(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 103457924 B
(45)授权公告日 2016.08.03

(21)申请号 201210183882.0

(22)申请日 2012.06.05

(73)专利权人 珠海市君天电子科技有限公司
地址 519015 广东省珠海市吉大景山路莲山巷8号金山电脑大厦

(72)发明人 邹敏 甘灿 潘建波 陈勇

(74)专利代理机构 北京市广友专利事务所有限责任公司 11237
代理人 祁献民

(51)Int.Cl.
H04L 29/06(2006.01)

(56)对比文件
CN 101534306 A,2009.09.16,
CN 101534306 A,2009.09.16,
CN 101211340 A,2008.07.02,
WO 2008/008339 A2,2008.01.17,
CN 101325495 A,2008.12.17,
Mona Ghotaish Alkhozae 等.Phishing Websites Detection based on Phishing Characteristics in the Webpage Source Code.《International Journal of Information and Communication Technology Research》.2011,第1卷(第6期),
张天红.网络钓鱼预警系统设计与分析.《中国优秀硕士学位论文全文数据库 信息科技辑》.2012,(第2期),
Ying Pan等.Anomaly Based Web Phishing Page Detection.《Computer Security Applications Conference, 2006. ACSAC 06. 22nd Annual》.2006,

审查员 徐苏宁

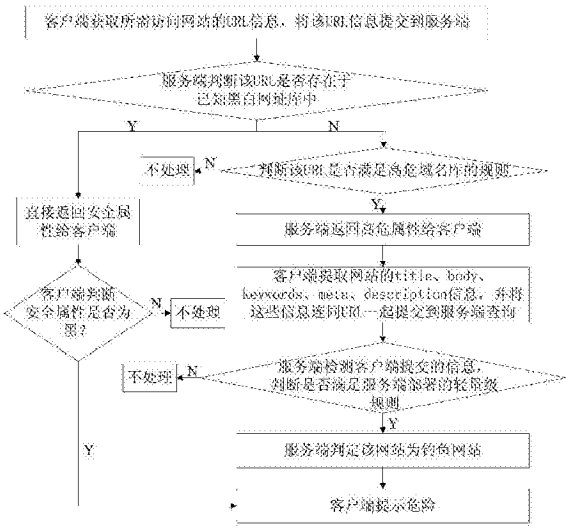
权利要求书3页 说明书5页 附图2页

(54)发明名称

检测点对点、瞬时生效性型钓鱼网站的方法及系统

(57)摘要

本发明提供一种检测点对点、瞬时生效性型钓鱼网站的方法及系统，该方法包括以下步骤：服务端接收客户端发送的所需访问网站的URL信息；判断该URL是否存在于已知黑白网址库中，若是则直接返回安全属性给客户端；若否则继续判断该URL是否满足高危域名库的规则；若满足则返回高危属性给客户端；服务端接收客户端提交的URL信息以及根据所述高危属性信息提取的网站的title、body、keywords、meta、description信息，判断是否满足服务端部署的轻量级规则，若是则判定该网站为钓鱼网站。本发明的方法及系统，能较准确识别出钓鱼网站，解决了服务端爬虫无法爬取点对点、瞬时生效类型钓鱼网站的问题。



CN 103457924 B

1.一种检测点对点、瞬时生效性型钓鱼网站的方法，其特征在于，包括以下步骤：

服务端接收客户端发送的所需访问网站的URL信息；

判断该URL是否存在于已知黑白网址库中，若是则直接返回安全属性给客户端；若否则继续判断该URL是否满足高危域名库的规则，所述高危域名库由从已知钓鱼网站域名中筛选出的高危的域名后缀列表构成；

若判断得出所述URL满足所述高危域名库的规则，则返回高危属性给客户端；

服务端接收客户端提交的URL信息以及客户端根据高危属性信息提取的网站的title、body、keywords、meta、description信息，判断是否满足服务端部署的轻量级规则，若是则判定该网站为钓鱼网站。

