(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 102752303 B
(45)授权公告日 2015.06.17

(21)申请号 201210232858.1
(22)申请日 2012.07.05
(73)专利权人 北京锐安科技有限公司
地址 100044 北京市海淀区中关村南大街乙56号方圆大厦9层
(72)发明人 梁源 史延涛
(74)专利代理机构 北京君尚知识产权代理事务所(普通合伙) 11200
代理人 余功勋
(51)Int.Cl.
H04L 29/06(2006.01)
H04L 29/08(2006.01)

(56)对比文件
CN 101247432 A,2008.08.20, 说明书第11页第5段至最后一段、附图7.
CN 101115004 A,2008.01.30, 说明书第3页.

审查员 石璐

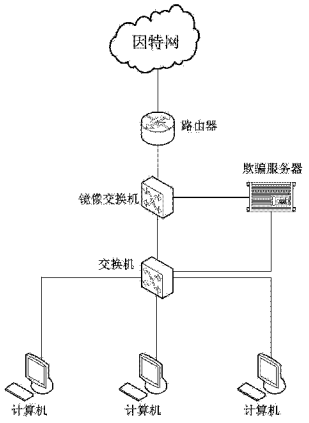
权利要求书2页 说明书5页 附图4页

(54)发明名称

一种基于旁路的数据获取方法及系统

(57)摘要

本发明提供一种基于旁路的数据获取方法及系统。在一网域出口设置一镜像交换机,并以旁路接入的方式设置一欺骗服务器;镜像交换机对发出网域的网络访问请求数据包进行镜像,并传输至欺骗服务器;欺骗服务器向目标计算机发送欺骗数据包,将目标计算机访问的页面替换成仿制页面;仿制页面对目标计算机在该仿制页面中提交的数据进行加密,并发送出所述网域;欺骗服务器通过镜像交换机获得所述加密数据并进行解析还原。本发明可获取基于HTTPS加密协议的网站的登录数据,为公安和政府机关对犯罪嫌疑人的取证提供帮助。



CN 102752303 B

1. 一种基于旁路的数据获取方法，其步骤包括：

1) 在一网域的出口设置一镜像交换机，并以旁路接入的方式设置一欺骗服务器，使所述欺骗服务器连接所述镜像交换机和所述网域；所述欺骗服务器内存储真实网站的仿制页面，所述仿制页面基于真实网页的源代码构建，并在其中插入不影响页面显示效果的控制代码；

2) 所述镜像交换机对目标计算机发出所述网域的网络访问请求数据包进行镜像，并将镜像数据传输至所述欺骗服务器；

3) 所述欺骗服务器接收所述镜像数据，按照仿制页面和所述网络访问请求数据包构造欺骗数据包，并使用目标计算机所访问的真实网站的源 IP 地址和源端口，基于数据链路层向目标计算机发送欺骗数据包，将所述目标计算机访问的页面替换成仿制页面；

4) 所述仿制页面对所述目标计算机在该仿制页面中提交的数据进行加密，通过该仿制页面中的控制代码构造一个独立的链接，并通过该独立的链接将加密数据发送出所述网域；

5) 所述镜像交换机对所述加密数据进行镜像并将镜像的数据传输至所述欺骗服务器，所述欺骗服务器对所述加密数据进行解析还原。

2. 如权利要求 1 所述的方法，其特征在于，所述网域为局域网或广域网。

3. 如权利要求 1 所述的方法，其特征在于，所述欺骗服务器按照所述仿制页面和目标计算机访问请求数据包的报文首部字段构造所述欺骗数据包。

4. 如权利要求 1 所述的方法，其特征在于，所述欺骗数据包采用 TCP/IP 协议，包括：

IP 报文首部，其源 IP 地址是所述网络访问请求数据包中 IP 报文首部的目的地址，其目的 IP 地址是所述网络访问请求数据包中 IP 报文首部的源地址；

TCP 报文首部，其源端口是所述网络访问请求数据包中 TCP 报文首部的目的端口，其目的端口是所述网络访问请求请求数据包中 TCP 报文首部的源端口。

5. 如权利要求 1 所述的方法，其特征在于，步骤 4) 采用可逆的加密算法进行所述加密。

6. 如权利要求 1 所述的方法，其特征在于，所述欺骗服务器根据解析还原后的数据进行统计分析，所述解析还原后的数据包括：访问网站的 URL、访问时间、源 IP 地址、目的 IP 地址、浏览器和操作系统类型。

