Evaluation der interaktiven NoPhish Präsenzschulung

Benjamin Berens
Karlsruhe Institute of Technology
Karlsruhe, Karlsruhe
Benjamin.Berens@kit.edu

Lukas Aldag
Karlsruhe Institute of Technology
Karlsruhe, Germany
Lukas.Aldag@kit.edu

Melanie Volkamer
Karlsruhe Institute of Technology
Karlsruhe, Germany
Melanie.Volkamer@kit.edu

ABSTRACT

Phishing Angriffe stellen nach wie vor eine große Bedrohung für Privatpersonen und Unternehmen, Vereine und öffentliche Einrichtungen dar. Es gibt bereits viel Forschung zur Effektivität von Security Awareness-Maßnahmen und insbesondere im Kontext von Phishing Angriffen. Die meisten Paper messen den Effekt unmittelbar nach der Durchführung der Maßnahme. Nur wenige Paper untersuchen, wie lange der Effekt hält, sprich wann eine Auffrischungsmaßnahme durchgeführt werden sollte. Ziel dieses Papers ist es, zu bestätigen, dass der Effekt der NoPhish Präsenzschulung vier Monate anhält. Hierzu wurde eine entsprechende Studie im Rahmen einer freiwilligen Präsenzschulung mit 19 Teilnehmener/innen bei der Polizei durchgeführt und konnte die bisherigen Ergebnisse nicht bestätigen.

1 EINLEITUNG

Menschen benutzen sowohl im geschäftlichen als auch im privaten Bereich zunehmend elektronische Kommunikation wie E-Mail oder Messenger. Kriminelle machen sich dies zu Nutzen und verschicken Phishing-Nachrichten, um an sensible Informationen zu gelangen oder Schadsoftware zu verteilen. Laut Bundeskriminalamt (BKA) umfasste der entstandene Schaden alleine durch gemeldete Computerbetrugsfälle (u.a. Phishing) im Jahr 2019 88,7 Millionen Euro, was einem Anstieg von 44,4% zum Vorjahr entspricht. Trotz stetiger Weiterentwicklung technischer Schutzmaßnahmen, werden Phishing-Nachrichten — u.a. mit gefährlichen Link oder Anhang – immer wieder zugestellt. Somit müssen die Empfänger/Innen diese selbst erkennen. Forschungsergebnisse zeigen, dass damit noch Probleme bestehen [19]. Unternehmen haben diese Problematik erkannt und damit begonnen, ihre Mitarbeitenden mit Hilfe von Awareness-Maßnahmen fortzubilden, zum Beispiel anhand von Flyern, Postern, Infokarten, Videos und Präsenzveranstaltungen.

Es gibt bereits Studien, die einen positiven Effekt von Awareness-Maßnahmen hinsichtlich der Unterscheidung betrügerischer und legitimer Nachrichten nachweisen [11, 20, 24] und diesen über einen längeren Zeitraum nachweisen [7, 10, 25]. Andere Ergebnisse zeigten, dass eine Schulung von Kindern nicht zur gewünschten Langzeitwirkung führt [12]. Die Studien mit dem Ziel, die Dauer der Effektivität zu messen, wurden meist über einen kürzeren Zeitraum von maximal einem Monat durchgeführt. Eine Studie über einen längeren Zeitraum wurde von Reinheimer et al. [19] durchgeführt, in der der Effekt einer Awareness-Maßnahme nach vier, sechs und zwölf Monaten evaluiert wurde.

Ziel dieses Papers ist es, zu bestätigen, dass der Effekt von vier Monaten auch in einem weiteren Kontext nachzuweisen ist. Hierzu wurde eine entsprechende Studie im Rahmen einer freiwilligen Präsenzschulung in Kooperation mit dem Polizeipräsidium Südhessen für die dortigen Angestellten durchgeführt. Hierfür verwendeten wir die „NoPhish“ Schulungsunterlagen, die in Form von Unterlagen zum Selbststudium bereits auf ihre Effektivität - unmittelbar nach dem Studium der Unterlagen - getestet wurde [10]. Die Inhalte wurden so umgewandelt, dass sie als Foliensatz für einen interaktiven Vortrag im Rahmen eines Workshops verwendet werden konnten.