2.根据权利要求1所述的检测点对点、瞬时生效性型钓鱼网站的方法，其特征在于：

在所述服务端接收客户端发送的所需访问网站的URL信息之前，还包括步骤：客户端获取所需访问网站的URL信息，并将该URL信息发送给服务端；

和/或

所述服务端返回高危属性给客户端之后、服务端接收客户端提交的URL信息以及根据所述高危属性信息提取的网站的title、body、keywords、meta、description信息之前，还包括步骤：客户端接收到高危属性信息后，提取网站的title、body、keywords、meta、description信息，并将这些提取的信息连同URL一起提交到服务端。

3.根据权利要求1所述的检测点对点、瞬时生效性型钓鱼网站的方法，其特征在于，所述服务端接收客户端提交的信息、判断是否满足服务端部署的轻量级规则的过程具体包括：服务端接收到客户端提交的信息后，判断客户端提交的URL、title、body、keywords、meta、description信息是否满足预先部署好的危险URL规则库、危险内容规则库中的规则；该危险URL规则库、危险内容规则库的部署过程包括：从已知钓鱼网址库中筛选出高危URL关键字，存入危险URL规则库中；以及从已知钓鱼网站域名后缀对应网站的title、body、keywords、meta、description中筛选出具有代表性的关键字，存入危险内容规则库中。

4.根据权利要求2所述的检测点对点、瞬时生效性型钓鱼网站的方法，其特征在于，所述客户端提取网站的title、body、keywords、meta、description信息的过程具体包括：客户端采用BHO技术获取浏览器IHTMLDocument3接口，通过该接口的getElementsByTagName来获取title、body、keywords、meta、description标签的内容。

5.根据权利要求1-4任意一项所述的检测点对点、瞬时生效性型钓鱼网站的方法，其特征在于：

在所述判定该网站为钓鱼网站之后，还包括步骤：服务端将判定结果发送给客户端，并在客户端提示危险；以及

在所述服务端直接返回安全属性给客户端之后，还包括步骤：客户端接收到服务端直接返回的安全属性后，判断所述安全属性是否为黑，若是则提示危险。

6.一种检测点对点、瞬时生效性型钓鱼网站的系统，其特征在于，包括设置于服务端上的信息接收模块、第一判断模块、安全属性返回模块、第二判断模块、高危属性返回模块以及检测与判定模块；

所述信息接收模块用于接收客户端发送的所需访问网站的URL信息；

所述第一判断模块用于接收到所述URL信息后，判断该URL是否存在于已知黑白网址库

中；

所述安全属性返回模块用于当判断出所述URL存在于已知黑白网址库中时，直接返回安全属性给客户端；

所述第二判断模块用于当判断出所述URL不存在于已知黑白网址库中时，继续判断该URL是否满足高危域名库的规则，所述高危域名库由从已知钓鱼网站域名中筛选出的高危的域名后缀列表构成；

所述高危属性返回模块用于当判断得出所述URL满足所述高危域名库的规则时，返回高危属性给客户端；

所述检测与判定模块用于接收客户端提交的URL信息以及客户端根据高危属性信息提取的网站的title、body、keywords、meta、description信息，判断是否满足服务端部署的轻量级规则，若是则判定该网站为钓鱼网站。

7.根据权利要求6所述的检测点对点、瞬时生效性型钓鱼网站的系统，其特征在于，还包括设置于客户端上的URL获取与发送模块、信息提取与发送模块；

所述URL获取与发送模块用于获取所需访问网站的URL信息，并将该URL信息发送给服务端；

所述信息提取与发送模块用于接收到高危属性信息后，提取网站的title、body、keywords、meta、description信息，并将这些提取的信息连同URL一起提交到服务端。