7. 一种采用权利要求 1 所述方法的基于旁路的数据欺骗系统，其特征在于，包括一镜像交换机和一欺骗服务器；

所述镜像交换机设于一网域的出口处，用于对目标计算机发出所述网域的网络访问请求数据包进行镜像，并将镜像数据传输至所述欺骗服务器；

所述欺骗服务器以旁路接入的方式连接所述网域和所述镜像交换机，所述欺骗服务器内 存储真实网站的仿制页面，所述仿制页面基于真实网页的源代码构建，并在其中插入不影响页面显示效果的控制代码；所述欺骗服务器接收所述镜像数据，按照仿制页面和网络访问请求数据包构造欺骗数据包，并使用所述目标计算机所访问的真实网站的源 IP 地址和源端口，基于数据链路层向发出网络访问请求数据包的目标计算机发送欺骗数据包；所述欺骗数据包将所述目标计算机访问的页面替换成仿制页面；所述仿制页面对所述目标计算机在该仿制页面中提交的数据进行加密，并通过该仿制页面中的控制代码构造一个独立的链接，通过该独立的链接将加密数据发送出所述网域；所述欺骗服务器通过所述镜像交

换机获得所述加密数据并进行解析还原。

8. 如权利要求 7 所述的系统，其特征在于，所述镜像交换机设于所述网域的出口的路由器和交换机之间 。

9. 如权利要求 7 所述的系统，其特征在于，所述欺骗服务器连接所述网域的路由器或者交换机。

一种基于旁路的数据获取方法及系统

技术领域

[0001] 本发明属于计算机网络安全技术领域，具体涉及一种数据获取方法及系统，从旁路对网络内部的计算机发起攻击，并在其访问特定网页时使用网络钓鱼的方法获取特定数据，便于公职部门进行调查取证。

背景技术

[0002] 对于传输敏感数据的网络，一般采用一定的措施来保证数据安全，例如对网络中传输的数据内容进行监控，避免敏感数据的泄露，或者避免某些特定的敏感信息被截取。这种监控措施根据接入网络的方式，采用的方法有：1. 在网络出口将监控设备串联进网络，对流经监控设备的数据进行解析还原；2. 在网络出口使用数据镜像的方法将网络内数据接入至监控设备，对接入监控设备的数据进行解析还原。这两种方法的缺点是：无法对加密数据或加密协议进行实时的解析还原。

[0003] 一个局域网内部的计算机是通过路由器来访问因特网，所有进出局域网内部的通讯全部要经过路由器或者与之相连的交换机。监控设备通过对进入局域网的数据包进行解析还原，获取并记录数据包中传输的内容。如图 1 所示，底侧表示局域网内及网上的用户，其间通过交换机与路由器相连接；上侧表示局域网通过路由器访问因特网，路由器与交换机中间的设备及线路表示使用串联的方式将监控设备接入到网络中，所有进出局域网的数据全部都要流经监控设备；右侧的线路及设备表示使用旁路的方式将监控设备接入到网络中，通过使用交换机的镜像功能，将交换机与路由器的通讯是数据镜像并接入至监控设备，监控设备可得到全部进出局域网的数据。监控设备可抓取全部进出局域网的数据，并对非加密数据进行实时的还原解析，但是对于使用强加密算法加密过的数据，无法实时或在短周期内对其进行破解，即便最终破解了加密数据并得知其内容，也为时已晚。

[0004] 为了能够获取这些加密数据的内容，采用的方式一般有两种：一是在网络出口处架设一个中间人代理，代理局域网内计算机的全部访问行为，并截取其通讯内容，这种方法的缺点在于，中间人在代理基于 HTTPS 协议连接时，需要一个数字证书以进行交互，这个证书通常是伪造的，浏览器会对此进行显式的提示，隐蔽性较差，如果使用真实的证书，则制造证书所花费的成本又过于巨大，这又是难以承受的。另一种方法是架设一个钓鱼网站，并对一些特定的真实网站进行仿制，局域网内计算机访问特定的真实网站时，强制其访问钓鱼网站，以骗取其提交的数据，这种方法的缺点除了上述中间人方法的假证书问题外，还非常容易暴露钓鱼网站的位置，并且还存在 COOKIE 的无法跨域的问题，其会导致计算机在向钓鱼网站提交数据后，无法正常登录真实网站，同样隐蔽性较差。

发明内容

[0005] 本发明的目的是针对上述问题，提供一种基于旁路的数据获取方法及系统，主要针对基于 HTTP 协议、HTTPS 协议的访问链接，对访问特定网页的计算机进行攻击，并通过网络钓鱼的方式获取指定的数据，并实时记录。公安部门可通过系统对这些数据进行分析对

比，如查找其中的敏感词，判别网络中用户是否具有威胁，对犯罪嫌疑人进行取证；同时这些数据，如网站的登录口令等，可被提供给公安干警登录网站，进一步获取犯罪人员的遗留信息。