Wir haben die Effektivität der in dieser Form geänderten Unterlagen unmittelbar nach dem Workshop und nach einem Zeitraum von vier Monaten evaluiert. Dafür wurde den Teilnehmer/Innen zu drei unterschiedlichen Zeitpunkten ein Quiz vorgelegt, in welchem sie verschiedene E-Mails beurteilen sollten, ob es sich um eine Phishing E-Mail handelt oder nicht. Diese Form der Evaluierung wurde bereits ähnlich in anderen Benutzerstudien angewendet [9, 13, 19, 23].

Zusammenfassend haben wir folgende wissenschaftliche Beträge geleistet: (1) Evaluation einer Präsenzschulung auf Basis des NoPhish Konzepts direkt nach der Schulung und vier Monate später mit dem Ergebnis, dass die Erkenntnis aus [19] bzgl. einer Auffrischung erst nach sechs Monaten nur indirekt bestätigt werden kann. (2) Anders als in vorherigen Studien, wurden die spezifischen Phishingangriffe aus dem Quiz individuell betrachtet und analysiert. So konnten Phishingangriffe identifiziert werden, die für Teilnehmer/Innen trotz der Schulung noch Probleme darstellten. Diese Ergebnisse geben Aufschluss zu weiteren Verbesserungsansätzen für die Awareness-Maßnahme selbst.

2 HINTERGRUND

Die NoPhish Schulung von Neuman et al. [18] beschäftigt sich mit verschiedene Arten von betrügerischen Nachrichten – oft auch Phishing-Nachrichten genannt. Neumann et al. unterscheiden drei Formen von betrügerischen Nachrichten: (1) unplausible betrügerische Nachrichten, (2) plausible betrügerische Nachrichten mit gefährlichen Links und (3) plausible betrügerische Nachrichten mit gefährlichen Anhängen. Die Inhalte der NoPhish Schulung von Neumann et al. basieren wiederum auf den Inhalten und Erkenntnissen aus anderen Papern, die Phishing-Awareness-Maßnahmen vorgeschlagen haben (u.a. Reinheimer et al. [19] und Stockhardt [22]). Die NoPhish Schulung von Neumann et al. [18] startet mit einer allgemeinen Einführung zu Phishing und möglicher Konsequenzen, wenn man Opfer eines Phishing Angriffs wird. Damit sollen die Teilnehmer/Innen motiviert und für die Wichtigkeit des Themas sensibilisiert werden.

Mensch und Computer 2021, Workshopband, Workshop on 7. Usable Security und Privacy Workshop

https://doi.org/10.18420/muc2021-mci-ws14-394

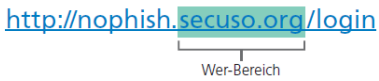

Figure 1: URL mit markiertem Wer-Bereich aus den NoPhish Materialien von [17].

Im Anschluss daran werden die drei Phishing-Kategorien vermittelt. Diese Kategorien werden dabei noch in thematisch passende Abschnitte unterteilt. Jeder Abschnitt beginnt mit einer Erklärung des zugehörigen Phishing Tricks und wie dieser erkannt werden kann. Anschließend wird mit Beispielen das Erkennen der jeweiligen Phishing-Tricks geübt. Im Folgenden wird zusammen gefasst, welche Inhalte für jede der drei Kategorien enthalten sind:

