Phishing Detection and Loss Computation Hybrid Model

A Machine-learning Approach

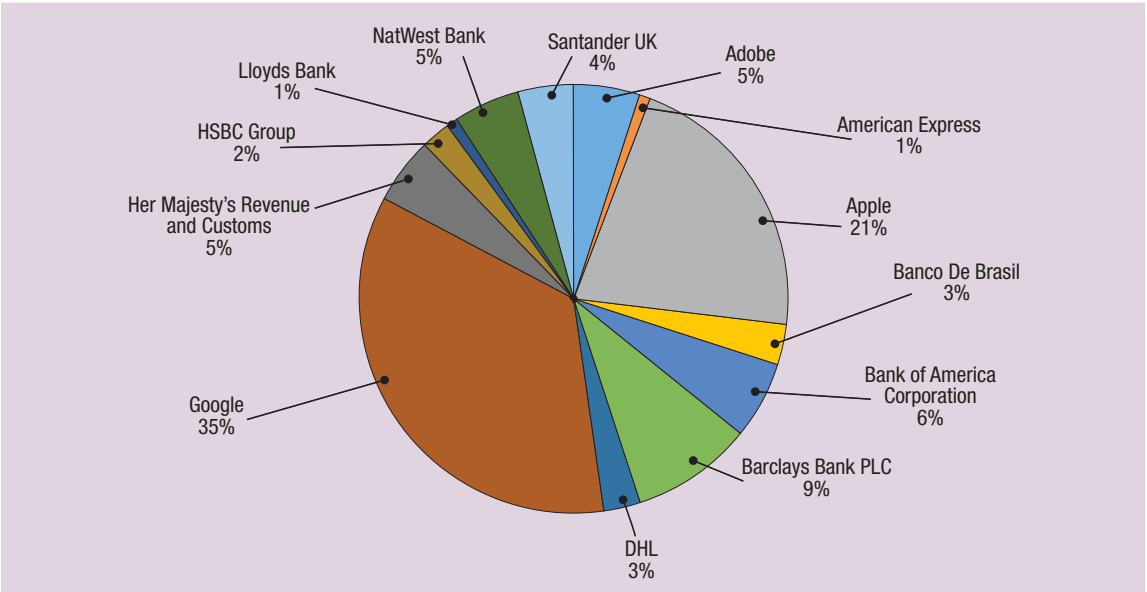
Do you have something to say about this article?
Visit the *Journal* pages of the ISACA® website *(www.isaca.org/journal)*, find the article and click on the Comments link to share your thoughts.

Phishing involves social engineering of data over the Internet to acquire personal or business information from unsuspecting users. The 2015 Internet Crime Report from the US Federal Bureau of Investigation (FBI) Internet Crime Complaint Center (IC3) states that chief executive officer (CEO) email scams, also known as business email compromise (BEC), cost US firms US $246 million in 2015. Affected firms have reported more than 7,833 BEC complaints to the FBI IC3.[1] In contrast, identity and credential theft costs were lower, at USD $57 million, with 22,000 reported cases in 2015.

Phishing attacks are aimed at naive users to trick them so they unintentionally divulge critical information, such as usernames; social network passwords; and banking, financial and credit card details. Phishing attackers use spam emails, corrupt web URLs and multimedia messages to target users and lure them to fake web pages. For example, Dridex phishers sent targeted emails that had malware attached in the form of Microsoft Office macros to users in English-speaking nations in efforts to steal their banking credentials.[2] **Figure 1** shows the top targeted firms in 2016 from various industries.[3]

In light of these events, a hybrid model can be considered to compute the probability of a URL being malicious and the expected loss for the first

Figure 1—Percentage Share of Top Targeted Firms Based on PhishTank Archive in 2016



Source: B. Biswas, A. Mukhopadhyay. Reprinted with permission.

