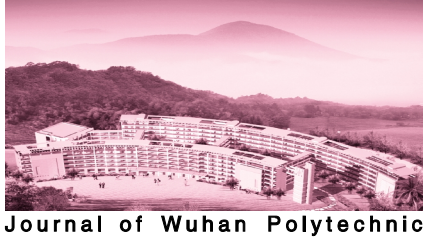

Journal of Wuhan Polytechnic

基于大数据分析的钓鱼网站监测研究与应用

周小松

（武汉职业技术学院 文化与传媒学院，湖北 武汉 430074）

摘 要：钓鱼网站通常是指伪装成银行及电子商务等网站，主要危害是窃取用户提交的银行账号、密码等私密信息。在移动互联网时代，伪基站盛行、集团客户管理宽泛、即时通 APP 应用泛滥，为钓鱼网站提供了温床。针对钓鱼网站盛行的现状，分析了钓鱼网站的特征，指出去传播路径，提出了基于大数据分析的钓鱼网站主动监测与基于 IP 地址封堵的钓鱼网站抑制方法。

关键词：钓鱼网站；用户隐私；即时通软件；伪基站；DNS

中图分类号：TP393.081　　**文献标识码**：A　　**文章编号**：1671-931X（2016）05-0079-03

一、引言

所谓“钓鱼网站”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，或者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。近年来，随着电子商务技术的快速发展，钓鱼诈骗行为日益猖獗。不法分子利用伪基站、行业网关的短彩信接口、以微信为代表的即时通讯软件等信息媒介传播钓鱼诈骗网站，通过领奖、积分兑换、优惠活动等方式，利用“人性好贪、便宜要占”的弱点，诱导用户登录虚假网站填写个人信息，造成个人信息泄露、财产损失。随着国家与社会对网络诈骗关注程度的加大，净化网络环境，促进互联网行业的健康发展已经成为电信运营商的重要社会责任。

钓鱼案事件高发，国内共发现钓鱼网站数量达到 10 万级别，且每年都在增长，花样不断更新，调查显示 30.4%的用户遭遇过网络钓鱼或网络欺诈。在移动互联网时代，受经济利益的驱使，以及互联网的开放性，犯罪分子瞄准互联网，传统的犯罪逐渐向互联网渗透，“钓鱼网站”已经成为继病毒和木马之后的“第三大网络毒瘤”。钓鱼网站数量急剧攀升，网站制作越发精良，对互联网金融环境造成极大危害，恶化了电子商务的生态环境，给网民造成严重的经济损失，同时也给公司网络的运行安全带来负面影响，阻碍了公司持续健康发展。由此可见，对网络钓鱼攻击的检测防御研究迫在眉睫。

针对钓鱼网站越整越泛滥、缺乏行之有效发现机制的局面，本项研究提出了基于大数据的钓鱼网站主动监测方法，旨在通过 DNS 大数据和垃圾短信监控系统，实现主动监测钓鱼诈骗网站，快速发现、及时封堵，实现监测、研判、封堵、上报全流程闭环处理，有效控制钓鱼诈骗网站的传播。

二、基于大数据分析的钓鱼网站监测

（一）钓鱼诈骗网站特征分析

钓鱼网站作为一种诈骗手段，最明显的特点就是使用一个看起来正常合法的链接，实则该链接指向另一个非法的钓鱼网站。欺骗用户访问钓鱼网站，

收稿日期：2016-08-10

作者简介：周小松（1978-），男，湖北武汉人，武汉职业技术学院文化与传媒学院讲师，研究方向：网络系统集成、云计算、大数据应用。

从而获取用户的信息。通过研究和分析，主要特点有：

1.高仿真网站。在不法分子申请的 URL 中，常利用障眼法，用外观字形容易混淆的字符来代替使用，达到迷惑用户的目的。另外经常使用官方网站的 LOGO、图表、文字、内容，只是在账号、密码等敏感数据处进行更改，骗取用户的信任，用以窃取用户敏感信息。

2.境外注册域名。不法分子通过互联网 IP 地址空间寻找存在漏洞活缺乏有效安全防护的主机，或租用 ISP 服务器空间，作为钓鱼诈骗的宿主主机，安装钓鱼诈骗网站页面内容，部署必要的后台程序来处理用户输入的信息。数据显示，钓鱼诈骗网站多来自于境外，香港、美国居多，以逃避网站监管。

