(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 101764748 A
(43)申请公布日 2010.06.30

(21)申请号 200910252775.7

(22)申请日 2009.12.16

(71)申请人 福建星网锐捷网络有限公司
地址 350002 福建省福州市仓山区金山大道618号桔园州工业园19#楼

(72)发明人 丁金生 余灿

(74)专利代理机构 北京同达信恒知识产权代理有限公司 11291
代理人 黄志华

(51)Int.Cl.
H04L 12/56(2006.01)
H04L 12/28(2006.01)

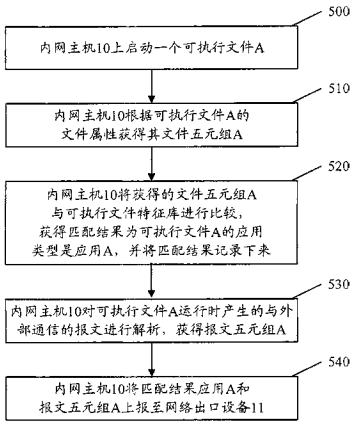
权利要求书 2 页 说明书 6 页 附图 4 页

(54)发明名称

一种应用程序识别方法、装置及系统

(57)摘要

本发明涉及互联网技术，公开了一种应用程序识别方法，用以在降低网络出口设备运行负荷的前提下，提高应用识别的准确性。该方法为：内网主机根据可执行文件的文件标识信息与预设的可执行文件特征库的匹配结果，确定所述可执行文件的应用类型，并解析出所述可执行文件产生的数据流的报文标识信息，并将所述应用类型和报文标识信息上报至网络出口设备，网络出口设备对应所述数据流的报文标识信息建立相应的流节点，并按照所述可执行文件的应用类型设置该流节点的优选级和流量控制策略。这样，既准确又快速地完成了数据流应用类型的判断，也极大地减轻了网路出口设备的运行负荷，本发明同时公开了一种用于应用程序识别的内网主机和局域网网络系统。



CN 101764748 A

1. 一种应用程序识别方法，其特征在于，包括：

内网主机启动可执行文件时，根据该可执行文件的文件属性获得其文件标识信息

所述内网主机将获得的文件标识信息与预设的可执行文件特征库进行匹配，获得匹配结果，所述可执行文件特征库用于记录文件标识信息与应用类型之间的对应关系；

所述内网主机根据所述匹配结果确定所述可执行文件的应用类型，以及在在所述可执行文件运行过程中产生与互联网交互的至少一种数据流时，解析出该至少一种数据流的报文标识信息，并将所述应用类型和所述报文标识信息上报至网络出口设备；

所述网络出口设备对应所述至少一种数据流的报文标识信息建立相应的流节点，并按照所述可执行文件的应用类型设置该流节点的优选级和流量控制策略。

2. 如权利要求 1 所述的方法，其特征在于，所述文件标识信息包括产品版本、产品名称、公司、文件版本和源文件名。

3. 如权利要求 1 所述的方法，其特征在于，还包括：所述网络出口设备预先将所述可执行文件特征库下发给所述内网主机，并且指示内网主机定期更新本地的可执行文件特征库。

4. 如权利要求 1 所述的方法，其特征在于，所述内网主机与所述网络出口设备归属于同一网段，或者，不归属于同一网段。

5. 如权利要求 1 ～ 4 任一所述的方法，其特征在于，还包括：所述内网主机在所述可执行文件结束执行后，通知所述网络出口设备删除所述相应的流节点。

6. 一种用于应用程序识别的内网主机，其特征在于，包括：

获取单元，在启动可执行文件时，根据该可执行文件的文件属性获得其文件标识信息；

匹配单元，将获得的文件标识信息与预设的可执行文件特征库进行匹配，获得匹配结果，所述可执行文件特征库用于记录文件标识信息与应用类型之间的对应关系；

解析上报单元，根据所述匹配结果确定所述可执行文件的应用类型，以及在所述可执行文件运行过程中产生与互联网交互的至少一种数据流时，解析出该至少一种数据流的报文标识信息，并将所述应用类型和所述报文标识信息上报至网络出口设备。

7. 一种用于应用程序识别的局域网网络系统，其特征在于，包括：

内网主机，用于在启动可执行文件时，根据该可执行文件的文件属性获得其文件标识信息，并将获得的文件标识信息与预设的可执行文件特征库进行匹配，获得匹配结果，所述可执行文件特征库用于记录文件标识信息与应用类型之间的对应关系，再根据所述匹配结果确定所述可执行文件的应用类型，以及在所述可执行文件运行过程中产生与互联网交互的至少一种数据流时，解析出该至少一种数据流的报文标识信息，并将所述应用类型和所述报文标识信息上报至网络出口设备；

网络出口设备，用于对应所述至少一种数据流的报文标识信息建立相应的流节点，并按照所述可执行文件的应用类型设置该流节点的优选级和流量控制策略。

8. 如权利要求 7 所述的网络系统，其特征在于，所述网络出口设备还用于预先将所述可执行文件特征库下发给所述内网主机，并指示内网主机定期更新本地的可执行文件特征库。