Baidyanath Biswas
Is a Ph.D. student in information technology and systems at the Indian Institute of Management (Lucknow, India). His research interests are privacy and risk issues in information systems, the economics of cyber security, and health care IT. He has worked as a senior software engineer for nine years with Infosys, IBM and Cognizant. He can be reached at fpm15005@iiml.ac.in.

Arunabha Mukhopadhyay, Ph.D.
Is an associate professor in information technology and systems at the Indian Institute of Management (Lucknow, India). He is the recipient of the Best Teacher in Information Technology Management Award in 2013 and 2011, and the 19th Dewang Mehta Business School Award. He can be reached at arunabha@iiml.ac.in.

24 hours after the phishing attack. The model also offers a set of strategies to help the C-suite make policy-level decisions and frame organizational security policies to minimize losses due to such phishing attacks.

Proposed Hybrid Model for Phishing Detection and Loss Computation

Figure 2 describes the hybrid model for phishing detection and loss computation for firms that regularly face phishing attacks. The hybrid model consists of three modules:

- Risk analysis to calculate the probability of a prospective URL that can lead to a phishing attack
- Loss computation to estimate the expected loss to stakeholders after the phishing attack
- Risk mitigation to offer techno-social recommendations to minimize losses arising from such an attack

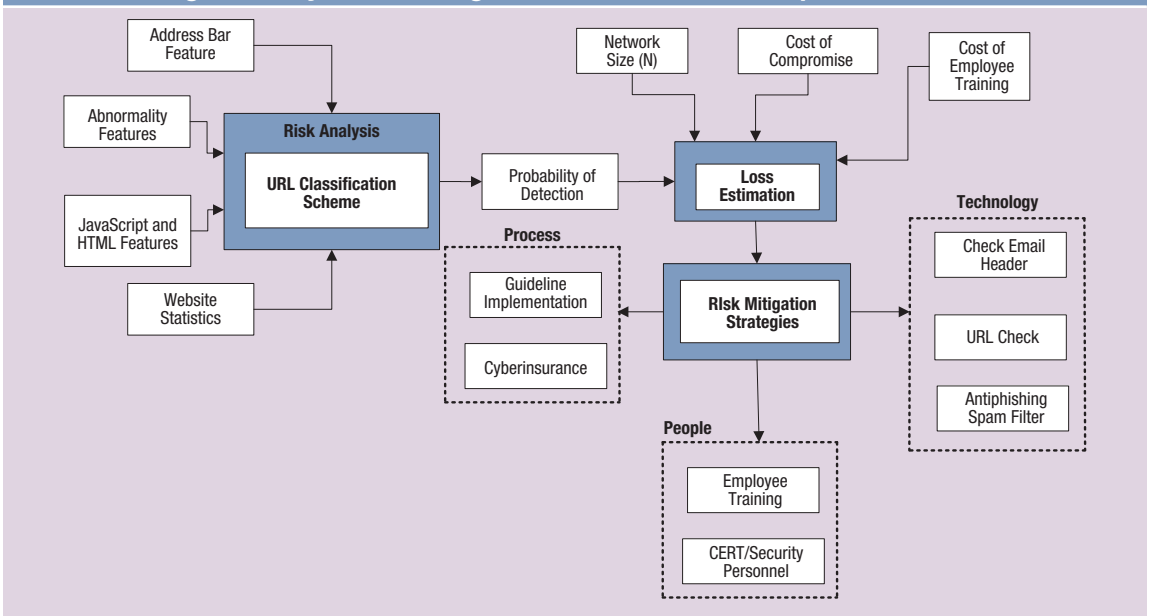
What Are Machine Learning Techniques?

Machine learning techniques consist of pattern recognition from data and learning algorithms that apply to practical applications such as intrusion detection systems (IDS) and antispam and antiphishing filters. **Figure 3** provides a schematic view of the classification and regression tree (CART)-based model that generates rules that apply to the data set of legitimate and corrupt websites. The CART uses the Tree Bagger method to ascertain the importance of the variables. Unknown URLs are predicted as legitimate or suspicious using the classifier and its rule set.

