A. Weisbecker, M. Burmester & A. Schmidt (Hrsg.): Mensch und Computer 2015 Workshopband, Stuttgart: Oldenbourg Wissenschaftsverlag, 2015, S. 647-655.

Über die Wirksamkeit von Anti-Phishing-Training

Simon Stockhardt, Benjamin Reinheimer, Melanie Volkamer

SECUSO, Technische Universität Darmstadt

Zusammenfassung

Phishing ist noch immer ein verbreitetes Problem im Internet. Konsequenzen von Phishing können sowohl finanzieller als auch persönlicher Natur sein. Phishingangriffe werden ausgefeilter und sind nicht mehr einfach anhand fehlerhafter Rechtschreibung oder Grammatik zu identifizieren. Somit ist es für Internetnutzer wichtig den Aufbau von URLs zu verstehen um sich gegen Phishingangriffe schutzen zu können. Das von uns entwickelte „NoPhish“ Training basiert auf der Idee Nutzern sowohl die notwendige Awareness als auch die notwendigen Fähigkeiten zur Identifikation von Phishingangriffen zu vermitteln. Wir haben NoPhish mit einer Nutzerstudie empirisch evaluiert und können eine signifikante Verbesserung der Teilnehmer in diesen Bereichen zeigen.

1 Einleitung

Betrüger versuchen im Internet die Identitäten von Nutzern zu stehlen. Hierfür werden gefälschte Nachrichten versendet die legitim aussehen, dies aber nur vortäuschen. Der Zweck der Nachrichten ist dabei Nutzer auf eine echt aussehende Website zu locken. Gibt ein Nutzer auf dieser Webseite Daten ein, landen diese Daten direkt bei den Betrügern. Mit diesen Daten können die Betrüger sich dann auf den jeweiligen Webseiten einloggen und dort z.B. im Namen ihrer Opfer Bestellungen tätigen. Diese Art von Internetbetrug wird als „Phishing“ bezeichnet. Es gibt bereits diverse technische Lösungen die diese Problematik direkt adressieren. So werden phishing Nachrichten und Webseiten (URLs) direkt nach ihrer Identifikation auf Blacklisten gesetzt (Ma, Saul, Savage, & Voelker, 2009; Prakash, Kumar, Kompella, & Gupta, 2015; Marchal, Francois, State, & Engel, 2012; Ramzan, 2010). Gemeinhin werden Nutzer von ihren Webbrowsern bzw. von ihren Mailingdiensten gewarnt, sollten sie eine URL erhalten, die sich auf einer Blacklist beendet. Neben Blacklisten gibt es andere technische Lösungen wie z.B. Whitelisten und Spam-Filter. Diese technischen Lösungen sind jedoch nicht ausreichend um den Nutzer vor Phishing zu schützen. Es gibt stets eine bestimmte Zeitspanne in welcher eine Webseite online ist bevor sie von den

angesprochenen technischen Lösungen erkannt wird. In dieser Zeitspanne sind die Nutzer auf sich selbst gestellt und können nur durch eigene Fähigkeiten verhindern Opfer eines Phishingangriffes zu werden. Durch Unwissen darüber, wie sie einen solchen Angriff erkennen können und vor allem die fehlende Awareness darüber, dass solche Angriffe überhaupt existieren werden Warnungen vor Phishingangriffen (und Sicherheitswarnungen generell) oft ignoriert (Wu, Miller, & Garnkel, 2006; Akhawe & Felt, 2013). In (Canova, Volkamer, Bergmann, & Borza, 2014) wird NoPhish als eine komplementäre Herangehensweise zur Bewältigung dieses Problems vorgeschlagen. Der Kerngedanke von NoPhish ist, den Nutzer zu befähigen Phishingangriffe selbstständig zu erkennen. Neben dem Wissen darüber, dass die URL einer Webseite das wichtigste zu überprüfende Element einer Webseite ist, wenn es darum geht einen Angriff zu identifizieren, benötigt der Nutzer dafür noch das Wissen darüber wie man eine URL findet, aus welchen Teilen sich eine URL zusammensetzt sowie Kenntnisse über die Tricks der Betrüger um URLs zu fälschen. Der Fokus des vorliegenden Papers liegt auf einer Nutzerstudie, welche die Wirksamkeit des NoPhish Trainings evaluiert. Die Studie wurde im Rahmen eines Volkshochschulkurses durchgeführt.