9. 如权利要求 7 所述的网络系统，其特征在于，所述内网主机与所述网络出口设备归

属于同一网段，或者，不归属于同一网段。

10. 如权利要求 7 ～ 9 任一所述的网络系统，其特征在于，所述内网主机还用于在所述可执行文件结束执行后，通知所述网络出口设备删除所述相应的流节点。

一种应用程序识别方法、装置及系统

技术领域

[0001] 本发明涉及互联网技术，特别涉及一种应用程序识别方法、装置及系统。

背景技术

[0002] 目前，在局域网网络环境中，出口带宽通常是有限的，而每个局域网内往往会存在一种或几种需要优选保证其运行流畅性的应用程序。例如，在企业局域网络中，需要优先保证企业管理软件、财务软件等应用程序的快速运转，以保证信息的快速传递；而在网吧局域网络中，需要优先保证网络游戏等应用程序的运行流畅性。

[0003] 可见，无论在何种局域网网络环境内，都需要尽量降低 P2P 应用对出口带宽的消耗，而为了实现此应用目的，就需要网络出口设备针对不同的应用程序采用不同的控制策略。目前，网络出口设备仅能根据应用程序运行时产生的不同数据流来对其进行识别，即网络出口设备必须将数据流和应用程序对应起来，才能准确识别出应用程序，才保证后续流程中对数据流的控制不会出错。

[0004] 现有技术下，一些常见的网络出口设备，例如，防火墙、路由器等等，主要通过模式匹配方式和端口识别方式来进行应用程序的识别。

[0005] 所谓模式匹配方式，即是将数据流携带的报文内容与系统已有的数据流特征库进行比较，从而确定数据流对应的应用程序（以下简称为应用）。具体为：通常情况下，一个应用可能包含多个数据流（如登录流，交互流等等），而每个数据流具有不同的特征（如，包含不同的特殊字符串）；因此，在采用模式匹配方式时，网络出口设备可以通过字符串匹配确定一个数据流对应的应用。采用模式匹配方式的优点是只需预先收集相关的数据集合形成数据流特征库即可以完成应用的识别，检测准确率较高。但是，采用模式匹配方式需要网络出口设备在报文转发的过程中，不断地分析每条数据流，匹配每条数据流的特征，从而大大加重了网络出口设备的运行负荷负担，降低了网络出口设备的转发性能；同时，为了保证识别的准确性，需要在建立数据流特征库时尽可能多地分析所有应用对应的所有数据流，找出每个数据流的特征，不能有所遗漏，这也加大了前期准备工作的负担。

[0006] 而所谓端口识别方式即是将数据流使用的源端口或目的端口与系统已有的端口数据库进行比较，从而确定数据流对应的应用。通常情况下，一个应用产生的数据流都是采用固定端口进行传输的，因此，采用端口识别方式，可以通过数据流的传输端口来确定该数据流对应的应用。采用端口识别方式的优点是只需要通过端口即可识别出数据流对应的应用，系统负担很小；但是，采用端口识别方式，网络出口设备对应用识别的准确性并不高，例如，针对通信端口不固定的情况网络出口设备不能够准确识别出应用（如 P2P 应用），又例如，对于非知名端口，以及一个端口对应多个应用的情况，网络出口设备也会出现严重的误判，不能够准确识别出应用。

发明内容

[0007] 本发明实施例提供一种应用程序识别方法、装置及系统，用以在降低网络出口设

备运行负荷的前提下，提高应用识别的准确性。

[0008] 本发明实施例提供的具体技术方案如下：

[0009] 一种应用程序识别方法，包括：

[0010] 内网主机启动可执行文件时，根据该可执行文件的文件属性获得其文件标识信息

[0011] 所述内网主机将获得的文件标识信息与预设的可执行文件特征库进行匹配，获得匹配结果，所述可执行文件特征库用于记录文件标识信息与应用类型之间的对应关系；

[0012] 所述内网主机根据所述匹配结果确定所述可执行文件的应用类型，以及在在所述可执行文件运行过程中产生与互联网交互的至少一种数据流时，解析出该至少一种数据流的报文标识信息，并将所述应用类型和所述报文标识信息上报至网络出口设备；

[0013] 所述网络出口设备对应所述至少一种数据流的报文标识信息建立相应的流节点，并按照所述可执行文件的应用类型设置该流节点的优选级和流量控制策略。

[0014] 一种用于应用程序识别的内网主机，包括：

[0015] 获取单元，在启动可执行文件时，根据该可执行文件的文件属性获得其文件标识信息；

[0016] 匹配单元，将获得的文件标识信息与预设的可执行文件特征库进行匹配，获得匹配结果，所述可执行文件特征库用于记录文件标识信息与应用类型之间的对应关系；

[0017] 解析上报单元，根据所述匹配结果确定所述可执行文件的应用类型，以及在所述可执行文件运行过程中产生与互联网交互的至少一种数据流时，解析出该至少一种数据流的报文标识信息，并将所述应用类型和所述报文标识信息上报至网络出口设备。