Bagger Algorithm for Decision Tree

Bootstrapping aggregation, also known as bagger, is an ensemble technique used in the CART algorithm. It generates multiple prediction trees and combines each model to improve accuracy and

Figure 2—Hybrid Phishing Detection and Loss Computation Model



Source: B. Biswas, A. Mukhopadhyay. Reprinted with permission.

reduce overfitting the original classifier. The input data are generated by randomly choosing records with replacement from the original training set. The error of the model is used as an estimator for the importance of a predictor variable. An ensemble model will have higher model error if the majority of the predictor variables are influential and *vice versa*.

Data

Google and Alexa Top 500 website rankings offer a list of legitimate sites.[4] Phishing sites that report through MillerSmiles and PhishTank archives deliver the malicious URLs.[5] The predictor variables in the data set are encoded:

+1 = legitimate URL

0 = suspicious URL

-1 = phishing URL

The target variable is encoded -1 for phishing and +1 for legitimate websites. Training and testing are performed in 80:20 ratios, with 8,844 records for testing and the remaining 2,211 for training.

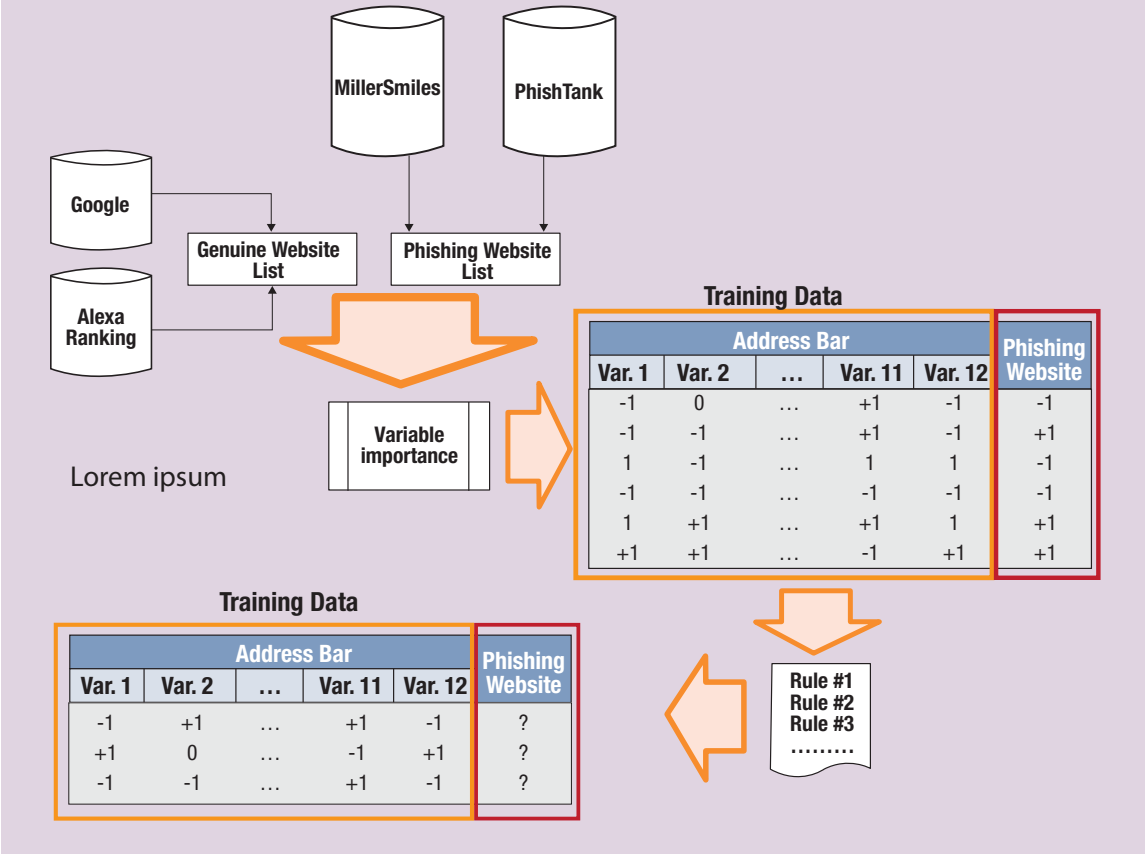
Methodology for the CART-based Hybrid Classifier

Figure 4 illustrates the steps of the CART-based hybrid classifier that focuses on the training data to create a rule set and run test data, as in **figure 3**. The classifier uses a bagger algorithm to create a list of the most significant variables from the total training set of the 30 encoded predictors.

