第32卷 第5期 南京邮电大学学报(自然科学版) Vol. 32 No. 5
2012年10月 Journal of Nanjing University of Posts and Telecommunications(Natural Science) Oct. 2012

基于文本特征分析的钓鱼邮件检测

彭富明[1],张卫丰[2],彭 寅[2]
(1. 南京理工大学 自动化学院,江苏 南京 210094
2. 南京邮电大学 计算机学院,江苏 南京 210023)

摘 要:提出了一种基于邮件文本特征的钓鱼邮件检测方法。首先利用邮件解析器将邮件中非文本部分内容剔除,然后提取邮件剩余部分中存在的网站链接及其他内容,并在此基础上提取10种特征。针对这些特征,利用机器学习方法对其进行训练和预测,将邮件分类为普通邮件和钓鱼邮件。我们改进了以往一些针对网站链接分析的检测方法,并结合钓鱼邮件发展的新趋势,提出了6种新的特征。实验证明,本方法结合了新的钓鱼邮件特征,有效地提高了钓鱼邮件检测的召回率以及精准率,同时误判率有所降低。并且,本方法稍加改进以后就能用于钓鱼网站的检测。

关键词:钓鱼检测;邮件;文本特征;网页链接
中图分类号:TP393.08 **文献标识码**:A **文章编号**:1673-5439(2012)05-0140-06

Detecting Phishing Email Based on Text Features Analysis

PENG Fu-ming[1],ZHANG Wei-feng[2],PENG Yin[2]
(1. Automatic School,Nanjing University of Science and Technology,Nanjing 210094,China
2. College of Computer Science & Technology,Nanjing University of Posts and Telecommunication,Nanjing 210023,China)

Abstract: A kind of phishing email detection based on text analysis is proposed. First,we deleted the non-text part of emails by email parsers. For the remaining part of the emails,we got the links and other contents,and extract ten features. According to the analysis of these features,the emails will be classified into ham and phishing by using machine learning method to train and forecast the emails. We improved the existed phishing detection which is based on the analysis of websites´links. By combining with the new trend of phishing email's development,we propose a method to extract some new features. The experiments shows that the proposed method demnonstrates a good performance in trems of recalling rate,false positive rate,and detection of phishing websites.

Key words: phishing detection; email; text feature; link

0 引 言

随着人们越来越多地依赖互联网工作、学习和生活,互联网诈骗已经成为一个越来越大的威胁,比如“网络钓鱼”。“网络钓鱼”的核心是“钓鱼网站”——犯罪分子做出的诈骗网站,通常此类网站与银行网站或其他正规网站外观几乎完全相同,欺骗使用者在“钓鱼网站”上提交他们的敏感信息(如:用户名、口令、帐号ID、ATM PIN码或信用卡详细信息等)[1-4]。最典型的网络钓鱼攻击过程如下:攻击者首先向用户发送一些伪造的“正规网站”的钓鱼邮件,引诱用户点击邮件中的链接(通常使用如用户的网银账户异常需要输入密码验证,电子商务网站有特大优惠等理由欺骗用户),从而将用户骗到攻击者精心设计的与正规网站几乎没有差别的钓鱼网站上来。继而要求用户输入敏感信息,例

收稿日期:2011-09-08
基金项目:江苏省青蓝工程、武汉大学软件工程国家重点实验室开放基金(BJ2110002)、桂林电子科技大学广西可信软件重点实验室开放基金(TJ211037)和苏州大学江苏省计算机信息处理技术重点实验室(KJS0714)资助项目
通讯作者:张卫丰 电话:13776678880 E-mail:zhangwf@njupt.edu.cn

如网银账号密码、电子商务网站账号密码等。由于网络钓鱼的存在,网络电子商务的推广受到了很大阻碍。从网民或者电子商务网站推广的角度,加强对网络钓鱼的源头——钓鱼邮件的检测,防范网络钓鱼,变得越来越重要和有必要。