8.根据权利要求6所述的检测点对点、瞬时生效性型钓鱼网站的系统，其特征在于，所述检测与判定模块包括危险URL规则库生成模块、危险内容规则库生成模块、结果判断模块；

所述危险URL规则库生成模块用于从已知钓鱼网址库中筛选出高危URL关键字，生成危险URL规则库中；

所述危险内容规则库生成模块用于从已知钓鱼网站域名后缀对应网站的title、body、keywords、meta、description中筛选出具有代表性的关键字，生成危险内容规则库中；

所述结果判断模块用于判断客户端提供的URL、title、body、keywords、meta、description信息是否满足所述危险URL规则库、危险内容规则库中的规则，若是则判定该网站为钓鱼网站。

9.根据权利要求7所述的检测点对点、瞬时生效性型钓鱼网站的系统，其特征在于，所述信息提取与发送模块包括BHO模块，用于采用BHO技术获取浏览器IHTMLDocument3接口，通过该接口的getElementsByTagName来获取title、body、keywords、meta、description标签的内容。

10.根据权利要求6-9任意一项所述的检测点对点、瞬时生效性型钓鱼网站的系统，其特征在于，还包括设置于服务端上的结果发送模块，以及设置于客户端上的黑白判断模块、危险提示模块；

所述结果发送模块用于当所述检测与判定模块判定网站为钓鱼网站之后，将判定结果发送给客户端；

所述黑白判断模块用于接收到服务端直接返回的安全属性后，判断所述安全属性是否为黑；

所述危险提示模块用于接收到所述结果发送模块发送的判定结果后在客户端提示危

险;以及当所述黑白判断模块判断得出所述安全属性为黑后在客户端提示危险。

检测点对点、瞬时生效性型钓鱼网站的方法及系统

技术领域

[0001] 本发明涉及电子商务安全技术领域，特别是涉及一种检测点对点、瞬时生效性型钓鱼网站的方法以及一种检测点对点、瞬时生效性型钓鱼网站的系统。

背景技术

[0002] 随着电子商务的日益普及，并且由于利益的驱使、法律法规的不完善等因素，导致虚假仿冒电子商务网站的钓鱼网站也逐渐猖獗起来。此类网站和正规电子商务网站基本相似，唯独所卖物品为虚假物品，一旦购买便会损失网银等相关财产。

[0003] 传统的钓鱼网站的检测方法，都是基于特征码匹配和图像相似度匹配，检测来源多是来自对搜索引擎的爬取、用户的举报等方式，由于所有这一切的检测都以来源数据有效为基础，因此当数据来源不可访问时，传统的钓鱼网站检测方法将失去作用。

[0004] 而且，当前的钓鱼技术在传统单一的虚假仿冒基础上已经增加了高强度的反检测技术。首先，钓鱼网站的传播开始趋向于点对点的传播，这使得第三方IP地址无法访问该网页，这导致服务端爬虫无法爬取相应的网页进行检测；其次，钓鱼网站的生效周期也越来越短。基本上是几个小时后就无法再访问，待到服务端爬虫有空闲去爬取该钓鱼网站时，该网站已经无法访问。

发明内容

[0005] 基于此，有必要针对上述服务端爬虫无法爬取点对点、瞬时生效类型钓鱼网站，导致无法准确检测出钓鱼网站的问题，提供一种检测点对点、瞬时生效性型钓鱼网站的方法及系统。

[0006] 一种检测点对点、瞬时生效性型钓鱼网站的方法，包括以下步骤：

[0007] 服务端接收客户端发送的所需访问网站的URL信息；

[0008] 判断该URL是否存在于已知黑白网址库中，若是则直接返回安全属性给客户端；若否则继续判断该URL是否满足高危域名库的规则，所述高危域名库由从已知钓鱼网站域名中筛选出的高危的域名后缀列表构成；

