

Detecting DNS-poisoning-based phishing attacks from their network performance characteristics

H. Kim and J.H. Huh

Most of the existing phishing detection techniques are weak against domain name system (DNS)-poisoning-based phishing attacks. Proposed is a highly effective method for detecting such attacks: the network performance characteristics of websites are used for classification. To demonstrate how useful the approach is, the performance of four classification algorithms are explored: linear discriminant analysis, naïve Bayesian, *K*-nearest neighbour, and support vector machine. Over 10 000 real-world items of routing information have been observed during a one-week period. The experimental results show that the best-performing classification method – which uses the *K*-nearest neighbour algorithm – is capable of achieving a true positive rate of 99.4% and a false positive rate of 0.7%.

Introduction: Over the years, many techniques have been developed to help identify phishing websites; some of these include whitelisting of legitimate sites, blacklisting of fraudulent sites and identifying common heuristics. Unfortunately, when it comes to detecting domain name system (DNS)-poisoning-based phishing attacks, such techniques tend to achieve low performance [1]: Abu-Nimeh and Nair introduced an attacking scenario in which a rogue access point (with stronger signal range) is setup at Starbucks, and a user connects to the Internet through this access point rather than Starbucks' hotspot [1]. As the local DNS in this access point is poisoned, when the user types a URL in her browser, she is directed to a phishing website hosted at the access point's local server. Here, her usual security toolbars and phishing filters will not provide any warning message. This is because most of these techniques merely check the domain names of websites and/or rely on remote verification servers to perform the necessary phishing detection tasks. In practice, however, the communication channel between client software and a remote server is not often secured, allowing attackers to easily forge the server's response. Keeping the IP addresses alongside the domain names could be effective only if the IP addresses were all static. Content-based phishing detection methods too have inherent weaknesses: phishing contents can be sophisticatedly created to avoid the (publicly known) properties that are being checked by heuristics.

We examine how a new heuristic, the 'network performance characteristics' of websites that a user has previously visited can be used to train classifiers and detect DNS-poisoning-based phishing attacks: classification results generated from real routing information collected during a one-week period are used to analyse the performance and suitability of different classification algorithms. A number of network performance characteristics (e.g. routing information) are highly sensitive to the network positions of the target website and client device [2]; the attackers would have to consider the physical locations of the web servers as well as the client devices to mimic the exact network performance characteristics. This would be a very expensive attack to perform.

Experimental results: In November 2010, during the course of a week, we collected 10000 items of routing information in total: 5000 from 50 highly targeted websites (100 per website) representing the legitimate samples; and the other 5000 from 500 phishing websites (10 per website) representing the DNS-poisoning-based phishing samples. The initial dataset for phishing websites was obtained from a community website called PhishTank (<http://www.phishtank.com>). The PC used for the experiments was equipped with a dedicated, non-congested 100 Mbit/s Ethernet connection to a LAN that was connected to the Internet via a gigabit-speed link.

For detecting phishing websites, we considered the following four aspects of routing information: (i) the use of a firewall, (ii) the mean of the RTT values of all hops, (iii) the standard deviation of the RTT values of all hops, and (iv) the total route length in hops. Although most websites are distributed, we can observe statistically significant differences in these features between legitimate and phishing websites; for example, only 19% of the phishing websites were running a firewall while 72% of the legitimate websites were running a firewall. This is not surprising since a high proportion of phishing websites is located on free web hosting servers, which do not provide sufficient security features.

For graphical interpretation, we reduced the dimensions of the data points (which represent the four aspects of routing information) to three using the kernel principal component analysis (PCA) algorithm with a linear kernel. Fig. 1 is the three-dimensional scatter plot of this reduced dataset; it shows three distinct clusters: the leftmost cluster of only the phishing websites, the middle cluster of only the legitimate websites and the rightmost cluster of both the legitimate websites and phishing websites.

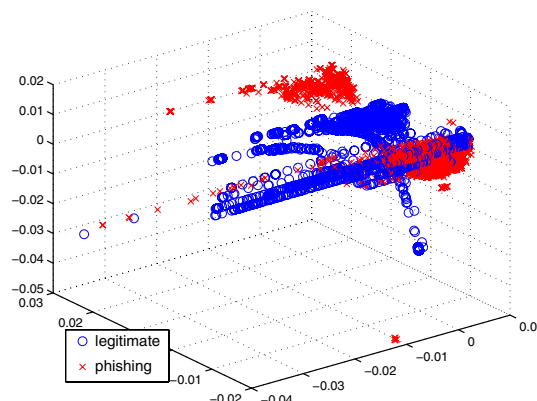


Fig. 1 Three-dimensional scatter plot of reduced dataset

During the setup phase, two-thirds of the data (6667 samples) was used to construct the classifiers and the rest (3333 samples) was reserved for out-of-sample testing. The classifiers were generated using this training dataset and the following four classification algorithms [3]: linear discriminant analysis (LDA), naïve Bayesian (NB), *K*-nearest neighbour (KNN), and support vector machine (SVM). The classifiers built by applying a classification method to the original and the reduced datasets are denoted as method(O) and method(R), respectively. NB(O), for example, represents the NB classifier with the original dataset.

