征向量所具有的性质，所有恶意文件都会被终端模糊扫描确认为嫌疑文件，但有可能正常文件也被误认为嫌疑文件，因此终端将 S 和∏发送给云服务器端进行确认。

2）云端精确扫描。云端在接收到终端发送的嫌疑文件集 S 和嫌疑特征坐标∏后进行精确扫描，根据∏中的特征坐标找到对应的特征匹配集合 M。

具体过程以 MD5 特征为例，对于任一嫌疑片段 $X \in S_{md5}$，其特征坐标为 $L(X) \in \prod_{md5}$，那么 X 的特征匹配集合 M_X 为{m | T_m=L(X),m 为 MD5 值特征码}。再将 X 与 M_X 中的每个 m 进行精确匹配，即比较两者的值是否相等，如果匹配成功，那么 X 被确认为恶意 MD5 特征，否则排除嫌疑。对于嫌疑文件碎片 $F \in S_{str}$，处理过程相同，在精确匹配时可采用一些经典的模式匹配算法，如 BM 和 AC 算法等。对 S 中的每个元素都进行精确匹配后，云服务端将匹配结果返回至终端，终端根据结果采取相应的安全措施。

本发明的技术方案具有如下优点：通过基于分段 BF 算法的特征模糊扫描，能大量减少匹配的特征数目，准确的定位嫌疑特征，有效的提高扫描的效率；在扫描过程中云服务端无需将特征码发送至终端，保护了服务提供商的核心利益；且终端只需将少量的文件碎片发送至服务端，保证了用户隐私泄露的风险较低。

附图说明

图 1 为本发明的恶意软件扫描架构示意图。

图 2 为本发明的云服务端模块的架构示意图。

图 3 为本发明的特征哈希模块的功能示意图。

图 4 为本发明的特征向量构成示意图。

图 5 为本发明的模糊扫描模块的结构示意图。

图 6 为本发明的精确扫描模块的结构示意图。

图 7 为本发明的结果反馈模块的结构示意图。

具体实施方案

本发明实施技术方案的基本构思是，针对终端现有恶意软件扫描技术无法兼顾扫描效率和用户隐私安全的问题提供一种扫描策略，既可以解决因传统安全扫描模式带来的运行效率低，资源耗费大的问题，也可以解决目前云计算扫描模式导致的用户和服务提供商隐私存在泄漏风险的问题，具有良好的应用前景。

下面结合说明书附图对本发明实施方案进行详细说明。

图 1 为本发明的恶意软件扫描架构示意图。本发明在提出的扫描策略的基础上，相应地提出了一种恶意软件扫描架构。架构由云服务端模块 101、特征哈希模块 102、模糊扫描模块 103、精确扫描模块 104 和结果反馈模块 105 组成。

其中，云服务端模块 101 用于维护和更新恶意特征数据库，记录恶意软件扫描日志，响应和处理终端的请求等。

特征哈希模块 102 用于根据分段 BF 算法将特征数据库中的特征哈希映射成为特征向量，组成特征，为模糊扫描模块提供特征数据匹配。

模糊扫描模块 103 用于将终端中的文件按照与特征哈希模块相同的方法映射成为文件向量，与特征中的特征向量进行匹配扫描。记录扫描结果，将匹配命中的文件碎片和特征向量发送至精确扫描模块。

精确扫描模块 104 用于根据模糊扫描结果，从特征数据库中提取出匹配命中的特征码，与文件碎片进行进一步的匹配确认，防止在模糊扫描中因分段 BF 算法假阳率造成的误报。

结果反馈模块 105 用于将精确扫描的结果记录在云服务端，并且发起更新特征数据库请求；将扫描结果返回给终端，并对确认为被恶意软件感染的文件进行清除、隔离或者粉碎等查杀动作。

扫描过程也由这些模块按顺序进行执行。其中 101、102、104 模块都由云服务端完成，103 模块由终端完成，105 模块则由云服务端和终端共同完成，结果最后返回至终端。终端与云服务端可以采用无线网络、移动互联网、短信或者彩信的方式进行通信。

图 2 为本发明的云服务端模块架构示意图，该模块由请求响应子模块 201、特征码维护子模块 202、恶意扫描记录子模块 203 组成。

请求响应子模块 201 用于处理来自终端和扫描过程中发出的请求，主要包括终端的连接请求、恶意软件扫描请求，扫描过程中产生的更新特征数据库请求，记录恶意扫描记录请求以及其他一些服务信息发布推送请求。

