



NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

A DEFENSE-IN-DEPTH APPROACH TO PHISHING

by

David S. Barnes

September 2006

Thesis Co-Advisors:

Craig H. Martell
Neil C. Rowe

Approved for public release; distribution is unlimited.

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			<i>Form Approved OMB No. 0704-0188</i>	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.				
1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE September 2006	3. REPORT TYPE AND DATES COVERED Master's Thesis	
4. TITLE AND SUBTITLE A Defense-in-depth Approach to Phishing			5. FUNDING NUMBERS	
6. AUTHOR(S) David S. Barnes				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000			8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A			10. SPONSORING/MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.				
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.			12b. DISTRIBUTION CODE	
13. ABSTRACT (maximum 200 words) Phishing is a form of crime in which identity theft is accomplished by use of deceptive electronic mail and a fake site on the World Wide Web. Phishing threatens financial institutions, retail companies, and consumers daily and phishers remain successful by researching anti-phishing countermeasures and adapting their attack methods to the countermeasures, either to exploit them, or completely circumvent them. An effective solution to phishing requires a multi-faceted defense strategy. We propose a model for phishing. We report on a survey we conducted of user detection of phishing. We also report on experiments to assess the success of automated methods for assessing clues to phishing email. We present recommendations for a defense-in-depth strategy to prevent phishing.				
14. SUBJECT TERMS Phishing, Social Engineering, Pharming, Fraudulent Web site, Anti-Phishing Working Group			15. NUMBER OF PAGES 89	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UL	

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited.

A DEFENSE-IN-DEPTH APPROACH TO PHISHING

David S. Barnes
Lieutenant, United States Navy
B.S., United States Naval Academy, 2000

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN COMPUTER SCIENCE

from the

**NAVAL POSTGRADUATE SCHOOL
September 2006**

Author: David S. Barnes

Approved by: Craig H. Martell
Thesis Co-Advisor

Neil C. Rowe
Thesis Co-Advisor

Peter J. Denning
Chairman, Department of Computer Science

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Phishing is a form of crime in which identity theft is accomplished by use of deceptive electronic mail and a fake site on the World Wide Web. Phishing threatens financial institutions, retail companies, and consumers daily and phishers remain successful by researching anti-phishing countermeasures and adapting their attack methods to the countermeasures, either to exploit them, or completely circumvent them. An effective solution to phishing requires a multi-faceted defense strategy. We propose a model for phishing. We report on a survey we conducted of user detection of phishing. We also report on experiments to assess the success of automated methods for assessing clues to phishing email. We present recommendations for a defense-in-depth strategy to prevent phishing.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

I.	INTRODUCTION.....	1
A.	SOCIAL ENGINEERING VIA PHISHING	1
II.	BACKGROUND	7
A.	FORMAL DEFINITION OF PHISHING	7
B.	CURRENT ANTI-PHISHING POSTURE.....	7
C.	PHISHING RESEARCH EXPERIMENTS	8
1.	Web Site Credibility Experiment	8
2.	Social Networking Experiment.....	11
3.	Security Toolbar Experiment	12
4.	Spear-Phishing Experiment	13
5.	Mailfrontier Email Experiment.....	13
III.	A PHISHING MODEL.....	15
A.	WHY MODEL PHISHING?	15
B.	PHASES OF ATTACK	15
1.	Target Determined.....	16
2.	Attack Developed	17
3.	Attack Performed.....	17
4.	Data Accumulated.....	18
5.	Fraud.....	18
C.	SAMPLE PHISHING ATTACK.....	18
IV.	PHISHING RESEARCH METHODS AND EXPERIMENTS	27
A.	OVERVIEW	27
B.	SURVEY	27
1.	Design	27
2.	First Run	33
3.	Second Run	33
C.	CONDITIONAL PROBABILITY STUDY	33
V.	RESULTS	35
A.	FIRST SURVEY	35
B.	SECOND SURVEY	37
C.	COMBINED RESULTS FOR BOTH SURVEYS	38
D.	DECEPTION AND CLUES IN PHISHING	40
E.	CONDITIONAL PROBABILITY STUDY	40
VI.	COUNTERMEASURES AND FUTURE WORK	45
A.	SUMMARY OF DEFENSE MECHANISMS AND COUNTERMEASURES	45
B.	EXPANDING CURRENT RESEARCH	49
	APPENDIX A: SURVEY	51
	APPENDIX B: UNIGRAM SAMPLE OUTPUT.....	65

APPENDIX C: BIGRAM SAMPLE OUTPUT	69
LIST OF REFERENCES.....	71
INITIAL DISTRIBUTION LIST	73

LIST OF FIGURES

Figure 1.	Model of a phishing attack.....	16
Figure 2.	(From APWG 2006) Fraudulent email to PayPal.com users	19
Figure 3.	(From APWG 2006) Fraudulent PayPal Web site.....	20
Figure 4.	(From APWG 2006) Web site properties	21
Figure 5.	(From APWG 2006) Page one of three sequential pages asking for personal information	22
Figure 6.	(From APWG 2006) Page two of three sequential pages asking for personal information	23
Figure 7.	(From APWG 2006) Page three of three sequential pages asking for personal information	24
Figure 8.	Legitimate PNC Bank email	29
Figure 9.	Web site associated with PNC Bank email.....	30
Figure 10.	Survey answer form	32
Figure 11.	Combined survey results	39
Figure 12.	Precision vs. Recall for unigram clues.....	42
Figure 13.	Precision vs. Recall for bigram clues.....	43
Figure 14.	Example of proper email correspondence (Wells Fargo Bank).....	46

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

Table 1.	(From Dhamija Survey 2006): Strategy employed by each site	10
Table 2.	(From Jagatic 2005) Results of social network phishing attack and control experiment.....	12
Table 3.	(From Jagatic 2005) Gender effects on experiment.....	12
Table 4.	Results for fraudulent email/site survey#1	35
Table 5.	Results for fraudulent email/site survey #2	37
Table 6.	Precision and Recall with given odds threshold for single word clues to identify phishing in our test set.....	41
Table 7.	Precision and Recall with given odds threshold for double word clues to identify phishing in our test set.....	42

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

Thank you to my thesis advisors Craig Martell and Neil Rowe, for your time, guidance, and dedication. And a special thanks to Mike Lowe, for being a sounding board throughout the thesis process.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. SOCIAL ENGINEERING VIA PHISHING

Millions of sophisticated social engineering attacks aimed at eliciting sensitive or confidential information target federal agencies, corporations, and consumers daily. Many such attacks are classified as phishing, also called carding or brand spoofing. For the purpose of this research, phishing is defined as the use of a deceptive email aimed at luring the recipient to follow a link to a fraudulent website. Once trust is established, the ultimate goal of a phisher is to obtain sensitive information, such as personal, financial, corporate, or network information (James, 2005). The Anti-Phishing Working Group, a global consortium of industrial and law enforcement groups, claims that the term "phishing" comes from an analogy that Internet scammers are using mass email lures to fish for unsuspecting Internet users to divulge sensitive information such as passwords, social-security numbers, and credit-card numbers. The term phishing derives its spelling from an earlier scam called phreaking, where scammers dialed free long-distance telephone calls. Phishing techniques are in a continual state of flux due to target exposure and security advances (Anti-phishing Working Group (APWG), 2006).

Phishing scams began over ten years ago with the advent of the World Wide Web and widespread use of the Internet. However, the original scams were often in a different form. Early examples used Internet Relay Chat and America Online (AOL) instant messaging. A typical scam involved an AOL user receiving a message from a phisher disguised as an AOL administrator. It alerted the user to an urgent service-related or billing-related problem, and requested password and credit-card information to remedy the situation. Today, phishing predominately attacks financial institutions and their customers. According to the Anti-Phishing Working Group, approximately 90 percent of all attacks target financial institutions, while the remaining 10 percent target Internet Service Providers, retail companies, and miscellaneous organizations (APWG, 2006). The scams are successful because the phishers use company names, logos, and disclaimers in emails sent to customers and many customers do not question an authentic-looking email. Phishers then use legitimate-sounding verbiage to lure customers to a fraudulent website where customers provide their personal information. Since 2003,

financial institutions have reported an estimated loss of one to two billion dollars annually as a result of phishing. However, the figures are likely higher due to the difficulty in associating stolen identity or money to a phishing scam.

The numbers and cleverness of phishing scams are escalating at an alarming rate. In the May 2006 Phishing Activity Trends Report, the following statistics regarding phishing scams are highlighted:

- Number of unique phishing emails reported in May: **20,109**
- Number of unique phishing sites discovered in May: **11,976**
- Number of brands hijacked by phishing campaigns in May: **137**
- Number of brands comprising the top 80% of phishing campaigns in May: **20**
- Country hosting the most phishing websites in May: **United States**
- Percentage containing some form of target name in URL: **46 %**
- Percentage without hostname, just IP address: **42 %**
- Percentage of sites not using port 80: **8 %**
- Average time online for site: **5.0 days**
- Longest time online for site: **31 days**

A year earlier, the Report said:

- Number of unique phishing emails reported in May: **14,987**
- Number of active phishing sites reported in May: **3326**
- Number of brands hijacked by phishing campaigns in May: **107**
- Number of brands comprising the top 80% of phishing campaigns in May: **7**
- Country hosting the most phishing websites in May: **United States**
- Percentage containing some form of target name in URL: **46%**
- Percentage without hostname, just IP address: **42%**
- Percentage of sites not using port 80: **8 %**
- Average time online for site: **5.8 days**
- Longest time online for site: **30 days**

In one year the number of email scams increased by over 5,000 and the number of phishing sites increased by almost 9,000. Nearly half of phishing websites contain some form of the target name in the URL (APWG, 2006). Internet users are thus easily

victimized by such deceptive tactics. A standard phishing email is sent to 100,000 recipients and approximately five percent of the recipients visit the phishing site. Of those 5,000 potential victims, phishers typically defraud 100 recipients. And of those victimized, most provide all requested information such as, name, address, phone number, Social Security number, account number, credit-card number, and passwords. Phishers use the information for identity theft, money laundering, or traditional theft (James, 2005).

Initially, phishers had little technical savvy and their scams predominantly targeted low-hanging fruit. Today, phishers continually hone their email scams to circumvent spam filters, as well as to appear more convincing and legitimate to email recipients. As anti-phishing groups discover and expose current phishing tactics, the phishers adapt and sometimes exploit the consternation created by phishing alerts by using security needs as a lure.

The sophistication of the more recent phishing scams has created a new cybersecurity threat for government as well. Government can be targeted to disclose national-security information or allow unauthorized access to classified information systems, and financial fraud is also a concern. Phishers have targeted federal information systems and personnel at the Federal Bureau of Investigation, Federal Deposit Insurance Corporation, and the Internal Revenue Service (Government Accountability Office (GAO), 2005). The Department of Defense information systems could be threatened by phishers associated with terrorism, foreign intelligence-gathering, and cyberwarfare.

A particularly dangerous method of attack on government is "spear phishing" (Message Labs, 2005). It is a scam appearing to originate from within an organization, often from the Information Technology (IT) department or senior personnel, and targets specific individuals. Spear phishing is a highly targeted scam aimed at individuals in certain key positions or who possess desired sensitive information, contrived to gain access to corporate or government networks and ultimately, sensitive or classified information.

No one strategy or methodology has adequately countered phishing attacks. Phishing has become more menacing and costly for both organizations and consumers. Previous defenses have generally focused on one aspect of the attack such as blocking

known phishing sites. This allows a phisher to employ another approach to bypass the countermeasure. Defense in depth, a strategy employing anti-phishing tactics at multiple phases of an attack, is imperative in adequately foiling phishing attacks.

Developing a defense-in-depth strategy requires a good model of a phishing attack. So we propose one here that details each phase of a phishing attack from determining a target to the ultimate fraud. One can use the model to design defense methods, properly safeguard vital stages in the attack, and highlight who should be responsible for safeguards at each stage in the attack. A defense-in-depth strategy hinders the phisher's ability to completely avoid security mechanisms while increasing the probability of user detection.

Unfortunately, previous research has not done experiments that include all phases of a phishing attack. For example, researchers have asked users to identify emails or websites appearing to be fraudulent, but have yet to conduct a combined experiment with both. So we do one here. Such an experiment is useful in exposing the factors of a phishing scam that participants focus on to judge authenticity. The experiment also reveals those factors that deceive the participants and produce a false negative classification.

Additionally, we perform experiments to verify the utility and accuracy in classifying emails as phishing based on word probabilities. The experiment is useful in determining conditional probabilities of words appearing in a phishing email while not in a legitimate email. The conditional probabilities reveal another tool in the defense-in-depth strategy.

Our research is focused on answering the following questions:

- How do we model phishing from an information-assurance perspective?
- What are the most successful types of phishing attacks?
- Why do certain types of phishing attacks succeed?
- Can phishing attacks be detected before any compromise of information?
- What are the long term prospects for phishing?
- How does the effectiveness of phishing change over time?
- Can awareness training be designed from phishing taxonomy?

- What guidelines can we formulate to focus our resources in fighting phishing?
- Who is responsible for defense against phishing attacks?
- What countermeasures can be taken to reduce the threat of phishing?

The remaining chapters of the thesis are outlined as follows; Chapter II provides the background information on previous experiments and defense strategies. Chapter III describes our application, taxonomy, and experiment setting, assumptions, constraints, and users. Chapter IV details our experiments. Chapter V discusses our research results and conclusions. Finally Chapter VI highlights areas of future work.

THIS PAGE INTENTIONALLY LEFT BLANK

II. BACKGROUND

A. FORMAL DEFINITION OF PHISHING

Phishing is a criminal activity employing social-engineering tactics to defraud Internet users of sensitive information and steal credentials, money, and/or identities. A phishing attack begins with a spoofed email masquerading as trustworthy electronic correspondence that contains hijacked brand names of banks, credit card companies, or e-commerce sites. The language of a phishing email is misleading and persuasive by generating either fear or excitement to ultimately lure the recipient to a fraudulent Web site. It is paramount for both the phishing email and Web site to appear credible in order for the attack to influence the recipient. “Without credibility, sites are not likely to persuade users to change their attitudes or behaviors – to embrace the site’s cause, register personal information, make purchases, click on ads or complete surveys” (Fogg, 2003). As with the spoofed email, phishers aim to make the associated phishing Web site appear credible. The legitimate target Web site is mirrored to make the fraudulent site look professionally designed. Fake third-party security endorsements, spoofed address bars, and spoofed padlock icons falsely lend credibility to fraudulent sites as well. The persuasive inflammatory language of the email combined with a legitimate looking Web site is used to convince recipients to disclose sensitive information such as passwords, usernames, credit card numbers, social security numbers, account numbers, and mother’s maiden name.