Identifying the Most Significant Variables

In the experimental data set,[6] there are 30 input variables, broadly categorized as address-bar

Figure 3—CART-based Phishing URL Classifier and Ruleset Generation



Source: B. Biswas, A. Mukhopadhyay. Reprinted with permission.

Figure 4—Steps to Implement CART-based Hybrid Classifier	
Step 1	Load the list of legitimate URLs based on website ranking (data set D1).
Step 2	Load the list of phishing and suspicious websites (data set D2).
Step 3	Load the input file after combining the two data sets (D1, D2).
Step 4	Identify the most significant predictor variables using the CART-based hybrid classifier algorithm.
Step 5	Train the Tree Bagger using the significant predictor variables only.
Step 6	Test with out-of-sample data and measure the accuracy of the CART-based hybrid classifier.

Source: B. Biswas, A. Mukhopadhyay. Reprinted with permission.

properties, abnormality features, HTML and JavaScript features, and website statistics.[7] **Figure 5** shows the plot of importance based on out-of-bag features for all 30 variables. The plot also indicates the top five significant variables in order of their importance, which are #8 (HTTPS in URL), #14 (URL of anchor), #26 (website traffic statistics), #15 (links in <Meta>, <Script> and <Link> tags) and #7 (URL subdomain). The classification technique generates rule sets based on all/some of these significant predictors only.

Figure 6 illustrates the general website URL-based predictor variables for probable phishing links and their attributes of identification.

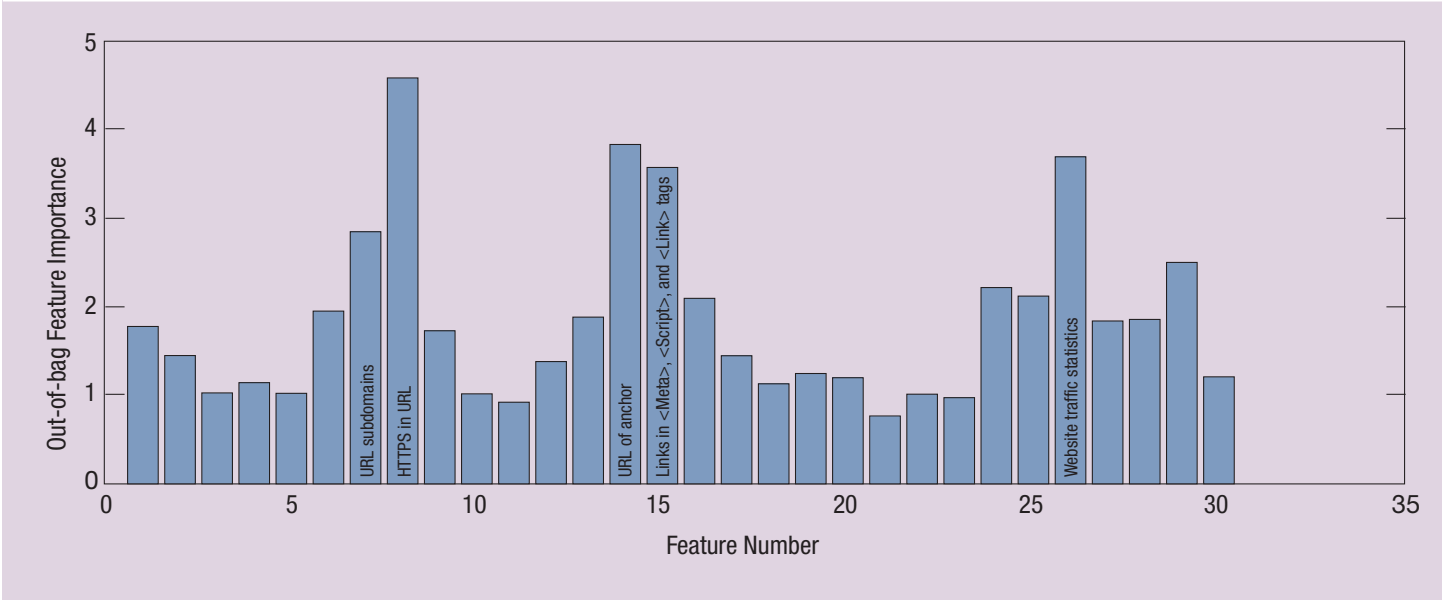
Loss Computation for Firms After a Phishing Attack