2 Aufbau des „NoPhish“ Trainings

Das Design von NoPhish kombiniert Serious Game Elemente, psychologische Erkenntnisse im Bereich des Lernens, sowie ein user centered design. Als Serious Game Elemente wurden Levels und Quiz-struktur eingebaut. Aus der Psychologie wurden Lernmethoden wie Wiederholungen, Effekt, Primacy und Intensität implementiert. Das User-centered Design wurde durch wiederholte Iterationen mit Nutzerstudien sichergestellt. Nachdem nun die prinzipiellen Designentscheidungen ausgeführt wurden, sollen im Folgenden der Awareness- sowie der Übungsteil des Trainings dargestellt werden.

2.1 Awareness-Teil

Im Awareness-Teil des Trainings werden den Teilnehmern allgemeine Informationen über Phishing vermittelt. Auch wird die Relevanz der URL bei der Identifikation von Phishing Webseiten betont. Dies wird anhand verschiedener Beispiele direkt gezeigt.

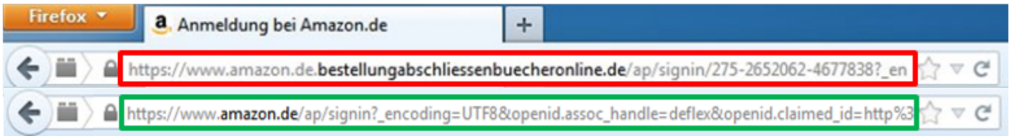

Abbildung 1: Beispiel für original und phishing Version der selben Webseite

2.2 Übungsteil

Der Übungsteil ist in neun Levels mit unterschiedlichen Methoden zur Fälschung von URLs eingeteilt (basierend auf Literatur (Sheng, et al., 2007; Aaron & Manning, 2014) und auf der Analyse von PhishTank URLs). Jedes Level besteht dabei aus zwei Teilen: einem einführendem Teil mit Informationen darüber, was in diesem Level neu ist, und dem Übungsteil. Die Übungen sind an ein Quiz angelehnt. Hierbei werden unterschiedliche URLs dargestellt und die Teilnehmer des Trainings werden jeweils nacheinander nach ihrer Einschätzung dazu, ob es sich dabei um eine Phishing Seite handelt oder nicht, gefragt. Im Folgenden sind die Inhalte der einzelnen Levels dargestellt.

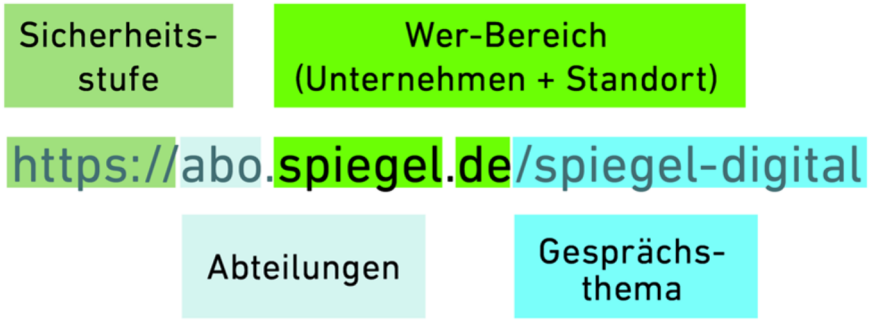

Abbildung 2: Erklärung der URL inkl. 'Wer-Bereich'

- Level 1: Aufbau der URL. In Level 1 wird den Teilnehmern der Aufbau einer URL erklärt.
- Da die Nutzung technischer Begriffe vermieden werden soll, wird dieser Teil der URL als „Wer-Bereich“ bezeichnet.
- Level 2: IP Adresse als Wer-Bereich
  Beispiel: https://33.58.198.82/
- Level 3: Unternehmensname kommt in der Webadresse nicht vor
  Beispiel: http://www.login.com/index.html
- Level 4: Ähnliche Namen im Bereich der Abteilungen
  Beispiel: http://csr.ebay.de.security.com/
- Level 5: Ähnliche Namen im Gesprächsthema-Bereich
  Beispiel: http://account-settings.de/facebook.com/
- Level 6: Ähnliche Namen im Wer-Bereich
  Beispiel: http://www.apple-support.com/