[0018] 一种用于应用程序识别的局域网网络系统，包括：

[0019] 内网主机，用于在启动可执行文件时，根据该可执行文件的文件属性获得其文件标识信息，并将获得的文件标识信息与预设的可执行文件特征库进行匹配，获得匹配结果，所述可执行文件特征库用于记录文件标识信息与应用类型之间的对应关系，再根据所述匹配结果确定所述可执行文件的应用类型，以及在所述可执行文件运行过程中产生与互联网交互的至少一种数据流时，解析出该至少一种数据流的报文标识信息，并将所述应用类型和所述报文标识信息上报至网络出口设备；

[0020] 网络出口设备，用于对应所述至少一种数据流的报文标识信息建立相应的流节点，并按照所述可执行文件的应用类型设置该流节点的优选级和流量控制策略。

[0021] 本发明实施例中，内网主机代替网络出口设备完成了针对可执行文件的应用类型的判断，并且内网主机是根据可执行文件 A 的文件属性所体现的文件标识信息与可执行文件特征库的匹配结果，对可执行文件的应用类型进行判断的。显然，通过在内网主机上提取文件标识信息以完成对可执行文件应用类型的匹配和识别，既准确又快速，极大地减轻了网路出口设备的运行负荷，保证了网络出口设备的性能不会下降，从而在整体上保证了网络性能不受影响。

附图说明

[0022] 图 1 为本发明实施例中文件五元组示意图；

[0023] 图 2 为本发明实施例中局域网网络环境示意图；

[0024] 图 3 为本发明实施例中内网主机功能结构图；

[0025] 图 4 为本发明实施例中内网主机网络注册流程图；

[0026] 图 5 为本发明实施例中内网主机对应用类型进行识别流程图。

具体实施方式

[0027] 在局域网网络环境中，为了在降低网络出口设备运行负荷的前提下，提高应用识别的准确性，本发明实施例中，内网主机启动可执行文件时，根据该可执行文件的文件属性获得其文件标识信息；所述内网主机将获得的文件标识信息与预设的可执行文件特征库进行匹配，获得匹配结果，所述可执行文件特征库用于记录文件标识信息与应用类型之间的对应关系；所述内网主机根据所述匹配结果确定所述可执行文件的应用类型，以及在在所述可执行文件运行过程中产生与互联网交互的至少一种数据流时，解析出该至少一种数据流的报文标识信息，并将所述应用类型和所述报文标识信息上报至网络出口设备；所述网络出口设备对应所述至少一种数据流的报文标识信息建立相应的流节点，并按照所述可执行文件的应用类型设置该流节点的优选级和流量控制策略。

[0028] 本发明实施例中，较佳地，采用文件五元组作为文件标识信息，所谓文件五元组是指 Windows 系统中可执行文件的版本属性中包含的多个信息，参阅图 1 所示，从文件五元组里取出其中几个关键信息用于对应用的识别，例如：产品版本、产品名称、公司、文件版本、源文件名。本发明实施例中，采用这五个信息唯一标志一个可执行文件，即一个应用。文件五元组提取过程简单，相对于根据数据流特征库对应用加以识别的模式匹配方式，可以大大节省前期准备的工作量。当然，还可以文件三元组、文件四元组等等作为文件标识信息，可以达到同样的技术效果，在此不再赘述。

[0029] 另一方面，较佳地，采用报文五元组作为报文标识信息，所谓报文五元组，即是指 IP 数据报文内包含的协议号，指明该报文的传输层协议是 UDP、TCP 或者是其他协议等等，如果是 TCP 或者 UDP，则 TCP/UDP 报文头还会包含源端口和目的端口两个字段。我们通常将 {IP 协议号，源 IP，源端口，目的 IP，目的端口 }5 个内容称之为一个 5 元组，在数据报文的转发过程中，一个 5 元组就可以标志出一条 TCP/UDP 的数据流。

[0030] 下面结合附图对本发明优选的实施方式进行详细说明。

[0031] 参阅图 2 所示，本发明实施例中，局域网网络中包括若干内网主机 10 和网络出口设备 11，其中，

[0032] 内网主机 10，用于在启动可执行文件时，根据该可执行文件的文件属性获得其文件标识信息，并将获得的文件标识信息与预设的可执行文件特征库进行匹配，获得匹配结果，所述可执行文件特征库用于记录文件标识信息与应用类型之间的对应关系，再根据所述匹配结果确定所述可执行文件的应用类型，以及在所述可执行文件运行过程中产生与互联网交互的至少一种数据流时，解析出该至少一种数据流的报文标识信息，并将所述应用类型和所述报文标识信息上报至网络出口设备 11；

[0033] 网络出口设备 11，用于对应所述至少一种数据流的报文标识信息建立相应的流节点，并按照所述可执行文件的应用类型设置该流节点的优选级和流量控制策略。

[0034] 参阅图 3 所示，本发明实施例中，内网主机 10 包括获取单元 101、匹配单元 102 和解析上报单元 103，其中，

[0035] 获取单元 101，在启动可执行文件时，根据该可执行文件的文件属性获得其文件标

识信息；

[0036] 匹配单元 102，将获得的文件标识信息与预设的可执行文件特征库进行匹配，获得匹配结果，所述可执行文件特征库用于记录文件标识信息与应用类型之间的对应关系；