(1) Unplausible betrügerische Nachrichten sind solche Nachrichten, die durch den unplausiblen Absender oder nicht plausible Inhalte (z.B. das Erfragen von Zugangsdaten zum Online-Banking durch das eigene Bankinstitut) erkannt werden können. Es wird darauf hingewiesen, dass die Aufforderung häufig mit psychologischen Tricks unterstrichen wird (z.B. weil Druck aufgebaut wird).
(2) Plausible betrügerische Nachrichten mit gefährlichen Links sind solche Nachrichten, die nur durch das Prüfen der URL hinter dem Link als solche erkannt werden können. Es wird vermittelt, welche Arten der Darstellung von URLs existieren (Tooltip und Statusleiste). Darauf folgt die Erklärung zum Aufbau einer URL (siehe Bild 1). Hier wird beschrieben, dass der Wer-Bereich (Domain) der wichtigste Bereich für die Erkennung von Phishing-URLs ist. Anschließen werden die folgenden URL basierten Phishing Tricks erklärt:
   (a) Der zu erwartende Wer-Bereich wird in der Subdomain (www.google.de.host547.com) oder dem Pfad (www.host547.com/google.de) verwendet. Es wird erklärt, dass diese Angriffsform häufig kombiniert wird mit vertrauenswürdig klingende Namen im Wer-Bereich benutzt.
   (b) Die URL weicht im Wer-Bereich leicht von dem legitimen Wer-Bereich ab. Dabei werden z.B. Buchstaben durch andere oder ähnlich aussehende Buchstaben ersetzt (www.arnazon.de), die Reihenfolge der Buchstaben (www.amaozn.de) wird verändert oder es werden Buchstaben hinzugefügt oder entfernt (www.googlle.de).
   (c) Zum legitimen Wer-Bereich werden noch vertrauenswürdige Wörter ergänzt (amazon-shop.de). Diese Tricks kann man nur erkennen, wenn man den Wer-Bereich des suggerierten Ziels tatsächlich kennt.
(3) Plausible betrügerische Nachrichten mit gefährlichen Anhängen sind solche mit Anhängen, bei denen gefährliche Dateiformate verwendet werden (.exe) oder Tricks angewandt werden z.B.eine doppelte Endung (.pdf.exe).

Außerdem werden noch Spezialfälle erklärt: Dass Phisher die URL im Text wie die legitime URL aussehen lassen aber dahinter dann die Phishing-URL versteckt ist (sogenannter Mismatch-Fall).

Da diese Materialien von [17] nur in Form von Unterlagen zum Selbststudium vorliegen, wurden die Inhalte für eine interaktive Workshop-Präsentation aufbereitet. Die interaktiven Elemente bestanden darin, dass per Handzeichen darüber abgestimmt werden konnte, ob eine URL bzw. ein Anhang bzw. eine ganze Nachricht legitim oder gefährlich ist.

3 METHODIK

Das Hauptziel dieses Papers ist es, die interaktive Präsenzschulung zu evaluieren. Die Teilnehmer/innen waren direkt nach der Schulung bei Neumann et al. [18] signifikant besser darin betrügerische Nachrichten zu erkennen. Ähnlich zu den Ergebnissen von Reinheimer et al. [19] zu einer NoPhish basierten Präsenzschulung mit Fokus auf Beispiele mit gefährlichen Links, in welcher die Teilnehmer/innen nach vier Monaten immer noch signifikant besser waren, erwarten wir, dass dies für die interaktive NoPhish Präsenzschulung, die bei der Polizei gehalten wurden, ebenso der Fall ist. Infolge dessen stellen wir folgende Hypothese auf: Die Teilnehmer/innen sind direkt und vier Monate nach dem Erhalt der Schulung signifikant besser darin legitime von betrügerischen Beispielen zu unterscheiden.

3.1 Rekrutierung und Ethik

Die Teilnehmer/innen wurden mit Hilfe des Fachberaters zu Thema Cybersicherheit der Polizei Südhessen rekrutiert. Per E-Mail wurden die Beamten/Innen über das Angebot informiert, d.h. dass im Rahmen der Kooperation die Schulung angeboten wird, die Teil einer Benutzerstudie ist. Sowohl die Schulung als auch die Teilnahme an der Erhebung nach vier Monaten war Teil der Arbeitszeit und sonst nicht vergütet. Die vor Ort beantworteten Quizze wurden auf Papier ausgefüllt und dann digitalisiert. Keine personenbezogenen Daten wurden erhoben.Das Retention-Quiz wurde online zur Verfügung gestellt über SoSci Survey[1]. Das Unternehmen ist konform mit der Datenschutz-Grundverordnung (DSGVO) und erhielt ebenfalls keine personenbezogenen Daten. Die Teilnehmer/innen wurden darüber informiert, dass sie jederzeit ihre Einwilligung zur Teilnahme ohne Angabe von Gründen und ohne negative Konsequenzen zurückziehen konnten. Dazu erhielten Sie weitere Kontaktinformationen für Fragen.