- Level 7: Tippfehler/Buchstabendreher im Wer-Bereich
  Beispiel: https://www.mircosoft.com/
- Level 8: Ähnliche Zeichen im Wer-Bereich
  Beispiel: https://www.vvetter.com/
- Level 9: Der Bereich der Sicherheitsstufe Hier geht es darum den Unterschied zwischen https und http zu vermitteln. Die Frage lautet anders als in den vorherigen Levels: Würden Sie auf dieser Seite sensible Daten eingeben?
  Beispiele: http://netbank.de bzw. https://netbank.de

3 Empirische Nutzerstudie zur Wirksamkeit von NoPhish

Im Folgenden wird auf die Motivation für die Durchführung der empirischen Studie, die Hypothesen der Studie, das verwendete Studiendesign, die Teilnehmer und die statische Auswertung eingegangen.

3.1 Motivation für eine Nutzerstudie

Für die Durchführung einer Studie ausserhalb des Labors sprechen verschiedene Gründe. Erstens gelten für eine Laborstudie bestimmte Einschränkungen wenn es um die Übertragbarkeit der Ergebnisse auf andere Situationen geht. Laborbedingungen ermöglichen eine optimale Kontrolle von Störvariablen. Dadurch entsteht jedoch auch eine Differenz zu realistischen Anwendungssituation. So kann externe Validität durch Studien, welche in einem von Störvariablen abgeriegelten Labor stattfinden, nur schwer nachgewiesen werden. Ein weiteres Problem von Laborstudien ist der Testeffekt. Studienteilnehmer werden durch die Laborsituation für die Testsituation sensibilisiert, was zu einer Verzerrung der Antworten führt.

Zusätzlich ist es unsere Motivation möglichst viele verschiedene Nutzergruppen mit NoPhish zu erreichen. Hiefür möchten wir NoPhish in verschiedenen Formaten testen. Mit dem Training hoffen wir auch ältere Internetnutzer und Nutzer ohne Smartphone zu erreichen. Das Studiendesign stellt sicher, dass jeder Teilnehmer den gleichen Input erhält. Auch sind jederzeit Rückfragen an die Trainings-Leiter möglich, falls bestimmte Dinge unklar sind.

3.2 Hypothesen

Die Hypothesen zielen hauptsächlich (3 von 4 Hypothesen) auf die detaillierte Messung der Verbesserung bezüglich der Wahrscheinlichkeit richtiger Entscheidung. Dabei gibt es verschiedene Aspekte die betrachtet werden sollen. Zum einen sollen beide Arten von Webseiten (Original & Phish) getrennt betrachtet werden. Zum anderen soll herausgefunden

werden, ob es einen Unterschied bezüglich alter (URLs bekannt aus Phase 1) oder neuer (noch nicht benutzer URLs) gibt.

Hypothese 1 | Phish: Es ist signifikant wahrscheinlicher, dass Teilnehmer nach dem Training Phishing-URLs erkennen.

Hypothese 2 | NoPhish: Nach dem Training hat sich die Erkennung von legitimen URLs signi_kant verbessert.

Hypothese 3 | Neue URL: Nach dem Training besteht ein signifikanter Unterschied bei der Erkennung zwischen neuen URLs (ausschließlich in Fragebogen Phase 2) und den bereits bekannten URLs (in Fragebogen Phase 1 enthalten).

Hypothese 4 | URL Verständnis: Nach dem Training stützen die Teilnehmer ihre Entscheidung signifikant häufiger auf die Domain.

3.3 Studiendesign

Für die Studie haben wir ein Innersubjekt-Design gewählt. Dazu wurde den Teilnehmern vor (Phase 1) und nach (Phase 2) des Trainings ein Fragebogen gereicht. Der Fragebogen in Phase 2 setzt sich aus den 15 Items aus Phase 1, sowie 7 weiteren neue Items zusammen. Die Aufteilung in legitime und phishing Screenshots folgte folgender Konzeption:

Phase 1: Fragebogen mit 15 Webseiten (7 legitim + 8 phishing).

Training: Auf NoPhish basierendes Training.

Phase 2: Fragebogen mit 22 Webseiten (11 legitim + 11 phishing).