Consider a corporate network of N = 10,000 users, and assume that the network traffic saturates as more users join in following a logistic diffusion curve.[8] **Figure 7** illustrates the multiple stages of a phishing attack and the probability of user decisions and actions.

The stages are:

- Attackers spam the network with infected emails.
- Attackers wait for a naive user to open the infected email.
- Users read the email(s).

Figure 5—Variable Importance for All Phishing Predictors



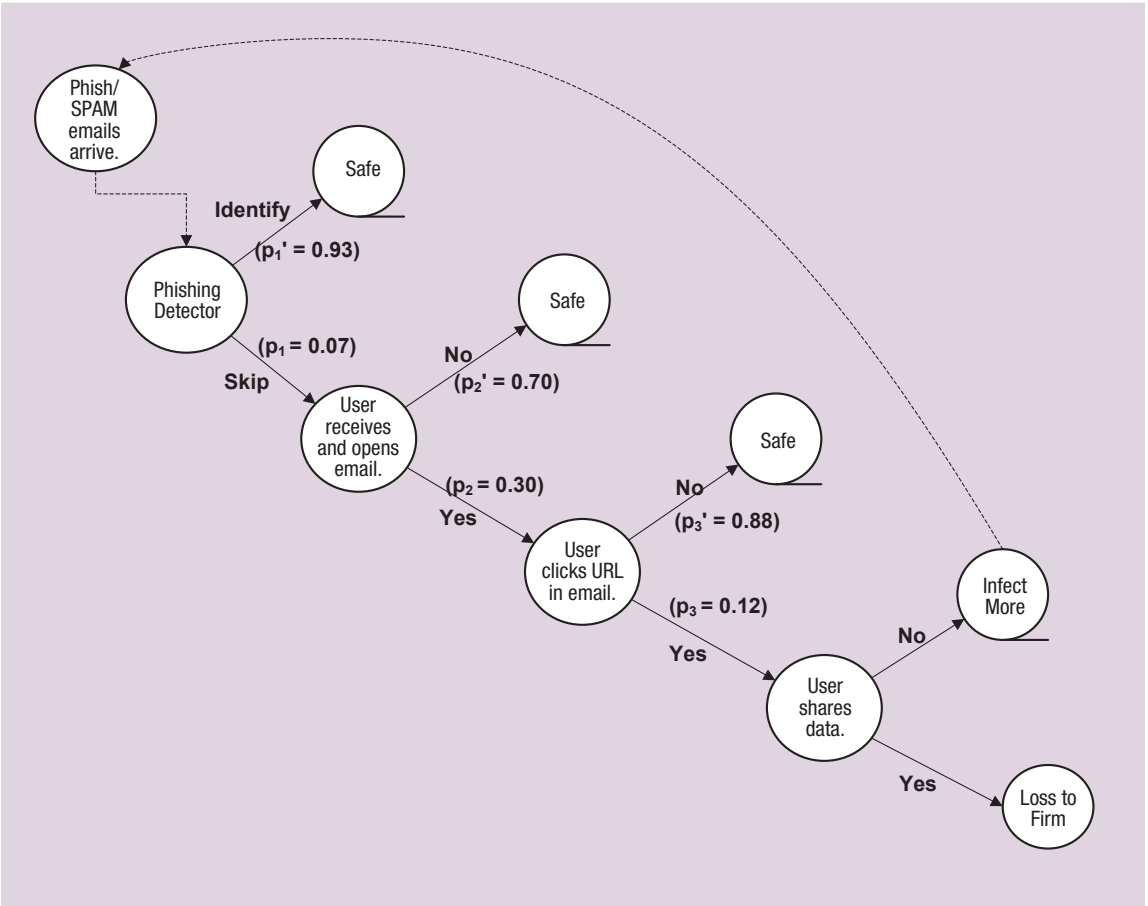
Source: B. Biswas, A. Mukhopadhyay. Reprinted with permission.