[0006] 为实现上述目的，本发明采用如下技术方案：

[0007] 一种基于旁路的数据获取方法，包括以下步骤：

[0008] 1）在一网域的出口设置一镜像交换机，并以旁路接入的方式设置一欺骗服务器，使所述欺骗服务器连接所述镜像交换机和所述网域；

[0009] 2）所述镜像交换机对发出所述网域的网络访问请求数据包进行镜像，并将镜像数据传输至所述欺骗服务器；

[0010] 3）所述欺骗服务器接收所述镜像数据并向发出所述网络访问请求数据包的目标计算机发送欺骗数据包，将所述目标计算机访问的页面替换成仿制页面；

[0011] 4）所述仿制页面对所述目标计算机在该仿制页面中提交的数据进行加密，并将加密数据发送出所述网域；

[0012] 5）所述镜像交换机对所述加密数据进行镜像并将镜像的数据传输至所述欺骗服务器，所述欺骗服务器对所述加密数据进行解析还原。

[0013] 进一步地，所述网域为局域网或广域网。

[0014] 进一步地，所述仿制页面基于真实网页的源代码构建，并在其中插入不影响页面显示效果的控制代码；所述控制代码构造一个独立的链接，实现将加密数据发送出所述网域。

[0015] 进一步地，所述欺骗服务器按照所述仿制页面和目标计算机访问请求数据包的报文首部字段构造所述欺骗数据包；所述欺骗数据包采用 TCP/IP 协议，包括：IP 报文首部，其源 IP 地址是所述网络访问请求数据包中 IP 报文首部的目的地址，其目的 IP 地址是所述网络访问请求数据包中 IP 报文首部的源地址；TCP 报文首部，其源端口是所述网络访问请求数据包中 TCP 报文首部的目的端口，其目的端口是所述网络访问请求请求数据包中 TCP 报文首部的源端口。

[0016] 进一步地，步骤 4）采用可逆的加密算法进行所述加密。

[0017] 进一步地，所述欺骗服务器根据解析还原后的数据进行统计分析，所述解析还原后的数据包括：访问网站的 URL、访问时间、源 IP 地址、目的 IP 地址、浏览器和操作系统类型。

[0018] 一种基于旁路的数据欺骗系统，包括一镜像交换机和一欺骗服务器；

[0019] 所述镜像交换机设于一网域的出口处，用于对发出所述网域的数据进行镜像，并将镜像数据传输至所述欺骗服务器；

[0020] 所述欺骗服务器以旁路接入的方式连接所述网域和所述镜像交换机，用于接收所述镜像数据并向发出网络访问请求数据包的目标计算机发送欺骗数据包；所述欺骗数据包将所述目标计算机访问的页面替换成仿制页面；所述仿制页面对所述目标计算机在该仿制页面中提交的数据进行加密，并将加密数据发送出所述网域；所述欺骗服务器通过所述镜像交换机获得所述加密数据并进行解析还原。

[0021] 进一步地，所述镜像交换机设于所述网域的出口的路由器和交换机之间

[0022] 进一步地，所述欺骗服务器连接所述网域的路由器或者交换机。

[0023] 本发明主要用于公安系统以及政府网络安全监控部门，提供对基于 HTTPS 加密协议的网站的登录数据的获取，通过对这些数据分析对比，为公安和政府机关对犯罪嫌疑人的取证提供帮助，使其能够更快的掌握嫌疑人传播的信息，并防止非法信息的进一步扩散。本发明能够提供极高的隐蔽性和安全性，使得被欺骗目标难以发现和进行溯源，保障了公安系统以及政府网络安全监控部门取证的隐蔽性和对取证类系统的安全性要求。如配合其他监控系统的使用，可极大的弥补其他监控系统对基于加密协议传输的数据无法获取的缺陷。具体来说，本发明的优点和积极效果如下：

[0024] 1）本发明中使用旁路接入的方式，只处理镜像后的进出网络的数据，即便系统损坏也不会对接入的网络造成影响。

[0025] 2）欺骗服务器对目标计算发送的欺骗数据包，源 IP 地址和源端口使用的不是欺骗服务器的 IP 地址和端口，而是目标计算机所访问真实网站的；并且欺骗服务器是基于数据链路层进行发包的，本身无需绑定 IP 地址，在网络中无法对欺骗服务器进行搜索，也无法通过欺骗数据包进行溯源，极大地提高了系统的隐蔽性。

[0026] 3）所使用的仿制的假页面在向外发送数据时，使用的是独立的链接，不影响页面的原始链接。即目标计算机可正常使用网页所提供的功能，也不影响网页原有的后续跳转流程，加强了骗取过程的隐蔽性。由于页面是预先仿制的，可定制抓取数据的种类，方式灵活多变，仿制的假页面发送的独立链接，在目标计算机向真实网站提交网页数据的瞬时发出，保证了系统解析数据的实时性。