B. CURRENT ANTI-PHISHING POSTURE

Phishing has threatened financial institutions, retail companies, and consumers for the past decade; however, security professionals and anti-phishing vendors did not begin focusing their efforts on phishing until roughly three years ago. Phishers have remained successful over the past decade by researching anti-phishing “solutions” and quickly adapting their attack method, either to exploit, or completely circumvent the supposed solutions. The prevailing defense mechanism against phishing has been to educate the user on known phishing methodologies and warn against divulging sensitive information to unauthorized people (APWG, 2006). As organizations or anti-phishing groups become

aware of specific phishing incidents, warnings and descriptions of the attack are issued to consumers. Such reactionary methodologies primarily attempt to counter a phishing attack by thwarting a single aspect of it. Phishers can often adjust their tactics to stay one step ahead of security professionals. But a reactionary approach is not a good solution because some users must first be victimized to protect others, solutions have little generality, and the warnings to users may provide the phishers with methods to scam users. “Most of the solutions I see address a solution today, but not tomorrow, and you end up wasting your money on proprietary solutions that may not work in the long run” (James, 2005).

Discovering and understanding which attack methodologies are successful and why is helpful in developing tools to guard users and systems against phishing attacks. But ethical issues arise with phishing experiments including use of false information to fool users, disclosure of sensitive information, and handling of sensitive information once collected. Many researchers propose performing actual phishing attacks as an experiment; however, the above issues have prevented any Human Subjects Committee from authorizing experiments. Ethical constraints may also alter the participants’ perception of the experiment, which can give unrealistic and inaccurate results. But some researchers have achieved an acceptable balance between the requirement to be ethical and the need to be accurate (Jakobsson, 2006).

C. PHISHING RESEARCH EXPERIMENTS

1. Web Site Credibility Experiment

(Dhamija, 2006) focused on what makes phishing sites credible. The goal of this research was to understand why certain phishing attacks were successful and to determine what proportion of users is deceived. They analyzed 200 known phishing websites and found that phishing strategies fell into three classes of exploitation: Lack of knowledge by the victim (both system and security knowledge), visual deception of the victim, and lack of attention by the victim to security indicators. The research included a usability study in which 22 participants were asked to view 20 websites and identify fraudulent ones. The test websites included financial institutions and e-commerce companies.

Participants were also asked to describe the factors they used to make their determination:

- Five out of twenty-two participants relied only on the content of the webpage to make their decision.
- Eight out of twenty-two participants used the webpage content as well as the domain and uniform resource locator.
- Two participants used the presence of “https” in the address.
- Five out of twenty-two participants used the previous mentioned factors and the presence of a padlock icon.
- Two participants used all previously mentioned factors as well as security certificates.

Results of the fraudulent-site survey are summarized in Table 1.

Website	Real?	Phishing or Security Tactic Employed	% Correct	%Incorrect
Bank of the West	Spoof	bankofthevvest.com, padlock, Verisign logo and certificate validation seal	9	91
PayPal	Spoof	XUL simulates browser chrome w/fake address, status, and SSL indicators	18	81
Etrade	Real	3rd party URL, SSL, simple design, no graphic for mobile users	23	77
PayPal	Spoof	paypal-signin03.com, padlock in content	41	59
PayPal	Spoof	IP address in dotted-decimal notation, padlock in content	41	59
Capital One	Real	3rd Party URL, SSL, dedicated login page, simple design	50	50
PayPal	Spoof	Screenshot of legitimate SSL protected Paypal page	50	50
Ameritrade	Spoof	ameritrading.net	50	50
Bank of America	Spoof	Rogue popup window on top of legitimate BOFA homepage, padlock	64	36
Bank of the West	Spoof	IP address in dotted-decimal notation, urgent anti-fraud warnings	68	32
USBank	Spoof	IP address in dotted-decimal notation, security warning, identity verification	68	32
Ebay	Spoof	IP address in dotted-decimal notation, account verification	68	32
Yahoo	Spoof	center.yahoo-security.net, account verification	77	23
NCUA	Spoof	IP address in dotted-decimal notation, account verification, padlock	82	18
Ebay	Real	SSL protected login page, TRUSTe logo	86	14
Bank of America	Real	Login page on non-SSL homepage, padlock in content	86	14
Tele-Bears	Real	SSL protected login page	91	9
PayPal	Real	Login page on non-SSL homepage, padlock in content	91	9
Bank One	Real	Login page on non-SSL homepage, padlock in content	100	0

Table 1. (From Dhamija Survey 2006): Strategy employed by each site

General observations about the fraudulent-site survey were:

- Good phishing websites fooled 90% of participants.
- Phishers using almost identical domain names to those of legitimate sites have a very high success rate, as in a site with two v's instead of a w (bankofthevest.com versus bankofthewest.com).
- Current anti-phishing browser alerts are ineffective: 23% of study participants failed to look at the address bar, status bar, or the security indicators.
- On average, the participants misjudged the websites 40% of the time.
- Warnings about fraudulent security certificates were ineffective: 70% of participants ignored the popup warning.
- Participants were equally vulnerable to victimization regardless of education, age, sex, previous experience, or hours of computer use.

The researchers are currently testing an anti-phishing solution that involves a remote server identifying itself in a way that is easy for a user to verify, but difficult for a phisher to spoof (Dhamija, 2006).

2. Social Networking Experiment

(Jagatic, 2005) investigated how easily and how effectively can an attacker exploit social network data to increase the return of a phishing attack. The research showed it could be done very easily and very effectively, since it found that users are four times more likely to be victimized by a phishing attack that appears to originate from an acquaintance. The initial phase of the study harvested publicly-available acquaintance data from online social networks such as myspace.com, friendster.com, facebook.com, orkut.com, and linkedin.com. Then researchers sent spoofed emails to the subjects appearing to be from a close friend. The emails directed the subjects to a phishing website that required their secure university credentials, then used university network services to authenticate the passwords of subjects without storing the password, thereby avoiding ethical issues with collecting sensitive information. Results are summarized in Table 2.

	Successful	Targeted	Percentage	95% C.I.
Control	15	94	16%	9-23%
Social	349	487	72%	68-76%

Table 2. (From Jagatic 2005) Results of social network phishing attack and control experiment

More than 80% of the study subjects followed the URL link believed to be sent to them by a friend, while 72% of the subjects entered their secure university credentials at the fake website. Female subjects were more likely to be victimized, and, the success rate was higher if the spoofed message was from someone of the opposite gender. However, the effect was more evident with males.

	To Male	To Female	To Any
From Male	53%	78%	68%
From Female	68%	76%	73%
From Any	65%	77%	72%

Table 3. (From Jagatic 2005) Gender effects on experiment

3. Security Toolbar Experiment

(Wu, 2006) examined the usability of security toolbars in guarding against phishing attacks. Security toolbars are integrated into web browsers to provide security alerts to users visiting fraudulent websites. The researchers examined whether security toolbars prevent users from being duped into disclosing personal information. The researchers conducted two user studies of three security toolbars. The toolbars included in the study were SpoofStick, Netcraft Toolbar, Trustbar, eBay's Account Guard, the System-Decision toolbar, and SpoofGuard. Thirty participants were presented with 20 emails and asked to read and process the emails as a personal assistant, while protecting sensitive information. Five of the twenty emails were attacks, including a similar-domain-name attack, dotted-decimal IP attack, hijacked-server attack, popup-window attack, and a PayPal "account misused" attack. Even when alerted of a possible phishing attack, participants disregarded the toolbar and divulged sensitive information 34% of the

time. The participants, like typical users, were focused on accomplishing their tasks and security became a secondary concern. Many failed to look at the toolbars and others explained away the security warnings if they felt the website appeared to be legitimate. Upon completion of the study, the researchers presented the following conclusions.

- Warnings are not effective at stopping users trying to accomplish a goal; users will take risks and ignore warnings.
- Security should be integrated into a critical path, making it impossible for the user to ignore.
- A safe mode should enable users to choose to finish their tasks
- Internet companies need to better distinguish their websites from malicious sites. Companies should use a single domain name representative of their company, use SSL (Secure Socket Layer) to encrypt every web page on their website, and provide certificates from well-known and trusted Certificate Authorities.

The study demonstrates many Internet users lack understanding of the sophistication of phishing attacks (Wu, 2006).

4. Spear-Phishing Experiment

In June of 2004, a West Point system administrator performed a spear-phishing study on the cadets at West Point (Ragucci, 2006). The system administrator sent a spoofed email to 512 of the 4,200 cadets instructing them to confirm their grade reports online. The spoofed email claimed to be from a fictitious Colonel and requested prompt action. Over 80% of the cadets were fooled by the email and disclosed their secure login credentials. The victim rate is rather high most likely due to the obedience-to-orders mentality cadets have; however, the study does indicate that spear-phishing email appearing to originate from inside an organization have a high likelihood of success.

5. Mailfrontier Email Experiment

In March of 2005, Mailfrontier released a study reporting participants correctly identified fraudulent emails 83% of the time, while correctly identifying legitimate emails 52% of the time (Mailfrontier, 2005). Mailfrontier concluded people assume an email is

fraudulent when in doubt. Another reason explaining the results is that participants have an increased awareness when specifically looking for phishing emails.

III. A PHISHING MODEL

A. WHY MODEL PHISHING?

An effective solution to phishing requires a multi-phase defense strategy. Such a strategy provides greater assurance phishers will not be able to easily alter attack methodologies to circumvent counterphishing tools. Constructing an accurate model of an attack provides valuable insight to a phisher's methodology, and suggests a blueprint for devising a defense-in-depth strategy.

B. PHASES OF ATTACK

We propose a model for phishing via a deceptive email which provides a link to a fraudulent Web site. The model contains the following five phases: Target Determined, Attack Developed, Attack Performed, Data Accumulated, and Fraud. Figure 1 summarizes the life-cycle of a phishing attack.

Phase 1	Phase 2	Phase 3	Phase 4	Phase 5
<u><i>Target Determined</i></u>	<u><i>Attack Developed</i></u>	<u><i>Attack Performed</i></u>	<u><i>Data Accumulated</i></u>	<u><i>Fraud</i></u>
<ul style="list-style-type: none"> • Footprint target 	<ul style="list-style-type: none"> • Mirror Web site 	<ul style="list-style-type: none"> • Bulk mail scam 	<ul style="list-style-type: none"> • Via email 	<ul style="list-style-type: none"> • Sell information
<ul style="list-style-type: none"> • Get account 	<ul style="list-style-type: none"> • Acquire domain 	<ul style="list-style-type: none"> • Provide link to phishing site 	<ul style="list-style-type: none"> • Via web form 	<ul style="list-style-type: none"> • Use information in attack
<ul style="list-style-type: none"> • Generate plausible scam 	<ul style="list-style-type: none"> • Harvest addresses / set up bulk mailing • Test filtering technique • Set up collection database 	<ul style="list-style-type: none"> • Install malware or spyware 	<ul style="list-style-type: none"> • Malware or spyware 	<ul style="list-style-type: none"> • Use information to gain access / steal

Figure 1. Model of a phishing attack.

1. Target Determined

A typical phishing scenario starts with the phisher registering online for an account with the target organization or impersonating a user to gain access. The phisher needs to see the actual log in process and the internal webpage that appears. The goal is to understand their online procedures and to receive legitimate emails from the organization. As an authorized user, the attacker can see exploitable weaknesses in online and email procedures.

2. Attack Developed

The next step is to construct the attack by acquiring a domain name, harvesting addresses, and configuring a bulk mailing tool. Performing email tests to discover spam filtering techniques allows the phishers to circumvent filters and reach more recipients. Mimicking the target Web site well is a vital step in this phase of the attack. The attacker can host the Web site on a valid or a compromised server that is not suspicious. Graphics can be copied from a legitimate Web site, including fake security certificates. Generating a fraudulent Web site that appears to be legitimate is a fairly easy task for an attacker. Mirror a Web site can be done using two commands, *wget* and *get*. For those images the attacker chooses not to mirror, he can link back to the original site and use the images from the legitimate Web site. The Web site source code can be used to create a phishing site and is readily accessible to the attacker by clicking on the 'View' drop down menu of a browser and selecting 'source.' It is also easy for a phisher to just copy the graphics of a Web site and use them to make a phishing site appear as the target site. To copy a graphic, the phisher right clicks on the image and selects 'Save Picture As.' The phisher then owns that image. Once the phisher creates the phishing Web site, a database is established to collect the desired information from the fraudulent Web site.

3. Attack Performed

To implement the attack, spam email is sent to millions of potential victims, providing a link to a fraudulent Web site. The email uses enticements or threats to encourage the user to visit a Web site. The phishing email must look authentic, just as the phishing Web site must. Phishers typically use target company logos or seals and hyperlinks disguising the actual hyperlink address in the phishing emails. Language of the email is as important as the appearance of the email. Current language used by phishers to lure victims falls into two general categories. The first category, fear, involves language such as:

- WARNING: Security Alert
- IMPORTANT – Account Verification
- Unauthorized Account Access
- Update Your Account Now
- Protect Yourself and Your Account Now

- Account Compromised
- Secure Your Account Now
- FINAL NOTICE: Account Will Be Deactivated

The second category, hope/excitement, involves language such as:

- Reward Notification
- Congratulations You Won
- One Time Deal
- Reply Now And Win

The phishing emails begin with the alarming or exciting information then provides 'easy' instructions to correct the situation or benefit from the exciting deal. The victim is always led to believe a prompt response is imperative. If the email successfully lures a victim, then the fraudulent Web site collects victim data and credentials.

4. Data Accumulated

Victims may divulge sensitive information in several ways. The spam email may suggest logging onto to the Web site from the email or suggest responding to the email with the requested information. The scam may require the victim to visit the fraudulent Web site and submit information in a web form. Or malware may be installed on the victim's computer to silently collect information for the attacker.

5. Fraud

Data that is collected may be sold to hackers or thieves; the phisher may use the information as part of another attack such as breaking into a network or hijacking a server; or the phisher may directly steal and launder the victim's money.

C. SAMPLE PHISHING ATTACK

The following is an example of a real phishing attack, including the fraudulent email and its associated Web site. The attack is targeting PayPal.com users.