(1) Einwilligungserklärung: Vor der Durchführung wurden die Teilnehmer gebeten eine Einwilligungserklärung zu unterschreiben.

(2) Webseiten Fragebogen Phase 1: In diesem Teil der Studie erhielt jeder Teilnehmer den Fragebogen Phase 1 mit 15 ausgedruckten Webseiten-Screenshots. Jede Seite beinhaltet den Screenshot einer Webseite und dazu fünf Fragen. Auf jedem Screenshots sind real existierende Seiten abgebildet. Bei acht Screenshots ist die URL durch eine Phishing Adresse ersetzt. Es ist sichergestellt, dass jeder Manipulationstyp zumindest einmal repräsentiert ist. Für jeden Screenshot werden die Teilnehmer gefragt, ob sie sich auf dieser einloggen würden. Zusätzlich sollen sie den Bereich des Screenshots markieren, der für die Entscheidung maßgeblich war. Außerdem sollen sie die Sicherheit ihrer Entscheidung auf einer 5-stufigen Likert-Skala einordnen. Abschließend wurden die Teilnehmer gefragt, ob sie die Marke kennen bzw. ob sie auf der Webseite einen Account besitzen.

(3) NoPhish Vorstellung: Hier werden die Inhalte von NoPhish, wie in 'Kapitel 2 - Inhalte von NoPhish' besprochen, vermittelt.

(4) Webseiten Fragebogen Phase 2: In diesem Abschnitt der Studie erhält jeder Teilnehmer den Fragebogen Phase 2 mit 22 ausgedrucktenWebseiten Screenshots. Diese setzen sich aus elf legitimen Screenshots, sowie elf Phishing Screenshots zusammen. Auch die weiteren

Aufgaben (Markierung eines Bereichs für die Entscheidung, Sicherheit der eigenen Entscheidung, Bekanntheit der Marke, Besitzen eines Accounts) gleichen dem des Phase 1 Fragebogens.

(5) Demographischer Fragebogen: Anhängend an den Fragebogen Phase 2 sollen die Teilnehmer noch fünf demografische Angaben machen: Jahrgang, Geschlecht, Farbenfehlsichtigkeit, der berufliche Abschluss sowie der Bereich in dem zurzeit gearbeitet/studiert wird.

(6) Aufbau der Fragebögen Phase 1 und Phase 2: Table 1 zeigt eine Auflistung aller genutzten URLs. Die Namen der Originale sind grün geschrieben. Phishing URLs in der Farbe Rot.

Marke	**URL**
Google	https://plus.google.com/u/0/me
Facebook	https://www.facebook.com.signing.com/Raumzeit
Amazon	http://www.amazon.de/Angebote/b/ref=cs_top_nav_gb27?ie=UTF8&r
Wikipedia	http://130.83.162.6/wiki/Wikipedia:Hauptseite
eBay	http://www.ebay.de/rpp/Deals/reisen-gutscheine/stadte-kultur/
Web.de	https://web.de.myponyfarm.com/
Focus	http://abo.net/www.focus.de/digital
GMX	http://www.gmx.net/produkte/mail/promail
Yahoo	http://de.yahoo.com/?p=us
Otto	https://www.0tto.de/damenmode/kategorien/anzuege-kostueme
Microsoft	http://windows.mircosoft.com/de-de/windows/products
Twitter	https://badcat.com/mobile.twitter.com/session/new
LinkedIn	https://touch.www.linkedin.com/login.html
Spiegel	http://m.spiegel.de/panorama/leute/a-937125.html#spRedire
Paypal	https://www.paypal-sicher.com/webapps/merchantboarding/web
	Zusätzlich in Phase 2:
GuteFrage	http://www.gutefrage.net.events-ma/tag/freizeit/1
T-Online	http://www.t-online.de/wetter/europawetter/64077226
Immobilienscout24	http://www.immobilienscout25.de/de/finden/wohnen/index.jsp
Wordpress	https://130.83.162.6/signup/
Bild	http://epaper.bild.de
Welt	http://www.welt.de/sonderthemen/mittelstand/forschung
Xing	https://blog.xing.com/category/german/

Tabelle 1: Legitime und manipulierte URLs.