[0027] 4）本发明的数据获取方法，并非是对加密后的数据进破解，而是在目标计算机提交数据前，即在数据被加密之前将页面数据放入独立的链接并使用私有加密算法加密后发送出去。而这些数据之后仍会随网页的原始链接，如基于 HTTPS 协议的链接等，发送至真实的网站，系统解析的是假页面所发出的私有加密协议加密过的数据，而非原始数据。故系统既能够获取基于 HTTPS 协议的加密链接的传输数据，也能够获取基于 HTTP 协议的非加密链接的传输数据，还能够应对将数据先用可逆或不可逆加密算法加密后再进行传输的情况，即达到了欺骗并获取网络中计算机访问特定网站数据的效果。

附图说明

[0028] 图 1 是现有技术中将监控设备接入到网络中进行监控的示意图

[0029] 图 2 是本发明实施例的基于旁路的数据获取系统的网络架构示意图。

[0030] 图 3 是本发明实施例的基于旁路的数据获取方法的步骤流程图。

[0031] 图 4 是本发明实施例的欺骗数据包结构示意图。

具体实施方式

[0032] 下面通过具体实施例并配合附图，对本发明做详细的说明。

[0033] 图 2 本实施例的基于旁路的数据获取系统的网络架构示意图，上部结构表示局域网通过路由器接入因特网，底部为局域网用户，局域网用户通过路由器访问因特网。如该图所示，该系统包括一台镜像交换机和一台欺骗服务器，镜像交换机架设在局域网出口处的路由器和交换机之间，用于对发出局域网的网络访问请求进行镜像，并传输至欺骗服务器；欺骗服务器以旁路接入的方式连接镜像交换机，并接入局域网，可架设在局域网中的任意

位置，用于接入镜像交换机镜像的数据，并向发出网络访问请求的目标计算机发送欺骗数据包，将所述目标计算机访问的页面替换成仿制页面。仿制页面通过私有加密算法对所述目标计算机在该仿制页面中提交的数据进行加密，并将加密数据发送出所述局域网。然后镜像交换机对加密数据进行镜像并传输至所述欺骗服务器；欺骗服务器对加密的数据进行解析还原，并记录。

[0034] 上述的欺骗服务器，采用可以发送和接收原始 TCP/IP 数据包的服务器，能够任意修改网络数据包的中信息和接收任意网络数据包。对于接入的数据，必须在网络的出口处将数据镜像，并接入欺骗服务器，只有在网络出口处才能将网络内部计算机上网的数据接全。这里的网络可以是局域网，也可以是广域网等任何一个网域。

[0035] 上述的欺骗服务器，要尽可能得靠近目标计算机，即目标计算机所访问的网站返回的响应数据包要比欺骗服务器向目标计算发送的欺骗数据包更晚到达。现实中，一般是使欺骗服务器和被欺骗目标的物理线路更短来实现。欺骗的网站一般都在网络外部，而欺骗服务器是在网络内部（发包线连在网络内部），所以欺骗服务器距离目标会更“近”（数据传输的距离更短）。除此之外，也可以通过提高线路的传输速度来实现，比如是用光纤代替普通使用的网线等。

[0036] 图 3 是本发明实施例的基于旁路的数据获取方法的步骤流程图。具体说明如下：

[0037] 1）局域网出口处的路由器和与之相连的交换机之间设置一台镜像交换机，将镜像交换机的镜像口，即输出镜像数据的接口与欺骗服务器的网络接口相连，并使用镜像交换机的镜像功能将进出局域网的数据进行镜像，发送至欺骗服务器；欺骗服务器可设置在局域网中任意位置，除了与镜像交换机相连接，还需使用另一个网络接口连接至局域网内，无需绑定 IP 地址。因为欺骗服务器上接入镜像数据线路的网卡只用于抓包，发包是从另一块网卡发出，所以需要将欺骗服务器的发包网卡用另一条线接到局域网中，用于发送欺骗包到目标计算机。具体地，欺骗服务器可通过与路由器或者交换机相连而接入局域网，图 2 所示为欺骗服务器与交换机连接的方式。

[0038] 2）欺骗服务器以真实的网页源码为基础，在其中插入特定的不影响页面显示效果的控制代码，仿制特定的网站的网页，并将仿制后的假页面预先存放在欺骗服务器上，欺骗服务器会将这些假页面加载至内存中随时备用。需要做仿制页面的网站，可由客户来指定。

[0039] 3）局域网内目标计算机访问因特网特定网站的特定网页，访问网页的请求数据包经过镜像交换机时，被镜像并发送至欺骗服务器。

[0040] 4）欺骗服务器会解析还原上述请求数据包，如果欺骗服务器中没有目标计算机所访问网页的假页面，则不进行任何动作，如果有，并且请求数据包请求的是预先指定的需要进行欺骗的网站，则使用请求数据包中的一部分数据（报文首部字段）和预先仿制假页面的数据构造欺骗数据包。