特征码维护子模块 202 用于更新和维护特征码数据库，该数据库主要包括两种类型的特征码：MD5 特征和字符串特征，其中 MD5 特征占总特征的 85%，字符串特征占 15%。特征哈希时将会对两种特征都进行哈希映射处理成为特征向量。

恶意扫描纪录子模块 203 用于纪录恶意扫描的结果信息，维护一段时间内的扫描历史记录。当收到终端发来的嫌疑文件片段时，如果命中恶意扫描纪录时则可跳过精确扫描步骤，直接返回扫描结果。

图 3 为本发明的特征哈希模块 102 的功能示意图。该模块的主要功能是将特征库中的

特征码映射成为特征向量，进而构成特征。由于特征库中包括 MD5 特征和字符串特征，所以对两种特征采取不同的方法进行映射。

图 4 为本发明恶意软件特征向量构成示意图，由 256 个 MD5 特征向量、256 个字符串特征向量组成，每个特征向量的长度为 2^{16} 位。

图 5 为本发明的模糊扫描模块 103 的结构示意图。该模块由文件哈希子模块 501、MD5 特征匹配子模块 502、字符串特征匹配子模块 503 和文件过滤子模块 504 共四个子模块构成。

文件哈希子模块 501 用于将待扫描的文件哈希成为文件特征坐标。由于特征分为 MD5 和字符串两类，所以也将文件哈希成为两种特征坐标，即 MD5 文件特征坐标和字符串文件特征坐标。对于 MD5 文件特征坐标，是采用 MD5 哈希算法，将文件映射成为 16 个字节的哈希值，相当于文件的摘要；再用发明内容中介绍的 MD5 特征向量生成的方法将 MD5 文件值映射为文件特征坐标。对于字符串特征坐标，则也采取相同的方法进行文件切割，得到若干个长度规整的文件片段，然后将这些片段通过发明内容中介绍的字符串特征映射方法映射成为字符串文件特征坐标。同时，模块 501 还建立起文件与文件特征坐标之间的对应关系，即文件映射记录。通俗的说，记录一个文件向量中哪几位是由哪个文件映射而来的，为之后溯源提供准备。

MD5 特征匹配子模块 502 用于完成 MD5 文件特征坐标与 MD5 特征向量 V_{md5} 的匹配。根据文件哈希子模块 501 的结果，将每个 MD5 文件特征坐标与模块 102 发送的 MD5 特征向量 V_{md5} 逐一进行匹配，若对应的值为 1，则说明该文件为嫌疑文件，需要进行下一步的精确扫描，否则为正常文件

字符串匹配子模块 503 用于完成字符串文件特征坐标与字符串特征向量 V_{str} 的匹配。首先对字符串文件片段进行短字符特征集预扫描，若匹配则直接确认为恶意文件并将结果提交至结果反馈模块 105，否则继续与 V_{str} 中的特征向量匹配，具体的过程与模块 502 类似。

文件过滤子模块 504 用于产生文件匹配的结果。对于 MD5 结果，根据模块 501 的映射记录，溯源出嫌疑的 MD5 和 MD5 值对应的文件。而对于字符串结果，则过滤出对应的文件片段和原始文件，从而筛选出嫌疑文件集 $S=\{S_{md5},S_{str}\}$ 以及其对应的嫌疑特征坐标集合 $\prod=\{\prod_{md5},\prod_{str}\}$，然后将 S 和 $\prod$ 发送至云服务端的精确匹配模块 104。这里值得说明的是，嫌疑的文件片段多为恶意软件的特征片段，由于分段 BF 假阳率的存在，可能将终端的用户隐私信息误报为嫌疑片段而进行发送。但是采用多个哈希函数进行映射后假阳率低

至可以忽略，从而使得用户的隐私能够得到很好的保护。

图 6 为本发明的精确扫描模块 104 的结构示意图。该模块主要由 MD5 文件精确扫描子模块 601 和字符串文件精确扫描子模块 602 构成。

MD5 文件精确扫描子模块 601 用于对模块 504 发至的 S_{md5},$\prod_{md5}$ 进行精确匹配，进一步确认嫌疑 MD5 值的恶意性。即对于任一嫌疑片段 $X \in S_{md5}$，其特征坐标为 $L(X) \in \prod_{md5}$，得出 X 的特征匹配集合 M_X 为{m | T_m=L(X),m 为 MD5 值特征码}。再将 X 与 M_X 中的每个 m 进行精确匹配，即比较两者的值是否相等，如果匹配成功，那么 X 被确认为恶意 MD5 特征，否则排除嫌疑。