3.群发“善意”短信。不法分子在钓鱼网站发布阶段，往往以伪基站群发短信等形式散播，引诱用户访问假冒网站，通过后台脚本随时监听用户输入的信息，在获得用户账户、密码等信息后，通过特定的工具快速转移用户资金。

钓鱼诈骗网站传播方式主要有三种：

(1)伪基站传播：即通过伪基站假冒 95588、10010 等银行、电信运营商等知名端口发送诈骗短信，该方式不需要接入网络，伪造发送方号码。

(2)集团客户传播：利用行业端口、短信群发器、普通手机设备等，向用户手机终端发送诈骗短信，该方式需要接入通信网络。

(3)新型 OTT 软件传播：使用微信、飞信等即时通信软件向用户终端发送诈骗短信，该方式需要接入互联网。

(二)基于 DNS 大数据的技术监测方案

钓鱼检测方法一般使用二值“真—假”分类，当前主要的钓鱼网站识别可以通过计算机识别和人工识别这两种不同方式同时进行。用户收到包含钓鱼网站域名的短彩信并点击链接之后，会向 DNS 服务器发送一条域名解析请求报文，DNS 服务器将该域名记入日志。根据钓鱼网站和真实网站域名的相似性，可以利用字符的字形或者语义的相似性编制探测程序，从每天监测 1T 数据量的 DNS 服务器中筛选可疑的钓鱼网站，通过自动化爬虫系统，二次过滤钓鱼网站，经人工判定确认的钓鱼网站，给予封堵处理。

1. DNS 日志分析系统

钓鱼诈骗网站分析系统是通过分析 DNS 日志，模糊查询获取所有包含指定关键词的域名，作为二次筛选数据源，用以实现以下功能：

(1)模糊查询。对钓鱼诈骗网站域名分析得出关键词，可指定多个关键词(逻辑或)实现模糊查询。例如假冒中国移动的钓鱼诈骗网站关键词可以设置为 10086 等。

(2)结合请求次数查询及下载。由于包含 10086 的域名解析记录数量较多，人工审核难度大，可优先关注用户访问多、影响范围大的钓鱼诈骗网站。例如，可提取请求次数在 1000 次以上的域名进行重点关注。

(3)白名单管理。部分域名虽然包含 10086 但经过确认后不是钓鱼诈骗网站，则加入白名单，后续查询时，不再命中该域名，以减小人工送审量。

(4)黑名单管理。已审核确认为钓鱼诈骗网站的域名，加入黑名单，后续查询时，不再命中该域名，以减少人工送审量。

(5)数据下载管理。为降低人工工作量，需通过定时任务等方式实现定时自动查询，查询条件可提前定制，查询完毕后将查询结果导出至文件，进入深层次处理。

2.域名自动化审核

从 DNS 分析系统选出 TOP1000，人工逐个拨测效率低，可采用“爬虫”技术实现钓鱼网站监测的自动化审核程序，通过 HTTP Get 方法将网站页面内容作为一个字符串返回，使用正则表达式设置匹配规则，将可疑域名与普通域名通过不同标记区分，每日筛选一定数量(如 100 个)非常可疑的域名，可有效提高了审核效率。

3.人工再次拨测封堵

人工根据自动审核筛选后的域名，逐一进行拨测，及时对疑似网站进行域名拼写核对、备案信息查询、内容风格比对等审核工作，并对违规网站进行取证，完成二次确认，上报集团进行全网统一封堵。

(三)基于海量短信的大数据分析，控制钓鱼网站传播

钓鱼诈骗行为主要通过发送包含钓鱼网站链接的短彩信以达到传播的目的。客户收到此类短信后，通常采用转发的形式，告诫或者询问亲朋好友短信的真实性。因此，可利用这一特点在短信监控系统部署相应策略，提取钓鱼网站传播域名，在用户登录网站上当受骗之前快速给予封堵，实现被动防御到主动防御，事后安全到事前预警的积极转变。