Figure 6—Common URL-based Features in Phishing URLs

Feature	Example Link (Source: PhishTank Archive)
Redirect Using //	*http://www.tasteofthewest.co.uk/images//wsecure/ap5c/*
Extremely Long URL	*https://docs.google.com/a/valpo.edu/forms/d/17zrMsBmbTzz4tvu3VqcXM3huxNwnxfeyuU0Bc9iTKZc/viewform?usp=send_form*
@ Symbol in the URL	*http://imessage-audits.org/profile/?email=abuse@example.com*
HTTPS (Hypertext Transfer Protocol with Secure Sockets Layer)	*https://accounts.google.com/ServiceLogin?continue=https://drive.google.com/st/auth/host/0Bz9pzRUAjfXaT3RXengxQXV3dlU/*
- Separator	*http://irstax.wap-ka.com/index.xhtml*
Sub/Multisub Domains	*http://www.grandimperial.com.my/v2/en/*
Nonstandard Port	*http://www.belcotech.com:32000/mail/wait.html*
IP Address in the URL	*http://194.78.154.195/CFIDE/services/labanquepostale.html*
HTTPS within URL	*http://www.roma.md/templates/system/https:/www2.itau.com.br/atendimento/*

Source: B. Biswas, A. Mukhopadhyay. Reprinted with permission.

Figure 7—Process Flow Diagram for User Action(s) After Phishing Attack



Source: B. Biswas, A. Mukhopadhyay. Reprinted with permission.

- Users click on the malicious URL.
- Users share their credentials through the fraudulent URL.[9]

The following equation gives the loss per hour after the phishing attack:

> Loss per hour = N (size of the network) * (Number of skipped URLs) * Prob (open_email) * prob (click_URL) * prob (share_ info) * (monetary impact of phishing)

Results

Figure 8 shows that out of 980 test records of phishing URLs, the classifier can pick up 876 records with a phishing URL, with a true positive (TP) rate of 89.29 percent [876/(876+105)]. The model-identified good websites are at a true negative (TN) rate of 94.24 percent [1,179/(1,179+72)]. The classifier works with an overall accuracy of 92.94 percent [(876+1,179)/2,211] in predicting phishing and legitimate websites. Out of 100 test URLs assigned to the rule-based model, 93 URLs were marked as legitimate, suspicious or phishing. Therefore, the probability of correctly identifying a phishing website is 0.9294 for the hybrid model described in **figure 2**.

The following example demonstrates loss computation. In 2016, a payment card firm was targeted by 29 percent of 1,000 URLs, which equals 290 phishing URLs. Out of this 29 percent, the probability of successful prediction by the classifier is 92.94 percent, and the dilemma of decision making for the firm's management may arrive from the remaining 7.06 percent of 290 URLs, which is approximately 21 URLs. In the next step, the estimated loss is calculated from the equation described previously.

