基于可信域名的网络钓鱼治理机制研究①

李海灵[1,2,3]　王　伟[1,2]　毛　伟[1,2]　(1.中国科学院计算机网络信息中心 北京 100190；
2.中国互联网络信息中心 北京 100190；3.中国科学院研究生院 信息学院 北京 100190)

摘　要： 反钓鱼技术是近年的研究热点。讨论了域名、域名系统与网络钓鱼攻击的联系，提出了一个基于可信域名的网络钓鱼治理框架。从三个不同区域采取措施保证域名及其应用的可信要素。介绍了框架内的关键技术、核心子系统，并与基于邮件的反钓鱼框架进行了比较。分析表明从基础资源着手解决网络钓鱼问题有着检测信息来源广泛、响应速度快、管理审计便捷的优势，对当前的互联网不良应用治理有着启发和借鉴意义。

关键词： 网络钓鱼;域名;可信要素;可信域名;反钓鱼框架

Anti-Phishing Method Based on Trustworthy Domain Names

LI Hai-Ling[1,2,3] , WANG Wei[1,2] , MAO Wei[1,2]

(1. Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China;
2. China Internet Network Information Center, Beijing 100190, China;
3. Graduate University of Chinese Academy of Sciences, Beijing 100190, China)

Abstract: Anti-phishing technology has been a research hotspot in recent years. In this paper, the relationship of domain names, domain name system and phishing is discussed. A novel anti-phishing framework based on trustworthy domain names is put forward with methods to secure the trustworthy factors of domain name from three domains. Key technologies and subsystems are described subsequently, and a comparison is done with an email-based phishing webpage detection system. Analysis shows that address phishing issues from basic resources have the advantage of a wider range of information resources, a faster response time, a more convenient management system and an audit, which is also instructive for bad Internent governance.

Keywords: phishing; domain name; trustworthy factors; trustworthy domain names; anti-phishing framework

1　引言

2009 年底中国网民规模达 3.84 亿[1]，居世界第一位。截至 2009 年 11 月，CN 域名数的注册量达 1368 万个，占全国域名总数的 80%。然而，由于互联网安全缺陷和利益驱使，欺诈、垃圾、低俗信息涌现，给互联网的健康发展和有效管理带来了较大挑战。CNCERT/CC 的安全报告[2]显示我国的安全事件中垃圾邮件、网络钓鱼(Phishing)所占比例高达 64.05%，其中网络钓鱼事件占 27.04%。随着钓鱼攻击日益猖獗，反钓鱼技术成为安全领域的研究热点之一。

本文首先介绍了网络钓鱼攻击，然后分析了域名、域名系统与网络钓鱼之间的联系；并根据已有的可信网络研究总结了互联网应用的可信要素，展开了对网络钓鱼攻击的可信分析。本文主要对基于域名的反网络钓鱼治理机制进行了研究。

① 基金项目:国家发展改革委员会 CNGI 项目(CNGI-09-03-04)
收稿时间:2010-02-02;收到修改稿时间:2010-03-04

2 网络钓鱼与域名

2.1 网络钓鱼与域名

网络钓鱼攻击是指向受害者发送电子邮件或即时消息欺骗他们登录网站提供个人身份等隐私信息的行为[3]。网络钓鱼攻击手段从最初的社交工程、网址网页模仿，逐渐与黑客手段结合，如结合恶意代码、SQL注入、域名劫持等，向着智能化、隐蔽化的方向发展。

2.2 域名、域名系统

《中国互联网络域名管理办法》中规定：域名是互联网络上识别和定位计算机的层次结构式的字符标识，与该计算机的互联网协议地址相对应。域名是连接主机底层资源和广泛的互联网应用的枢纽，在互联网的普及和发展中起着非常重要的作用。