[0037] 解析上报单元 103，根据所述匹配结果确定所述可执行文件的应用类型，以及在所述可执行文件运行过程中产生与互联网交互的至少一种数据流时，解析出该至少一种数据流的报文标识信息，并将所述应用类型和所述报文标识信息上报至网络出口设备 11。

[0038] 基于上述网络架构，本发明实施例，可以在网络出口设备 11 上预设一个用于识别应用程序的可执行文件特征库，该可执行文件特征库用于记录每个应用与文件五元组之间的映射关系。可执行文件特征库由网络出口设备 11 下发到内网主机 10 上，内网主机 10 每启动一个可执行文件前都要提取出该可执行文件的文件五元组并与可执行文件特征库进行匹配。内网主机 10 和网络出口设备 11 可以在同一个网段，也可以不在同一个网段，只要能正常通信即可。具体包括：内网主机 10 每次启动一个新的可执行文件前，提取该可执行文件的文件五元组，并与可执行文件特征库进行匹配，将匹配结果和该可执行文件使用到的数据流的报文五元组信息通告给网络出口设备 11。网络出口设备 11 根据内网主机 10 上报的报文五元组建立数据流节点，以及根据内网主机 10 上报的匹配结果设定数据流的应用类型和应用优先级，为后续应用控制提供依据。而当某个可执行文件（即某个应用）结束执行后，内网主机 10 通知网络出口设备 11，删除与该可执行文件相关的数据流节点和优先级设定。

[0039] 为了实现上述交互，本发明实施例中，在内网主机 10 上需要设置专用的客户端才能保证与网络出口设备 11 的通信。如果某台内网主机 10 上运行的客户端未连接至网络出口设备 11，但是网络出口设备 11 却收到该内网主机 10 发送的数据流，则网络出口设备 11 立即进行警告提示。内网主机 10 上使用的可执行文件特征库来自网络出口设备 11 的下发，这样既可以保证内网主机 10 上的可执行文件特征库的时效性，同时也简化了网络管理员升级特征库时的工作量。

[0040] 基于上述交互过程，本发明实施例中，假设网络出口设备 11 的 IP 为 IP_{router}，内网主机 10 的 IP 为 IP_{PC}，管理员在内网主机 10 上安装好指定软件后，设置其服务端 IP 为 IP_{router}，那么，参阅图 4 所示，本发明实施例中，内网主机 10 开机启动后进行网络注册的详细流程如下：

[0041] 步骤 400：内网主机 10 启动后，获取本地保存的可执行文件特征库的版本号，并根据预设的服务端 IP 向网络出口设备 11 通告上线信息和本地的可执行文件特征库的版本号。

[0042] 步骤 410：网络出口设备 11 对内网主机 10 的 IP，即 IP_{PC} 进行注册。

[0043] 步骤 420：网络出口设备 11 将内网主机 10 上报的可执行文件特征库的版本号与本地最新的可执行文件特征库版本号进行比较，如果两者不相同，则执行步骤 430；如果两者相同，则执行步骤 450。

[0044] 步骤 430：网络出口设备 11 向内网主机 10 通知内网主机 10 更新可执行文件特征库；接着执行步骤 440。

[0045] 步骤 440：内网主机 10 从网络出口设备 11 上下载最新版本的可执行文件特征库对本地的可执行文件特征库进行更新，并向更新结果通知网络出口设备 11；接着，进行步

骤 350。

[0046] 步骤 450 :网络出口设备 11 向内网主机 10 通告注册成功。

[0047] 基于上述实施例,参阅图 5 所示,本发明实施例中,内网主机 10 完成网络注册后,在启动一个可执行文件时,对该可执行文件的应用类型进行识别的详细流程如下 :

[0048] 步骤 500 :内网主机 10 上启动一个可执行文件,即开启了一个应用程序,本实施例中,将其称为可执行文件为 A。

[0049] 步骤 510 :内网主机 10 根据可执行文件 A 的文件属性获得其文件五元组 A。

[0050] 步骤 520 :内网主机 10 将获得的文件五元组 A 与可执行文件特征库进行比较,获得匹配结果为可执行文件 A 的应用类型是应用 A,并将匹配结果记录下来。

[0051] 步骤 530 :内网主机 10 对可执行文件 A 运行时产生的与外部通信的报文进行解析,获得报文五元组 A。

[0052] 步骤 540 :内网主机 10 将匹配结果应用 A 和报文五元组 A 上报至网络出口设备 11。

[0053] 本实施例中,网络出口设备 11 接收到内网主机 10 上报的应用 A 和报文五元组 A 后,会对应该报文五元组 A 建立新的流节点 A,并对应流节点 A 记录内网主机 10 上报的可执行文件 A 的应用类型,即应用 A,以及根据应用 A 设置相应的控制优先级,并关联相应的流量控制策略,从而开始对可执行文件 A 产生的数据流进行流量控制。