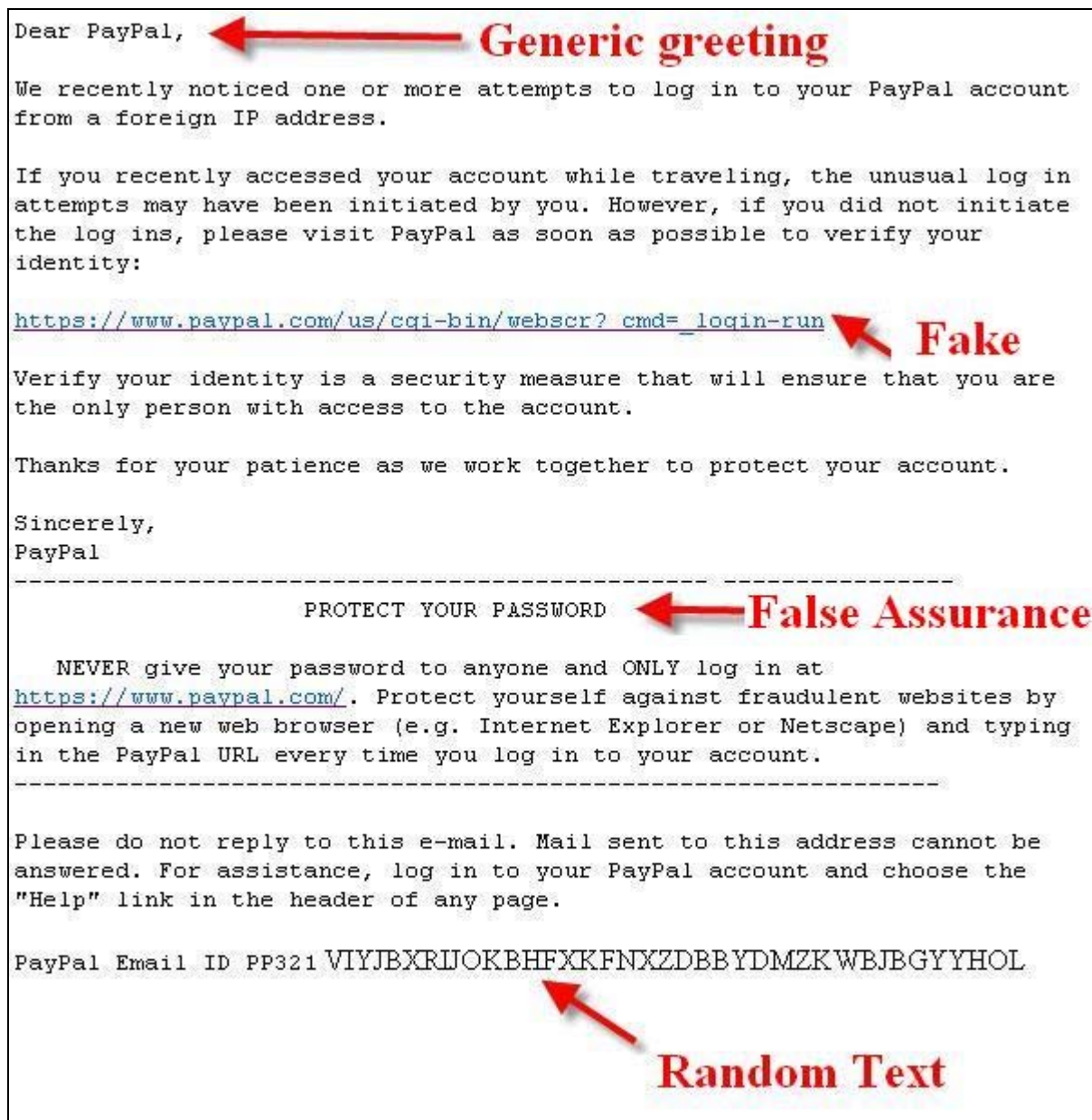


Figure 2. (From APWG 2006) Fraudulent email to PayPal.com users

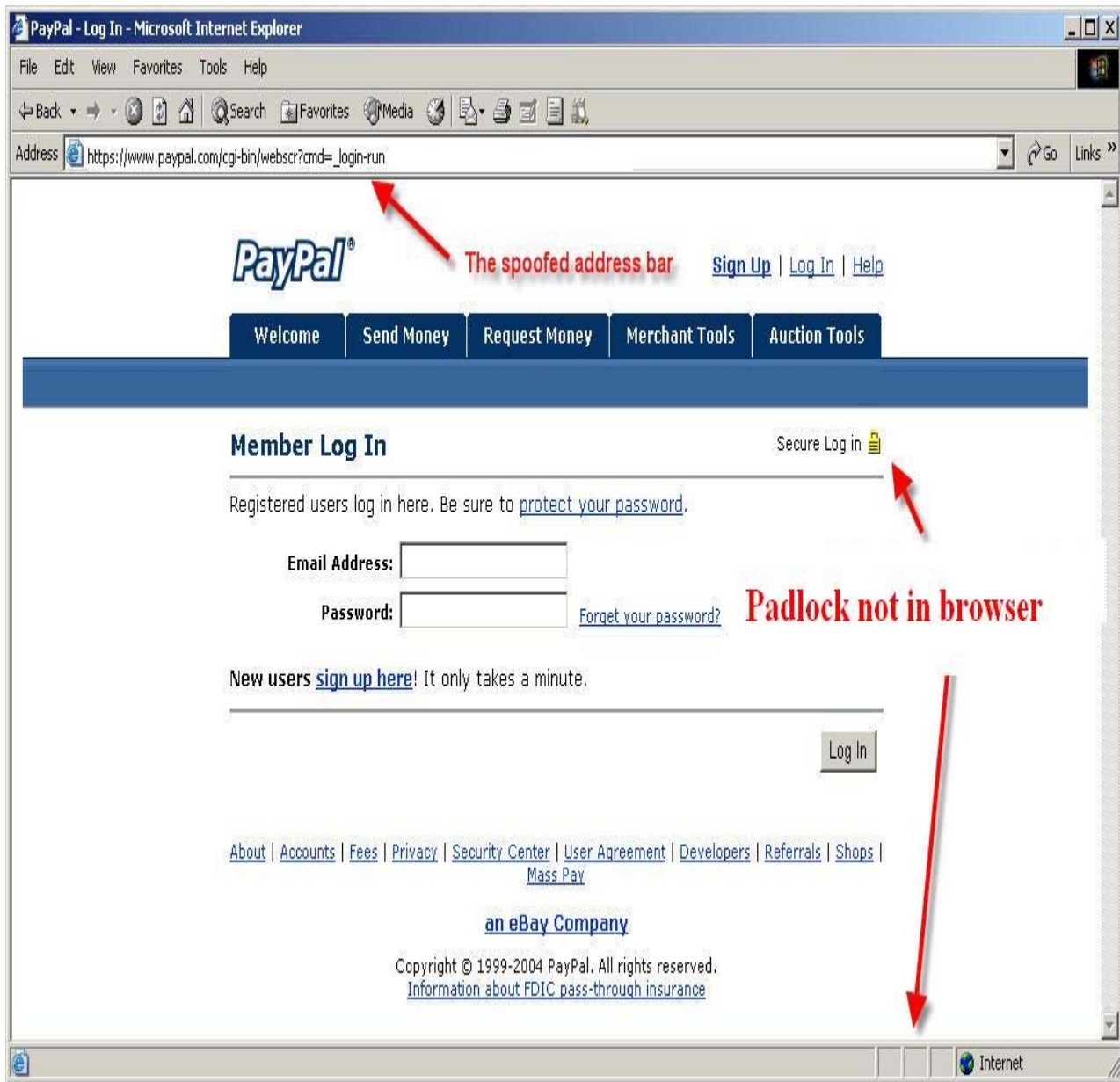


Figure 3. (From APWG 2006) Fraudulent PayPal Web site

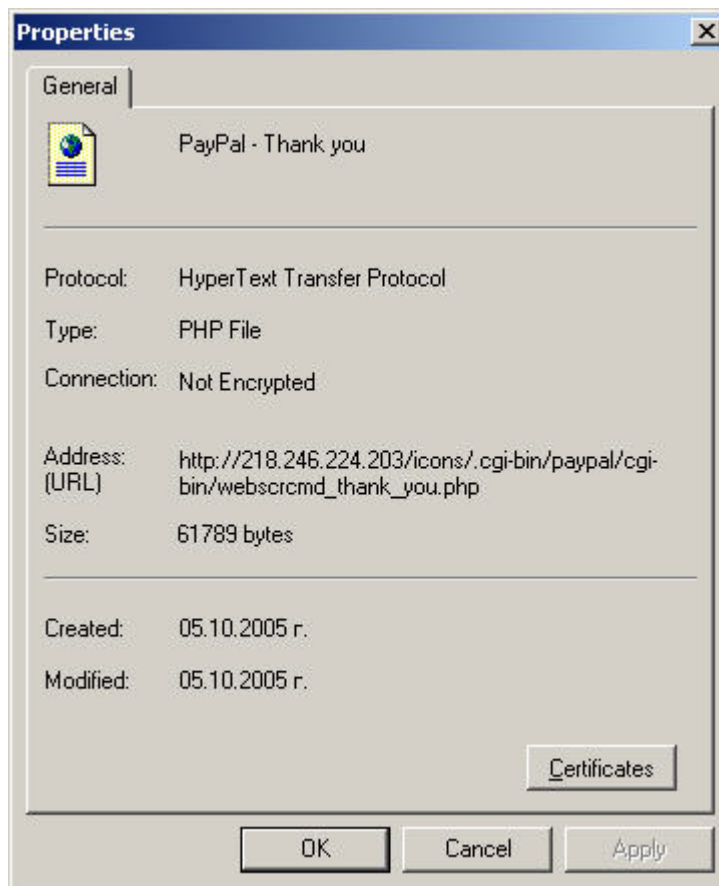



Figure 4. (From APWG 2006) Web site properties


[Log Out](#) | [Help](#)






My Account
Send Money
Request Money
Merchant Tools
Auction Tools

Overview
Add Funds
Withdraw
History
Profile

Update Your Credit Card or Debit Card

Debit Cards (also called check cards, ATM cards or banking cards) are accepted if they have a Visa or MasterCard logo.

First Name:
Last Name:
Card Type:
Card Number:
Expiration Date:
Card Verification Number: (On the back of your card, find the last 3 digits)
ATM PIN Number:

[Help finding your Card Verification Number](#) | [Using AmEx?](#)

Billing Address

Enter the address where you receive monthly billing statements for this card:

Enter your address as billing address

Address 1:
Address 2:
(optional)
City:
State:
Zip Code: (5 or 9 digits)
Country:
Home Telephone:
Work Telephone:
(optional)


For your protection, we verify credit card and debit card billing addresses.
The process normally takes about 30 seconds, but it may take longer during certain times of the day.
Please click **Submit** to update your information. When your card has been successfully update, you will go to the next verification page.

Submit

Cancel

PayPal VISA® CARD

Get Clear Choices!



Apply Now!
30-second Response

Your Choice of Card Designs













Figure 5. (From APWG 2006) Page one of three sequential pages asking for personal information



[Log Out](#) | [Help](#)

My Account

Send Money

Request Money

Merchant Tools

Auction Tools

Overview

Add Funds

Withdraw

History

Profile

Personal Identification Information

Please complete your information below. It's a secure process and your personal information is safe. Transfer of your information is protected by secure 128-bit encrypted SSL.

Social Security Number:

- -

Mother Maiden Name:

Date Of Birthday:

Month

Day

19

Driver's License Number:

Select One

e.g. A189764530

You understand that by clicking on the **Submit** button below, you are providing "written instructions" to PayPal under the Fair Credit Reporting Act authorizing PayPal and its service partners to obtain information from your personal credit profile from a credit bureau on PayPal's behalf. You authorize PayPal and its service partners to obtain such information solely to confirm your identity to avoid fraudulent transactions in your name. If you wish to "opt out" of sharing your personal information with PayPal and its service partners, DO NOT click on the Submit button.

Submit

Cancel

[Mobile](#) | [Mass Pay](#) | [Money Market](#) | [ATM/Debit Card](#) | [BillPay](#) | [Referrals](#) | [About Us](#) | [Accounts](#) | [Fees](#) | [Privacy](#) | [Security Center](#) | [User Agreement](#) | [Developers](#) | [Shops](#) | [Gift Certificates/Points](#)

[an eBay Company](#)

Copyright © 1999-2004 PayPal. All rights reserved.
[Information about FDIC pass-through insurance](#)

Figure 6. (From APWG 2006) Page two of three sequential pages asking for personal information


[Log Out](#) | [Help](#)

[My Account](#)
[Send Money](#)
[Request Money](#)
[Merchant Tools](#)
[Auction Tools](#)

[Overview](#)
[Add Funds](#)
[Withdraw](#)
[History](#)
[Profile](#)

Update Bank Account (U.S. Bank Accounts Only)

The safety and security of your bank account information is protected by PayPal. We protect against unauthorized withdrawals and will notify you by email whenever you deposit or withdraw funds from this bank account.

Bank Name:

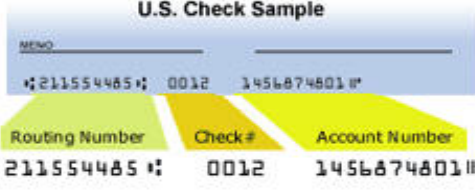
Account Type: ☒ Checking
☐ Savings

Routing Number: (Is usually located between the symbols on your check.)

Account Number: (Typically comes before the symbol. Its exact location and number of digits varies from bank to bank.)

Retype Account Number:

U.S. Check Sample



Routing Number: 211554485
Check #: 0012
Account Number: 1456874801

[Mobile](#) | [Mass Pay](#) | [Money Market](#) | [ATM/Debit Card](#) | [BillPay](#) | [Referrals](#) | [About Us](#) | [Accounts](#) | [Fees](#) | [Privacy](#) | [Security Center](#) | [User Agreement](#) | [Developers](#) | [Shops](#) | [Gift Certificates/Points](#)

an eBay Company

Copyright © 1999-2004 PayPal. All rights reserved.
[Information about FDIC pass-through insurance](#)

Figure 7. (From APWG 2006) Page three of three sequential pages asking for personal information

The language of the email creates a sense of fear in the recipient. The subject of the email is “Unauthorized Account Access.” The ‘From’ line is spoofed and reads, PayPal <service@paypal.com>. Figure 2 shows a generic greeting, disguised hyperlink, and random text at the bottom of the email, all indicative signs of a phishing email. Additionally, the email provides false assurance by advising the recipient on password

security and how to avoid fraudulent Web sites. Figure 3 shows the fraudulent website and demonstrates the use of a spoofed address bar. The absence of the padlock icon in the browser chrome indicates the “https” in the address bar is false and the Web site is fraudulent. Figure 4 shows Properties menu of the Web site and indicates the connection is not encrypted and the actual address of the site is http://218.246.224.203/icons/.cgi-bin/paypal/cgi-bin/websrcmd_login.php. Figures 5-7 show three sequential pages requesting a significant amount of sensitive information, a common characteristic of phishing scams.

THIS PAGE INTENTIONALLY LEFT BLANK

IV. PHISHING RESEARCH METHODS AND EXPERIMENTS

A. OVERVIEW

Our background research highlights the need for a defense-in-depth strategy to effectively counter phishing attacks. We performed a survey of user detection of phishing and a study to associate conditional probabilities to phishing emails.