3.2 Evaluation mittels Quiz

Für die Messung der Erkennungsrate von betrügerischen bzw. legitimen Nachrichten wurde auf eine Art Quiz zurückgegriffen. In diesem Quiz sahen die Teilnehmer/innen verschiedene Screenshots. Auf die genaue Zusammensetzung der Screenshots gehen wir in Kapitel 3.3 ein. Pro Seite bekamen sie einen Screenshot und sollten entscheiden, ob es sich hierbei um eine Phishing-Nachricht handelt oder nicht. Die Screenshots stellten dabei statische E-Mails dar, bei denen die Maus bereits über dem Link platziert war und somit die URL zu sehen war (ähnlich zu [18, 19]). Während der Erhebung auf Papier gab es die Möglichkeit zwischen allen Screenshots zu springen. In der Erhebung nach vier Monaten war das nicht möglich.

3.3 Auswahl der Phishing Tricks für das Quiz

Grundsätzlich sollten alle Tricks aus der Schulung (siehe Kapitel 2) abgedeckt werden – analog zu dem Vorgehen in [18, 19]. Im

[1] https://www.soscisurvey.de/

Folgenden werden alle verwendeten Beispiele präsentiert und die Namen der Beispiele werden in Klammer dargestellt, die im weiteren Verlauf des Textes zur Vereinfachung verwendet werden.

Es gab *zwei Phishing E-Mails vom Typ unplausible betrügerische Nachricht*: Die erste unplausiblen betrügerische Nachricht war eine, bei der die Absender-Adresse unplausibel war (Absender). Bei der zweiten unplausiblen betrügerischen Nachricht war der Inhalt nicht legitim in dem Sinne (Nonsense), dass nach sehr vielen persönlichen Daten gefragt wurde (z.B. Name, Adresse und Bankverbindung).

- (1) Absender Phish: amazon@host547.ru - Legitim: versandbestaetigung@amazon.de
- (1) Nonsense Phish: Kundendaten Abfrage - Legitim: Vertragsbenachrichtigung ohne Abfrage

Bei den *plausiblen betrügerischen Nachrichten mit gefährlichen Links wurden sieben E-Mails* verwendet:

- (2) Random-URL: www.hisoliajio.host547.com... - Legitim: www.lufthansa.de...
- (2a) Subdomain Phish: www.amazon.de.kolwerg.com... - Legitim: www.amazon.de...
- (2a) Path Phish: www.qpglljhjotqgg.com/vodafone.com... - Legitim: www.vodafone.com...
- (2b) Substitute Phish: clhl.de... - Legitim: dhl.de...
- (2b) Typo Phish: www.luftthansa.de... - Legitim: www.lufthansa.de...
- (2b) Swap Phish: www.papyal.de... - Legitim: www.paypal.de
- (2c) Erweiterung Phish: www.amazon-shopping.de... - Legitim: www.amazon.de...

Außerdem wurde noch ein Spezialfall der plausiblen betrügerischen Nachrichten mit gefährlichen Links ergänzt:

- (Special) Missmatch Phish: www.host.547.com... - Legitim: www.dhl.de...

Insgesamt drei Beispiele entfielen auf die Kategorie 3 *plausible betrügerische Nachrichten mit gefährlichen Anhängen*:

- (3) Anhang EXE Phish: Rechnung.exe - Legitim: Rechnung.pdf
- (3) Anhang DOC Phish: Newsletter Anmeldung.doc - Legitim:Newsletter Anmeldung.pdf
- (3) Anhang PDF.EXE Phish: Informationen.pdf.exe - Legitim: Informationen.pdf

3.4 Erstellung der Screenshots für das Quiz

Analog zu dem Vorgehen in [18, 19] war das Ziel genauso viele legitime wie Phishing-Nachrichten zu verwenden und der Fokus war auf E-Mail Nachrichten.