字符串文件精确扫描子模块 602 用于对模块 504 发至的 S_{str}，$\prod_{str}$ 进行精确匹配，进一步确认嫌疑文件片段的恶意性。其过程与 601 模块的过程类似，采用典型的模式匹配算法 BM 和 AC 对嫌疑文件片段进行精确匹配。

图 7 为本发明的结果反馈模块 105 的结构示意图。该模块由特征码更新子模块 701 和结果返回子模块 702 构成。

特征码更新子模块 701 用于根据精确扫描的结果向云服务端提出特征数据库更新请求。在精确扫描后的文件片段中，可能片段中的若干个字节为恶意特征，其他字节为正常字节或者恶意特征的变种。对于这两种情况，本发明都视为新的恶意特征而更新到特征数据库中。

结果返回子模块 702 用于将精确扫描的结果返回至终端，并对被恶意软件感染的文件进行清除、隔离或者粉碎等查杀动作。

最后应说明的是，以上实施例仅用于说明本发明的技术方案而非限制。尽管参照实施例对本发明进行了详细说明，本领域的普通技术人员应当理解，对本发明的技术方案进行修改或者等同替换，都不脱离本发明技术方案的精神和范围，其均应涵盖在本发明的权利要求范围当中。

附图

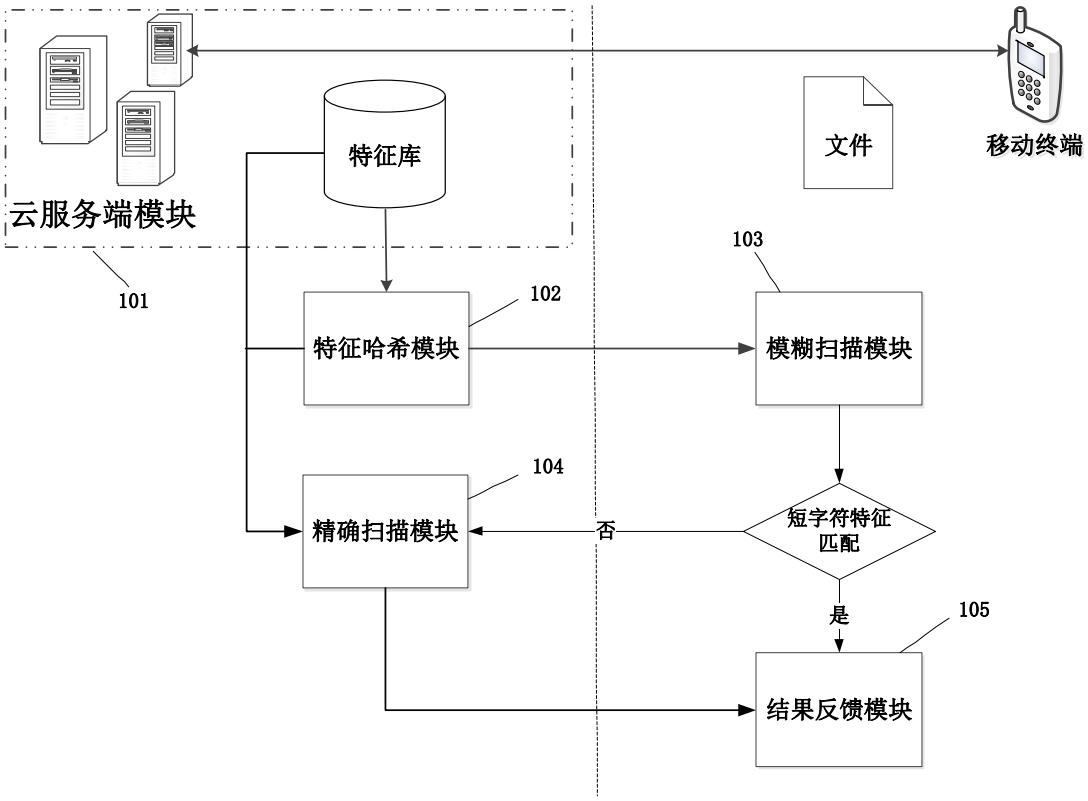
特征库
文件
移动终端
云服务端模块
101
103
102
特征哈希模块
模糊扫描模块
104
精确扫描模块
否
短字符特征匹配
是
105
结果反馈模块