B. SURVEY

1. Design

In developing and designing our first experiment, we collected phishing emails along with the associated Web sites. They were collected from personal accounts and a honeypot email account set up to collect spam and phishing email. We compiled a list of common characteristics and tactics used by phishers in both the emails and Web sites. We used the list to ensure a proper representation of phishing methodologies in our experiment.

We then conducted a survey containing ten emails and the associated Web sites that the emails aim to lure recipients to visit. The survey was performed on two groups. The goal of our survey is to reveal the factors of a phishing scam that participants focus on to judge authenticity. We are also interested in those factors that deceive the participants and produce a false negative or false positive classification. The survey asks the participants to examine both an email and its associated Web site to determine if the email/Web site pair appears to be fraudulent. The survey also asks the participants to indicate which factors they used in making their determination. Five email/Web site pairs were legitimate and five were fraudulent. The pairs were presented to the participants in random order. The legitimate pairs included Amazon, USAA, myPay, Bank of America, and PNC Bank. The fraudulent pairs represented Chase Bank, PayPal, and myPay. Both the legitimate and fraudulent email/Web site pairs in our survey employed methodologies we observed in our assortment of phishing attacks:

- Use of fear: Unauthorized Account Access, FINAL NOTICE Account Will Be Deactivated, WARNING – Security Alert, Secure Your Account Now
- Use of hope/excitement: Congratulations You Won, Reward Notification, Reply Now and Win
- Use of disguised hyperlinks like Click Here instead of a web address
- Use of a similar domain name: bankofthevest.com instead of
- Use of a visible dotted decimal IP address
- Use of third-party hosting of Web sites
- Use of a spoofed sender
- Use of victim context such as the phisher's use recent purchase history or eBay bidding history of recipient
- Use of security features such as padlock icon, https, VeriSign logo, certificates
- Use of fake security certificates
- Use of logos and graphics
- Use of generic greetings such as Dear Customer or Valued PayPal Member
- Use of general security and password security advice
- Use of information concerning fraudulent sites and how to avoid them

Our background research demonstrated legitimate companies oftentimes send emails to their customers that could be mistaken as fraudulent. Additionally, companies can employ third-party Web hosting services which can falsely alarm customers, or worse, give credibility to fraudulent sites. Therefore, we included samples in our survey as seen in Figures 8 and 9.

From: **Individualized BankCard Services** <<mailto:IBS@email.cardsatisfaction.net>>
Date: Apr 28, 2006 4:46 PM
Subject: Use your PNC Bank credit card today.
To: jackmcdowell@comcast.net



RE: Your account number
ending in 3272

**Consolidate your balances
into one payment .**

Dear Jack D. McDowell,

Enjoy the power of extra cash this spring with a balance transfer. With a click, you can transfer higher-rate balances to your PNC Bank Visa® credit card account. With just one monthly payment, it's the perfect opportunity to:

- Get rid of those department store balances.
- Make improvements to your home for spring.
- Plan a summer get-away.
- Join a gym.

Why not? Just click to open up a world of new possibilities for yourself. Go to <http://links.cardsatisfaction.net/ajtk/servlet/JJ?H=25bwbv&R=1571356816&P=www.pncnetaccess.com> to transfer balances, or visit your local bank to get a cash advance.

Your credit line is
\$27,500!

[Click Here To Start Your Balance Transfer](#)

Consolidate balances.
Make just one monthly
payment.*

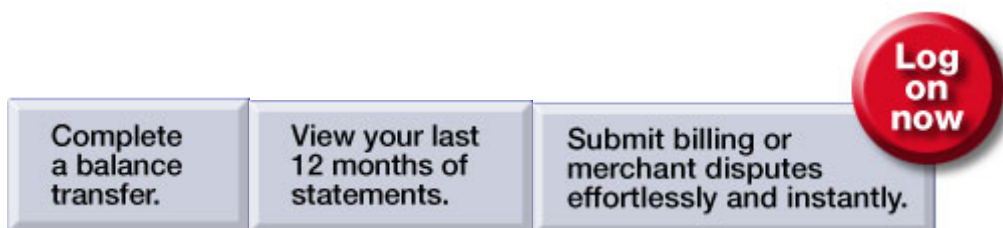


Figure 8. Legitimate PNC Bank email

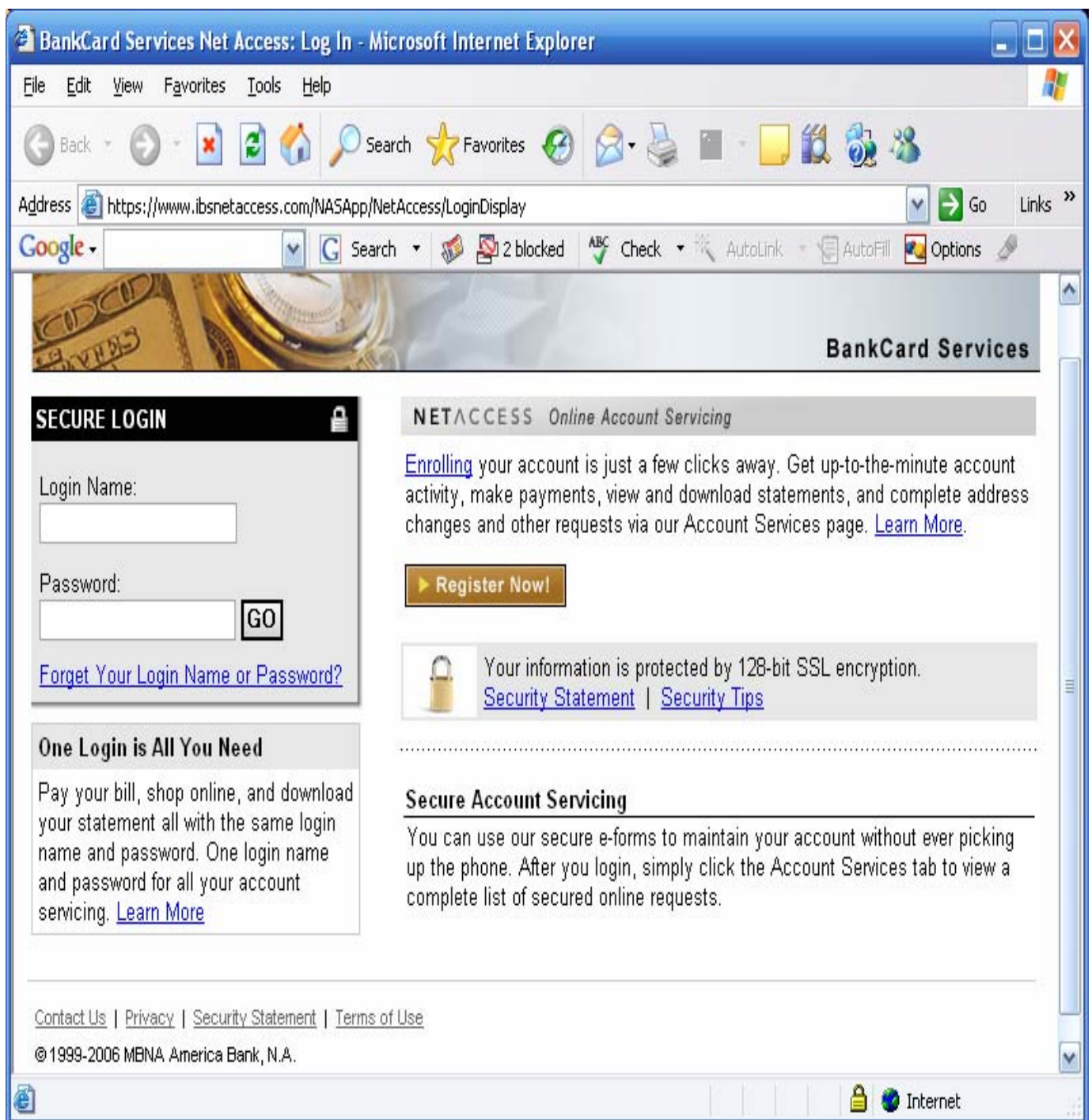


Figure 9. Web site associated with PNC Bank email

The email and Web site from PNC Bank appeared suspicious and initially we had trouble determining authenticity. We were interested in the survey participants' reactions to the factors we found equivocal. First, the email is from Individualized BankCard

Services, IBS@email.cardsatisfaction.net, not from a PNC Bank domain address. Second, the email uses context awareness by including the last four numbers of the account, which could easily be obtained by a phisher. Third, the email contains disguised hyperlinks behind the text of “Click here” and “Log on Now”. Fourth, the hyperlink that is displayed bears no resemblance to a PNC Bank domain name. Fifth, the language of the email is alluring, as it promises quick and easy debt consolidation. Lastly, the Web site address is of a third party. It does not match the hyperlink displayed in the email or a PNC Bank domain.

Our survey began with a general questionnaire designed to assess the awareness level of the participants with respect to security and phishing attacks. It included questions such as:

- Do you view Web site security certificates when prompted?
- Have you ever used any tool to determine who owns an IP address?
- Do you open emails if you do not recognize the sender?
- Do you view the full header information of an email?
- Do you use clickable link in emails?

We collected our data using an answer form following each email/Web site pair (Figure 10).

1. Please indicate whether the email/website pair appears to have fraudulent intent or not.

Fraudulent	Not Fraudulent
-------------------	-----------------------

2. If fraudulent, then on a scale of 1-5, which factors lead you to believe the pair to be fraudulent? (Scale: 1-Not at all, 5-Key identifying factor)

Text of email:

1	2	3	4	5
----------	----------	----------	----------	----------

Web site address (e.g. www.nps.edu):

1	2	3	4	5
----------	----------	----------	----------	----------

What the Web site looks like:

1	2	3	4	5
----------	----------	----------	----------	----------

Where the email came from:

1	2	3	4	5
----------	----------	----------	----------	----------

Clickable link in email:

1	2	3	4	5
----------	----------	----------	----------	----------

Believability of claims in email:

1	2	3	4	5
----------	----------	----------	----------	----------

Please circle any other factors here you use to determine an email/Web site is fraudulent.

-Spelling/Grammar	-Webpage Source code	-Other:_____
-Consistency	-Security features	
-Unusual Pronunciation	-Specificity	

Figure 10. Survey answer form

Besides judging the legitimacy of each email/Web site pair, if the participants judged a pair to be fraudulent, they were asked to rate the influence of six specific factors on their decision making process. The six factors were the text of the email, the Web site address, what the Web site looks like, where the email came from, clickable link in the email, and believability of the claims in the email. Participants were asked to rate the factors on a scale of one to five, one indicating the factor had no influence and five indicating the factor was the key influence in making their decision. Participants were also asked to identify any factors they use other than the six targeted factors.

2. First Run

Ten subjects participated in the first run of our survey. The group was comprised of nine adults and one child. The adult ages ranged from 24 to 62 and the child was 16 years old. Aside from the child, all participants had at least a Bachelor's level education. Most were Master's level students and one was a doctoral student. Only one participant had extensive computer knowledge, a person was working on a Master's degree in Computer Science. All participants were volunteers, however, they were individually sought out to represent an inclusive field of users.

3. Second Run

Seventeen subjects participated in the second run of our survey, only adults. All participants were students in a Master's level Computer Security course. We hoped to determine if education or computer experience would significantly bias the survey results. The survey is given in Appendix A.

C. CONDITIONAL PROBABILITY STUDY

Our second experiment explored conditional probabilities associated with word clues in phishing emails and non-phishing emails. We were interested in the practicality and efficiency in identifying phishing attacks based on words used in the email. The experiment generated conditional probabilities for both unigram (one-word) and bigram (two-word) clues to phishing. Words such as "Viagra" or "Meds" are indicative of spam because these words are rarely seen in legitimate email. Phishing is a form of spam, so we designed our second experiment to determine if there are strong word clues associated with phishing email that are not associated with legitimate email.

In conducting our second experiment, we collected 400 examples of phishing email and 2656 examples of legitimate email, from which we randomly chose a training set of 350 phishing and 1817 non-phishing, and a test set of 50 phishing and 839 non-phishing. The phishing email was collected from personal email accounts of the authors, the Anti-Phishing Working Group, and a honeypot email account. The non-phishing email was obtained through a spam project as part of Professor Neil Rowe’s class. The program used to read and calculate the conditional probabilities was designed by Neil Rowe. From the training sets we calculated conditional probabilities of words appearing in phishing email. Capitalization, numbers, and 602 common words such as “we”, “do”, and “as” were ignored. Also, 2864 personal names were ignored excluding those with common English meanings like “Ball.” Message-header information was included in the analysis so words of the subject or originating sender could also be clues.

Our output included word-probability pairs for all words whose counts was greater than a minimum M and whose probability deviated from the expected value in a binomial model more than a certain multiple S of the standard deviation of $\sqrt{ne/(n+e)}$, where e is the number of occurrences of the word in phishing and n is the number of occurrences of the word in non-phishing in the training set. We also explored double word clues. For example, “verify” is a single word clue, where “verify account” is a double word clue. Our goal was to determine if single word or double word clues are adequately indicative of phishing.

V. RESULTS

A. FIRST SURVEY

The results of our first experiment with surveys are consistent with our hypotheses. Results are summarized in Tables 4 and 5. The top of each column lists the email/Web site pair and indicates whether the pair is fraudulent (F) or not fraudulent (NF). Each row represents a different participant in the survey. A (1) indicates the participant correctly identified the email/Web site pair and a (0) indicates the participant incorrectly identified the pair.

Participant	Amazon NF	USAA NF	Chase F	myPay NF	PayPal F	PayPal F	BofA NF	PNC NF	PayPal F	myPay F	Correct
#1	0	1	1	0	1	1	1	0	1	0	60%
#2	1	1	1	1	1	1	1	1	1	0	90%
#3	1	1	1	1	1	1	1	0	1	1	90%
#4	1	1	1	1	1	1	1	1	1	1	100%
#5	1	1	1	1	1	1	0	1	1	0	80%
#6	1	1	1	1	1	1	0	0	0	1	70%
#7	1	1	0	1	1	1	0	0	1	0	60%
#8	1	1	0	1	1	1	0	0	1	0	60%
#9	1	1	1	1	1	1	1	0	0	0	70%
#10	0	0	1	1	1	1	1	1	1	1	80%
Correct	80%	90%	80%	90%	100%	100%	60%	40%	80%	40%	76%
NF = Not Fraudulent							0 = incorrect ID				
F = Fraudulent							1 = correct ID				

Table 4. Results for fraudulent email/site survey#1

The results may contain bias due to the participants' heightened awareness as they were instructed to identify any fraudulent email/Web site pairs. However, we surmise the reason for results that appear overly cautious is that many legitimate company emails and Web sites look suspicious. The first survey had a 28% false positive rate and a 20% false negative rate. The PNC Bank email/Web site pair had the highest false positive rate at 60%. The Bank of America email/Web site pair had the second highest false positive rate at 40%. The fraudulent DFAS myPay email/Web site pair had the highest false negative rate at 60%. The overall high false positive rate along with the false positive rate for

PNC Bank and Bank of America suggests companies should develop less suspicious emails and Web sites. The high false negative rate for the DFAS myPay email/Web site pair reveals that use of similar domain names to real sites creates convincing scams.