Calculation for Expected Loss

1. The accuracy of the classifier: 92.94 percent (calculated)
2. Phishing URLs that may skip the filter: 1-92.94 percent = 7.06 percent
3. Out of 1,000 URLs sent in total, a firm targeted in approximately 29 percent of the cases received 29 percent of 1,000, which equals 290 URLs.
4. Combining (2) and (3), total phishing URLs that skip the filter are 7.06 percent of 290 = 21.
5. Given the probability of opening email equals 30 percent, probability of clicking URL equals 12 percent and probability of sharing info equals 12 percent. Average monetary impact of phishing in financial industry equals US $264.[10] Substituting values into the equation, the cumulative loss per hour = (N) * 21 * 30% * 12% * 12% * $264, where N increases exponentially with network diffusion rate equals 0.2, and total strength of the network equals 10,000.
6. The hourly calculation is shown in **figure 9** (also indicated by the blue graph in **figure 10**).

Based on the exponential rule of diffusion, after the users start clicking on the phishing URLs, the network starts blocking these sites. Gradually, the system is saturated and the phishing attackers cannot extract much of a financial impact and, thus, the loss begins to reduce. The nonlinear and diminishing nature of the loss curves (**figure 10**) attributes to this phenomenon. With a high probability state of {open, click, share} = {0.50, 0.20, 0.20}, the loss is greater than that of the medium state, which is {0.40, 015, 0.15}, and that of the low state, which is {0.20, 0.10, 0.10}.

Figure 8—Confusion Matrix for Classification Based on URL Predictors

	Predicted: Phishing		Predicted: Genuine	
Actual: Phishing	876	**TP**	105	**FN**
Actual: Genuine	72	**FP**	1179	**TN**

TP = true positive FP = false positive TN = true negative FN = false negative

Source: B. Biswas, A. Mukhopadhyay. Reprinted with permission.

Figure 9—Calculation for Expected Loss

Time (in Hours)	Hourly Network Strength	Cumulative Loss (US Dollars)	Loss Per Hour (US Dollars)
1	5,498	128,387	11,407
2	5,987	139,794	10,967
3	6,457	150,761	10,348
.....	.....	.....	.....
.....	.....	.....	.....
24	9,918	231,595	417
25	9,933	231,939	343
26	9,945	232,220	281

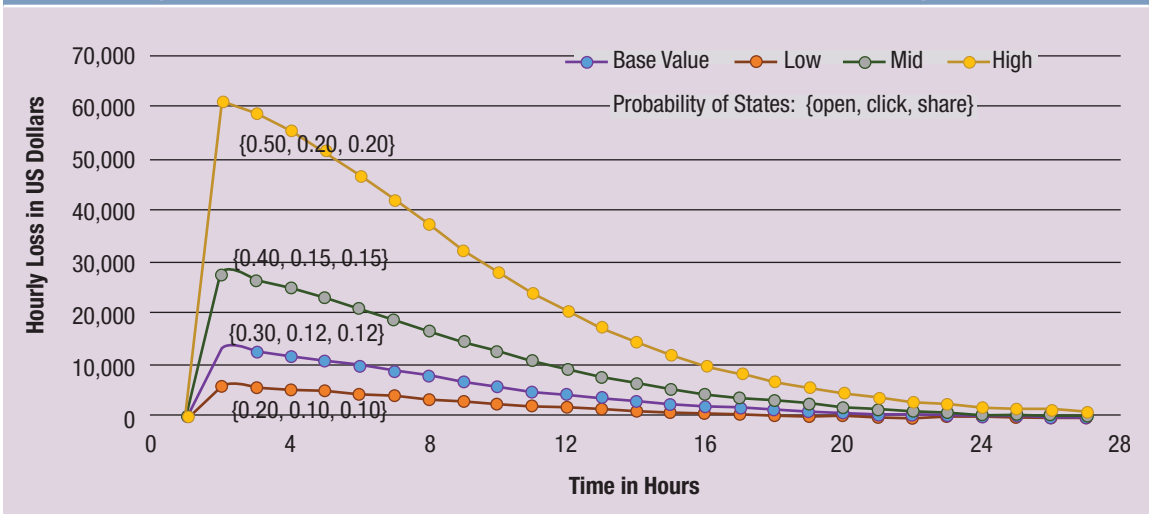
Source: B. Biswas, A. Mukhopadhyay. Reprinted with permission.