[0009] 若判断得出所述URL满足所述高危域名库的规则，则返回高危属性给客户端；

[0010] 服务端接收客户端提交的URL信息以及根据所述高危属性信息提取的网站的title、body、keywords、meta、description信息，判断是否满足服务端部署的轻量级规则，若是则判定该网站为钓鱼网站。

[0011] 一种检测点对点、瞬时生效性型钓鱼网站的系统，包括设置于服务端上的信息接收模块、第一判断模块、安全属性返回模块、第二判断模块、高危属性返回模块以及检测与判定模块；

[0012] 所述信息接收模块用于接收客户端发送的所需访问网站的URL信息；

[0013] 所述第一判断模块用于接收到所述URL信息后，判断该URL是否存在于已知黑白网址库中；

[0014] 所述安全属性返回模块用于当判断出所述URL存在于已知黑白网址库中时，直接返回安全属性给客户端；

[0015] 所述所第二判断模块用于当判断出所述URL不存在于已知黑白网址库中时，继续判断该URL是否满足高危域名库的规则，所述高危域名库由从已知钓鱼网站域名中筛选出的高危的域名后缀列表构成；

[0016] 所述高危属性返回模块用于当判断得出所述URL满足所述高危域名库的规则时，返回高危属性给客户端；

[0017] 所述检测与判定模块用于接收客户端提交的URL信息以及根据所述高危属性信息提取的网站的title、body、keywords、meta、description信息，判断是否满足服务端部署的轻量级规则，若是则判定该网站为钓鱼网站。

[0018] 通过以上方案可以看出，本发明的检测点对点、瞬时生效性型钓鱼网站的方法及系统，对于采用点对点方式或者是瞬时生效方式欺骗买家的钓鱼网站，由服务端对客户端所提取的title、body、keywords、meta、description信息进行二次检测，这样就能较准确识别出钓鱼网站，解决了服务端爬虫无法爬取点对点、瞬时生效类型钓鱼网站的问题，保证了用户财产的安全，具有较好的市场应用前景。而且本发明的方案对采用加密处理过的网站的检出也有一定的效果。

附图说明

[0019] 图1为一种检测点对点、瞬时生效性型钓鱼网站的方法的流程示意图；

[0020] 图2为一种检测点对点、瞬时生效性型钓鱼网站的系统的结构示意图。

具体实施方式

[0021] 下面结合附图以及具体的实施例，对本发明的技术方案做进一步的描述。

[0022] 参见图1所示，一种检测点对点、瞬时生效性型钓鱼网站的方法，包括以下步骤：

[0023] 步骤S1，服务端接收客户端发送的所需访问网站的URL(Uniform Resource Locator，统一资源定位符，即网页地址)信息。

[0024] 作为一个较好的实施例，在所述服务端接收客户端发送的所需访问网站的URL信息之前，还可以包括步骤S0：当客户端需要访问到某一网站时，首先由客户端获取该所需访问网站的URL信息，并将该URL信息发送给服务端。

[0025] 步骤S2，服务端判断该URL是否存在于已知黑白网址库中，若是则直接返回安全属性给客户端；若否则继续判断该URL是否满足高危域名库的规则。

[0026] 需要说明的是，上述的黑白网址库采用的是业内公知的黑白网址库，如baidu.com是白网址，cu.cc是黑网址……

[0027] 另外，所述高危域名库是由从已知钓鱼网站域名中筛选出的高危的域名后缀列表构成的。我们可以对已知钓鱼网站域名进行分类，筛选出分布较集中的域名后缀，如：.cu.cc、.3322.org等，将其收录保存即可生成高危域名库。本实施例中，高危域名库中收录了以下域名后缀：