Positives indicate legitimate websites and Negatives indicate phishing websites; true positive (*TP*), false positive (*FP*), true negative (*TN*) and false negative (*FN*) are defined as below:

- *TP* – legitimate websites correctly classified as legitimate websites;
- *FP* – legitimate websites incorrectly classified as phishing websites;
- *TN* – phishing websites correctly classified as phishing websites;
- *FN* – phishing websites incorrectly classified as legitimate websites.

Table 1 shows the performance of the classification algorithms using the following three measurements: accuracy ($(TP + TN)/P + N$), specificity ($TN/(TN + FP)$) and sensitivity ($TP/(TP + FN)$). To show the efficiency of the classification algorithms, we also measured the running time of the classifiers. These classifiers were implemented using the built-in MATLAB library functions. The PC we used for the experiments was equipped with an Intel quad-core 2.4 GHz CPU and 64-bit Windows operating system.

Table 1: Accuracy, specificity and sensitivity values for different classifiers

		Accu.	Spec.	Sens.	Train	Test
LDA	(O)	74.87%	79.79%	69.94%	0.000s	0.176s
	(R)	75.42%	79.88%	70.96%	0.000s	0.020s
NB	(O)	90.86%	95.19%	86.36%	1.166s	4.420s
	(R)	91.40%	96.51%	86.15%	0.037s	2.248s
KNN	(O)	99.34%	99.31%	99.37%	0.000s	1.448s
	(R)	96.43%	96.93%	95.94%	0.000s	0.800s
SVM	(O)	74.91%	79.79%	70.03%	68.901s	0.419s
	(R)	75.42%	79.88%	70.96%	10.803s	0.398s

The KNN classifier produced the best results: all the accuracy, specificity and sensitivity values are significantly higher compared to other classification methods. Since the performance of KNN is primarily determined by the choice of *K*, we tried to find the best *K* by varying it from 1 to 5; and found that KNN performs best when *K* = 1. In Fig. 1, it seems hard to correctly classify the rightmost cluster but we found that each data point can find the nearest neighbour that is in the

same class. The accuracy of KNN(O) is 99.34%, the specificity is 99.31%, and the sensitivity is 99.37%. This implies that it can accurately detect about 99% of the DNS-poisoning-based phishing attacks, while misclassifying less than 1% of the legitimate websites. This result outperforms existing heuristics [4, 5] that achieve true positive rates between 90.06 and 91.40%, and false positive rates between 1.95 and 5.98%. Considering that its running time of 1.448 s is also relatively fast, we strongly recommend using KNN(O). Unlike our expectations, SVM, which involves an expensive tuning phase, did not outperform other algorithms. Therefore, SVM is not our recommendation.

Another interesting observation is that all the classification methods except KNN worked better on the linearly reduced dataset than on the original dataset. This implies that KNN can only properly handle the nonlinear properties of the network performance characteristics.

Conclusions and future work: The proposed approach would incur a high cost to the DNS poisoning attackers since they would have to consider the network characteristics of original websites on top of the usual visual imitations. For performance evaluation, we measured a total of 10 000 routing information from 50 highly targeted websites and 500 of their phishing websites. The measured values were experimented with several classification methods: linear discriminant analysis, naïve Bayesian, K -nearest neighbour and support vector machine algorithms, demonstrating high accuracy in detecting phishing websites. The K -nearest neighbour classification algorithm produced the best results and is our recommendation: accuracy, specificity and sensitivity were all above 99%. The primary aim of this work was to demonstrate the feasibility of using network performance characteristics for detecting phishing websites. Some of the features (e.g. the RTT to each hop in the network route), which could significantly improve the performance, have not been considered in depth. Future work may look at further exploring these features to improve the performance of classification.

Also, we used a PC with a dedicated 100 Mbit/s Ethernet connection to a LAN that was connected to the Internet via a gigabit-speed link. Experimenting with different access networks such as ADSL or a cable model could be another interesting future investigation.

© The Institution of Engineering and Technology 2011
14 February 2011

doi: 10.1049/el.2011.0399

One or more of the Figures in this Letter are available in colour online.

H. Kim (*University of Cambridge – Computer Laboratory, Cambridge, CB3 0FD, United Kingdom*)

E-mail: hk331@cl.cam.ac.uk

J.H. Huh (*Information Trust Institute, University of Illinois at, Urbana – Champaign, USA*)

References

- 1 Abu-Nimeh, S., and Nair, S.: 'Circumventing security toolbars and phishing filters via rogue wireless access points', *Wirel. Commun. Mobile Comput.*, 2010, **10**, pp. 1128–1139
- 2 Padmanabhan, V.N., and Subramanian, L.: 'An investigation of geographic mapping techniques for internet hosts'. SIGCOMM '01: Proc. ACM 2001 Conf. on Applications, Technologies, Architectures, and Protocols for Computer Communications, New York, NY, USA, 2001, pp. 173–185
- 3 Duda, R.O., Hart, P.E., and Stork, D.G.: 'Pattern classification' (Wiley, 2001, 2nd edn)
- 4 Xiang, G., and Hong, J.I.: 'A hybrid phish detection approach by identity discovery and keywords retrieval'. WWW '09: Proc. 18th ACM Int. Conf. on World Wide Web, New York, NY, USA, 2009, pp. 571–580
- 5 Zhang, Y., Hong, J.I., and Cranor, L.F.: 'Cantina: a content-based approach to detecting phishing web sites'. WWW '07: Proc. 16th Int. Conf. on World Wide Web, New York, NY, USA, 2007, pp. 639–648