[0054] 下面以一个具体的实施列对网络出口设备 11 执行的后续操作进行详细介绍。

[0055] 假设内网主机对可执行文件 A 的应用类型进行识别后,可执行文件 A 在运行过程中产生与外部通信的数据流,网络出口设备 11 到可执行文件 A 对外发起的第一个数据流,获取内网主机 10 报的针对该第一个数据流的报文五元组,查找发现并没有对应该五元组建立的流节点,则建立新的五元组节点 A_1 接着,网络出口设备 11 接收到可执行文件 A 对外发起的第二个数据流,获取内网主机 10 报的针对该第二个数据流的报文五元组,查找发现并没有对应该五元组建立的流节点,则建立新的五元组节点 A_2 ;网络出口设备 11 根据内网主机 10 上报的第一数据流和第二数据流的应用匹配结果,获知其应用类型均为应用 A,则设置五元组节点 A_1 和 A_2 的应用类型为应用 A,并设置相应的执行优先级和对应的数据流控制策略,以对第一数据流和第二数据流进行流量控制。可执行文件 A 运行结束后,内网主机 10 上报网络出口设备 11,网络出口设备 11 会将五元组节点 A_1 和 A_2 进行删除。

[0056] 本实施例中,当网络出口设备 11 接收确定上述内网主机 10 下线时,会对其注册信息进行注销处理,并删除对应该内网主机 10 建立的流节点 A。如,根据某个数据流的报文五元组确定该数据流的源 IP 已非注册的内网主机 10 的 IP,则记录相关信息,并终止该数据流。另一方面,当网络出口设备 11 在设定时间内未收到某个内网主机 10 上报的保活报文,则也会对该内网主机 10 的注册信息进行注销。

[0057] 当然,在步骤 530 中,若可执行文件 A 没有立即产生与外部通信的报文,则内网主机 10 也可以仅仅先将应用 A 上报至网络出口设备 11,网络出口设备 11 会对应应用 A 先建立相应的流节点,设置对应的优选级和流量控制策略,并在后续接收到可执行文件 A 产生的数据流时,根据已建立的流节点对接收的数据流进行控制。

[0058] 另一方面,网络出口设备 11 也需要维持最新的可执行文件特征库,在内网主机 10 上线过程中,当可执行文件特征库有所更新时,需要随机通知内网主机 10 进行同步更新。

[0059] 通过上述实施例,内网主机 10 代替网络出口设备 11 完成了针对可执行文件 A 的

应用类型的判断，并且内网主机 10 是根据可执行文件 A 的文件属性所体现的文件五元组 A 与可执行文件特征库的匹配结果，对可执行文件 A 的应用类型进行判断的。显然，通过在内网主机 10 上提取文件五元组 A 以完成对可执行文件 A 应用类型的匹配和识别，既准确又快速，极大地减轻了网路出口设备 11 的运行负荷，保证了网络出口设备 11 的性能不会下降，从而在整体上保证了网络性能不受影响；在后续流程中，已被识别的可执行文件 A 产生新的数据流时，内网主机 11 只需要对网络出口设备 11 上报新数据流的报文五元组即可，无需进行重复识别。本发明实施例提供的技术方案简单易用，特别适合在类似于企业、网吧等环境的局域网网络中使用，可以得到很好的应用效果。

[0060] 显然，本领域的技术人员可以对本发明中的实施例进行各种改动和变型而不脱离本发明的精神和范围。这样，倘若本发明实施例中的这些修改和变型属于本发明权利要求及其等同技术的范围之内，则本发明中的实施例也意图包含这些改动和变型在内。