[0028] ".bij.pl",".ch.tf",".345.pl",".vv.cc",".rr.nu",

[0029] ".moo.no",".downsc.net",".dnsla.net",".imblog.in",

[0030] ".c.la",".ch.tf",".sg.tf",".edu.tf",".mu.la",

[0031] ".com.uue",".uue.osa",".cp.cx",".pc.to",".ne.to",

[0032] ".joa.to",".ivy.to",".ce.to",".th.tc",".us.tc",

[0033] ".tk",".co.cc",".cn.im",".gv.vg",".osa.pl",

[0034] ".ec.tc",".ce.ms",".co.tv",".co.be",".sg.tf",

[0035] ".dfha.net",".cu.cc",".ss.la",".cz.cc",

[0036] ".cx.cc",".bname.us",".co.in",".cn.la",".3322.net",

[0037] ".5166.info",".oicp.net",".usa.cc",".vicp.cc",".mir3fb.com",

[0038] ".hk.tc",".usa.cc",".pee.pl",".cnla.in",".ibiz.cc",

[0039] ".3322.org",".cc.ai",".c0m.li",".ax.lt",".cc.ai",

[0040] ".qc.to",".55.lt",".bee.pl",".uc.gd",".it.cx",

[0041] ".iosv.in",".publicvm.com",".imbbs.in",".hk.ms",".data-idc.com",

[0042] ".88dns.org",".nom.vg",".10dig.net",".321cc.cn",

[0043] ".fr.ms",".de.ms",".au.ms",".cn.ms",".cn.mu",".assexyas.com",

[0044] ".51vip.biz",".itemdb.com",".host11.net",".xindns.org",".ocry.com",

[0045] ".vr.lt",".ka.hn",".555hsf.com",".ze.tc"……

[0046] 步骤S3,若服务端判断得出所述URL满足所述高危域名库中的规则时,返回高危属性给客户端;当然若服务端判断得出所述URL并不满足所述高危域名库中的规则时,则可以不进行处理。

[0047] 事实上作为一个较好的实施例,在所述步骤S3服务端返回高危属性给客户端之后、步骤S4之前,还可以包括步骤S31:客户端接收到高危属性信息后,提取网站的title(网页标题)、body(网页内容)、keywords(网站的关键字信息,主要方便搜索引擎爬取)、meta(用来描述keywords、description的标签)、description(网站的描述信息,主要方便搜索引擎爬取)信息,并将这些提取的信息连同URL一起提交到服务端进行二次查询。

[0048] 作为一个较好的实施例,本步骤S31中客户端提取网站的title、body、keywords、meta、description信息的过程具体可以包括S311:客户端采用BHO技术获取浏览器IHTMLDocument3接口,通过该接口的getElementsByTagName来获取title、body、keywords、meta、description标签的内容。

[0049] 步骤S4,服务端接收客户端提交的URL信息以及根据所述高危属性信息提取的网站的title、body、keywords、meta、description信息,判断这些信息是否满足服务端预先部署的轻量级规则,若是则判定该网站为钓鱼网站。

[0050] 作为一个较好的实施例,步骤S4具体可以包括步骤S41:服务端接收到客户端提交的信息后,判断客户端提交的URL、title、body、keywords、meta、description信息是否满足预先部署好的危险URL规则库、危险内容规则库中的规则,若否则不进行处理;若是则判定该网站为钓鱼网站。上述危险URL规则库、危险内容规则库的部署过程可以包括步骤S411和步骤S412:

[0051] 步骤S411,从已知钓鱼网址库中筛选出高危URL关键字,存入危险URL规则库中。

[0052] 对已知钓鱼网站域名进行分类,筛选出分布较集中的host关键字,将其收录并保存以生成危险URL规则库。例如本实施例中可以将以下内容收录进危险URL规则库中:

[0053] "alipay","taobeo","taotao","taodao","ta0ba0","tabao","taoibao","taobac","taobao","ailqay";

[0054] "pay.com","allqay","alibaba","ailibaba","zhifubao","alipay","allpay";

[0055] "member/mini_login.asp","wap/mini_login.asp","ssyy/login.php";

[0056] "auction/buy_nowsw.asp";

[0057] "auction/sw_buy_now.htm.asp","auction/buy_nowxn.asp";