3.4 Rekrutierung und Belohnung

Die Studienteilnehmer wurden aus dem Volkshochschulkurs rekrutiert und bekamen als Incentive für die Studienteilnahme die für den Volkshochschulkurs enstandenen Kosten zurückerstattet.

3.5 Demografische Daten

Insgesamt haben 12 Teilnehmer an unserer Studie teilgenommen. Das Durchschnittsalter beträgt 53.33 Jahre mit einer Standardabweichung von 9.178 und einem Bereich von 35 bis

66. In der analysierten Stichprobe befanden sich 8 Frauen und 4 Männer. Lediglich eine Person erklärt einen Bezug zur IT.

3.6 Hypothesentests

Um unsere Hypothesen zu analysieren wurden die Wahrscheinlichkeiten für richtige Antworten berechnet. Die Antwort wurde als korrekt gewertet, wenn auf legitime Webseiten „Ja“ als Antwort auf die Frage „Würden Sie sich auf dieser Webseite einloggen?“ gewählt wurde und „Nein“ auf Phishing-Webseiten. Außerdem wurden die Teilnehmer gebeten, die für ihre Entscheidung relevante Stelle auf dem Screenshot zu markieren. Jede Markierungen von mehreren Bereichen oder einem größeren Bereich als der Domain, wurde als URL gewertet. Die Markierung der gesamten Domain oder Bereiche der Domain (z.B. Position eines Fehler) wurde weiterhin als Domain gezäahlt. Für die Betrachtung der Hypothese 2, wurde Domain als richtig und die anderen Bereiche (z.B. Favicon, Schlosssymbol, URL oder Inhalt der Webseite) als falsch eingetragen.

Im Zuge der Hypothesen-Testung interessiert uns zunächst, ob unsere Intervention bei der Erkennung von Phishing und Originalen hilft. Die getesteten Variablen sind auf dem Intervall-Skalenniveau. Außerdem handelt es sich um gepaarte Daten, für deren Differenz die Hypothese der Normalverteilung getestet wurde. Hier erreicht keiner der Test Signifikanz, weshalb die Hypothese der Abweichung von der Normalverteilung verworfen wird (H1: Z =.636, p =.814; H2: Z =.358, p =1.0; H3: Z =.603, p =86; H4: Z =.451, p =.987). Der innersubjekt Faktor (unabhängige Variable) der Studie war die Phase (1 vs. 2) und die abhängige Variable waren die Wahrscheinlichkeiten für richtige Antworten in Bezug auf Phishing Webseiten, Originale Webseiten, neue alte Webseiten und den markierten Bereich. Anschließend wurde für jede Hypothese ein t-Test für abhängige Stichproben durchgeführt. Grafik 3 gibt einen Überblick über die deskriptiven Daten der Hypothesen und deren Analyse.

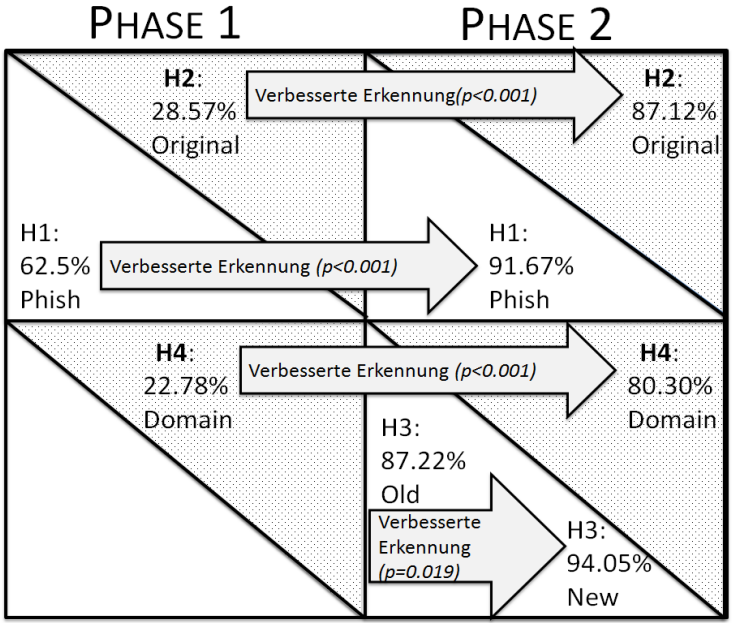

Abbildung 3: Wahrscheinlichkeiten für richtige Antworten

Hypothese 1 | Phish: Nach dem Training geben die Teilnehmer signifikant höher richtige Antworten in Bezug auf Phishing-URLs ($T = -5.625$; $p < 0.001$; $\eta^2 = 0.569$).