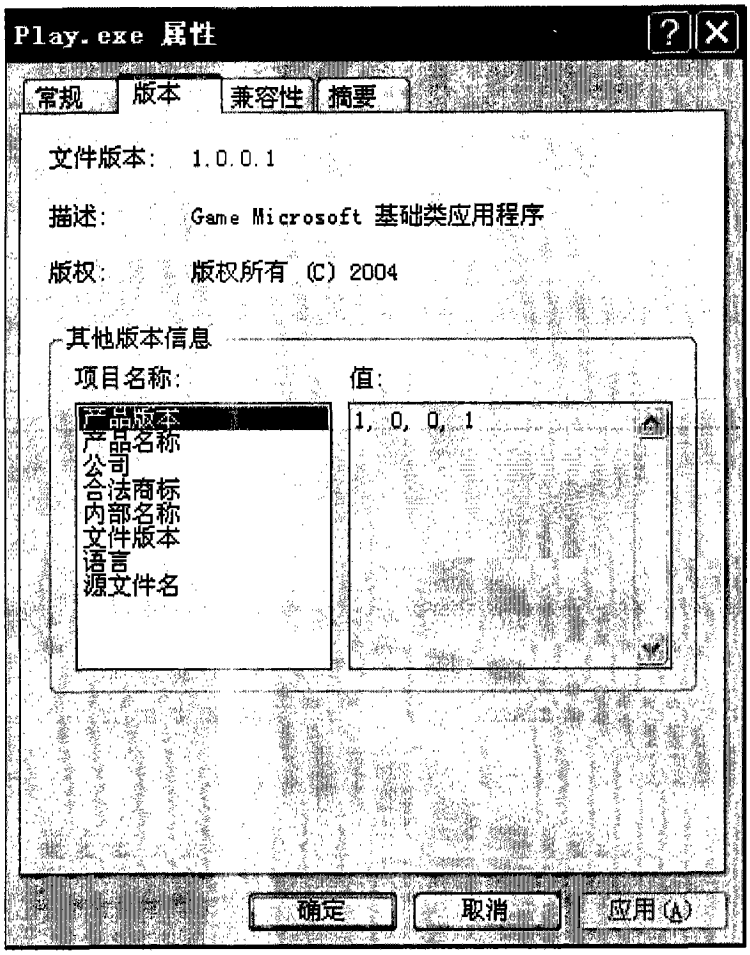

Play.exe 属性
常规
版本
兼容性
摘要
文件版本: 1.0.0.1
描述: Game Microsoft 基础类应用程序
版权: 版权所有 (C) 2004
其他版本信息
项目名称:
值:
产品版本
产品名称
公司
合法商标
内部名称
文件版本
语言
源文件名
1, 0, 0, 1
确定
取消
应用(A)