[0058] "shop","item";

[0059] "tao","auction/item_detail","intel/item_detail","item_htm.asp?id=","item.htm.asp?id=","item.htm.php?id=","item.asp?id="……

[0060] 步骤S412,从已知钓鱼网站域名后缀对应网站的title、body、keywords、meta、description中筛选出具有代表性的关键字,存入危险内容规则库中。

[0061] 对已知钓鱼网站的标题及内容等进行分类,可以筛选出使用频率较高的关键字,将其收录以生成危险内容规则库。例如本实施例中可以将以下内容收录进危险内容规则库中:"六合彩"、"官方"、"预测"、"福彩"、"双色球"、"足球预测"、"皇冠投注"、"皇冠足球"、"皇冠博彩"、"足球贴士"、"足球博彩"、"皇冠赔率"、"娱乐城"、"淘宝网"、"weibo"、"welbo"、"网上支付"、"账户监管"、"游戏"、"装备"、"交易"……

[0062] 作为一个较好的实施例,在所述判定网站为钓鱼网站之后,还可以包括步骤S5:服务端将判定结果发送给客户端,并在客户端提示危险。

[0063] 另外,在所述步骤S2服务端直接返回安全属性给客户端之后,还可以包括步骤S21:客户端接收到服务端直接返回的安全属性后,判断所述安全属性是否为黑,若否则不进行处理,若是则在客户端提示危险。

[0064] 与上述一种检测点对点、瞬时生效性型钓鱼网站的方法相对应的,本发明还提供一种检测点对点、瞬时生效性型钓鱼网站的系统,如图2所示,包括设置于服务端上的信息接收模块、第一判断模块、安全属性返回模块、第二判断模块、高危属性返回模块以及检测与判定模块;

[0065] 所述信息接收模块用于接收客户端发送的所需访问网站的URL信息;

[0066] 所述第一判断模块用于接收到所述URL信息后,判断该URL是否存在于已知黑白网址库中;

[0067] 所述安全属性返回模块用于当判断出所述URL存在于已知黑白网址库中时,直接返回安全属性给客户端;

[0068] 所述所第二判断模块用于当判断出所述URL不存在于已知黑白网址库中时,继续判断该URL是否满足高危域名库的规则,所述高危域名库由从已知钓鱼网站域名中筛选出的高危的域名后缀列表构成;

[0069] 所述高危属性返回模块用于当判断得出所述URL满足所述高危域名库的规则时,返回高危属性给客户端;

[0070] 所述检测与判定模块用于接收客户端提交的URL信息以及根据所述高危属性信息提取的网站的title、body、keywords、meta、description信息,判断是否满足服务端部署的轻量级规则,若是则判定该网站为钓鱼网站。

[0071] 本发明的检测方案可以由客户端和服务端程序配合在一起进行工作,因此作为一

个较好的实施例，本发明的系统除了包括上述设置于服务端上的各模块之外，如图2所示，还可以包括设置于客户端上的URL获取与发送模块、信息提取与发送模块；

[0072] 所述URL获取与发送模块用于获取所需访问网站的URL信息，并将该URL信息发送给服务端；

[0073] 所述信息提取与发送模块用于接收到高危属性信息后，提取网站的title、body、keywords、meta、description信息，并将这些提取的信息连同URL一起提交到服务端。

[0074] 作为一个较好的实施例，所述检测与判定模块可以包括危险URL规则库生成模块、危险内容规则库生成模块以及结果判断模块；

[0075] 所述危险URL规则库生成模块用于从已知钓鱼网址库中筛选出高危URL关键字，生成危险URL规则库中；

[0076] 所述危险内容规则库生成模块用于从已知钓鱼网站域名后缀对应网站的title、body、keywords、meta、description中筛选出具有代表性的关键字，生成危险内容规则库中；