图 1

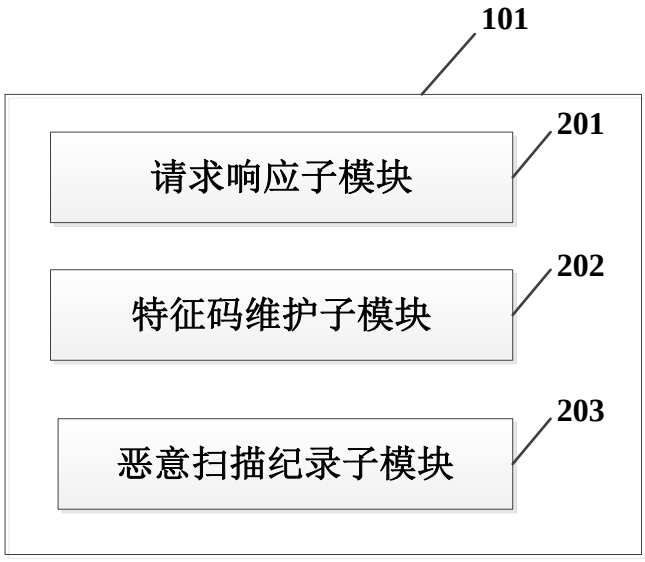
101
201
请求响应子模块
202
特征码维护子模块
203
恶意扫描纪录子模块

图 2

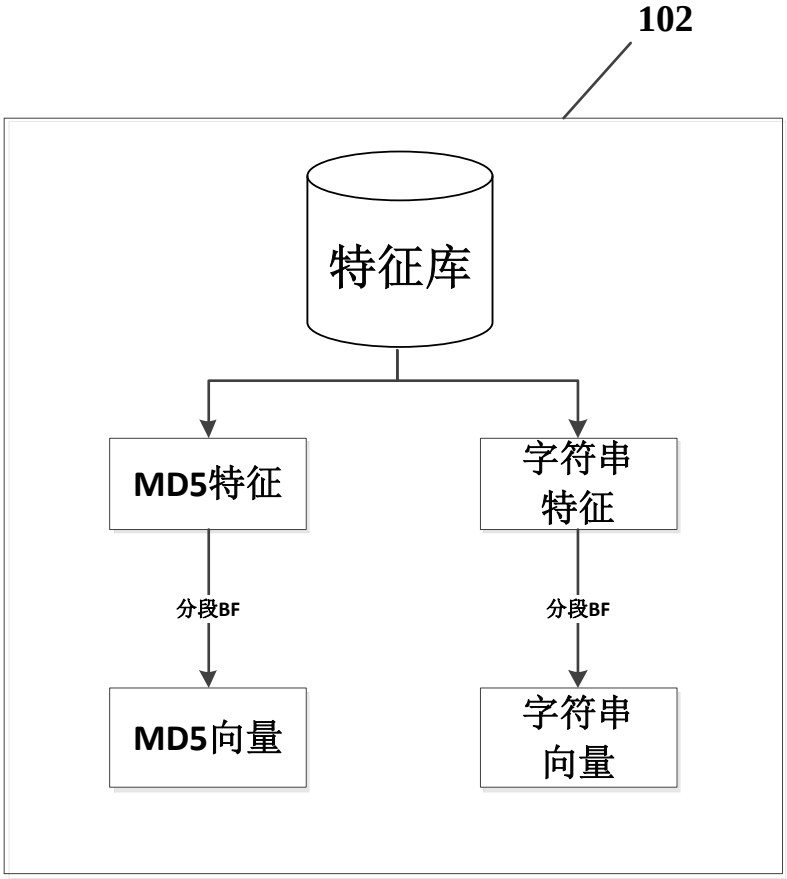
102
特征库
MD5特征
字符串特征
分段BF
分段BF
MD5向量
字符串向量