图 1

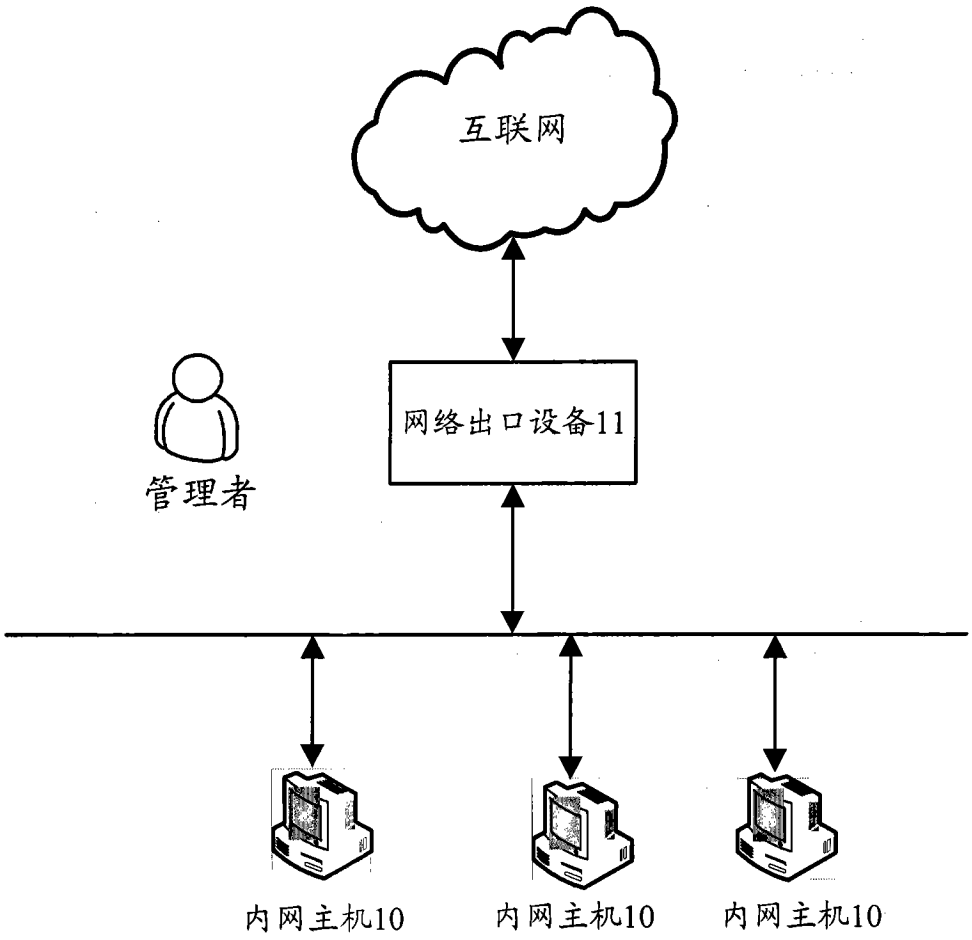

互联网
网络出口设备11
管理者
内网主机10
内网主机10
内网主机10

图 2

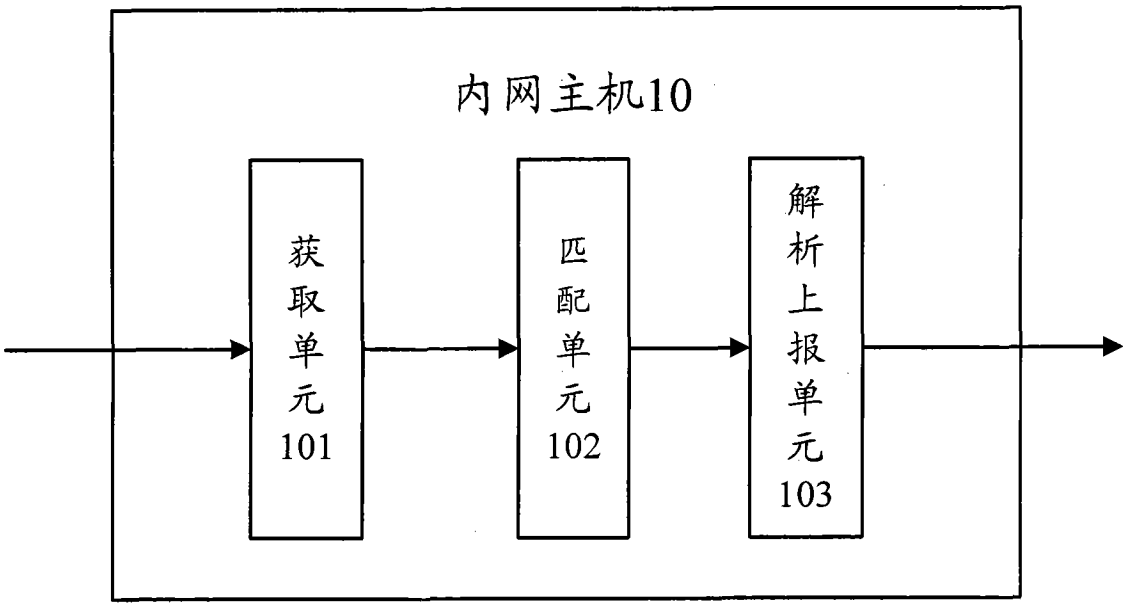

内网主机10
获取单元101
匹配单元102
解析上报单元103

图 3

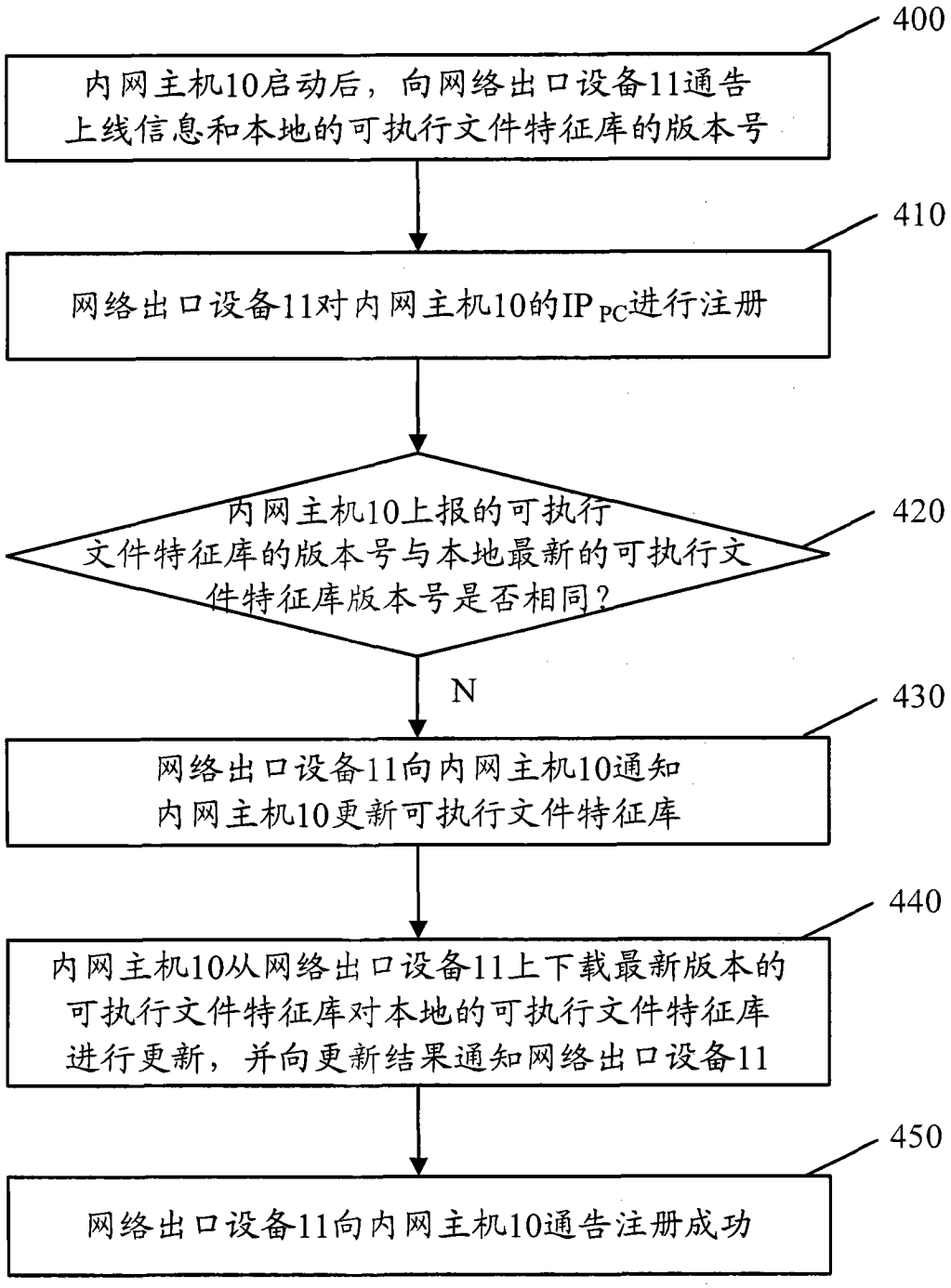

400
内网主机10启动后，向网络出口设备11通告上线信息和本地的可执行文件特征库的版本号
410
网络出口设备11对内网主机10的IP PC进行注册
420
内网主机10上报的可执行文件特征库的版本号与本地最新的可执行文件特征库版本号是否相同？
N
430
网络出口设备11向内网主机10通知内网主机10更新可执行文件特征库