[0041] 图 4 是欺骗数据包的结构示意图。正常的 TCP/IP 数据包结构包括 IP 报文首部、TCP 报文首部。IP 报文首部又包括 4 位版本号、4 位报头长度、8 位服务类型、16 位总长度、16 位标识等等。TCP 报文首部包括 16 位源端口、16 位目的端口、32 位序列号等等。欺骗数据包的结构和正常数据包的结构的区别在于：其 IP 报文首部中，源 IP 地址是上述请求数据包 IP 报文首部中的目的地址，即目标计算机所访问网站的 IP，目的 IP 地址是上述请求数据包中 IP 报文首部中的源地址，即目标计算机的 IP 地址；其 TCP 报文首部中，源端口是上

述请求数据包中 TCP 报文首部的目的端口，即目标计算机访问网站所使用的端口，目的端口是上述请求数据包中 TCP 报文首部的源端口，即目标计算机所访问网站的访问端口。其 TCP 报文数据段为经过压缩的仿制的假网页。

[0042] 5）欺骗服务器将上述结构的欺骗数据包发送至目标计算机。由于欺骗数据包 IP 报文首部和 TCP 报文首部的结构，与目标计算机所访问网站返回的响应数据包 IP 报文首部和 TCP 报文首部的结构几乎一致，并且欺骗服务器在局域网内部，而目标计算机访问的网站在局域网外部，目标计算机接收到网站返回的响应数据包的时间比其接收到欺骗数据包的时间要晚，根据 TCP 协议可靠传输的原理，目标计算机先接收了欺骗数据包，对欺骗数据包进行验证并认为其是网站返回的响应数据包，之后，在网站的响应数据包到达目标计算机时，会被直接丢弃，因此目标计算机最终显示的网页是欺骗服务器发送过来的仿制的假页面，而并非是其原本所请求的页面。

[0043] 6）目标计算机显示上述的假页面后，计算机使用者进行一系列的操作，在网页提交数据前的一瞬，假页面中的控制代码会将还未提交的数据复制并使用私有算法加密打标，然后放入一个单独的链接发送至局域网外部，并且不会对这个链接的响应做出任何动作，之后数据仍会被提交到它需要提交的网站上去，网页会根据网站的响应结果做出对应的动作，即控制代码不影响网页原本的功能和流程。

[0044] 上述加密算法可以使用任意可逆的加密算法，无论是对称还是非对称的，而不可逆的 MD5、SHA1 等是不能用的。由于使用公用的算法很容易就被第三方猜到并直接把内容解析了，所以可以采用特别设计的算法，比如：把字符还原成字符编码（如 ASCII 码值），根据该字符编码随机生成另一位数较长的字符串。可以通过下述方式生产位数较长的字符串：

[0045] 首先生成一个随机数 i（范围 1-9），然后按照顺序将每个字符还原成 ASCII 码，每个字符所在字符串中的位置记为 j；然后判断 j 是不是 2 的倍数，如果不是则将字符的 ASCII 码值减去 j 除以 i 的余数（即 ASCII 码值 -j%i），如果是 2 的倍数则加上 j 除以 i 的的余数（即 ASCII 码值 +j%i）；然后将变化后的 ASCII 码值的位数记为 Z，将变换后的值乘以 10 再加上 Z（例如，如果变化后的 ASCII 码值为 87 则按照算法得出结果为 872），此为这个字符最终变换后所得数字；最后一步是将这个数字转成字符，按照以上顺序将每个字符（字母）转成 ASCI I 码的变化字符再按顺序拼到一起，所得结果就是一串很长的数字字符串（例如 432562314533）。当然也可以设计成其它算法，其主要目的是防止数据被第三方解析并取走。

[0046] 7）网页提交的数据经过镜像交换机，被镜像并发送至欺骗服务器，欺骗服务器对数据进行解析还原，如果是上述打标的加密数据则对其进行解密并记录，如果不是则会按照欺骗流程中 1）所述的步骤，判别是否要发送欺骗数据包。

[0047] 另外，还可以在欺骗服务器解析还原进出局域网的数据包时，记录下访问网站的 URL、访问时间、源和目的 IP 地址、浏览器和操作系统类型等信息，并进行统计分析，总结出网络内计算机的上网规律和环境，为网络的管理提供更多的数据支撑和依据。

[0048] 以上实施例仅用以说明本发明的技术方案而非对其进行限制，本领域的普通技术人员可以对本发明的技术方案进行修改或者等同替换，而不脱离本发明的精神和范围，本发明的保护范围应以权利要求所述为准。