For those email/Web site pairs designated as fraudulent, the participants rated the influence of six specific factors on their decision-making process. The scale was one to five, five being the key influential factor.

<u>Factor</u>	<u>Rating</u>
• Web site address:	4.02
• Text of email:	3.15
• Believability of claims in email:	2.65
• Clickable link in email:	2.22
• Where the email came from:	2.09
• What the Web site looks like:	1.74

Of the six factors, Web-site address was most often used to determine authenticity. Phishers can exploit this factor either by using similar domain names or spoofing the address bar. Email text was the second most used factor to determine authenticity. Phishers continually alter scams to stay ahead of user education and awareness, and many times copy current customer alerts to make the email text appear legitimate. Appearance of the Web site was the least influential factor in determining authenticity. Users are apparently becoming more aware of the capabilities of phishers to replicate authentic Web sites. Aside from the six specific factors, other factors indicated by the participants included generic email greeting, spelling/grammar of email, unusual pronunciation, consistency, specificity, and security features.

The participants answered five general survey questions.

- 0% of subjects view security certificates when prompted
- 10% of subjects have used a tool to determine the owner of an IP address
- 30% of subjects open emails from unknown senders
- 40% of subjects view the full header information of an email
- 90% of subjects use clickable links in emails

A possibly worrisome response was that nine out of ten participants use clickable links in emails. All phishing emails have links of varying types, whether they are disguised, spoofed, or redirected as a man-in-the-middle attack. Three out of ten participants open emails from unknown senders. However, social relationships can be discovered and used to spoof email sender names [Jagatic 2005]. None of the participants view security certificates when prompted, a similar result to the Wu experiment with security toolbars and pop-up warnings [Wu 2006]. These tools have proven ineffective in defending against phishing.

B. SECOND SURVEY

Our second run of the survey produced a slightly different outcome, but overall yielded fairly similar results.

Participant	Amazon NF	USAA NF	Chase F	myPay NF	PayPal F	PayPal F	BofA NF	PNC NF	PayPal F	myPay F	Correct
#1	1	1	1	1	1	0	1	1	1	0	80%
#2	1	0	1	0	1	1	0	0	1	1	60%
#3	1	1	1	1	1	1	1	0	1	1	90%
#4	1	1	1	1	1	1	0	0	1	1	80%
#5	1	0	1	1	1	1	0	0	1	0	60%
#6	1	0	1	1	1	1	0	0	1	0	60%
#7	1	1	1	1	1	1	1	0	1	0	80%
#8	1	1	1	1	1	1	1	1	1	1	100%
#9	1	1	1	1	1	1	1	1	1	0	90%
#10	1	1	1	1	1	1	1	0	1	1	90%
#11	1	1	1	1	1	1	1	0	1	1	90%
#12	1	1	1	1	1	1	1	0	1	0	80%
#13	1	1	1	1	1	1	1	1	1	1	100%
#14	1	0	1	1	1	1	1	1	1	1	90%
#15	1	1	1	1	1	1	1	1	1	1	100%
#16	1	1	1	1	1	1	1	1	1	1	100%
#17	0	1	1	1	1	1	1	0	1	0	80%
Correct	94%	76%	100%	94%	100%	94%	76%	41%	100%	58%	78%

NF = Not Fraudulent
F = Fraudulent

0 = incorrect ID
1 = correct ID

Table 5. Results for fraudulent email/site survey #2

The second survey had a 33% false positive rate and an 11% false negative rate. As with the first survey, the PNC Bank email/Web site pair had the highest false positive rate at 59% and the Bank of America email/Website had the second highest false positive rate at 24%. The highest false negative was the same as in the first survey as well, DFAS myPay email/Web site at 41%. Only the false negative rates differ significantly between the two surveys. The lower false negative rate of survey number two indicates that Master's level students in a Computer Security course are more equipped to recognize a fraudulent site.

The ratings in survey two of the six influential factors are similar.

<u>Factor</u>	<u>Rating</u>
• Web site address:	4.23
• Text of email:	3.95
• Believability of claims in email:	2.82
• Clickable link in email:	2.47
• Where the email came from:	2.18
• What the Web site looks like:	1.90

Other factors denoted by the participants in survey two were identical to the first survey, generic email greeting, spelling/grammar of email, unusual pronunciation, consistency, specificity, and security features.

Responses to the general questions in survey two differed slightly with survey one responses. However, use of clickable links and opening emails from unknown senders remained high.

- 35% of subjects view security certificates when prompted
- 52% of subjects have used a tool to determine the owner of an IP address
- 30% of subjects open emails from unknown senders
- 24% of subjects view the full header information of an email
- 60% of subjects use clickable links in emails

C. COMBINED RESULTS FOR BOTH SURVEYS

The combined results of the first and second survey are presented in Figure 11.

<u>Survey 1 and 2 Combined</u>	
False positives: 31%	
False negatives: 14%	
PNC Bank email/website pair had the highest false positive rate: 59%	
Bank of America email/website had the second highest false positive rate: 30%	
The fraudulent myPay email/website pair had the highest false negative rate: 48%	
<u>Participant Determining Factors:</u>	<u>Rating</u>
Web site address:	4.15
Text of email:	3.65
Believability of claims in email:	2.75
Clickable link in email:	2.38
Where the email came from:	2.15
What the Web site looks like:	1.84
<u>Other factors:</u>	To generic customer Spelling/grammar of email Unusual Pronunciation Consistency Specificity Security features
<u>General survey questions:</u>	
22% of subjects view security certificates when prompted	
37% of subjects have used a tool to determine the owner of an IP address	
30% of subjects open emails from unknown senders	
30% of subjects view the full header information of an email	
70% of subjects use clickable links in emails	

Figure 11. Combined survey results

D. DECEPTION AND CLUES IN PHISHING

The survey results showed participants used the Web site address most often to determine authenticity of both the email and Web site. Phishers are aware of this fact and employ one of three deceptive tactics to trick Internet users. Phishers may include a hyperlink in the fraudulent email that appears to be authentic, www.paypal.com, however, the actual link destination is different from the text of the link and sends the user to www.iamstealingyourmoney.com. Phishers may disguise the link with text such as “Click here” or “Log on” and the actual destination is to a fraudulent Web site. Finally, phishers may combine the previous two techniques in a fraudulent email. In order to quantify the use of deception, we randomly chose 50 phishing emails from our collection and 50 legitimate emails from financial organizations and online retailers. Among the phishing emails, 50% had fake text hyperlinks, 32% had disguised hyperlinks, and 18% combined both deceptive tactics. Only 10% of the legitimate emails had disguised hyperlinks. Security logos such as the Verisign logo or the padlock icon on Web sites are both used by phishers as a deceptive tactic. Over 90% of the fraudulent Web sites had either fake security logos, padlock icons, or both. The email greeting was the other factor most often used by participants to determine authenticity. Among the phishing emails, 98% used generic customer greetings such as “Dear PayPal Customer” or “Dear Customer.” Among the non-phishing emails, 72% used personalized greetings, while 28% used generic greetings or no greeting at all.

E. CONDITIONAL PROBABILITY STUDY

The second experiment produced conditional probabilities of word clues associated with phishing email and non-phishing email. We considered a positive clue as one whose probability was two standard deviations from the value of 0.0128 that represented the fraction of words within phishing email in our training sample. Example positive word clues for phishing were “renew”, “verify”, “bank”, “alert”, “paypal”, “chase”, “notify”, and “immediate.” Example negative word clues for phishing were “course”, “student”, “week”, “data”, “question”, “technology”, “width”, and “style.” Example positive two-word clues for phishing were “renew immediate”, “from paypal”, “nt access”, “verify click”, “chase bank”, “immediate pay”, “of Oklahoma”, and “ship

alert.” Example negative two-word clues for phishing were “computer science”, “image delivery”, “any question”, “next week”, “date submit”, and “to meet.” Our experiment proved there are single and double word clues that are indicative of phishing.

We used the unigram and bigram clues to calculate precision and recall capabilities. Recall is defined as the fraction of the phishing in the test set that was identified. Precision is defined as the fraction of actual phishing in the set of email identified as phishing. We also calculated the f-score, which is the harmonic mean of precision and recall. Recall and precision were definitely not as good as those above 95% reported for spam-detection systems. This is because phishing email attempts to appear legitimate, therefore judging legitimacy on text clues alone yields poor results. For these experiments, we used the *odds* formulation of conditional probability. It is given by the following equation: $O(\text{phish} \mid \text{word clues}) = (O(\text{phish} \mid \text{word 1}) / O(\text{phish})) * (O(\text{phish} \mid \text{word 2}) / O(\text{phish})) * \dots * O(\text{phish}))$. The odds terms are generated by first computing the associated probability, then dividing the probability by one minus the probability: $O(x) = P(x) / (1 - P(x))$. The odds formulation was used to determine an appropriate threshold for classifying an email as phishing. We experimented with odds-thresholds from one tenth to five and calculated precision and recall for each threshold. The f-score was maximized at a threshold of one for both unigram clues and bigram clues. The results for unigrams and bigrams are presented in Tables 6 and 7 respectively. Precision versus recall is presented in Figures 12 and 13.

Threshold	Precision	Recall	f-score
0.1	0.14	0.1988	0.165
0.5	0.1621	0.21	0.183
1	0.2666	0.16	0.198
1.5	0.2333	0.14	0.174
2	0.2333	0.14	0.174
5	0.2001	0.12	0.15

Table 6. Precision and Recall with given odds threshold for single word clues to identify phishing in our test set.

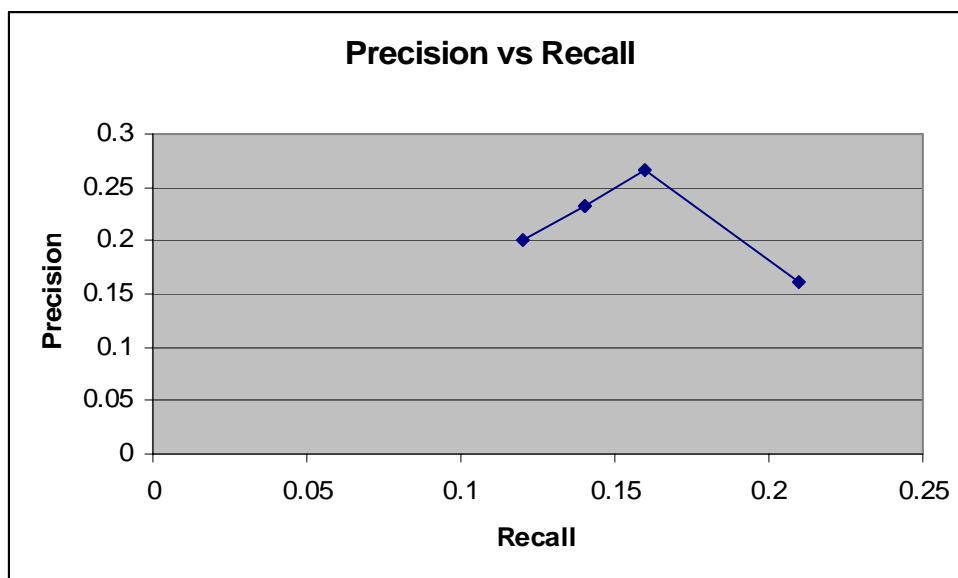


Figure 12. Precision vs. Recall for unigram clues

Threshold	Precision	Recall	f-score
0.1	0.1148	0.3521	0.1732
0.5	0.2061	0.2416	0.2224
1	0.5625	0.18	0.2727
1.5	0.4375	0.14	0.2121
2	0.4375	0.14	0.2121
5	0.375	0.12	0.1818

Table 7. Precision and Recall with given odds threshold for double word clues to identify phishing in our test set.

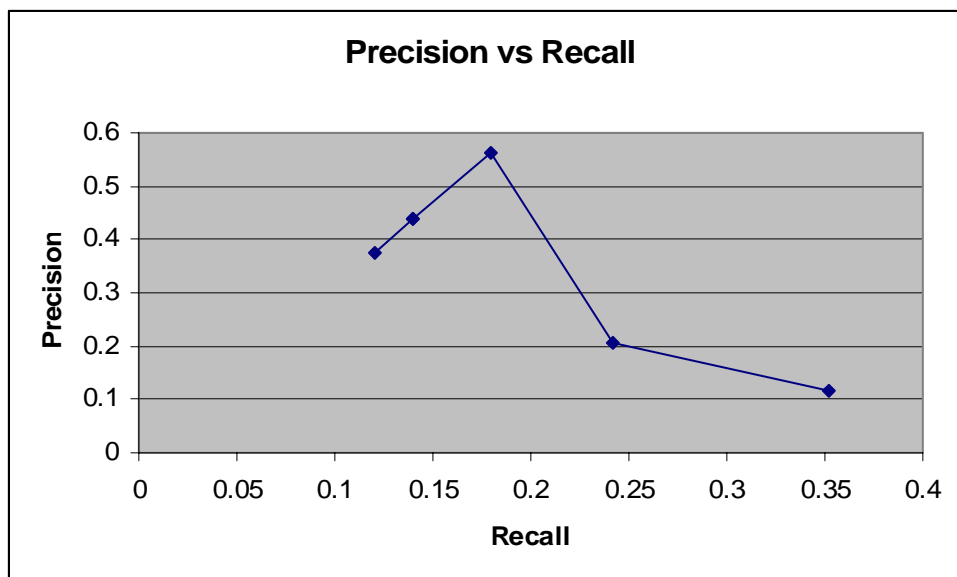


Figure 13. Precision vs. Recall for bigram clues

THIS PAGE INTENTIONALLY LEFT BLANK

VI. COUNTERMEASURES AND FUTURE WORK

A. SUMMARY OF DEFENSE MECHANISMS AND COUNTERMEASURES

Based on our analysis and experiments, we provide here suggestions for countermeasures to prevent phishing attacks. The countermeasures are organized by whom or what should be responsible for implementing the countermeasure. Financial institutions and e-commerce organizations should implement the following countermeasures.

- Personalize all greetings in email correspondence. Phishing emails are typically addressed as “Dear Customer” and not individually. Except for instances of spear phishing, a personalized greeting would indicate that the email is not a bulk phishing email and is most likely legitimate.
- Ensure online correspondence with customers does not look suspicious or provide material for phishers to take advantage of for future scams. Avoid using inflammatory language and urgently requesting sensitive information in an email. Email containing language that uses fear or excitement and which attempts to rouse a sense of urgency should alarm the recipient. Rather, inform the Internet user that they have a message in their account inbox and direct them to log on the company Web site to retrieve the message. All requests for information and submission of information should occur on the company’s secure Web site.
- Never direct customers to submit sensitive information via a Web form in an email. Web forms in email are a common phishing practice. Companies should have users log on to a secure Web site to submit all requested information.
- Maintain valid security certificates from reputable well-known third parties.
- Monitor or pay for services that monitor domain-name registries to prevent phishers from establishing fraudulent Web sites with similar domain names, such as www.bankofthevest.com.