然而,随着钓鱼邮件过滤技术不断提高,钓鱼邮件发送者不断地改变钓鱼邮件的特征,希望绕过邮件过滤器。而且针对现有的邮件过滤技术而发展的新特征也越来越多,现有的邮件检测和过滤技术已经比较难以应对。以往的人工识别[1]采用黑名单机制,用户对某个网站进行举报,通过人工识别是否为钓鱼网站,检测速度太慢。因此,Ian Fette[3]在 2006 年首先提出针对邮钓鱼网站检测的机器学习方法,在提取去网站链接的相关特征后分别用随机树(PILFER)、支持向量机、决策树、动态贝叶斯等分类器进行训练和测试。发现随机树分类器可以单独使用,也可以结合现有的垃圾邮件(spam)过滤器使用,后者效果更佳。但是随着钓鱼邮件的演化,其中有些特征检测效果较低,而且该方法在处理域名增添字符类的链接时,误判率较高。Bergholz 则认为使用邮件黑名单并不能起太大作用,因为黑名单总是滞后于钓鱼邮件的产生,因此在 Ian Fette 的基础上,除了一些基本钓鱼邮件特征外,提出一种由经过训练的马尔可夫链和潜在的主题等级(Class-Topic)模型产生的邮件特征,然后用分类器进行分类[5];另一种是针对钓鱼者的 Salting 策略,通过隐藏的 Salting 模拟系统提取 OCR 文本,然后基于坚实文本对照技术,利用 2 个不同文本训练得到一个分类器[6]。就基本特征而言,Bergholz 没有 Fette 的性能好,但加上两个基于模型的特征后发现误判数减少了 2/3。也有学者采用分类器分类,Abu-Nimeh 在 2007 从钓鱼网页传播的角度提出了一种特征提取方法,主要比较了下列 6 种机器学习方法在邮件特征分类上的效果: Logistic Regression(LR)、Classification and Regression Trees(CART)、Bayesian Additive Regression Trees (BART)、Support Vector Machines (SVM)、Random Forests (RF)、and Neural Networks (N-Net)。该方法拓展了钓鱼网页的特征,在一定程度上提高了钓鱼网页检测的精度,但抽取钓鱼网页特征时仍然只是采用单个网页的信息,因而容易被钓鱼攻击者欺骗[7-11]。

针对以上问题,我们提出一种基于文本特征分析的钓鱼邮件检测方案,该方法在 Ian Fette 提取的特征的基础上,结合目前钓鱼邮件的发展新趋势,提出了几种新的特征,同时考虑到钓鱼网站多个特征结合的情况,解决了在域名增添字符类链接精度低和误判率较高的问题,同时在精确度和召回率上依然能保证较高的数值。

1　基于文本特征分析的钓鱼邮件检测

文中基于文本特征分析的钓鱼邮件检测,主要是由以下两个部分组成的。

1.1　邮件的文本特征提取

一些垃圾邮件分类器使用上百个特征来检测不需要的邮件。针对这些特征我们做了相应的比较,然后选取了几项特征。同时针对目前钓鱼邮件的演化特征,提出了几种新的钓鱼邮件的特征,然后将这些特征用于邮件分类器。主要特征有以下 10 种:

(1) 基于 IP 类型的网站链接[3]

最早的一些钓鱼网站是由个人 PC 作为主机的,他们没有 DNS 解析,所以最简单的方法就是将网站链接设置成为 IP 地址类型的链接。相比较主流网站的链接特点,文中认为含有 IP 地址类型的网站链接更有可能是潜在的钓鱼网站。比如,如果邮件中出现了类似 htttp://192.168.0.1/taobao.cgi?_account 的网站链接时,就认定该邮件是一个钓鱼邮件。当然这种钓鱼网站出现的时间比较早,不过仍然是一种比较有用的特征。主要提取步骤如下。算法 1 为网站链接提取步骤,算法 2 为 IP 类型链接判定步骤。