域名系统(Domain Name System)为因特网上的主机分配域名地址和IP地址。域名系统采用C/S结构，DNS 客户端发出域名解析请求，DNS 服务器做出响应。域名服务是一项互联网基础资源服务，互联网应用必须通过查询基础资源服务获得相关资源信息后才能进行数据通信和互联互通。因特网的域名空间是一个树状结构，树的顶层是根域的服务器，根下有若干顶级域组成，每个域划分成若干子域，子域又被进一步划分，以此类推。每级域服务器对本域名直接管辖的下级域名给予权威认定。

2.3 域名与网络钓鱼的联系

域名作为互联网资源标识，为 Web 访问、电子邮件、FTP 等应用服务提供资源的寻址和定位。由于域名注册相对自由，用户能很快地注册并使用一个域名。并且，域名系统在协议设计、软件实现、使用配置上存在一定的脆弱性[4]，成为黑客的目标。目前发现的利用域名或域名系统进行钓鱼攻击的手段包括但不限于以下几类：

① 恶意注册域名搭建钓鱼网站。根据中国反钓鱼联盟 2009 年上半年收集的钓鱼网站数据分析，钓鱼域名分为以下四类：英文仿冒域名，如用 tsobao-aution.cn 、baiduts.cn ；子域域名仿冒，如 qq.ok673.cn；随机无意义的域名，如 zxszk.cn。此外，随着国际化域名(International Domain Name)的兴起，IDN 仿冒域名钓鱼攻击随之出现，如用 pà ypal.com 仿冒 paypal.com。

② 域名劫持(DNS Hijacking)，指黑客拦截合法的域名解析请求，并将请求导向假的 IP 地址或不执行解析工作使得请求超时而失败。著名的事件有 2010 年 1 月 12 日发生的“百度被黑”事件。

③ DNS 缓存中毒(DNS Cache Poisoning)，又称作域欺骗（Pharming），指向 DNS 服务器注入非法域名地址，使其进入 DNS 服务器缓存，导致域名对应错误的主机 IP 地址。

④ Hosts 文件篡改(Desktop Phishing)，指修改位于 DNS 客户端的 Hosts 文件记录，修改或添加 DNS 记录，使其导向错误站点。

APWG(Anti-Phishing Working Group)的全球钓鱼调查报告[3]指出：2009 年上半年全球共有至少 55689 个被记录的钓鱼攻击，发生在 30131 个特定域名上。钓鱼者恶意注册的域名主要集中在五个顶级域的域名空间，分别是.com，.net，.org，.eu 和.ru。30131 个特定域名中有 13 个是利用 IDN 域名进行仿冒。IDN 域名仿冒将不会在未来发展；子域域名仿冒是未来的发展趋势。

综上分析，域名与网络钓鱼攻击联系密切。一方面，钓鱼者多利用精心注册的域名诱骗用户，域名系统也是常受攻击的对象；另一方面，钓鱼域名和普通的合法域名相比存在着一定的“仿冒特征”，能够使用技术手段进行筛查。并且，从域名处入手也是直接快速的阻断钓鱼攻击的方法。

3 可信域名和反钓鱼治理

3.1 互联网应用的可信要素

因特网不安全因素来自三个方面：先天设计侧重互联互通而非安全性；外在环境不安全；缺乏统一的系统安全标准[5]。过去的互联网研究以追求高效为目标，而现有和未来的网络应用要求互联网朝着高可信、可控、可管的方向发展。网络不仅关注系统和数据的安全，更要关注系统的内容和行为的可信[6]。

目前可信网络的发展尚处于初期阶段。ISO/IEC15408 标准提出“可信”的含义[7]：一个可信的组件、操作或过程的行为在任意操作条件下是可预测的，并能很好地抵抗应用程序软件、病毒以及一定物理干扰所造成的破坏。可信网络平台(TNP)[8]的概念指出如果网络中的行为与行为结果总是预期和可控的，网络就是可信的。也有学者指出新一代互联网应具有可控性、可信性、可扩展性的基本属性[9]。经过分析，我们提出了互联网应用的可信模型，并总结出信息交