440
内网主机10从网络出口设备11上下载最新版本的可执行文件特征库对本地的可执行文件特征库进行更新，并向更新结果通知网络出口设备11
450
网络出口设备11向内网主机10通告注册成功

图 4

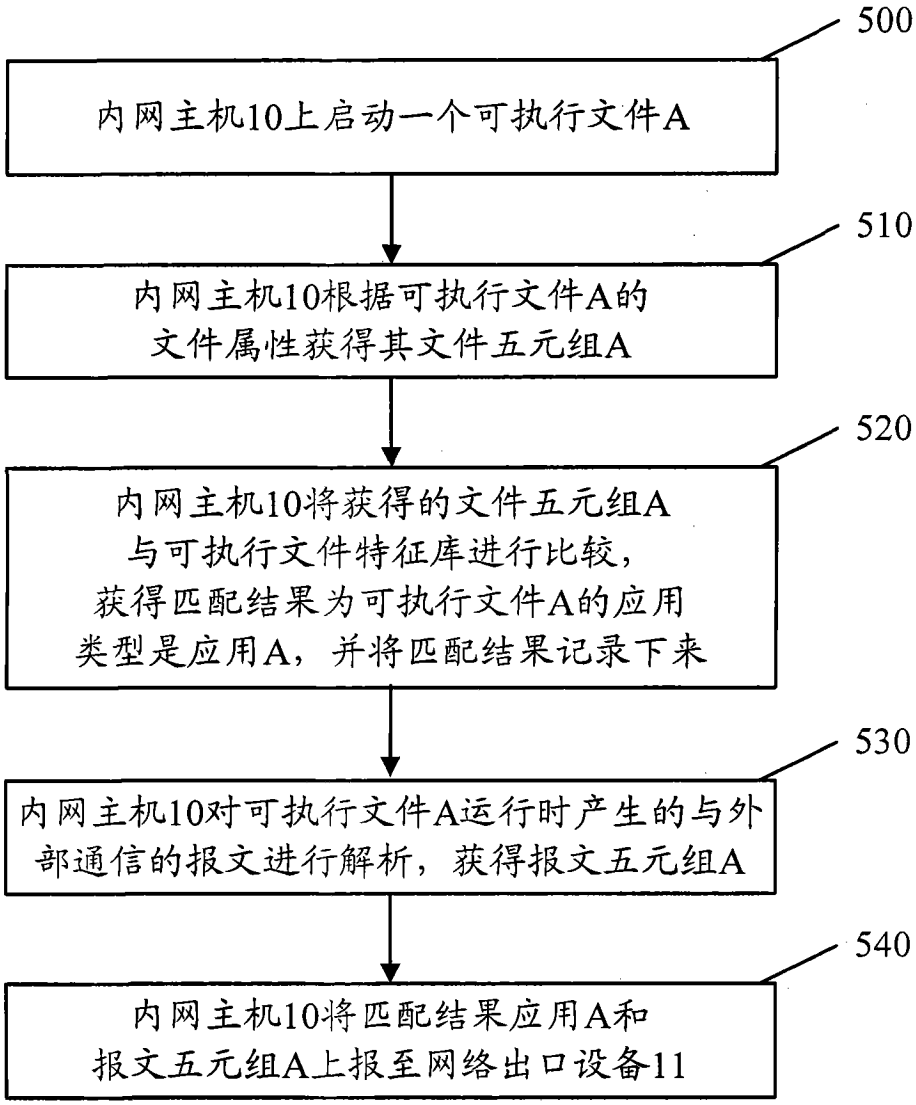

500
内网主机10上启动一个可执行文件A
510
内网主机10根据可执行文件A的
文件属性获得其文件五元组A
520
内网主机10将获得的文件五元组A
与可执行文件特征库进行比较，
获得匹配结果为可执行文件A的应用
类型是应用A，并将匹配结果记录下来
530
内网主机10对可执行文件A运行时产生的与外
部通信的报文进行解析，获得报文五元组A
540
内网主机10将匹配结果应用A和
报文五元组A上报至网络出口设备11

图 5