圆桌·钓鱼网站与反钓鱼技术剖析

钓鱼网站的简介阐述

钓鱼网站通常指伪装成银行及电子商务，窃取用户提交的银行帐号、密码等私密信息的网站。“钓鱼”是一种网络欺诈行为，指不法分子利用各种手段，仿冒真实网站的URL地址以及页面内容，或利用真实网站服务器程序上的漏洞在站点的某些网页中插入危险的HTML代码，以此来骗取用户银行或信用卡账号、密码等私人资料。

“钓鱼网站”的频繁出现，严重地影响了在线金融服务、电子商务的发展危害公众利益，影响公众应用互联网的信心。钓鱼网站通常伪装成为银行网站，窃取访问者提交的账号和密码信息。它一般通过电子邮件传播，此类邮件中一个经过伪装的链接将收件人联到钓鱼网站。钓鱼网站的页面与真实网站界面完全一致，要求访问者提交账号和密码。

有趣的是现在黑产中鱼站越来越强了，一次性换模板与域名，我必须的纠正一下我以前所说的：“给别人写鱼站的，都是菜B前端。”

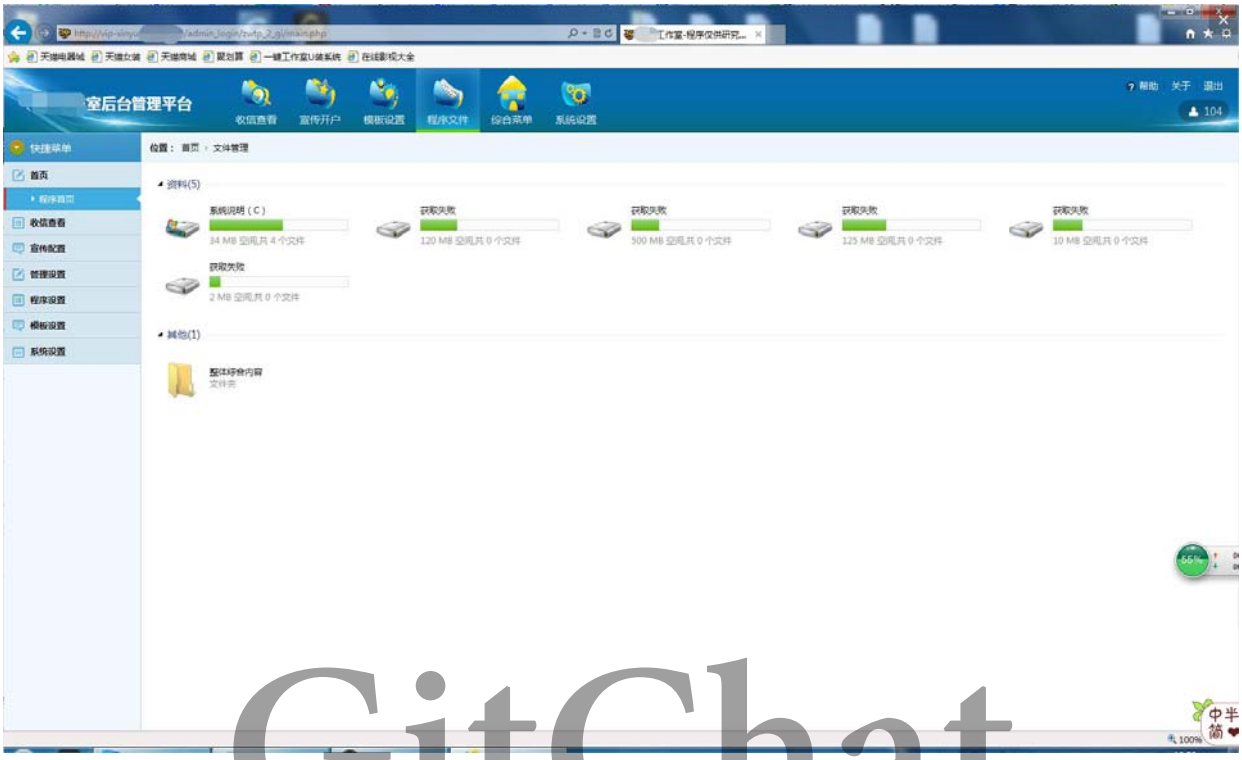
专业的鱼站，写的不仅仅好，功能齐全，而且还标注上“2017最新版防Xss防sql注入”。价格也高到一定程度，渔民租用鱼主的服务器和域名，一天一百块，想想还是有点利润的。