Als Basis wurden E-Mails verwendet, die genau so von den Institutionen verschickt werden. Zunächst wurden diese alle auf einen Namen geändert, sprich, dass die E-Mail sich immer an die gleiche Privat-Person richtet. Für die Phishing E-Mails wurden diese legitimen E-Mails dann ausschließlich bezogen auf die Umsetzung der einzelnen Tricks verändert. Entsprechend sind z.B. die Beispiele für die legitimen Fälle mit Links identisch zu den Phishing Beispielen, außer, dass die URLs hinter dem Link sich unterscheiden. Es sei

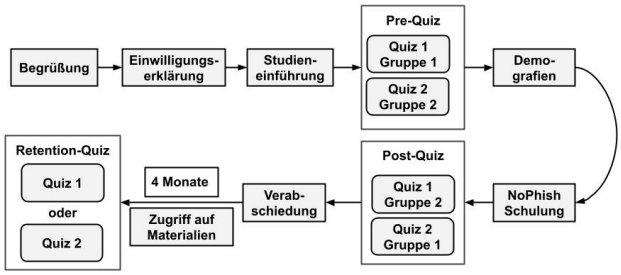

Figure 2: Darstellung des Studienablaufs.

darauf hingewiesen, dass Beispiele für legitime Fälle mit Anhängen ausschließlich Anhänge mit dem Format ".pdf" sind.

Anders als in früheren Studien (z.B. als in [18, 19]), wurden für alle legitimen und Phishing E-Mails aus der Kategorie 1 und 2 (gefährliche Links) zwei Screenshots erstellt: Einmal im Design von Thunderbird unter Windows und einmal mit der macOS Mail Umgebung. Diese beiden Umgebungen wurden als Repräsentanten für die Darstellung mit Statusleiste (Thunderbird) und Tooltip (Mail) gewählt, welche insbesondere für Kategorie 2 relevant ist. Dadurch folgen:

- Für Kategorie 1: 8 Screenshots
- Für Kategorie 2: 32 Screenshots
- Für Kategorie 3: 6 Screenshots

In Summe wurden also 46 Screenshots erzeugt. Um die Anzahl der Beispiele für die Teilnehmer nicht zu groß werden zu lassen, wurde entschieden, das Quiz in zwei Versionen zu unterteilen. Jede Version erhielt dabei alle Beispiele aus Kategorie 1 und 2 allerdings entweder in der Windows Thunderbird oder in der macOS Mail Umgebung. Jede Version des Quiz enthält damit einen Mix aus beiden Darstellungsformen[2]. Jede Variante enthält drei Screenshots aus der Kategorie 3[3]. Insgesamt waren zu jedem Zeitpunkt der Messung und in jeder Gruppe 23 E-Mails im Quiz enthalten.

3.5 Studienablauf

Die Studie wurde in Kooperation mit der Polizeidienststelle Südhessen durchgeführt und war in das Weiterbildungsangebot der Dienststelle als Präsenzschulung integriert. Die Schulung wurde an zwei Tagen angeboten. In Abbildung 2 ist der gesamte Studienablauf dargestellt und im Folgenden werden die einzelnen Aspekte genauer beschrieben.

Begrüßung. Die Schulung begann damit, dass die Teilnehmer/innen zuerst von dem Fachberater zu Thema Cybersicherheit und dem Leiter der Dienststelle begrüßt wurden und eine kurze Einführung in die aktuelle Cyber-Bedrohungslage. Nachfolgend wurde durch die Organisatoren der Dienststelle ein Überblick über die Schulung

[2]Hinweis: Jeder Teilnehmer/in erhielt über Pre- und Post-Quiz zusammen beide Versionen. Lediglich der Zeitpunkt wurde zufällig gewählt. Wobei darauf geachtet wurde, dass etwa gleich viele Versionen eins und zwei von den Teilnehmer/Innen zu jedem Zeitpunkt ausgefüllt wurden.