算法 1　邮件内网站链接的提取

输入: 去除非文本内容的邮件 mailContent

伪代码:

```
读入 mailContent
While( mailContent 非空)
do{ 查找正则表达式为( "href = □( [^\"] + ( \'?) ") 或者( "www\.) ( [^\s] +?) ( ( \ "|( /) ) ") 的字符串;
    while( 如果查找到相关匹配字符串) {
        if( 链接不重复)
        将所得字符串加入链接列表 Links; } }
```

输出: 邮件中所有的不重复的网站链接 Links

算法 2　IP 类型链接的提取

输入: 该邮件中所有不重复的网站链接 Links

伪代码:

```
逐个读入 Links 中的链接
do{ 查找是否含有正则表达式为( http://|https://) +
 + ( ( 2 [0-4] [0-9] |25 [0-5] | [01]? [0-9] [0-9]?)
\.) {3} ( 2 [0-4] [0-9] |25 [0-5] | [01]? [0-9] [0-9]?) )
{1,} ") ;
```

结果记录为 result}
if(result 为 1)
{跳出循环输出结果为 ture;}
else{输出结果 false;}
}
输出: true/false(含有 IP 类型链接为 true; 不含为 false)

(2) 链接中域名的注册时间[3]

相比较那些使用 IP 地址型的钓鱼网站,目前攻击者已经改变了策略,注册一个比较接近正规网站的域名,然后使用该域名的链接进行攻击。如果用户没有注意此类域名和正规网站域名之间的差异的话,就很容易被欺骗。此类钓鱼网站一般存在的时间比较短,通常在几天到几个小时不等。所以,钓鱼者一般在注册以后很短的时间就使用该域名,域名注册的时间很短。本文利用 WHOIS 查询每个链接中的域名。然后设定一个阈值,如果域名没有超过该阈值,则很有可能为钓鱼链接。本文中阈值设定为 50 天,判定过程如算法 3 所示。

算法 3 域名注册时间的判断

输入: 去除非文本内容的邮件 mailContent
遍历 mailContent;
While(mailContent 非空)
{查找其中开头为"Creation Date"的字符串;
获取邮件创建的时间 time_1;}
提取 Links 其中的主域名 domain;
访问 WHOIS 域名管理服务器查询 domain 的注册时间 time_2;
if((time_2 - time_1) >50 天)
{结果为 false};
else{结果为 true};
输出: true/false(不超过 50 天为 true; 否则 false)

(3) 链接中含有诱导点击的模块[3]

一般钓鱼攻击者为了诱导用户进入设计好的钓鱼网站,在邮件中会设置一些类似"Click"或者"Here"标题的标记语言模块。点击以后则将用户导向到钓鱼网站,从而骗取其个人敏感信息。因此,本文认为如果邮件中出现此类标记语言,很有可能为钓鱼链接。

算法 4 诱导点击模块的判断

输入: 去除非文本内容的邮件 mailContent
While(mailContent 非空) {
查找是否含有正则表达式为("< a. * href. * >. * (link|here|click) . * ") 的字符串内容;
if(如果查找到)
{输出结果 true;}
else{输出结果 false;}}
输出: true/false(含有诱导模块为 true; 否则为 false)

(4) 登录链接域名与邮件发送者邮箱域名不符

通常,正规网站给用户发送验证类的邮件时,会使用自己注册的域名邮箱发送。而钓鱼者无法获得此类域名特定 ID 的邮箱,只能通过其他网站的邮箱来发送钓鱼邮件。例如,收到一封伪造的淘宝的邮件,而邮件来源于 account. taobao@ tom. com。本文将邮件中所含的链接分为登录链接和非登录链接,将登录链接中的域名和邮件发送者的邮箱域名进行比较。若登录链接中的域名和发送者邮箱的域名不一致,则很有可能为钓鱼邮件攻击。