图 3

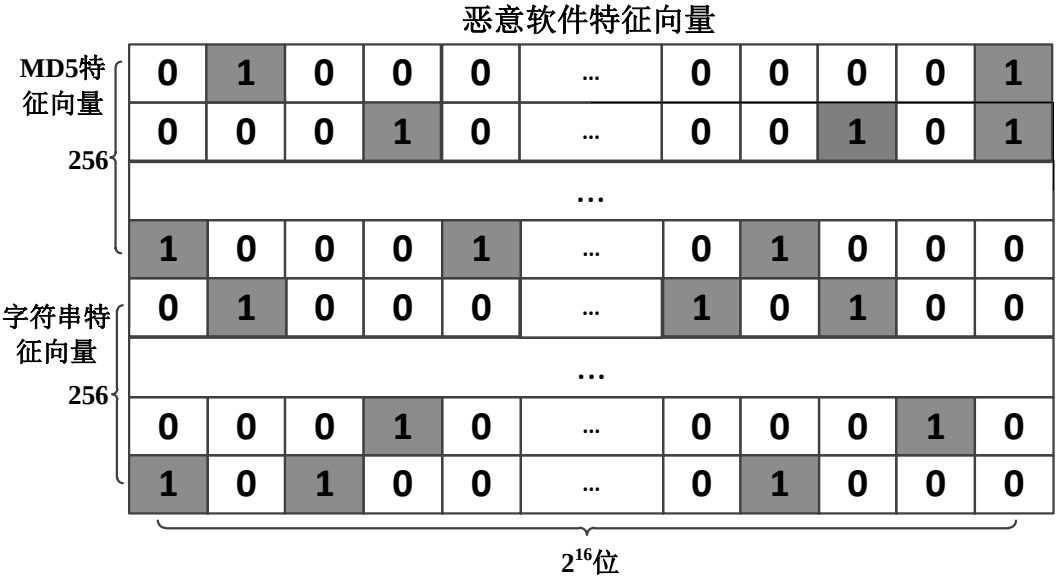
恶意软件特征向量
MD5特征向量
256
字符串特征向量
256
2^{16}位

图 4

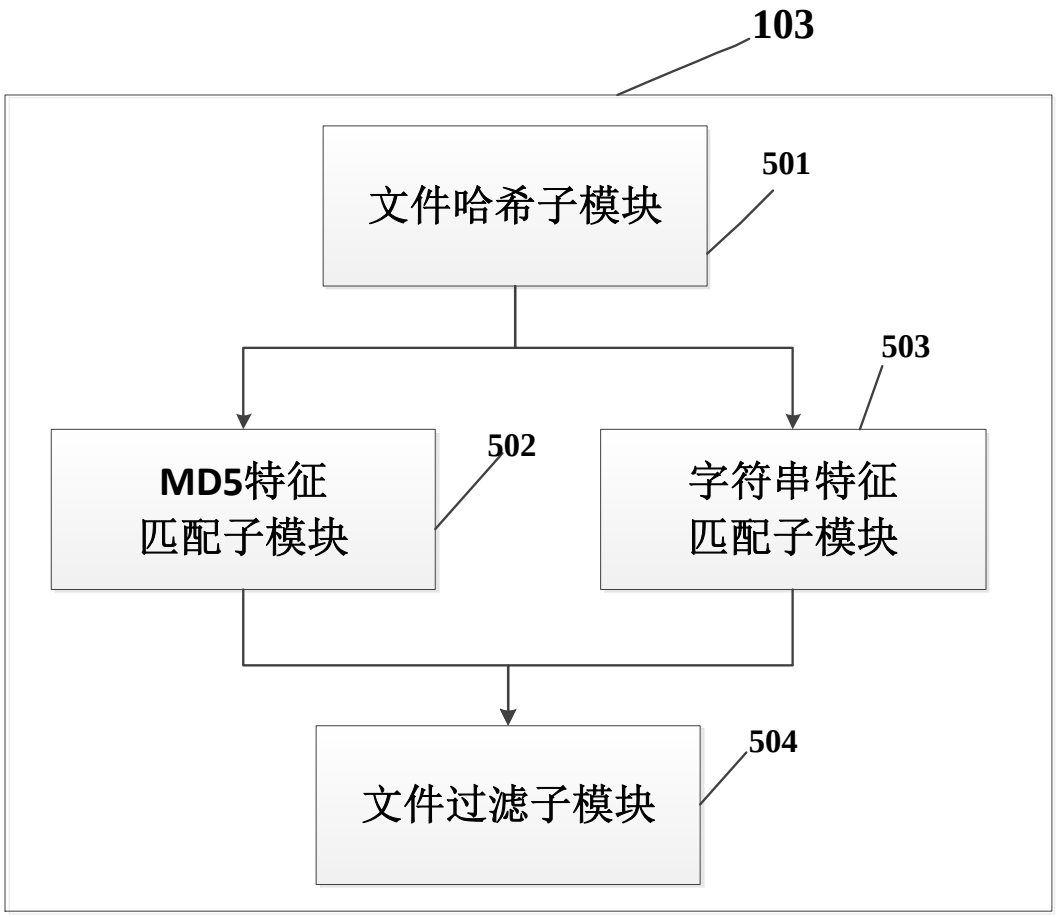
103
文件哈希子模块
501
503
MD5特征
匹配子模块
502
字符串特征
匹配子模块
文件过滤子模块
504