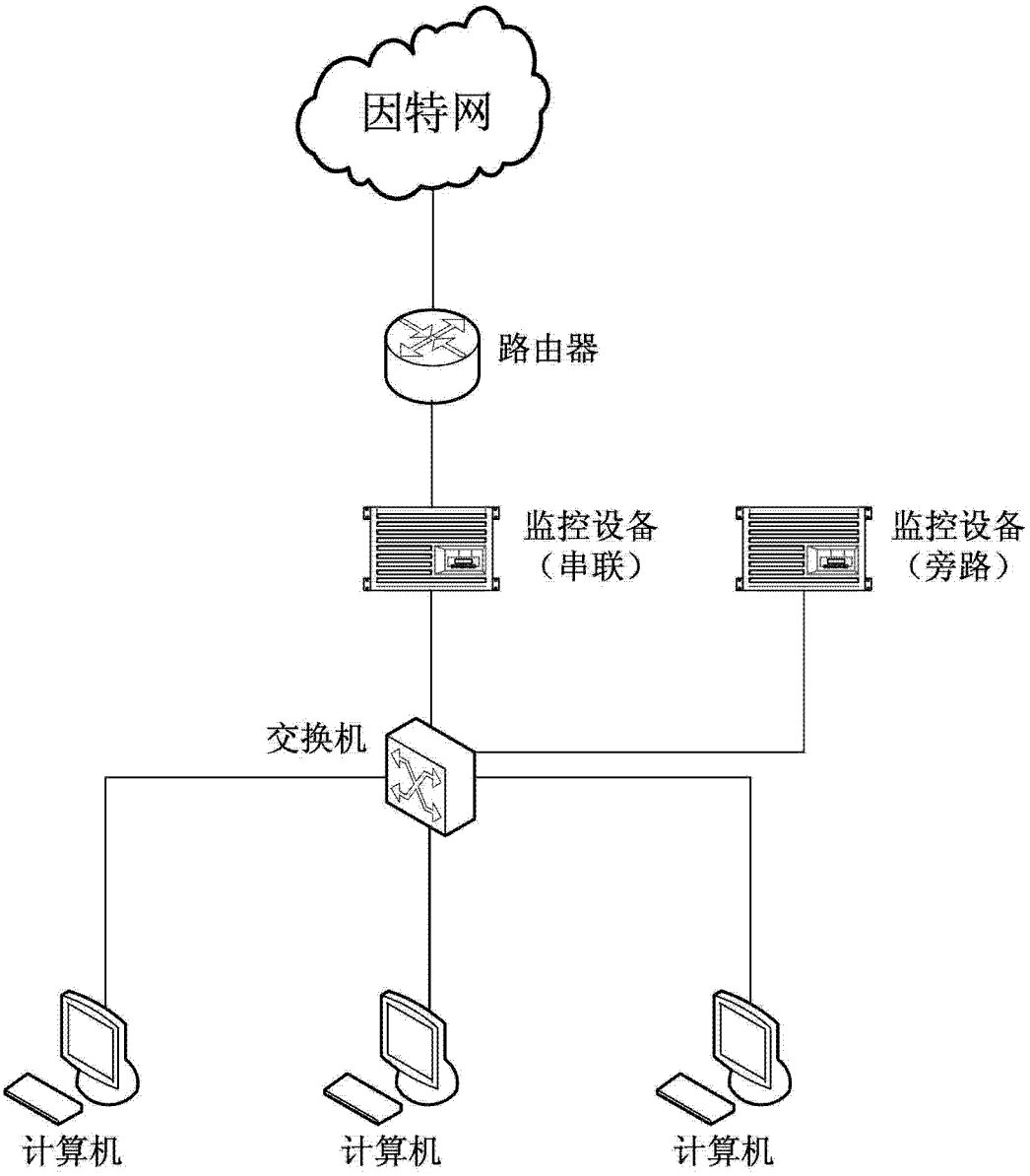


图 1

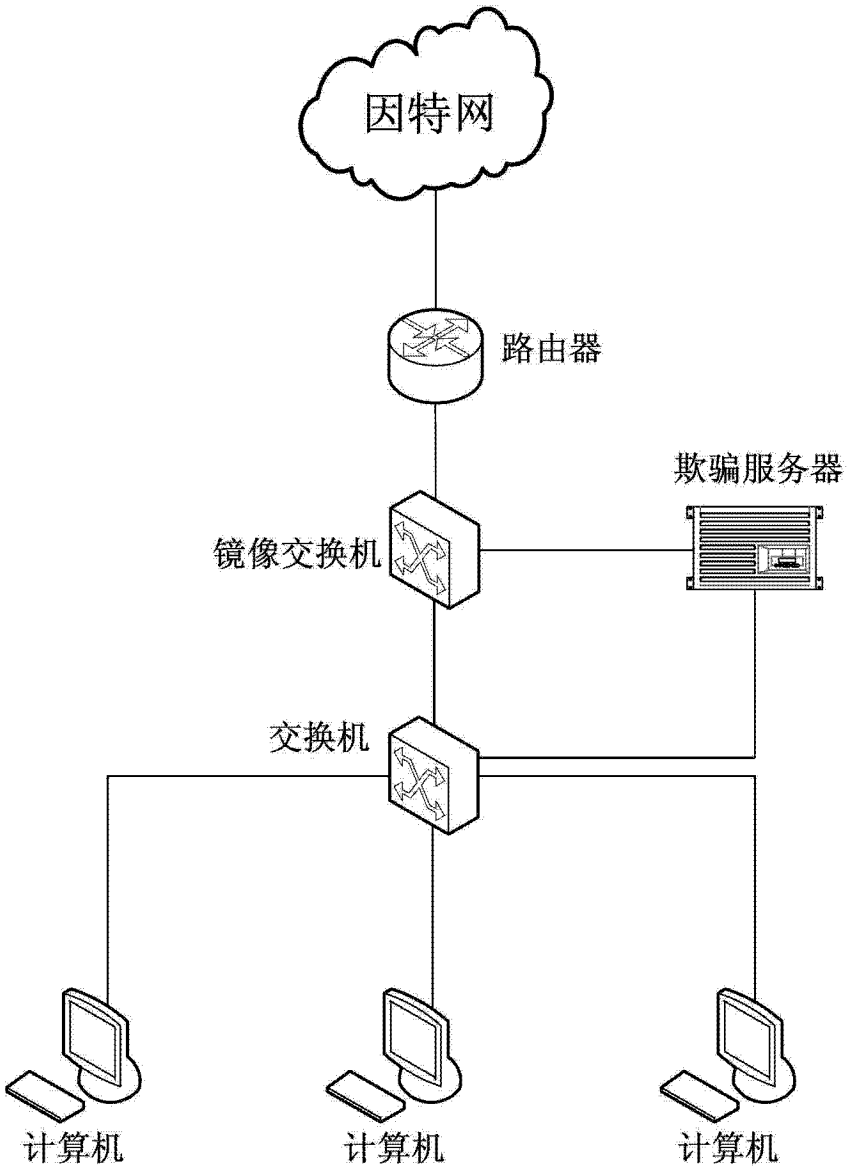


图 2

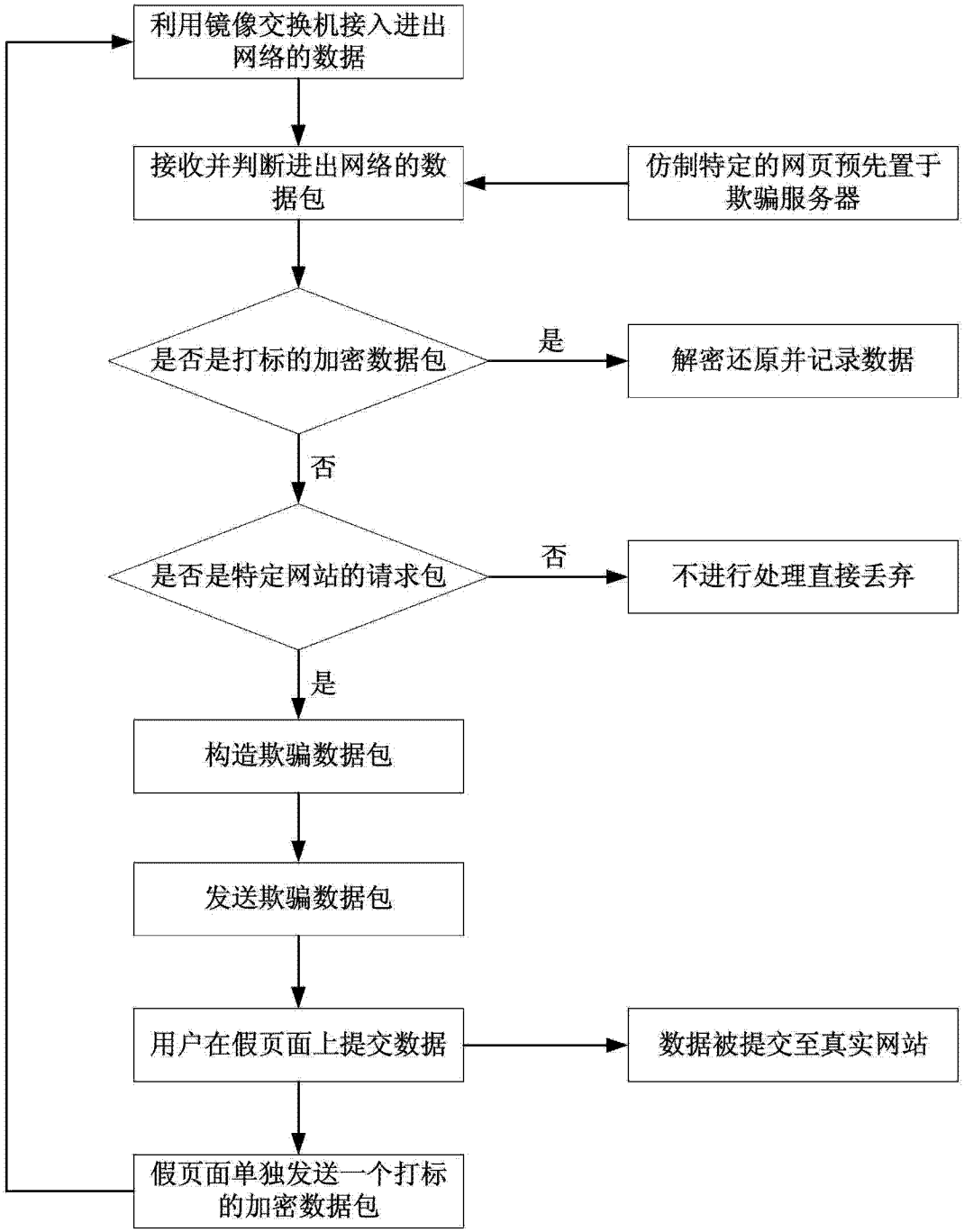

利用镜像交换机接入进出网络的数据
接收并判断进出网络的数据包
仿制特定的网页预先置于欺骗服务器
是否是打标的加密数据包
是
解密还原并记录数据
否