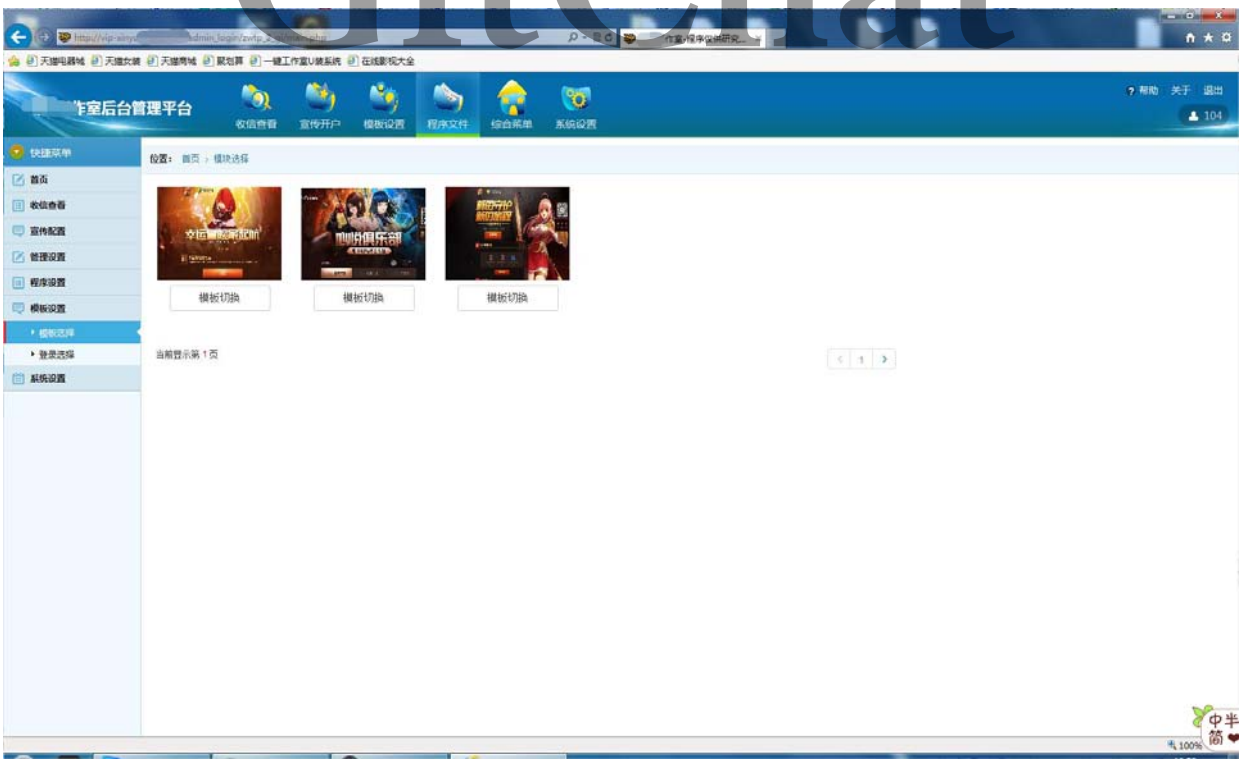
我和鱼民的聊天记录，他这里的不红，意思是 发在群里 QQ里不会被拦截，不提示危险网站。