互过程中五大可信要素。如图 1 所示。

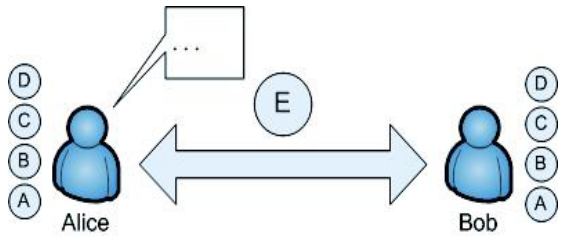

图 1　互联网应用可信模型示意图

ABCDE 五个要素分别是身份可信、内容可信、行为可信、可生存性以及传输和通讯可信。每个属性的涵义释义如表 1 所示。

表 1　互联网应用的可信要素

属性	涵义
A 身份可信	网络参与实体所声称的身份可以被准确鉴别，不被他人冒充。解决身份认证和可追溯性的问题。
B 内容可信	网络实体、服务提供的内容真实可靠。解决信息收集，识别以及控制等方面的问题。
C 行为可信	根据网络参与实体的行为历史，科学的判断其行为是否正常，对其行为进行实时控制，以及预测信用等级。
D 可生存性	对系统或者网络的可生存性进行建模和评估；提供系统或者网络发生意外时的可生存性解决方案。
E 传输和通讯可信	传输过程中是机密，完整、不可否认的。

3.2 用可信域名的思路解决网络钓鱼问题

网络钓鱼攻击行为不是简单的网络安全问题，它暴露出现有网络环境难以满足网络应用安全运行所需的可信要素。网络钓鱼事件主要涉及到可信网络中的“身份可信”、“内容可信”和“行为可信”三个要素。

3.2.1 域名关联信息

域名作为互联网上最重要的基础资源，是网络实体在名字空间的映射。一个域名所关联的信息可分为以下四类，如图 2 所示。

域名字面信息，即呈现给用户的网络资源标识符；域名注册相关信息，包括申请单位、管理联系人、技术联系人、缴费联系人、域名服务器等信息，部分信息可以通过 Whois 数据库进行查询；DNS 系统中的域名解析日志及 DNS 记录；域名对应的资源的内容，包括网页、邮件、图片、文档等。

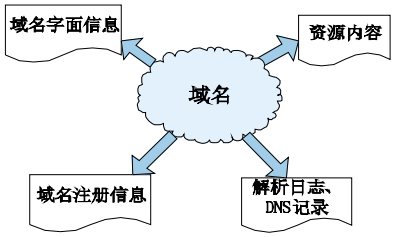

图 2　域名相关信息示意图

3.2.2 构建可信域名

可信域名是可信网络的重要组成部分。实现可信域名，要从域名产生、域名解析、域名应用等多个阶段来保证域名及域名系统的身份可信、内容可信和行为可信：

①域名注册信息的准确、真实、完整；

②域名字面的合法性、无害性及域名所挂载的互联网应用的真实、可用、安全；

③域名注册行为、解析行为的可审计。

在域名层面实现可信，众多建立在域名基础上的应用就可以依据域名这条线索进行监管。①②可以通过加强域名注册阶段进行保障；②可以通过对网页、邮件等典型的域名应用的内容进行检测来实现；③可以在内容、行为检测的基础上，通过对域名注册行为、域名解析行为、域名应用行为、信誉评估等历史数据的综合分析评估来保证。如果整个域名的产生和使用阶段能够保证自身的身份可信、内容可信和行为可信的可信要素，就能自底向上实现对网络不良应用的抵制。

4　基于可信域名的反钓鱼框架

本文围绕可信域名及域名应用服务保护，提出了一个基于域名的反钓鱼体系框架。以往的反钓鱼体系框架多从某个应用场景入手，综合采用多种网络安全技术手段和反钓鱼措施。本文提出基于域名的反网络钓鱼体系框架是一种全新的思路，围绕地址资源设置一系列保护措施治理网络钓鱼犯罪。