Hypothese 2 | NoPhish: Die Häufigkeit richtiger Antworten bezüglich legitimer Webseiten ist nach dem Training signifikant größer ($T = -7.262$; $p < 0.001$; $\eta^2 = 0.6683$).

Hypothese 3 | Neue URL: Die Häufigkeit der richtigen Antworten von neuen URLs ist signifikant höher ($T = -2.748$; $p = 0.019$; $\eta^2 = 0.09$).

Hypothese 4 | URL Verständnis: Die Domain wird signifikant höher als Grund für die Entscheidung ausgewählt ($T = -8.143$; $p < 0.001$; $\eta^2 = 0.638$).

4 Zusammenfassung und zukünftige Planung

Im Rahmen der vorliegenden Arbeit haben wir eine Methode zur Vermittlung des zum Schutz vor Phishing notwendigen Wissens erarbeitet. Die Identifikation von Phishinangriffen ist in ein Spiel eingebaut, das auf psychologischen Erkenntnissen, dem serious-game Ansatz sowie einem user-centered design basiert. Die Ergebnisse der Studie sind vielversprechend. Alle Hypothesen haben sich mindestens signifikant bestätigt. In Zukunft möchten wir die Studie mit einer höheren Teilnehmerzahl wiederholen. Auch möchten wir die Inhalte der Studie in verschiedenen Formaten testen. Ideen hierfür sind das hier vorgestellte Training mit einer existierenden Smartphone App (Canova, Volkamer, Bergmann, & Borza, 2014) sowie

einem Flyer mit den selben Inhalten zu testen.Parallel arbeiten wir daran den Inhalt des Trainings durch ein Webinterface zugänglich zu machen.

5 Literaturverzeichnis

Aaron, G., & Manning, R. (2014). *Phishing Activity Trends Report 2nd Quarter 2014*. Von http://docs.apwg.org/reports/apwg_trends_report_q2_2014.pdf abgerufen

Akhawe, D., & Felt, A. P. (2013). Alice in warningland. A large-scale field study of browser security warning effectiveness. *22nd USENIX Security Symposium* (S. 257-272). Washington, D.C.: USENIX.

Canova, G., Volkamer, M., Bergmann, C., & Borza, R. (2014). NoPhish: an anti-phishing education app. *In Security and Trust Management, vol. 8743*, S. 188-192.

Ma, J., Saul, L. K., Savage, S., & Voelker, G. M. (2009). Beyond blacklists: Learning to detect malicious web sites from suspicious urls. *15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (S. 1245-1254). New York: ACM.

Marchal, S., Francois, J., State, R., & Engel, T. (2012). Proactive discovery of phishing related domain names. In D. Balzarotti, S. Stolfo, & M. Cova, *Research in Attacks, Intrusions, and Defenses* (S. 190-209). Heidelberg: Springer Berlin.

Prakash, P., Kumar, M., Kompella, R., & Gupta, M. (2015). Phishnet: Predictive blacklisting to detect phishing attacks. *IEEE INFOCOM 2010* (S. 1-5). San Diego: IEEE.

Ramzan, Z. (2010). Phishing attacks and countermeasures. In P. Stavroulakis, & M. Stamp, *Handbook of Information and Communication Security* (S. 433-448). Heidelberg: Springer Berlin.

Sheng, S., Magnien, B., Kumaraguru, P., Acquisti, A., Cranor, L. F., Hong, J., & Nunge, E. (2007). Anti-phishing phil: The design and evaluation of a game that teaches. *3rd Symposium on Usable Privacy* (S. 88-99). New York: ACM.

Wu, R., Miller, R. C., & Garnkel, S. L. (2006). Do security toolbars acutally prevent phishing attacks? *SIGCHI Conference on Human Factors in Computing Systems* (S. 601-610). New York: ACM.

Kontaktinformationen

Simon Stockhardt, simon.stockhardt@secuso.org
Benjamin Reinheimer, benjamin.reinheimer@secuso.org
Melanie Volkamer, melanie.volkamer@secuso.org