就单单这个不红就能实现绕过之后，已经有黑产人员搭话：“测试一天群发，不红了，三百一天收你这个技术。”

当然是义正言辞的拒绝了。不过在下属实缺钱也没有和黑产勾肩搭背的想法。

贴两个鱼站截图，可以看到功能完善之强，这是我之前X进去的一个站台。





目前我接触到的有针对游戏钓鱼的，比如DNF，一天钓几千号没问题。洗号洗装备洗钱几千个号上万利润没问题。再比如，iPhoneID锁，一次钓几个号没问题，解开一个就赚一个，利润嘛，请问现在市场二手的iPhone7P是多少钱？

走正题了。

钓鱼网站常见的厂商检测

现在常见的安全软件厂商有，360，腾讯，百度，等厂商。在于钓鱼网站的检测力度来讲，腾讯反钓鱼网站力度是数一数二的，因为腾讯的QQ，微信，占据了社交软件的绝大部分市场。也是钓鱼网站转播的主要途径。当然在技术上腾讯的反钓鱼系统做的也是相当的不错，基本上能杜绝大部分钓鱼网站的伪装。

根据百度百科的介绍腾讯拥有一个全球最大的网站数据库，能敏锐鉴定网站的安全性，轻松识别假冒、诈骗、钓鱼等恶意网站。

钓鱼网站的检测方法

由于钓鱼网站的识别率很高，一开始没有理会检测方法和原理，通常钓鱼网站发出来只要几十秒就会报毒，可是当看到一个钓鱼网站连续几天都没有报毒，这时候我惊呆了。于是我开始着手了解反钓鱼的检测方法。在度娘上找到了一篇关于钓鱼网站的检测系统的研究论文。


首页 > 学位首页 > 计算机技术 > 基于AdaBoost算法的网络钓鱼检测系统的研究

基于AdaBoost算法的网络钓鱼检测系统的研究

目录

网络钓鱼通常利用邮件或者高度模仿的网页对用户进行欺诈犯罪。如今，网络钓鱼的攻击形式更加多元化，使得钓鱼攻击的防范和检测变得更加困难。据统计，近年来钓鱼攻击带来的损失呈现翻倍式的上升。因此，网络钓鱼已成为危害网络安全的最主要因素之一。它不仅降低了网络中人与人之间的信任度，而且严重阻碍了电子商务的发展。

当前常见的钓鱼检测技术存在检测后滞单一、信息获取不够全面等问题，因此，本文提出了一种URL黑白名单过滤结合机器学习(AdaBoost算法)分类的检测方法，主要工作如下：

对待检测的网站首先经过URL黑白名单过滤，若是匹配成功作为结果输出，若是匹配失败，则进行下一步分类器的检测。通过这种检测方法，可以快速的检测出时效性较差的钓鱼网站。新形式的钓鱼网站可以通过机器学习的方法进行检测。

分类器检测关键之处在于特征的如何提取。为了获得钓鱼网站足够的信息，本文从URL中提取了14个特征，从网页结构中提取了5个特征，从网页内容中提取了大量的特征用于分类器的训练和检测。

特征中可能参杂着大量的噪声，以及维度较高等问题，因此，加入数据预处理模块进行降维和去除噪声。

通过比较K-近邻算法、朴素贝叶斯算法、逻辑回归和AdaBoost算法的检测性能，最终选取AdaBoost算法作为本文的检测方法。由于钓鱼网站检测存在非平衡代价问题，提出一种改进算法AdaCostBoost,实验结果表明，改进的算法在保证检测精准性的同时降低了正规网站的误判率，减小了误判带来的影响，提高了其在实际应用中的可能性。

收起

在该论文中大概了解到，该检测系统的原理是是对url内容特征进行匹配，在原理的介绍上是通过，url白名单过滤，和网站内容特征进行检测判断的。我在关于“钓鱼网站”百度百科的介绍中看到腾讯也介绍了腾讯安全云库的一种url检测和过滤的原理，该原理是通过用户访问的网址在腾讯云库存储的钓鱼网站数据库和特征进行匹配，从而判断用户访问的网址是否为钓鱼网站。