4.1 整体框架概述

本反钓鱼框架的体系结构如图 3 所示。设计的理念是：从整体考虑，在域名参与互联网活动的各个位置布置防控，以域名为线索对网络犯罪活动进行检测和跟踪；利用已有反钓鱼技术，信息共享，减少重复建设；不同角色实体间建立快速的通信机制，缩短事

件的处理时间。

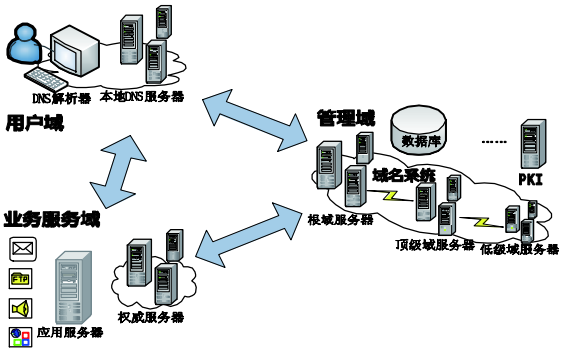

图 3 基于域名的钓鱼治理框架模型

该框架由来自三个域的参与者相互协作，围绕域名的产生和使用对现有的域名系统、用户、业务应用部分加强可信保护，并在适当位置结合已有的反钓鱼技术。

(1)管理域由政府机构、社会团体和第三方机构组成。本区域涉及到域名服务、IP 服务，PKI 服务等网络基础服务。针对域名服务系统，从域名注册、域名系统安全、域名使用等方面加强工作，保障域名、域名服务的身份内容行为可信。此外，要建立快速的钓鱼事件处理响应机制，协调各方制止犯罪行为，比如限制域名解析。

(2)业务服务域内的应用服务商要制定有效的安全管理方案，提供有效的服务；并维护本域内业务服务器、DNS 服务器的运行信息等关键信息，以备管理机关审计使用；建立快速的通信机制，及时向用户和管理机关通报安全威胁。

(3)用户域是包括请求互联网服务的用户主机和其本地 DNS 递归服务器。用户主机上安装多样的反钓鱼代理，与管理机构建立良好的通信渠道。DNS 解析器、本地 DNS 递归服务器上应采取安全措施保证相关 DNS 记录不被非法篡改，且日后可追踪审计。

4.2 相关技术与核心子系统

4.2.1 可信域名相关技术

基于域名的各项可信措施，除去人工干预的部分，需要结合多个领域的技术手段，如模糊集、信息检索、模式识别、神经网络、自然语言理解等领域。下面针对三个不同的可信要素介绍下相关的典型技术。

①域名身份可信包括两方面。域名的合法、无害性检测根据计算的特征的不同，分为基于字符、基于语义、基于统计关联三类方法[10]。域名与其归属人的关系可以通过授权、认证等技术来保证。

②域名内容可信可以通过内容检测技术、安全防护技术来实现。如分析钓鱼邮件[11]、网页文档对象模型[12]、网页转化为图片进行处理[13-14]等，根据特征匹配进行相似性检测或异常检测。

③域名的行为包括注册行为、解析行为和使用行为。对行为的监测是一个动态长期的过程[15]。本框架下审计和取证的主要对象为域名注册信息历史，用户主机的访问日志、应用服务器的工作日志、各级 DNS 服务器的解析日志。重点分析各类域名数量及访问量、域名对应主机 IP 地址的变化频度、站点生存时间；借助于 IP 地址库等资料，分析域名查询的来源及分布，对异常情况及时通报。

4.2.2 核心子系统建设

本框架内拥有控制模块、通信模块、数据库及三个核心子系统。控制模块是整个系统的管理核心，协调子系统之间的合作，调动通信模块完成子系统之间及子系统与外界间的信息互通。如图 4 所示。三个核心子系统通过各自的技术手段获取信息，并存储入数据库。数据库中的数据可在控制模块的调度下，根据不同的算法计算相关的信息提供给有需求者。