包含钓鱼诈骗网址链接的短信属于垃圾短信，目前，垃圾短信的发送已形成一套产业链条，群发设备研制者和销售者、短信发送者等各司其职，形成了关系密切的利益群体。山东移动已实现通过技术手段对垃圾短信进行监控和拦截，满足流量、关键词、黑名单的短信将送入监控系统。

利用短信监控系统实现钓鱼诈骗域名的预警，需要经过以下三个环节：

1. 基于通用不良信息监测系统的关键字监测：根据客户举报钓鱼诈骗网站的短信，提取严格的关键字逻辑与、逻辑或的组合，在垃圾短信监控系统中部署相应的关键词，若有用户发送可以匹配关键词

的短信或者发送流量超过一定的频次，将第一时间送入监控系统，实时监控拦截诈骗短信的传播。

2.基于后台数据库的域名自动化提取：监控系统中拦截的短信中是汉字、数字、域名的组合，如“中国移动提醒：您的积分已满足兑换 268 元现金条件，请登录移动商城 wap.10086ykj.com 根据提示安装领取。【中国移动】”，需人工逐条提取域名，工作效率极低，因此本项目自主开发了域名自动化提取程序，利用正则匹配算法自动识别非汉字的组合，支持批量导入包含短信内容的 excel 文件，自动输出每条短信对应的域名，供人工拨测确认，界面简单，易于操作，极大地提高了工作效率。

3.基于客户感知提升的人工审核上报：人工对提取后的纯域名进行逐一审核拨测，及时对疑似网站进行域名拼写核对、备案信息查询、内容风格比对等审核工作，筛选钓鱼诈骗网站，并取证留存，上报集团，在网络出口防火墙上实施 IP 地址全网封堵。

（四）研究创新点

结合实际中的使用，基于大数据分析的钓鱼网站监测研究与应用，具有如下创新点：

首先，通过信息收集与大数据分析，对钓鱼网站的特点、传播途径等进行了全面分析与总结。

其次，提出了一套钓鱼网站监测的闭环管理流程，该流程包括基于 DNS 日志系统的钓鱼网站分析、基于海量历史数据的钓鱼网站对比与判研、基于“用户至上、精诚服务”的人工拨测分析。

第三，提出了基于不良信息监测系统的钓鱼网站发现与基于 IP 地址封堵的抑制方法。不良信息监测系统是各运营商均具备的系统，IP 地址封堵也是运营商常用的方法，这样的实现，代价最低。

三、结束语

基于 DNS 大数据和短信监控系统的钓鱼诈骗网站监测方案具有可操作性、易于理解性，它的实施不仅提高了钓鱼诈骗网站的治理水平，有效净化了通信网络环境，更增强了企业信息安全管理能力，维护了企业良好形象。该项研究在湖北省的安全管理工作中，获得了广泛的使用，取得了良好的效果，下一步将从两个方面开展工作。一方面，继续对现有的方法继续进行改善与优化，全面提升系统的稳定性、容错性和可持续使用性能。另一方面，面向全省所有地市分公司进行推广，结合地市分公司的实际特征，对系统进行个性化配置，以满足区域化需求。

参考文献：

[1] 郭萍.钓鱼网站的危害及防范策略研究[J].网络安全，2012，(2).

[2] 周燕新.防范钓鱼网站，保卫网上银行[J].现代经济信息，2011，(18).

[3] 程元斌.一种防范钓鱼网站的方法[J].网络安全技术与应用，2011，(8).

[4] 唐海腾.钓鱼网站分析及其对策研究[J].信息与电脑(理论版)，2014，(9).

[责任编辑：胡大威]

Research and Application of Fishing Website Monitoring Based on Large Data Analysis

ZHOU Xiao-song

(Wuhan Polytechnic, Wuhan430074,China)

Abstract: Fishing sites usually refer to those which disguised as banks and e-commerce sites. The main harm is to steal the user's bank account number, password and other private information. In the mobile Internet era, the pseudo-base stations are popular, group customer management is extensive, real-time APP application is flooded, which provide a breeding ground for phishing sites. According to the status quo of fishing websites, this paper analyzed the characteristics of fishing websites and pointed out the path of transmission. Then put forward the solution to active monitoring of fishing websites based on large data analysis and fishing website blocking based on IP addresses.

Key words: fishing site; user privacy; instant messaging software; pseudo base station; DNS