众所周知，钓鱼网站实际是一种网络欺诈行为，是不法分子利用各种手段仿照真实网站的网址及页面内容，或者利用真实网站服务器上的漏洞，在网站的某些页面中插入危险的HTML代码，以此来骗取用户银行和信用卡账号、密码等私人资料。钓鱼网站传播途径最主要为即时通讯工具和社交网络，这些钓鱼网站不仅页面制作精良，同正规官网相似度极高，而且惯用一些极易混淆的字符做域名。

如何辨别钓鱼网站首先应该是多留心网址，看看是不是与官方网站一致。这种方法对有一定网络知识的网民来说并不难，但是对很多不太了解网络的网民则比较困难。腾讯电脑管家云安全中心已经建立了恶意网址安全云查杀平台，通过强大的云检测引擎来识别并有效打击钓鱼网站。

海联达公司产品总监蔡政强表示，当用户访问一个网址的时候，我们（安全路由器）会将这个网址发送到腾讯的安全云库里去查询，在0.5秒之内就能完成整个查询、识别过程。

（海联达公司产品总监蔡政强接受央视采访）

腾讯云安全网址检测如上图。

反钓鱼系统进行测试实验

（由于其他原因具体详情不放出，故告知。）

为了论证该技术的可行性我自行搭建了一个钓鱼网站：

在我搭建好的第一时间发给了一个朋友进行检测测试，发现在刚刚发布的几十秒内并没有报毒。

而在几分钟后我发给我另一个朋友，得到的回复是已经被拦截。随后我在腾讯管家安全检测查询，的确被拦截判断为钓鱼网站，最关键的是紧紧在发布了30秒左右就检测到了。这让我感到管家的技术强大。然而我开始了我的实验之路。

根据上述的图片原理我进行了实验，首先url白名单过滤，我找了一个新域名，所以不在白名单内，也不在钓鱼网站库里，然后第一个理论我就顺利通过url过滤判断，其次在上述的原理中提到的对页面的内容特征进行检测，这个是判断钓鱼网站的核心，因为url检测域名的相似度导致误判会很高，所以电脑管家会检测网页上的源代码特征，是否跟已在安全库中出现的钓鱼网站的页面特征进行匹配，找到相似处，从而判断是否是钓鱼网站。然后为了躲过页面特征被检测的风险，我处理了一下页面源代码的特征。然后开始测试。结果出乎我的意料。我发给朋友测试，管家检测并没有跟我第一次测试的时候，在30秒内检测出是钓鱼网站。一开始觉得是发布的人数较少没被检测出来，然后我在群里群发了一下网站，但是结果还是没有被检测出是钓鱼网站，接着我来到管家网址检测，对网站再次进行检测，结果还是没有被检测出来，进过10几分钟的等待再次检测一遍，依然没有被检测出来。由此可以断定，管家的检测方法也跟上述我找到的论文方法一致。

经过几番测试，对与我处理过页面源代码特征的页面也一直没被反钓鱼系统检测出来。

复现过程到此结束，因为很多原因不能放出具体详情过程和贴代码，只能贴个论文地址了。

AdaBoost算法的网络钓鱼检测系统的研究

我直接贴个论文地址吧，我是看的这个论文，内容太多也不好引用，请读者自查。

> https://wenku.baidu.com/view/ff36c5e94b35eefdc9d333b6.html

来自江西理工大学的三位同志研究论文。

关于钓鱼网站的识别与防范

对应钓鱼网站的识别与防范，大家只要记住，看url链接的地址的是不是你要访问的网站域名，还有就是不点来历不明链接。

也请记住以下五种方法防范办法：

- 第一、查验“可信网站”
- 第二、核对网站域名
- 第三、比较网站内容
- 第四、查询网站备案
- 第五、查看安全证书

It’s your turn to speak, my friend.