- Avoid providing hyperlinks in email correspondence to customers.
Eliminating hyperlinks in email does inconvenience users and businesses;
however, users will be assured of the legitimacy of the email.

As an example, Wells Fargo Bank uses smart, safe, and authentic looking email correspondence as in Figure 12. The only aspect of the email that could be criticized is the greeting. The greeting should be personalized, but the email provides no reason to suspect fraud and no means for a phisher to capitalize on the method of correspondence. There are no clickable links, no inflammatory language, no HTML, and the email does not request any sensitive information.



Figure 14. Example of proper email correspondence (Wells Fargo Bank)

Financial institutions and e-commerce organizations should be responsible for providing training to their users on their standard online business practices as well as countermeasures for users to implement. Internet users should be responsible for implementation of the following countermeasures.

- Avoid using hyperlinks in emails regardless of the sender.

- Log onto a Web site directly by typing the web address in the browser address bar. Typing an address directly into the browser address bar is a trustworthy method except in rare instances of pharming or DNS cache poisoning which will be discussed later.
- Avoid disclosing any sensitive information when requested in an email and never submit sensitive information via HTML forms in email. Phishing emails typically request personal sensitive information such as usernames, passwords, social security numbers, credit card numbers, and mother's maiden name.
- Check Web site security, ensure that the Secure Hypertext Transfer Protocol (https) is employed in the browser address bar and the padlock icon is shown in the bottom right portion of the browser chrome. Verify security certificates are valid and from a trusted third party.
- Ensure web browser is patched with all security related patches. For Internet Explorer users, Microsoft has a patch for vulnerabilities that certain phishing scams utilize <http://www.microsoft.com/security/>.

Automated countermeasures may provide a viable defense as well.

- Countering fake hyperlinks: Integrate a tool into web browsers that compares the hyperlink text to the actual hyperlink when the user mouses over the link browser, and disable any links that do not match. Our background research as well as our own research shows that a displayed warning is typically ignored or explained away, so automatically disabling the link is most effective.
- Spoofed address bar: Integrate a tool into web browsers that compares the spoofed address to the actual address and have the browser close the Web page and display an explanatory message if the addresses do not match. Usability is always a concern, but where phishing is concerned, automated programs must be forceful and not allow users to easily ignore security warnings.

- Disguised or undisguised fraudulent hyperlinks: Disable all hyperlinks in an email.
- Spoofed email header information: Before email is delivered to an inbox, look at the full header information to analyze the route an email message took to arrive. If the “from” and “by” fields of the received header are inconsistent, that is indicative of a forged email header. Also, look for inconsistency between the sender address domain and the site domain. This is a good measure of legitimacy except for companies using a third party for online correspondence and transactions like the PNC Bank example.
- Use a hash table to compare Web site data to detect similar appearing Web sites, which may be phishing.
- Offensive countermeasure: Enter false information on the Web site forms. If companies are receiving charge attempts or transactions with bad credentials, this would indicate that phishers are trying to defraud users and the company. If large numbers of forms are submitted with obviously false information, this provides a kind of "denial of service" attack on the phishing site as well as making it much easier for victimized credit-card companies and other businesses to detect phishing on their accounts and respond quickly. However, some phishing Web sites have started checking the validity of credit card numbers with a publicly available formula that does not require a connection to a credit-card server.
- Both the unigram and bigram word clues provided useful probabilities for classifying phishing email. The unigram and bigram word clues may be used to filter phishing email or may be used in conjunction with other defensive indications to positively identify a phishing attack. An effective means of using word clues would be to flag an email as suspicious and combine the suspicious classification with other defensive clues to accurately determine legitimacy. The automated tool would have a pre-defined threshold probability for phishing. The tool would look for clues

like generic greetings, fake hyperlinks, and word clues. Each clue would have an associated probability and if the combined probabilities reach the pre-defined threshold, then the tool would alert the user that the email and Web site are fraudulent.

- **Pharming:** The only current defense against pharming is the nslookup tool which looks up data on a given site. Designing an automated tool that performs “nslookup” and warns the user from visiting fraudulent Web sites may be a viable defense.

B. EXPANDING CURRENT RESEARCH

Our survey involving detection of fraudulent email/Web site pairs provided useful information, but more people should be surveyed. The survey could be converted to a Web format to allow a greater number and range of participants and more data. Also, continually updating the examples used in the survey with new forms of attack would provide fresh data and new defense ideas.

Our second experiment involving conditional probabilities of single word and double word clues in phishing email suggested there is a good possibility that phishing emails can be detected with word clues, but it is harder than detecting spam in general. This experiment could be enhanced by using a broader and larger experimental set of email. The unigrams and bigrams can be used to develop an automated filter to prevent attacks.

An important area our research did not focus on is pharming, for which many of the countermeasures discussed in this thesis may be appropriate. Pharming is a new phishing tactic that could allow phishers to operate undetected. Pharming involves a fraudulent Web site used to collect sensitive information from Internet users. It aims to redirect traffic from legitimate Web sites to fraudulent Web sites while the client believes to be in a trusted domain. Pharming is achieved in one of two ways. First, the attacker can exploit a vulnerable DNS server; the server is “poisoned” by changing the IP address of the target Web site to the IP address of the fraudulent Web site. Internet users can enter the correct legitimate Web site address in the address bar, but be directed to the

fraudulent site, while the address bar still displays the legitimate Web site address. Second, the attacker can use malware to corrupt a client's host file which contains its own local name to IP address mapping. There is currently no practical defense mechanism for pharming.

APPENDIX A: SURVEY

The following survey involves ten emails and the associated websites the emails direct the recipient to visit. The survey asks for you to determine if the email/website pair appears to be fraudulent and indicate which factors led you to believe it to be fraudulent. Before you look at the first sample please answer the questions below.

1. Do you view website security certificates when prompted?

Yes No

2. Have you ever used any tool to determine who owns an IP address?

Yes No

3. Do you open emails in which you do not recognize the sender?

Yes No

4. Do you view the full header information of an email?

Yes No

5. Do you use clickable links in emails?

Yes No

1. Please indicate whether the email/website pair appears to have fraudulent intent or not.

Fraudulent

Not Fraudulent

2. If fraudulent, then on a scale of 1-5, which factors lead you to believe the pair to be fraudulent? (Scale: 1-Not at all, 5-Key identifying factor)

Text of email:

1 2 3 4 5

Website address (e.g. www.nps.edu):

1 2 3 4 5

What the website looks like:

1 2 3 4 5

Where the email came from:

1 2 3 4 5

Clickable link in email:

1 2 3 4 5

Believability of claims in email:

1 2 3 4 5

Please circle any other factors here you use to determine an email/Web site is fraudulent.

-Spelling/Grammar

-Webpage Source code

-Other:_____

-Consistency

-Security features

-Unusual Pronunciation

-Specificity

From:	"PayPal" <admin@paypal.com>
Subject:	Notification of Limited Account Access
Date:	Sun, 5 Mar 2006 15:19:01 +0200



We recently reviewed your account, and we need more information about your business to allow us to provide uninterrupted service. Until we can collect this information, your access to sensitive account features will be limited. We would like to restore your access as soon as possible. We apologize for the inconvenience.

Why is my account access limited?

Your account access has been limited for the following reason(s):

We have reason to believe that your account was accessed by a third party. Because protecting the security of your account is our primary concern, we have limited access to sensitive PayPal account features. We understand that this may be an inconvenience but please understand that this temporary limitation is for your protection.

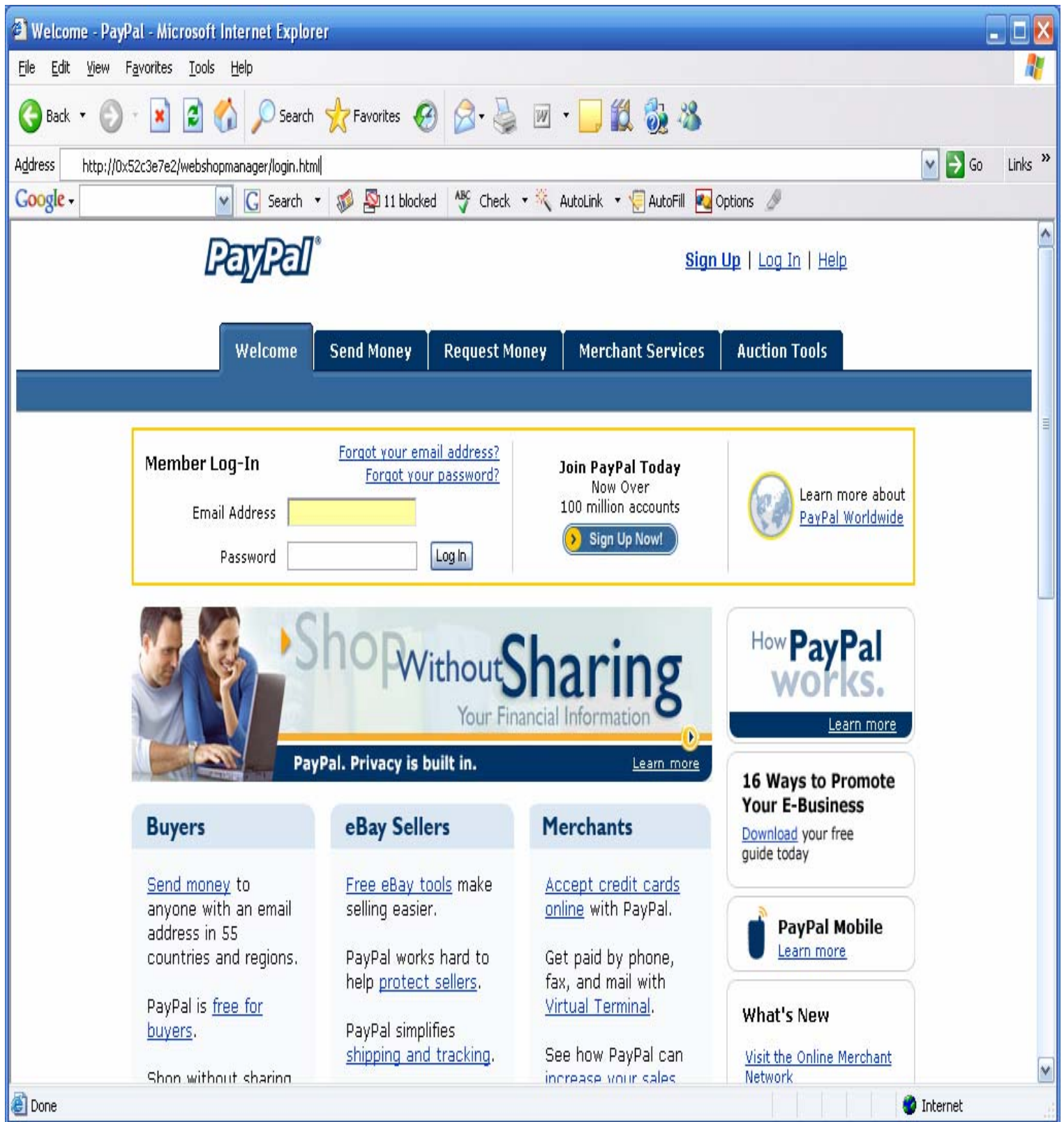
(Your case ID for this reason is PP-136-124-102.)

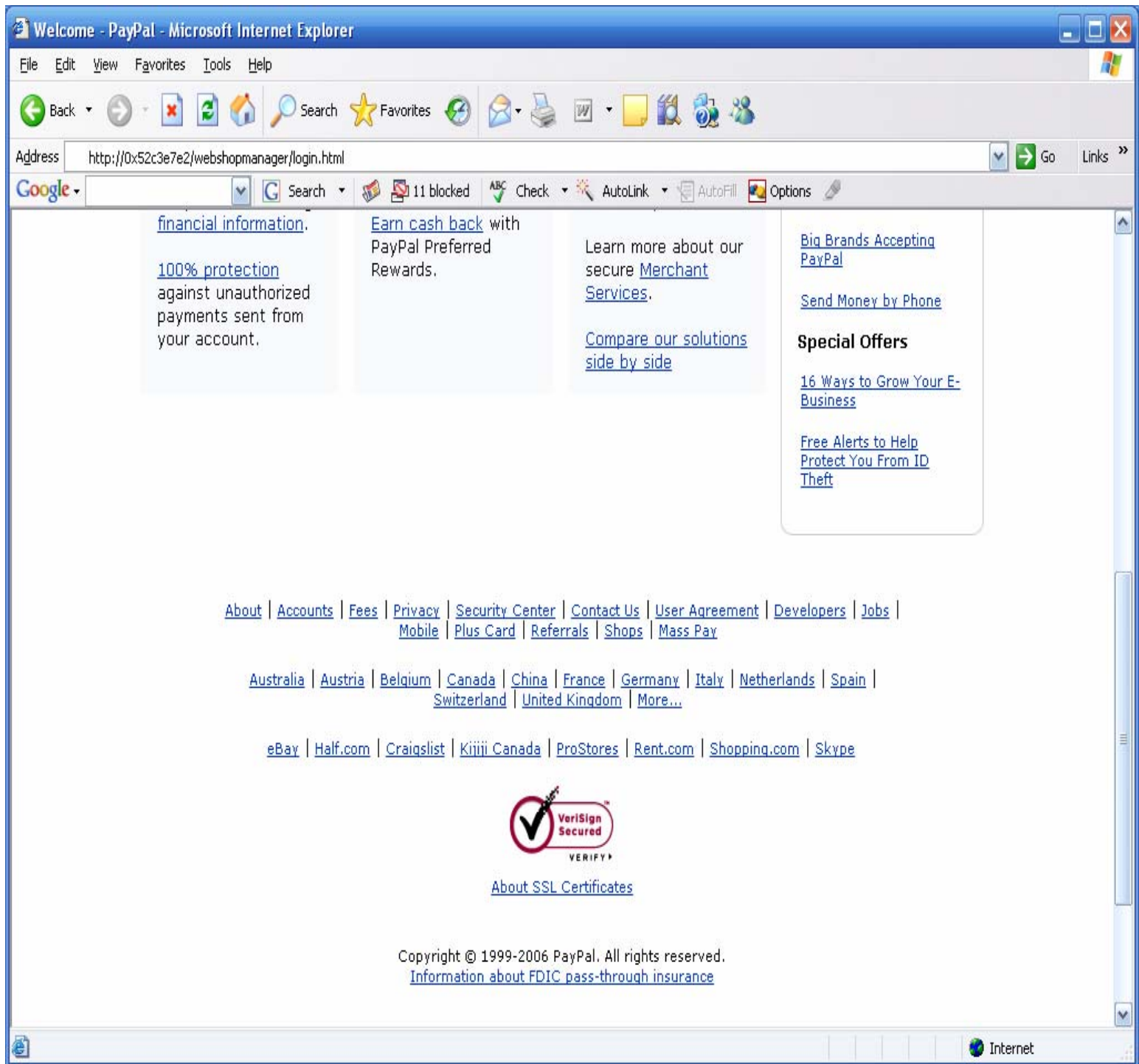
How can I restore my account access?