[0077] 所述结果判断模块用于判断客户端提供的URL、title、body、keywords、meta、description信息是否满足所述危险URL规则库、危险内容规则库中的规则，若是则判定该网站为钓鱼网站。

[0078] 作为一个较好的实施例，所述信息提取与发送模块可以包括BHO模块，用于采用BHO技术获取浏览器IHTMLDocument3接口，通过该接口的getElementsByTagName来获取title、body、keywords、meta、description标签的内容。

[0079] 另外，本发明的系统还可以包括设置于服务端上的结果发送模块，以及设置于客户端上的黑白判断模块、危险提示模块；

[0080] 所述结果发送模块用于当所述检测与判定模块判定网站为钓鱼网站之后，将判定结果发送给客户端；

[0081] 所述黑白判断模块用于接收到服务端直接返回的安全属性后，判断所述安全属性是否为黑；

[0082] 所述危险提示模块用于接收到所述结果发送模块发送的判定结果后在客户端提示危险；以及当所述黑白判断模块判断得出所述安全属性为黑后在客户端提示危险。

[0083] 本发明的一种检测点对点、瞬时生效性型钓鱼网站的系统的其它技术特征与上述一种检测点对点、瞬时生效性型钓鱼网站的方法完全相同，此处不予赘述。

[0084] 由以上方案可以看出，本发明的一种检测点对点、瞬时生效性型钓鱼网站的方法及系统，对于采用点对点方式或者是瞬时生效方式欺骗买家的钓鱼网站，由服务端对客户端所提取的title、body、keywords、meta、description信息进行二次检测，这样就能较准确识别出钓鱼网站，解决了服务端爬虫无法爬取点对点、瞬时生效类型钓鱼网站的问题，保证了用户财产的安全，具有较好的市场应用前景。另外本发明的方案对采用加密处理过的网站的检出也有一定的效果。

[0085] 以上所述实施例仅表达了本发明的几种实施方式，其描述较为具体和详细，但并不能因此而理解为对本发明专利范围的限制。应当指出的是，对于本领域的普通技术人员来说，在不脱离本发明构思的前提下，还可以做出若干变形和改进，这些都属于本发明的保护范围。因此，本发明专利的保护范围应以所附权利要求为准。