(5) 登录链接中的域名与 B_Name 不符

钓鱼攻击都想让收件人相信这封邮件是一封合法的邮件,所以在邮件中可能会多次使用合法网站的域名,我们称之为 B_Name。比如一封伪造的淘宝邮件,邮件中会多次出现"taobao"的字样。我们使用 tf_idf 算法将这类词提取出来,作为 B_Name,然后将其与登录链接的域名进行比较。如果不同,则该邮件很有可能是钓鱼邮件。

算法 5 提取 B_Name 的步骤

输入: 去除非文本内容邮件 mailContent
While(mailContent 非空) {
截取内容中的各个单词;
if(单词首字母为大写字母) {
单词加入字符串列表 Words;
该单词词频 +1;}
比较 Words 中单词的词频并返回词频最高的单词 word;
B_Name = word;}
提取邮件头部中的发送人域名 domain;
If(判断 B_Name 与 domain 一致) {
return fasle;}
else{return true;}
输出: true/fasle(一致为 fasle; 不一致为 true)

(6) 含有 html 语言[3]

邮件根据 MIME 协议可以分为纯文本、纯 html 和两者混合的 3 种类型。钓鱼邮件中很多时候必须使用 html 语言(虽然普通邮件中也可能含有 html 语言) 。如果不使用 html 语言,则钓鱼攻击者很难进行钓鱼攻击。因此,邮件中如果含有 html 语言,则有可能为钓鱼邮件。

算法 6 html 特征的提取

输入: 去除非文本内容邮件 mailContent
While(mailContent 不为空) {
读取邮件头部信息;
if(含有 html 或者 html/text) {
结果为 true;}
else{结果为 false;}
输出: true/fasle(含有 html 语言为 true; 不含则为 false)

(7) href 中链接域名与网页的展开字符串(Display String)不符

钓鱼攻击者在选择了使用 html 语言后,会制造一个类似正规网站的钓鱼网站链接。通常,此类链接的形式为 < a href = "http://www.taobao1932.com" > taobao.com < /a > 。此链接所导向的网页 Display String 为 taobao,但是导给用户的是一个域名为 taobao1932 的网站。所以,本文认为这是一个钓鱼攻击的特征。

算法 7　链接导向网页是否与 Display String 相符

输入: Links 中不同的 href 链接
　逐个读入 href 链接;
　查找是否含有正则表达式为(< a + href = \" + (. *) \" + > (. *) < /a >) 的字符串;
　提取 href 标记之后链接中的 domain_1;
　提取 DisplayString 中的 domain_2;
　if(domain_1 = domain_2) {
　　返回结果 true; }
　Else{ 结果 false; }
输出: true/false(导向正确为 true; 导向错误为 false)

(8) 链接中点号分隔符的个数[3]

钓鱼攻击者为了让钓鱼网站的链接看起来和正规网站链接很像,会想尽办法把真的域名隐藏起来,不容易让用户看到。这样链接的长度必然会很长,直接的结果就是链接中的". "分隔符的个数会比较多。

(9) 链接中斜杠分隔符的个数

此特征的原理和(8)中的点号个数原理一样。

算法 8　链接中点号分隔符和斜杠分隔符的提取

输入: 不同链接的列表 Links
　遍历 Links 中每个 Links;
　查找是否含有 "/") 的字符串
　　{ while(查找到的是"/"&& 不是"//") count_1 + + ; }
　查找是否含有 ". ") {
　　While(查找到) count_2 + + ; }
输出: count_1; count_2(count_1 为斜杠号的个数; count 为点号的个数)

(10) 链接中 http 协议使用的次数

钓鱼链接中有时会多次使用 http 协议簇,然后改变链接导向,将用户导向设计好的钓鱼网站中去。例如如下链接 http://www.sina.com.cn/url? q = http://www.s289iwb53.com. 该链接看起来似乎是导向新浪主页,而事实上当用户点击时会被重定向到后面的伪造网站上去。因此,本文认为多次使用 http 协议的链接,很有可能是钓鱼网站链接。