[3]Die Anzahl der legitimen und Phishing-Nachrichten ist dadurch nicht genau identisch sondern eine Variante hat eine Legitime Nachricht mehr und die andere eine Phishing-Nachricht.

und die integrierte Studie gegeben. Zudem wurden erste Fragen zum Ablauf der Schulung und Studie beantwortet.

Einwilligungserklärung. Nach der Begrüßung durch die Studienleitung wurden die Teilnehmer/innen über die freiwillige Teilnahme an der Schulung, sowie der Studie aufgeklärt. Zudem erhielten alle Teilnehmer/innen eine schriftliche Einwilligungserklärung, in welchem alle wichtigen Punkte zu der Studie aufgeführt wurden. Die Teilnehmer/innen erhielten erneut Zeit Fragen zu der Studie zu stellen, falls Bedenken aufgekommen sein sollten.

Studieneinführung. Das Quiz fängt mit einer kurzen Einleitung an, in welcher erklärt wurde, dass im folgenden Beispiel-Screenshots von E-Mails gezeigt werden.Die Bewertung erfolgte indem die Teilnehmer/innen eines von zwei möglichen Feldern ankreuzten mit „Ja, betrügerisch." oder „Nein, nicht betrügerisch.". Nach der Erklärung der Aufgabe wurde den Teilnehmer/innen ein Szenario präsentiert, das wie folgt zu lesen war:

Die E-Mails gehen alle an Martin Müller. Stellen Sie sich vor, dass Herr Müller Ihnen die E-Mail zeigt und Sie persönlich um Hilfe bittet, ob es sich hierbei um eine vertrauenswürdige E-Mail handelt. Bitte beachten Sie: Herr Müller fragt Sie, weil die E-Mail für ihn plausibel ist, weil er z. B. bei dem Dienst ein Benutzerkonto hat und kürzlich damit interagiert hat (z. B. Bestellungen getätigt hat). Er fragt Sie, weil Ihn auch das Gefühl das etwas nicht stimmen könnte.

Nach dem Lesen des Szenario wurden die Teilnehmer/Innen gebeten einen individuellen Code zu erzeugen. Dieser sollte garantieren, dass die Angaben anonym gemacht werden können, jedoch über den Verlauf der Studie, Pre, Post und Retention zueinander geordnet werden kann.

Pre-Quiz. Nach Erstellung des Codes, sollten die Teilnehmer/innen 23 Beispiel E-Mails bewerten. Das Pre-Quiz ist auf Papier vorab schon an die Plätze der Teilnehmer/Innen verteilt worden.

Demografie. Die Teilnehmer wurden nach Alter, Geschlecht und ihren Phishing/Security Vorkenntnissen gefragt.

NoPhish Schulung. Anschließend an dem Pre-Quiz wurde die eigentliche Schulung durchgeführt. Jedes Thema der Schulung wurde auf vergleichbare Weise vorgetragen, erst gab es einführende Informationen, die mit einer Übung und Fragerunde abgeschlossen wurden. So konnten die Teilnehmer/innen das erworbene Wissen direkt anwenden.

Post-Quiz. Im Anschluss wurden die Teilnehmer/Innen gebeten das Post-Quiz auszufüllen. Hierzu wurde ihnen dies auf Papier verteilt. Um eine mögliche Verwechslung der Quizze auszuschließen, wurden die Versionen getrennt voneinander aufbewahrt. Auf der ersten Seite wurde erneut nach dem Code gefragt. Die Teilnehmer und Teilnehmerinnen erhielten im Post-Quiz immer das jeweilige Gegenstück zu ihrer Pre-Quiz Version.

Verabschiedung. Mit Beendigung des Post-Quiz wurden die Teilnehmer/innen verabschiedet. Den Teilnehmer/innen wurde für die Teilnahme gedankt und falls weitere Fragen bestehen, würden diese noch beantwortet werden. Es wurde auf die Erhebung nach vier Monaten hingewiesen.