是否是特定网站的请求包
否
不进行处理直接丢弃
是
构造欺骗数据包
发送欺骗数据包
用户在假页面上提交数据
数据被提交至真实网站
假页面单独发送一个打标的加密数据包

图 3

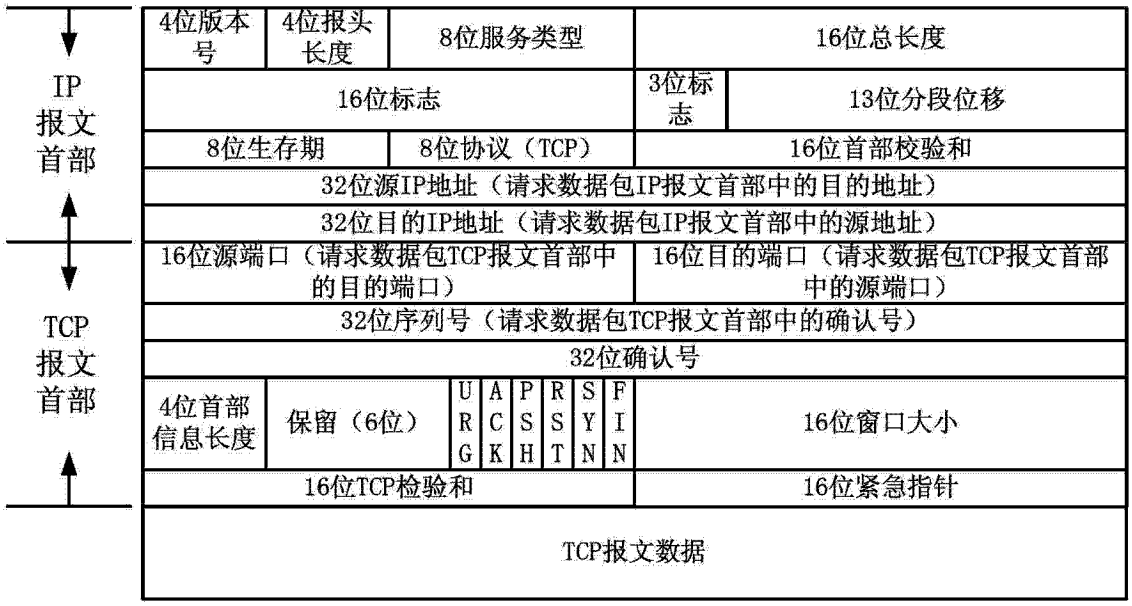

IP
报文
首部
TCP
报文
首部
4位版本号
4位报头长度
8位服务类型
16位总长度
16位标志
3位标志
13位分段位移
8位生存期
8位协议（TCP）
16位首部校验和
32位源IP地址（请求数据包IP报文首部中的目的地址）
32位目的IP地址（请求数据包IP报文首部中的源地址）
16位源端口（请求数据包TCP报文首部中的目的端口）
16位目的端口（请求数据包TCP报文首部中的源端口）
32位序列号（请求数据包TCP报文首部中的确认号）
32位确认号
4位首部信息长度
保留（6位）
URG
ACK
PSH
RST
SYN
FIN
16位窗口大小
16位TCP检验和
16位紧急指针
TCP报文数据

图 4