Please visit the [Resolution Center](#) and complete the "Steps to Remove Limitations."

Completing all of the checklist items will automatically restore your account access.

Copyright © 1999-2006 PayPal. All rights reserved.





1. Please indicate whether the email/website pair appears to have fraudulent intent or not.

Fraudulent

Not Fraudulent

2. If fraudulent, then on a scale of 1-5, which factors lead you to believe the pair to be fraudulent? (Scale: 1-Not at all, 5-Key identifying factor)

Text of email:

1 2 3 4 5

Website address (e.g. www.nps.edu):

1 2 3 4 5

What the website looks like:

1 2 3 4 5

Where the email came from:

1 2 3 4 5

Clickable link in email:

1 2 3 4 5

Believability of claims in email:

1 2 3 4 5

Please circle any other factors here you use to determine an email/Web site is fraudulent.

-Spelling/Grammar

-Webpage Source code

-Other:_____

-Consistency

-Security features

-Unusual Pronunciation

-Specificity

From: **Individualized BankCard Services** <<mailto:IBS@email.cardsatisfaction.net>>
Date: Apr 28, 2006 4:46 PM
Subject: Use your PNC Bank credit card today.
To: jackmcdowell@comcast.net



RE: Your account number
ending in 3272

**Consolidate your balances
into one payment .**

Dear Jack D. McDowell,

Enjoy the power of extra cash this spring with a balance transfer. With a click, you can transfer higher-rate balances to your PNC Bank Visa® credit card account. With just one monthly payment, it's the perfect opportunity to:

- Get rid of those department store balances.
- Make improvements to your home for spring.
- Plan a summer get-away.
- Join a gym.

Why not? Just click to open up a world of new possibilities for yourself. Go to <http://links.cardsatisfaction.net/ajtk/servlet/JJ?H=25bwbv&R=1571356816&P=www.pncnetaccess.com> to transfer balances, or visit your local bank to get a cash advance.

Your credit line is
\$27,500!

[Click Here To Start Your Balance Transfer](#)

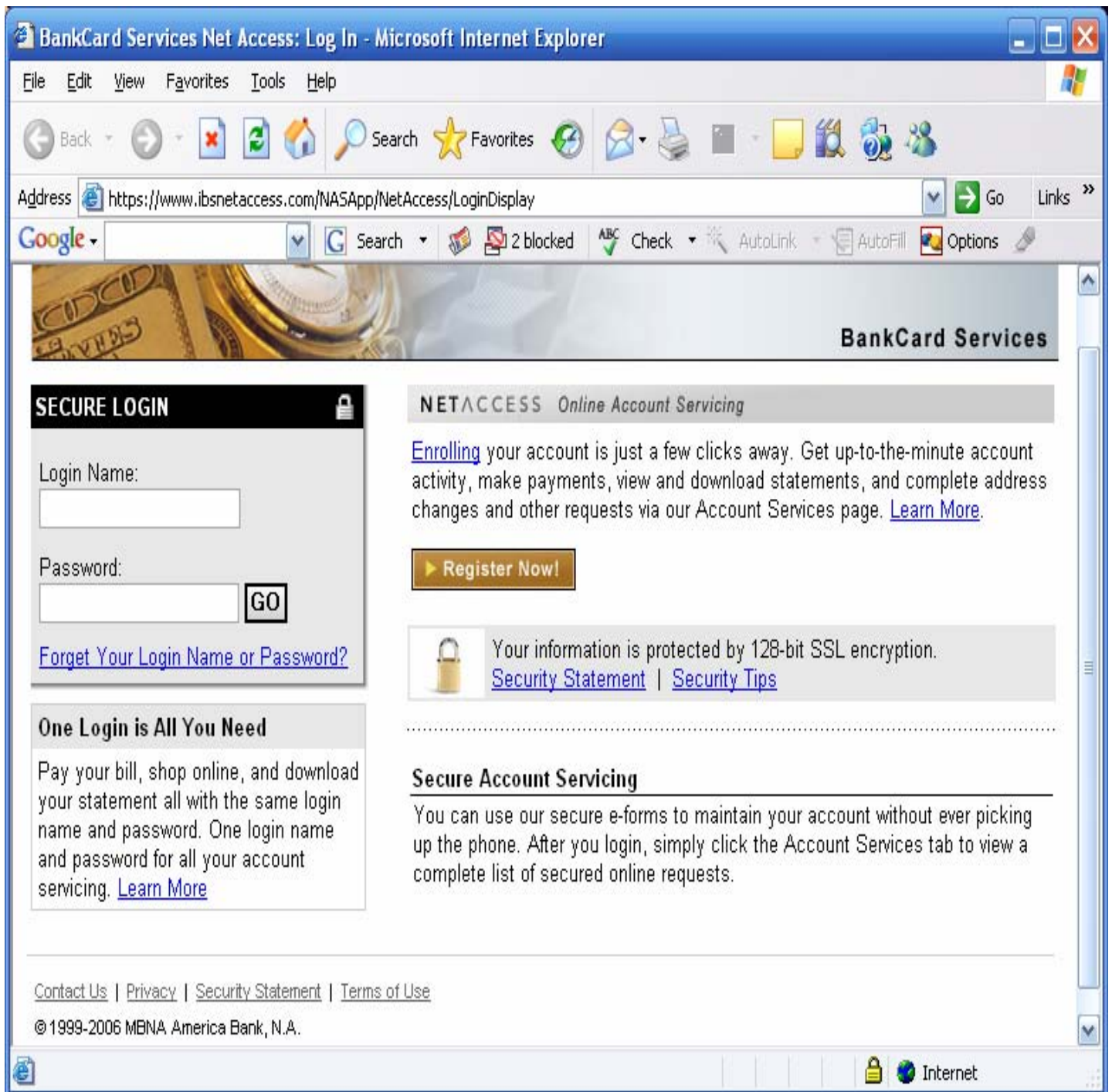
Consolidate balances.
Make just one monthly
payment.*

Complete
a balance
transfer.

View your last
12 months of
statements.

Submit billing or
merchant disputes
effortlessly and instantly.

Log
on
now



1. Please indicate whether the email/website pair appears to have fraudulent intent or not.

Fraudulent	Not Fraudulent			
<p>2. If fraudulent, then on a scale of 1-5, which factors lead you to believe the pair to be fraudulent? (Scale: 1-Not at all, 5-Key identifying factor)</p>				
<p>Text of email:</p>				
1	2	3	4	5
<p>Website address (e.g. www.nps.edu):</p>				
1	2	3	4	5
<p>What the website looks like:</p>				
1	2	3	4	5
<p>Where the email came from:</p>				
1	2	3	4	5
<p>Clickable link in email:</p>				
1	2	3	4	5
<p>Believability of claims in email:</p>				
1	2	3	4	5
<p>Please circle any other factors here you use to determine an email/Web site is fraudulent.</p>				
-Spelling/Grammar	-Webpage Source code	-Other:_____		
-Consistency	-Security features			
-Unusual Pronunciation	-Specificity			

Date: 21 Mar 2006 02:52:13 -0800

To:	"John Crawford" <JohnnyC@yahoo.com>
From:	"Amazon Marketplace" <marketplace-messages@amazon.com>
Subject:	Rate Your Transaction (058-2178846-0981142) at Amazon.com



Dear John Crawford,

Congratulations on your purchase from textbookexec2 on 02/18/2006 05:40 PM PST.

Leave Feedback About This Transaction

Please take a moment to rate this transaction (058-2178846-0981142). It's easy--just pick a rating and add any helpful comments. Your input will help Amazon.com and our sellers continually improve the customer experience. Your rating will apply to the following item:

1 of Phishing Exposed by Lance James [Paperback]

 (Need Help?)

Need

help?

If you have trouble with the "Leave feedback" button above, you can rate your transaction by completing the following steps:

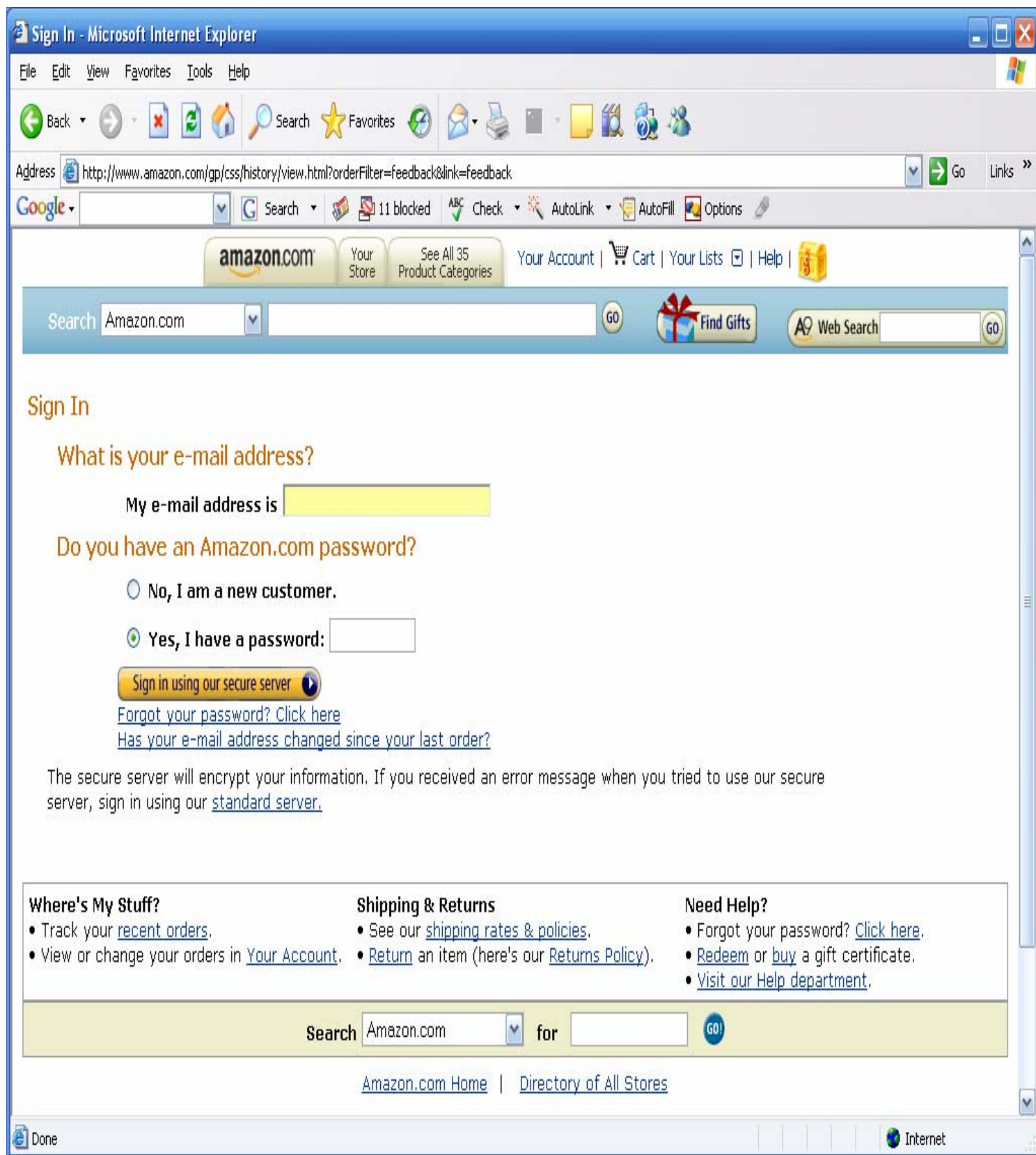
1. Go to <http://www.amazon.com/feedback>.
2. You'll be prompted for a log-in.
3. After logging in, you'll see a list of all of the orders that need feedback.
4. Find the order on the list and click the "Leave seller feedback" button on the right.

If you have questions regarding your transaction, we advise you to check with your seller directly at textbookexec@hotmail.com for details. To review your latest transactions, view [Your Account](#).

Please note that this message was sent to the following e-mail address:

JohnnyC@yahoo.com

Copyright 2005 Amazon.com, Inc. All rights reserved.



1. Please indicate whether the email/website pair appears to have fraudulent intent or not.

Fraudulent

Not Fraudulent

2. If fraudulent, then on a scale of 1-5, which factors lead you to believe the pair to be fraudulent? (Scale: 1-Not at all, 5-Key identifying factor)

Text of email:

1 2 3 4 5

Website address (e.g. www.nps.edu):

1 2 3 4 5

What the website looks like:

1 2 3 4 5

Where the email came from:

1 2 3 4 5

Clickable link in email:

1 2 3 4 5

Believability of claims in email:

1 2 3 4 5

Please circle any other factors here you use to determine an email/Web site is fraudulent.

-Spelling/Grammar

-Webpage Source code

-Other:_____

-Consistency

-Security features

-Unusual Pronunciation

-Specificity

Date:

Tue, 4 Apr 2006 14:09:51 +0200

To: ds_barnes@yahoo.com
From: customercare@chase.com
Subject: NOTICE FROM CHASE BANK



Dear Chase Bank Client,

This is your official notification from Chase Bank that the service(s) listed below will be deactivated and deleted if not renewed immediately. Previous notifications have

been sent to the Billing Contact assigned to this account. As the Primary Contact, you

must renew the service(s) listed below or it will be deactivated and deleted.

[Renew Now](#) your Chase Bank Bill Pay and Services.

If you are not enrolled in Online Banking, please enter your checking account number as

Sign-in Username and Social Security Number as Password.

SERVICE : Chase Bank Bill Payment.

EXPIRATION: April 9, 2006

Thank you, sincerely,

Lionel Wood

Customer Service

=====
IMPORTANT CUSTOMER SUPPORT INFORMATION
=====

Document Reference: (88856431).