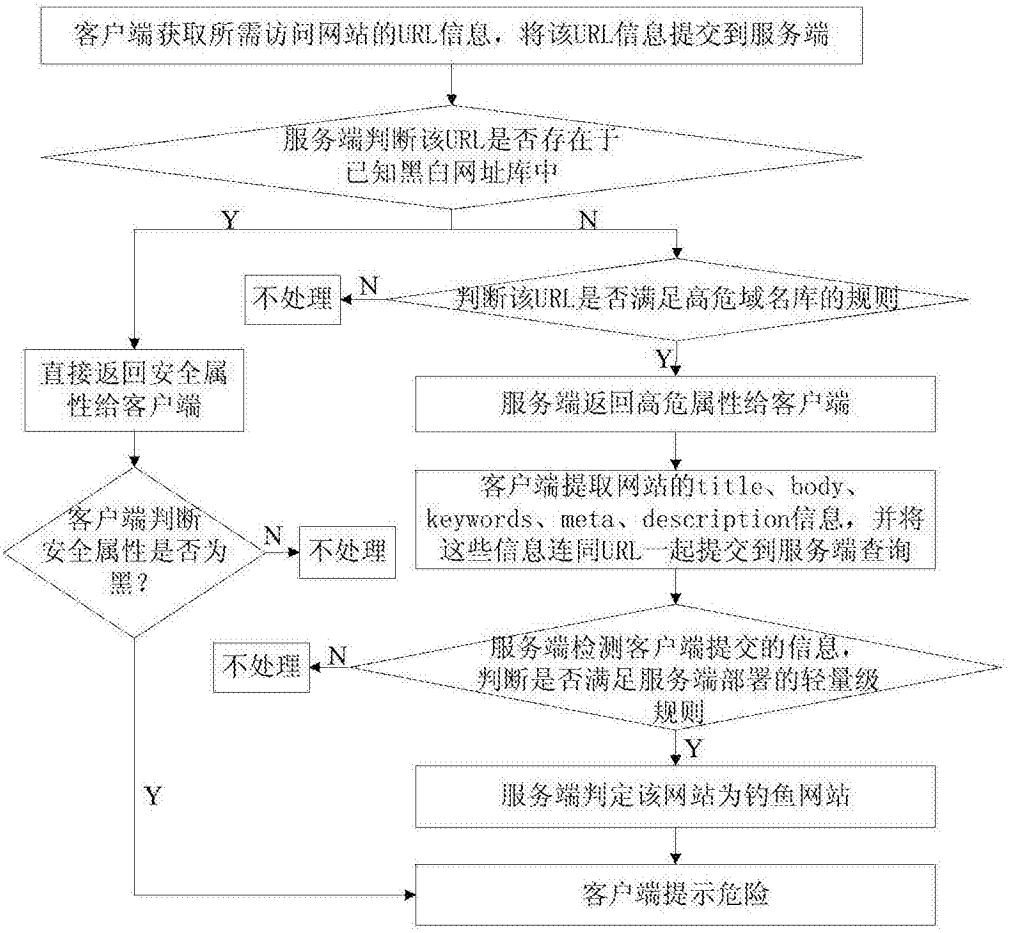

客户端获取所需访问网站的URL信息，将该URL信息提交到服务端
服务端判断该URL是否存在于已知黑白网址库中
Y
N
不处理
N
判断该URL是否满足高危域名库的规则
Y
直接返回安全属性给客户端
服务端返回高危属性给客户端
客户端判断安全属性是否为黑？
N
不处理
客户端提取网站的title、body、keywords、meta、description信息，并将这些信息连同URL一起提交到服务端查询
不处理
N
服务端检测客户端提交的信息，判断是否满足服务端部署的轻量级规则
Y
Y
服务端判定该网站为钓鱼网站
客户端提示危险

图1

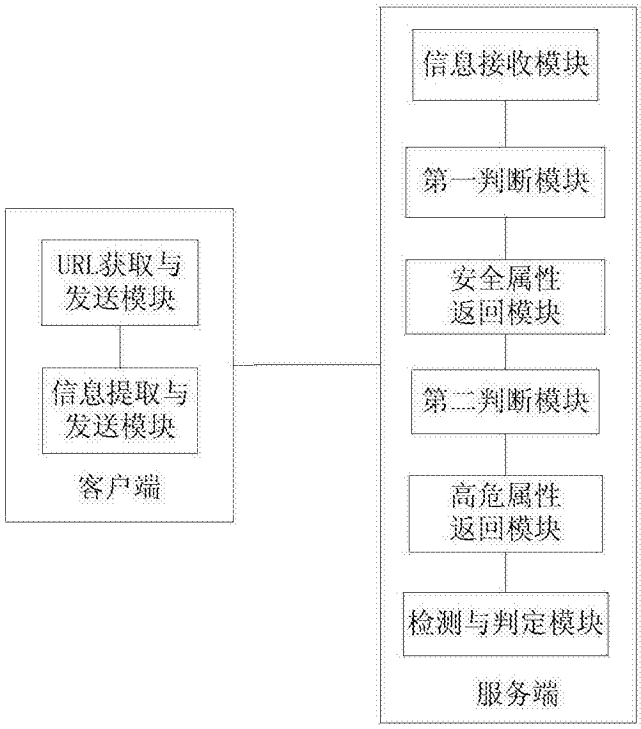

信息接收模块
第一判断模块
安全属性
返回模块
第二判断模块
高危属性
返回模块
检测与判定模块
服务端
URL获取与
发送模块
信息提取与
发送模块
客户端

图2