邮件特征的主要提取过程如图 1。

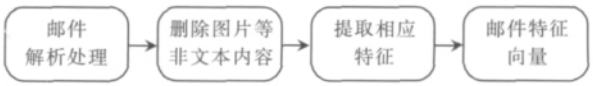

图 1　邮件特征向量提取过程

则我们得到的邮件特征向量可以表示为: $E_i = (a_1, a_2, \cdots, a_{10})$ 。

1.2　基于文本特征的邮件检测

本部分的主要思想是,先利用标记好的部分邮件作为训练集,利用训练集训练出特定的分类器模型。然后通过该模型来测试余下部分的邮件,对其进行分类,得出结果后再完善该分类器模型。重复此过程数次,即可得到所需要的分类器模型。主要过程如图 2 所示。

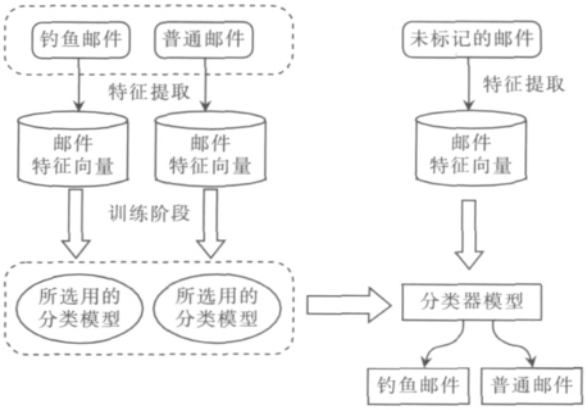

图 2　钓鱼邮件分类检测的过程

如图 2 所示,先将一部分的邮件(包含普通邮件和钓鱼邮件)标记好类别作为分类器的训练集,然后对它们进行相应的文本特征提取(对应 1.1 中提出的 10 项基本特征),得到该训练集的邮件特征向量。然后选择分类器,利用这些标记好的特征向量进行训练,从而得到所需的分类器模型。接着,对未标记的邮件提取邮件特征,并利用此前训练好的分类器模型进行预测分类,最终将该邮件分为钓鱼邮件或者普通邮件。至此,整个邮件的分类检测过程完成。

2　实　验

第 1 部分介绍了实验方法,下面将结合实际的邮件数据进行实验。实验代码的编写环境为 MyEclipse7.5; 数据处理的硬件环境 CPU 是主频 2.0 GHz 的 AMD Turlon64,内存 1 GB,操作系统为 Windows XP Professional SP2; 结果分析软件使用的是 weka 中自带的几种分类器。以下为实验具体内容:

2.1 实验准备

实验数据集主要来自 http://monkey.org/~jose/wiki/doku.php?id=PhishingCorpus/(一个钓鱼邮件检测组织的网站),普通邮件来自 http://spamassassin.apache.org/publiccorpus/,所提供的邮件数据,已经通过邮件解析器解析完成。本文实验数据中的钓鱼邮件数量为500封,标记为phishing;普通邮件的数量为500封,标记为legal。将这1000封邮件的特征向量提取出来,作为本次实验所使用的特征向量集。

通过参考大量国内外有关网络钓鱼检测的实验研究,总结出主要的评价指标形式有TPR、FPR、Precision(精确率)、AUC值和准确率等。通过比较,本文认为此处使用TPR、FPR和Precision更有意义。它们计算公式如下:

$$\mathrm{TPR}=\frac{Phish_{\mathrm{phish}}}{Phish_{\mathrm{phish}}+Phish_{\mathrm{ham}}}$$

$$\mathrm{FPR}=\frac{Ham_{\mathrm{phish}}}{Ham_{\mathrm{phish}}+Ham_{\mathrm{ham}}}$$

$$\mathrm{Precision}=\frac{Phish_{\mathrm{phish}}}{Phish_{\mathrm{phish}}+Ham_{\mathrm{phish}}}$$