2006 Chase Bank

Chase Personal Banking Investments Credit Cards Home Auto Commercial Small Business Insurance - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Reload Home Search Favorites Print Mail New Tab

Address <http://storkexpress.co.za/designs/chaseonline/CheckSession.php> Go Links

Google Chase bank online Search 11 blocked Check AutoLink AutoFill Options Chase bank online

[Find ATM / Branches](#) | [Contact Us](#) | [Site Map](#) |

Start banking online now
Get a User ID

Returning Users: Log On ⓘ

User ID:

Password:

☐ Remember my User ID
[Forgot User ID/Password?](#)

Consolidate your high interest bills with Chase and save.
0% APR for up to 12 Months*

The Chase Platinum Visa® Card
[Learn More](#)

Personal Banking

- ▶ Checking
- ▶ Credit Cards
- ▶ Savings
- ▶ CDs
- ▶ Online Banking & Bill Pay

Business

- ▶ Small Business Banking
Revenues up to \$10MM
- ▶ Commercial Banking
Revenues over \$10MM

Personal Lending

- ▶ Home Equity
- ▶ Mortgage
- ▶ Auto/Vehicle Loans
- ▶ Student Loans

Insurance & Investing

- ▶ Insurance
- ▶ Investing
- ▶ Retirement Planning

Tell me about...

- ▶ **Premier Platinum Banking**
Exclusive banking and investment benefits for clients with higher balances

News & Announcements

- ▶ **Circuit City Customers**
Chase is notifying a segment of Circuit City credit card account holders that computer tapes containing their personal information were mistakenly discarded.
- ▶ **U.S. Armed Forces Overseas**
Please contact us if you need assistance with your Chase or Bank One accounts.
- ▶ **Chase offers Zero-Fees!**
For academic year 2006-2007 the origination fee will be paid on all

Security Center Highlights

Chase helps keep you safe and informed.

- ▶ [Scams involving advance fees and cashier's checks](#)
- ▶ [Other online fraud and e-mail scams](#)
- ▶ [Ways we protect you](#)
- ▶ [How you can protect yourself](#)

You could win a trip to **Universal Orlando® Resort.** [Details](#)

New street. New house. Fresh start. [Learn More](#)

Internet

APPENDIX B: UNIGRAM SAMPLE OUTPUT

4832 examples, 372542 nonexamples, 0.012804273744349108 probability.

Negative Clues:

5.324813631522897E-4 0 938 computer
5.701254275940707E-4 0 876 program
5.720823798627002E-4 0 873 cellpad
5.733944954128441E-4 0 871 school
5.74052812858783E-4 0 870 cellspac
5.88235294117647E-4 0 849 faculty
6.527415143603133E-4 0 765 verdana
6.666666666666666E-4 0 749 paper
6.711409395973154E-4 0 744 bgcolor
6.729475100942127E-4 0 742 network
7.173601147776184E-4 0 696 include
7.621951219512195E-4 0 655 question
7.645259938837921E-4 0 653 technology
7.680491551459293E-4 0 650 science
7.763975155279503E-4 0 643 develop
7.937767899666614E-5 0 6298 font
8.237232289950577E-4 0 606 naval
8.403361344537816E-4 0 594 netflix
8.547008547008547E-4 0 584 head
8.771929824561404E-4 0 569 week
8.787346221441124E-4 0 568 project
8.849557522123894E-4 0 564 left
9.433962264150943E-4 0 529 postgraduate
9.671179883945841E-4 0 516 title
0.001004016064257028 0 497 data
0.0010183299389002036 0 490 conference
0.0010204081632653062 0 489 submit
0.0010330578512396695 0 483 discuss
0.0010504201680672268 0 475 university
0.0010548523206751054 0 473 space
0.0010615711252653928 0 470 3dmsnorm
0.0011037527593818985 0 452 fund
0.0011135857461024498 0 448 issue
0.0011135857461024498 0 448 plan
0.0011394712853236098 2 2191 color
0.0011467889908256881 0 435 current
0.0011574074074074073 0 431 write
0.0011655011655011655 0 428 topic
0.0011682242990654205 0 427 internets
0.001182033096926714 0 422 nextpart
0.0012254901960784314 0 407 engineer
0.0012285012285012285 0 406 security
0.0012345679012345679 0 404 thesis
0.0012355848434925864 1 1212 strong
0.0012376237623762376 0 403 new
0.0012468827930174563 0 400 request
0.0012594458438287153 0 396 search
0.0012594458438287153 0 396 webprnew
0.001272264631043257 0 392 describe
0.001272264631043257 0 392 harvard

0.0012755102040816326 0 391 view
 0.001282051282051282 0 389 colspan
 0.001288659793814433 0 387 require
 0.0012919896640826874 0 386 llpx
 0.0012919896640826874 0 386 delivery
 0.0012953367875647669 0 385 gmail
 0.0013020833333333333 0 383 background
 0.0013089005235602095 0 381 professor
 0.0013157894736842105 0 379 test
 0.0013192612137203166 0 378 message-id
 0.0013192612137203166 0 378 teach
 0.0013477088948787063 0 370 account
 0.0013736263736263737 0 363 body
 0.0014005602240896359 0 356 esmtp
 0.0014124293785310734 0 353 charset
 0.0014124293785310734 0 353 update
 0.0014367816091954023 0 347 shape
 0.0014534883720930232 0 343 smtpsvc
 0.001483679525222552 0 336 alumni
 0.0015015015015015015 0 332 bfi0
 0.0015151515151515152 0 329 newsletter
 0.0015290519877675841 0 326 design
 0.0015479876160990713 0 322 sensor
 0.001567398119122257 0 318 model
 0.001597444089456869 0 312 rect
 0.001607717041800643 0 310 committee
 0.0016181229773462784 0 308 cs-announc
 0.001639344262295082 0 304 format
 0.0016835016835016834 0 296 academic
 0.0016891891891891893 0 295 10px
 0.0016891891891891893 0 295 return-path
 0.0017064846416382253 0 292 |hour 9
 0.001718213058419244 0 290 |hour 1
 0.0017361111111111111 0 287 author
 0.0017361111111111111 0 287 copy
 0.0017421602787456446 0 286 10pt
 0.0017482517482517483 0 285 print
 0.0017543859649122807 0 284 3dhttp
 0.0017605633802816902 0 283 product
 0.0017667844522968198 0 282 0pt
 0.0017667844522968198 0 282 coord
 0.0017730496453900709 0 281 |hour 2
 0.0017793594306049821 0 280 quoted-print
 0.0017921146953405018 0 278 line-height
 0.0017985611510791368 0 277 redshift
 0.0018050541516245488 0 276 found
 0.0018050541516245488 0 276 iso-8859-1
 0.0018050541516245488 0 276 result
 0.0018115942028985507 0 275 a676
 0.0018181818181818182 0 274 media
 0.0018315018315018315 0 272 content-class
 0.0018315018315018315 0 272 horn
 0.0018315018315018315 0 272 suggest
 0.0018450184501845018 0 270 bold
 0.0018656716417910447 0 267 tigerdirect
 0.0018796992481203006 0 265 hear
 0.0018796992481203006 0 265 opportune

0.0018867924528301887 0 264 en-u
 0.001893939393939394 0 263 base
 0.0019011406844106464 0 262 document
 0.0019011406844106464 0 262 epson
 0.0019083969465648854 0 261 game
 0.0019083969465648854 0 261 localhost
 0.0019157088122605363 0 260 language
 0.0019157088122605363 0 260 public
 0.0019230769230769232 0 259 communicate
 0.0019455252918287938 0 256 workshop
 0.001953125 0 255 create
 0.001953125 0 255 solve
 0.001976284584980237 0 252 spawar
 0.0020242914979757085 0 246 institute
 0.0020325203252032522 0 245 educate
 0.00205761316872428 0 242 future
 0.00205761316872428 0 242 train
 0.00205761316872428 0 242 |hour 4
 0.002066115702479339 0 241 suntop
 0.002074688796680498 0 240 college
 0.0020833333333333333 0 239 akamaitech
 0.0020833333333333333 0 239 firewall
 0.0021008403361344537 0 237 technic
 0.002109704641350211 0 236 redir
 0.002109704641350211 0 236 tool
 0.002127659574468085 0 234 schedule
 0.002136752136752137 0 233 bottom
 0.002136752136752137 0 233 origin
 0.002145922746781116 0 232 compute
 0.0021551724137931034 0 231 cost
 0.0021551724137931034 0 231 final
 0.0021551724137931034 0 231 purchase
 0.002183406113537118 0 228 subscribe
 0.0022026431718061676 0 226 award
 0.0022026431718061676 0 226 commune
 0.0022026431718061676 0 226 hour
 0.0022123893805309734 0 225 feature
 0.0022222222222222222 0 224 control
 0.0022222222222222222 0 224 delivered-to
 0.002242152466367713 0 222 do
 0.002242152466367713 0 222 participate
 0.0022727272727272726 0 219 ieee
 0.0022727272727272726 0 219 sponsor
 0.00228310502283105 0 218 chair
 0.0022935779816513763 0 217 Prof
 0.0022935779816513763 0 217 ientri
 0.0022935779816513763 0 217 ientrymail
 0.0022935779816513763 0 217 rowspan
 0.0022935779816513763 0 217 source
 0.002304147465437788 0 216 machine
 0.0023148148148148147 0 215 navy
 0.002325581395348837 0 214 defense
 0.002347417840375587 0 212 digit
 0.002347417840375587 0 212 value
 0.0023584905660377358 0 211 window
 0.002369668246445498 0 210 store
 0.002369668246445498 0 210 try

0.002403846153846154 0 207 edit
0.002403846153846154 0 207 unsubscribe
0.002407704654895666 1 621 applicate

Positive Clues:

0.19293478260869565 35 148 limit
0.19879518072289157 16 66 protect
0.203125 6 25 anymore
0.203125 6 25 sensitive
0.2073170731707317 8 32 delay
0.2080745341614907 33 127 notify
0.20833333333333334 12 47 apology
0.21014492753623187 14 54 transact
0.22058823529411764 7 26 foreign
0.22727272727272727 22 76 immediate
0.2288135593220339 13 45 obtain
0.24008810572687225 54 172 yahoo
0.2403846153846154 12 39 claim
0.24152542372881355 28 89 money
0.2661290322580645 16 45 delete
0.2973568281938326 67 159 paypal
0.3046875 19 44 verify
0.37272727272727274 20 34 Card
0.5092592592592593 27 26 renew
0.7442748091603053 97 33 bank
0.8582474226804123 166 27 paypal
0.8854166666666666 42 5 chase
0.953125 30 1 mber
0.9848484848484849 32 0 riti
0.9848484848484849 32 0 thro
0.9901960784313726 50 0 bject
0.9971751412429378 176 0 acco

APPENDIX C: BIGRAM SAMPLE OUTPUT

4784 examples, 372070 nonexamples, 0.012694571372467852 probability.

Strong Negative Clues:

3.952569169960474E-4 0 1264 PM send
4.3975373790677223E-4 0 1136 font size
4.6816479400749064E-4 0 1067 AM send
9.671179883945841E-4 0 516 postgraduate school
9.98003992015968E-4 0 500 font color

0.0013157894736842105 0 379 computer science
0.0019305019305019305 0 258 face are
0.002008032128514056 0 248 area shape
0.002109704641350211 0 236 School naval
0.002136752136752137 0 233 rect coord
0.002173913043478261 0 229 g akamaitech
0.0022026431718061676 0 226 A href
0.0022727272727272726 0 219 style 3d'margin-left
0.002304147465437788 0 216 redir internets
0.002304147465437788 0 216 www netflix
0.0023584905660377358 0 211 span class
0.0024154589371980675 0 206 width 100%

Strong Positive Clues:

0.4125 16 23 the service
0.5078125 32 31 to address
0.515625 16 15 and delete
0.6770833333333334 32 15 address book
0.9285714285714286 32 2 mobile alert
0.9517543859649122 108 5 can call
0.9517543859649122 108 5 go ahead
0.9517543859649122 108 5 help link
0.9539473684210527 72 3 email notify
0.9539473684210527 72 3 of limit
0.9539473684210527 72 3 the offer
0.955026455026455 180 8 the primary
0.9587765957446809 360 15 and delete
0.9601769911504425 108 4 help center
0.9601769911504425 108 4 this form
0.9601769911504425 108 4 to release
0.9666666666666667 72 2 limit access
0.9666666666666667 72 2 of charge
0.9666666666666667 72 2 please verify
0.9674329501915708 252 8 please confirm
0.9848484848484849 32 0 Add mobile
0.9848484848484849 32 0 paypal acco
0.9852941176470589 33 0 chase bank
0.9901960784313726 50 0 S bject
0.9930555555555556 71 0 r acco

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

1. Anti-Phishing Working Group(APWG), *Committed to Wiping Out Internet Scams and Fraud*. Retrieved May 10, 2006 from <http://www.antiphishing.org/index.html>
2. Dhamija, R., Tygar, J., & Hearst, M. (2006). Proceedings of the Conference on Computers and Human Interaction: *Why Phishing Works*. Montréal, QB, Canada.
3. Government Accounting Office. (2005). *Information Security: Emerging Cybersecurity Issues Threaten Federal Information Systems*. Retrieved December 1, 2005 from www.gao.gov/cgi-bin/getrpt?GAO-05-231
4. Fogg, B.J. (2003). *Persuasive Technology: Using Computers to Change What We Think and Do*. San Francisco: Morgan Kaufmann Publishers.
5. Jagatic, T., Johnson, N., Jakobsson, M., & Menczer F., (2005) Draft Preprint for Communications of the ACM: *Social Phishing*. Retrieved March 16, 2006 from <http://www.indiana.edu/~phishing/social-network-experiment/phishing-preprint.pdf>
6. Jakobsson, M. & Ratkiewicz, J. (2006). Proceedings of the Conference on the World Wide Web: *Designing Ethical Phishing Experiments*. Edinburgh, Scotland.
7. James, L. (2005). *Phishing Exposed*. Rockland, MA: Syngress Publishing, Inc.
8. Mailfrontier. Press Release: *Phishing IQ Test*. (March 30, 2005). Retrieved Jan 10, 2006 from http://www.mailfrontier.com/press/press_uk_phishingtest.jsp
9. MessageLabs Intelligence. (2005). *Annual Security Report: Cyber-criminals Narrow Their Focus*. Retrieved January 26, 2006 from www.messagelabs.com/Threat_Watch/Intelligence_Reports.
10. Ragucci, J.W. & Robila, S.A. (2006). *Societal Aspects of Phishing*. Retrieved May 16, 2006 from http://pages.csam.montclair.edu/~robila/RSL/Papers/istas06_2.pdf
11. Wu, M., Miller, R.C., & Garfinkel, S.L. (2006). Proceedings of the Conference on Computers and Human Interaction: *Do Security Toolbars Actually Prevent Phishing Attacks?* Montréal, QB, Canada.

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
Ft. Belvoir, Virginia
2. Dudley Knox Library
Naval Postgraduate School
Monterey, California
3. Craig Martell
Naval Postgraduate School
Monterey, California
4. Neil Rowe
Naval Postgraduate School
Monterey, California
5. David S. Barnes
Naval Postgraduate School
Monterey, California