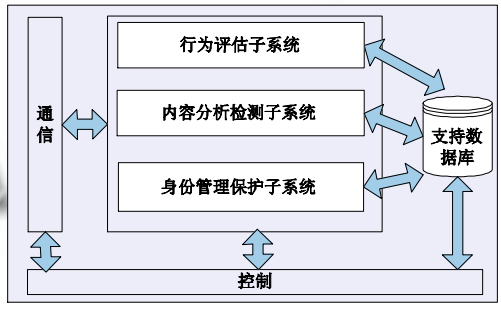

图 4 框架内各个系统之间的配合

(1)身份管理保护子系统

该系统的功能主要是保证域名的身份真实性。具体的工作包括：域名合法无害性的检查，域名注册服务机构资质的检查，域名注册申请信息是否属实的检查；记录域名注册人的注册行为，如是否曾虚假注册，是否经常变更个人信息或 NS 记录等。信息的审核常常需要一些权威数据的支持，比如信誉数据库、公民身份信息库、ICP 备案信息库等。域名的钓鱼倾向检测包含在无害性检测当中。该系统主要由预置关键词

列表，过滤规则、过滤器三部分组成。

域名注册阶段对不合法不真实的域名请求一律拒绝，不完整不准确的退回修改。对发现的具有不良应用倾向的域名建立可疑名单重点监控，起到事前预防的作用。一旦发生网络犯罪，可立即通知有关机构响应处理。

(2)内容分析检测子系统

系统根据域名获取相关应用的服务内容，通过内容分析确定该服务是否包含有害内容，包括钓鱼、黄色、垃圾信息等。系统可设计为 C/S 结构，同时在服务器端设立有害历史信息库，数据库中至少包含域名、类别、发现时间和取证信息等数据项。提取不在数据库中的域名对应的应用数据，然后利用内容检测算法对获取的数据加以分析，如网页上的文本、图片、链接、视频等。检测的结果放入历史库中存储，为后续工作指导。

由于钓鱼等犯罪手段不断地变更，有关的算法和规则要不断完善，数据库也要不断更新。对确认的域名犯罪事件，通过身份管理保护子系统中的域名注册信息查找相关责任人，并通报管理机关处理。

(3)行为评估子系统

对地址资源进行信誉评级。综合 IP 地址、域名和 URL 等不同资源的内容和行为记录，汇集来自于授权客户和第三方联盟的信息反馈、安全产品以及研究团队的风险预警，与目标资源的历史信息进行整合，建立针对互联网领域的长期信誉追踪机制。评估机制一种主动预防机制，对钓鱼攻击起到事前预防的作用。

本系统拥有广泛的数据来源，通过信息分析统计建立信誉信息库。这些数据信息与有害历史信息库中的数据想结合，并进一步挖掘，能为内容分析检测子系统提供支持。同样地，信誉信息库也需要及时更新。

4.3 与其他反钓鱼框架的比较

基于邮件的钓鱼网页检测架构[14]以邮件为入手点，设计了邮件服务器、网络钓鱼分析节点、网络钓鱼控制中心三级防范措施。该架构能实现钓鱼检测信息从检测控制中心-->钓鱼目标-->政府机构的单向通报。本文将从入手点、参与者、信息来源、技术手段、数据库建设等多个方面与本反钓鱼体系结构进行比较，详见表 2。

表 2　反钓鱼体系结构对比

	钓鱼网页检测体系架构	基于域名的反钓鱼体系结构
入手点	电子邮件	域名
参与者	政府机构，钓鱼目标，用户	政府机构，业务应用提供商(不仅仅是钓鱼目标)，用户
信息来源	电子邮件内容、网页内容	域名注册信息、域名解析信息、举报信息、钓鱼电子邮件、网页内容
技术手段	邮件过滤、URL 检测和网页检测	身份认证技术、域名分析检测、内容检测技术、信用评估技术
技术部署位置	邮件服务器	应用服务器(包括网页服务器、邮件服务器等)、DNS 服务器、用户主机
数据库建设	黑白名单	黑白名单、不良信息库、信誉数据库
处理阶段	事中控制、事后审计	事前预防、事中控制、事后审计
报警途径	逐级上报，单向	直接的通信机制，两两可通信