其中,$Phish_{\mathrm{phish}}$指原本为钓鱼邮件被判定为钓鱼邮件的个数,Ham_{phish}指原本为普通邮件被误判为钓鱼邮件的个数,Ham_{ham}指原本为普通邮件被判定为普通邮件的个数,$Phish_{\mathrm{ham}}$指原本为钓鱼邮件被误判为普通邮件的个数。

2.2 实验过程和结果

实验过程:先利用标记好的一部分邮件作为训练集,训练出特定的分类器模型。然后通过该模型预测余下部分的邮件,对其进行分类,得出结果后再完善该分类器模型。最后分析实验结果数据。

本文在提取了邮件特征向量以后,使用了weka中现成的分类器来进行实验,为了通过比对找到合适的分类器,我们选择使用决策树分类器、LR(Logistic Regression)和贝叶斯分类器来进行比较。在利用10折交叉验证的基础上,所得的实验结果如表1所示。

表1 实验结果分析

分类器	TPR	FPR	Precision	Recall	ROC area
贝叶斯	0.958	0.366	0.724	0.958	0.963
决策树	0.952	0.036	0.964	0.952	0.983
LR	0.970	0.042	0.958	0.970	0.990

通过比较3种不同的分类器,可以发现,除了在使用贝叶斯分类器时FPR的数值过高外,其他评价指标都比较满意。尤其在使用LR和决策树分类器时,在FPR较低的情况下,TPR和Precision指标都令人满意。

我们还对已有几种钓鱼邮件检测的方法做了实验分析,首先选用I. Fette提出的10个特征检测算法(10 features)[3],接着选用了由Guang Xiang和Jason I. Hong基于IF-IDF提出的2种检测算法strategy I和II。通过实验,得到相应的实验结果如表2所示。

表2 其他检测方法实验结果

检测算法	TPR	FPR	Precision
10 features	0.914 0	0.059 8	0.908 3
Strategy I	0.933 1	0.023 6	0.941 7
Strategy II	0.900 6	0.019 5	0.921 5

通过对比实验结果,与10 features相比,本文的方法在TPR和Precision上有了比较明显的提高,同时在FPR上有所降低。在使用10 features方法进行试验时发现,由于钓鱼邮件制作者为了通过垃圾邮件过滤器的检测,已经有针对性的对邮件的特征进行改进,因此该方法所提出的部分特征的代表性已经不突出。比如10 features中认为钓鱼邮件中的网站链接会比较多,但是目前的钓鱼邮件目的很明确,就是要将受害者引导到钓鱼网站上来,而且此类网站一般页面不多,所以钓鱼邮件中的链接反而会比较少;另外该方法检测文本内容比较少的邮件和关键词有插入字符的邮件时效果比较差,从而导致Precision的数值不高。但是该方法提出的特征思路仍然值得借鉴。

而在使用Strategy I和II时,我们发现,这两种方法的结果过于依赖所用的搜索引擎。例如针对某个域名分别使用google和bing时,搜索排名差异比较明显,故实验结果很容易受到搜索引擎的影响,因而TPR不高并且结果不稳定。此外,对于附加符的多态性分析存在问题,例如"espaňol"会被解析为"espa"和"ol",从而导致检测不到phish对象。另外,有些钓鱼邮件中的钓鱼网站存在的时间可能会稍微长一点,会出现搜索排名较高的情况,此时使用这2种方法时,检测的效果较差。而使用我们的方法,不需要基于搜索引擎排名,故稳定性比上述3种方法高,结果也更加准确。

由于连接WHOIS服务器处理数据,会受到服务器响应时间、网络延迟等网络条件的影响。而Strategy I和II方法与上述两种方法相比,检测的主要步骤需要大量的网络访问,所以此处也不参与运行时