图 5

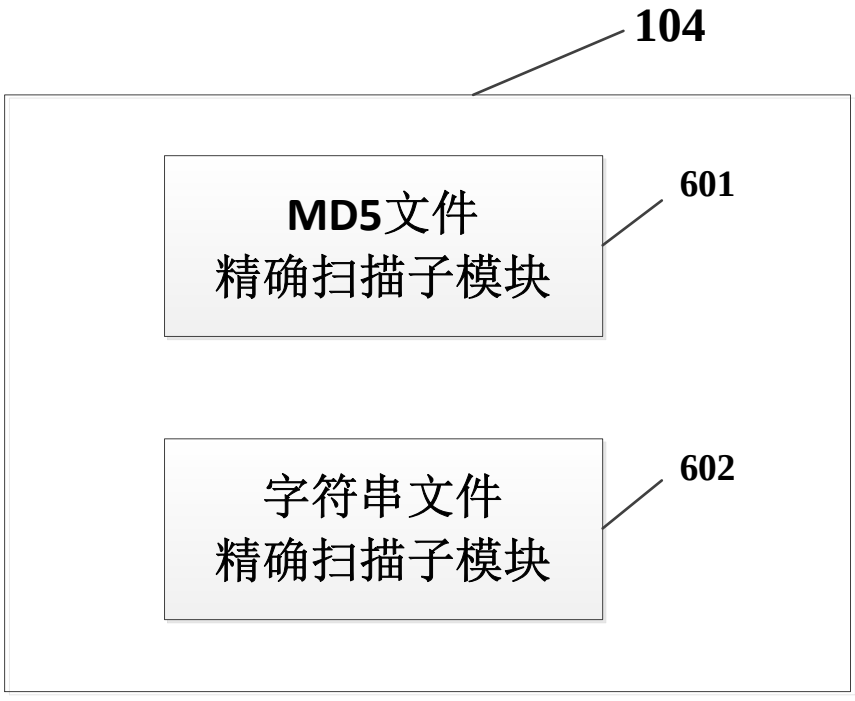
104
MD5文件
精确扫描子模块
601
字符串文件
精确扫描子模块
602

图 6

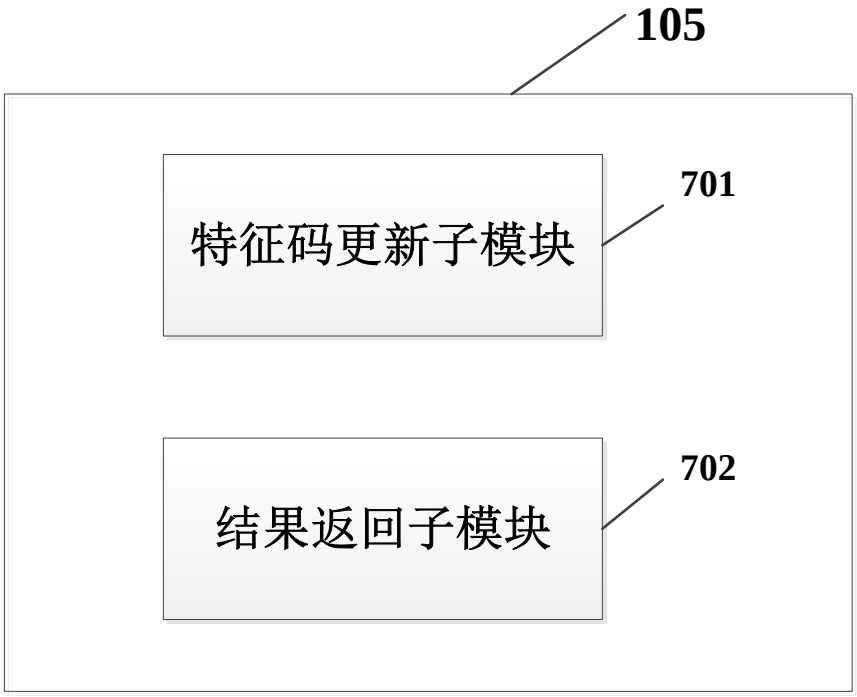

图 7