两个框架都涉及到三个领域的参与者，在技术手段尤其是在内容检测技术上有重叠。这是因为网络钓鱼事件涉及到三个方面的参与者，并且内容检测是相对研究成熟的反钓鱼技术，两个框架都进行了考虑。

然而，基于域名的反钓鱼框架从域名入手，具有其独特的优势。探测信息来源丰富，技术部署位置广泛，能实现钓鱼攻击事前预防；直接的通信机制保证了快速的响应处理速度；可通过域名处理来限制钓鱼网站的继续运行。

5　结语

本文分析了如何用可信域名的思路解决网络钓鱼问题，并开创性地提出了基于可信域名的反钓鱼治理机制，强调从互联网基础资源管理出发，深度打击互联网犯罪。

网络钓鱼攻击手段逐渐与域名、域名系统相结合，并呈不断上升的趋势。域名作为一种宝贵的网络资源，被网络钓鱼者的恶意注册和使用既是对网络环境的破坏，又是一种资源浪费。因此，从打造可信域名出发提出的反网络钓鱼框架对于保护互联网资源、快速制

止钓鱼行为具有重要意义，对治理其他网络不良应用也有着启发和借鉴意义。

下一步的工作是对核心子系统的开发及关键算法的改进，以提高检测不良域名和钓鱼网页的效率。

参考文献

1 中国互联网信息中心.第 25 次中国互联网信息中心发展状况统计报告.北京, CNNIC, 201001:3－4.

2 CNCERT/CC. CNCERT/CC2008 年上半年网络安全工作报告.北京, CNCERT/CC, 200811: 6－7.

3 罗广杰,贺敏伟. DNS 协议的安全解析.计算机系统应用, 2004,13(1):36－38.

4 APWG. Global Phishing Survey: Trends and Domain Name Use in 1H2009, 200909:5－6.

5 王功明,关永等.可信网络框架及研究.计算机工程与设计, 2007,28(5):1016－1018.

6 高铁杠,顾巧论,陈增强.可信网络的可信模型与算法设计研究.计算机应用研究, 2007,(06):142－144.

7 ISO/IEC.I Information technology--Security techniques --Evaluation criteria for IT security--Part 1: Introduction and general model. New York：Standard ISO/IEC 15408, 1999.

8 中国信息产业商会. 信息安全产业分会中国信息安全产业发展白皮书(2005－2010). http://www. itsec. gov.cn/webportal/document/baipishu.pdf.

9 林闯,任丰原.可控可信可扩展的下一代互联网.软件学报, 2004,15(12): 1815－1821.

10 章成志.基于多层特征的字符串相似度计算模型.情报学报, 2005,24(6):696－701.

11 陈涓,郭传雄.网络钓鱼攻击的在线检测及防治. 解放军理工大学学报（自然科学版）, 2007,8(2):133－138.

12 Pan Y, Ding XH. Anomaly Based Web Phishing Page Detection. Proc. of the 22nd Annual Computer Security Applications Conference, Washington. DC, USA, 2006:381－392.

13 Liu WY, Huang GL, Liu XY, Zhang M, Deng XT. Detection of Phishing Webpages based on Visual Similarity. Special Interest Tracks and Posters of the 14th International Conference on World Wide Web, Japan, 2005:1060－1061.

14 曹玖新,毛波,罗军周,刘波.基于嵌套 EMD 的钓鱼网页检测算法.计算机学报, 2009,(5):922－929.

15 林闯,田立勤,王元卓.可信网络中用户行为可信的研究.计算机研究与发展, 2008,45(12):2033－2043.