间的比较。比较 10 features 和本文提出的改进方法的实验运行时间时,发现后者的运行时间有少许下降。但对于实际检测钓鱼邮件时,因为待测邮件数量少,所以体现在检测时间上的差别不会很明显。

3　结束语

文中使用了针对邮件中文本内容的特征提取以及分类方法,利用特征向量训练了相应的分类器模型,然后利用分类模型对未标记的邮件进行分类。实验表明,在使用了合适的分类器建立分类模型的情况下,本方法的实验结果在 TPR 和 Precision 这 2 个指标上的数值都很高,同时误判率控制在可以接受的范围之内。因而,我们认为本方法在检测钓鱼邮件方面是有效可行的。同时在以后的工作中,主要的工作目标放在发掘钓鱼邮件发展的趋势以及改进相关特征提取方面,以提高邮件分类检测的准确性。

参考文献:

[1] CRANOR L,EGELMAN S,HONG J,et al. Phishing phish: An evaluation of anti-phishing toolbars [EB/OI]. http://www.cylab.cmu.edu/research/techreports/2006/tr_cylab06018.html.

[2] COLLIN J,SIMON D R,TAN D S,et al. An Evaluation of Extended Validation and Picture-in-Picture Phishing Attacks [C]//Proceedings of Usable Security (USEC'07). 2007.

[3] FETTE I,SADEH N,TOMASIC A. Learning to Detect Phishing Emails [EB/OI]. http://reports-archive.adm.cs.cmu.edu/anon/isri2006/abstracts/06-112.html.

[4] ABU-NIMEH S,NAPPA D,WANG X,et al. A Comparison of Machine Learning Techniques for Phishing Detection [C]//Proceedings of the anti-phishing working groups 2nd annual eCrime researchers summit. New York: ACM. 2007.

[5] BERGHOLZ A,CHANG J H,PAAβ G,et al. Improved Phishing Detection Using Model-based Features [C]//Proceedings of the Conference on Email and Anti-Spam (CEAS). 2008.

[6] BERGHOLZ A,PAAβ G,REICHARTZ F,et al. Detecting Known and New Salting Tricks in Unwanted Emails [C]//Proceedings Conference on Email an AntiSpam(CEAS). 2008.

[7] ZHANG Y,HONG J,CRANOR L. CANTINA: A Content-Based Approach to Detecting Phishing Web Sites [C]//Proceedings of the 16th International Conference on World Wide Web. 2007.

[8] BERGHOLZ A,BEER J D,GLAHN S,et al. New Filtering Approaches for Phishing Email [J]. Journal of Computer Security,2010,18(1):7-35.

[9] ALBRECHT K,BURRI N,WATTENHOFER R. Spamato—An Extendable Spam Filter System [C]//2nd Conference on Email and Anti-Spam (CEAS). Palo Alto,California,USA,2005.

[10] CHANDRASEKARAN M,KARAYANAN K,UPDAHYAYA S. Towards phishing e-mail detection based on their structural properties [EB/OL]. http://www.albany.edu/iasymposium/proceedings/2006/chandrasekaran.pdf.

[11] MIYAMOTO D,HAZEYAMA H,KADOBAYASHI Y. A proposal of the Ada Boost-based detection of phishing sites [C]//Proceedings of the Joint Workshop on Information Security. 2007.

作者简介:

彭富明(1965-),男,江苏宜兴人。南京理工大学自动化学院副教授。主要研究方向为 Spam 检测技术、汽车自动化技术等。

张卫丰(1975-),男,江苏南通人。南京邮电大学计算机学院教授,博士。(见本刊 2012 年第 4 期第 69 页)

彭　寅(1986-),男,江苏宜兴人。南京邮电大学计算机学院硕士研究生。主要研究方向为 Spam 检测技术、搜索引擎技术和移动电子商务。

(本文责任编辑: 潘雪松)