Risk Mitigation Strategies

Figure 11 shows that when mitigation strategies (people, process and technology) are low, the measured financial impact of phishing attacks is highest. When the mitigation plan is high for all the factors (people, process and technology), the loss due to phishing minimizes.

Risk reduction should begin with technology tools, for example, software checks for suspicious emails and web pages, and installing antispam and antiphishing filters across the network. Top management executives such as chief information security officers (CISOs) and chief technology officers (CTOs) should readily implement stringent security guidelines and system processes in the organization to be able to identify such scenarios. Appropriate training organized by human resources executives should follow so that employees remain cognizant of the behavior of phishing attacks and their categories. Organizations should maintain computer emergency response teams (CERT) and system administrators for their corporate networks to accurately scan assets and encourage employees to abide by the guidelines.

Figure 10—Expected Loss for the First 24 Hours After a Phishing Attack



Source: B. Biswas, A. Mukhopadhyay. Reprinted with permission.

Figure 11—Multilevel Mitigation Strategies and Loss Levels

Level of Phishing / Mitigation Strategy	Base Values (Verizon DBIR 2016)	Low-impact Phishing	Medium-impact Phishing	High-impact Phishing
People		High	Middle	Low
Process				
Technology				
Prob. (Open)	30%	20%	40%	50%
Prob. (Click)	12%	10%	15%	20%
Prob. (Share)	12%	10%	15%	20%

Source: B. Biswas, A. Mukhopadhyay. Reprinted with permission.

Conclusion

The three-level model proposed in this article can be used to compute the probability of phishing through corrupt URLs and the expected loss during the first 24 hours after an attack. This article presents multitier recommendations against phishing attacks for broad categories of businesses and their employees. The classification scheme (**figure 3**) considers significant variables to predict the target class—phishing or legitimate websites. The associated probability of the classifier is then applied to compute the estimated loss (**figure 8**) through a period of 24 hours, immediately after the firm has suffered a phishing attack. Recommendation strategies for people, process and technology should be applied in sync with each other so that the estimated loss arising due to phishing attacks is lessened.

Endnotes

1 Department of Justice, Federal Bureau of Investigation, "2015 Internet Crime Report," Internet Crime Complaint Center, USA, *https://pdf.ic3.gov/2015_IC3Report.pdf*

2 O'Brien, D.; *Dridex: Tidal Waves of Spam Pushing Dangerous Financial Trojan*, Symantec, 2016, *www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/dridex-financial-trojan.pdf*

3 APWG, *2016 APWG Phishing Attack Trends Reports*, 2016, *www.antiphishing.org/resources/apwg-reports/*

4 Alexa, "The Top 500 Sites on the Web," *www.alexa.com/topsites*

5 PhishTank Archives, *https://www.phishtank.com/developer_info.php*

6 Lichman, M.; "UCI Machine Learning Repository," 2013, *http://archive.ics.uci.edu/ml/*

7 Mohammad, R. M.; F. Thabtah; L. McCluskey; "Predicting Phishing Websites Based on Self-structuring Neural Network," *Neural Computing and Applications*, vol. 25, iss. 2, 2014, p. 443-458, *http://eprints.hud.ac.uk/19220/3/RamiPredicting_Phishing_Websites_based_on_Self-Structuring_Neural_Network.pdf*

8 Ransbotham, S.; S. Mitra; "Choice and Chance: A Conceptual Model of Paths to Information Security Compromise," *Information Systems Research*, vol. 20, iss. 1, 2009, p. 121-139

9 Verizon Enterprise, *2016 Data Breach Investigations Report*, 2016, *www.verizonenterprise.com/verizon-insights-lab/dbir/*

10 Ponemon Institute, *2016 Cost of Data Breach Study: United States,* 2016, *www-03.ibm.com/security/data-breach/*