Zugang zu Materialien: Damit die Teilnehmer/innen bei Interesse oder Fragen zum Thema Phishing ausreichend Informationen erhalten können, wurde ihnen Zugang auf den internen Servern zu den NoPhish Materialien gegeben. Diese enthalten die gleichen Informationen, die in der Schulung präsentiert wurden.

Retention Quiz. Die Teilnehmer/innen von der Schulung erhielten vier Monate nach Erhalt der Schulung eine Nachricht mit der Bitte erneut an dem Quiz teilzunehmen. Mit Bekanntgabe des Retention-Quiz, wurde der Zugang zu den Materialien verwehrt. Dadurch ist es für uns möglich zu kontrollieren, dass die Teilnehmer/innen nicht auf Grund der Retention Studie erneut die Materialien anschauen. Im Gegensatz zu dem Pre- und Post-Quiz erfolgte das Retention-Quiz über das Internet, da zu gegebener Zeit durch das Corona-Virus keine persönlichen Treffen möglich waren. Da die Teilnahme an dem Retention-Quiz ebenfalls freiwillig war, rechneten wir mit einer etwas geringeren Teilnahme. Deshalb wurde entschieden die Teilnehmer/innen nicht den gleichen Gruppen der Schulung zuzuweisen, sondern zufällig eine der beiden Versionen des Quiz zuzulosen.

4 ERGEBNISSE

4.1 Teilnehmer und Datenbereinigung

Von den insgesamt 39 Teilnehmer/Innen aus den beiden Veranstaltungen konnten 19 Datensätze für alle drei Zeitpunkte verwendet werden. Zum einen haben nicht alle das Retention-Quiz ausgefüllt, zum anderen haben wir fünf ausgeschlossen, weil sie bereits in der Vergangenheit an einer ähnlichen Schulung teilgenommen hatten. Außerdem konnten nur die Teilnehmer/innen betrachtet werden die anhand ihres Versuchspersonen Codes eindeutig zugeordnet werden konnten. Die Teilnehmer/Innen waren im Durchschnitt 39 Jahre alt. Sieben der Teilnehmer/Innen waren weiblich, neun männlich und drei haben keine Angabe gemacht.

4.2 Hypothesen Test

Die Fähigkeit zur Differenzierung zwischen betrügerischen und legitimen Beispielen wurde mit der Signal Detection Theory (SDT) [21] operationalisiert, welche bereits vielfach in der Forschung zu Phishing-Nachrichten verwendet worden ist [1–6, 8, 14–17, 20].

Hierbei werden die folgenden Kennzahlen berechnet: Treffer, falscher Alarm, Auslassung und korrekte Ablehnung. Diese Kennwerte werden im Rahmen der SDT in die Werte Sensitivität und Kriterium überführt. Sensitivität d' beschreibt dabei die Fähigkeit zwischen dem Signal (Phish) und dem Rauschen (Legitim) zu trennen. Ein großes d' bedeutet dabei eine bessere Fähigkeit zur Trennung der beiden. Das Kriterium beschreibt dazu die Antwort Tendenz, also ob jemand besonders vorsichtig ist und somit eine Tendenz eher Phish zu antworten hat oder ob jemand viel Risiko eingeht und eher ein Beispiel als legitim bewertet. Hier ist ein Wert von Null das Ziel, da hier Neutralität besteht sprich keine Tendenz in einer der beiden Richtungen.

Zur Evaluation der Hypothese aus Kapitel 3 wurde eine ANOVA mit Messwiederholung durchgeführt. Bevor wir die eigentlich Analyse mit der ANOVA Rechnung starten konnten, mussten zunächst

die Voraussetzungen z.B. Abhängigkeit der Messungen, Intervallskalierung abhängige Variablen, Nominalskalierung Innersubjektfaktoren, Normalverteilung abhängige Variable und Ausreißer Kontrolle. Die Sphärizität wurde im Rahmen der Berechnung in R direkt mit der Greenhouse-Geisser Korrektur adressiert.

4.3 Ergebnisse

Folgende Hypothese stand im Zentrum der Analyse:
H-Komplett: Die Teilnehmer/innen sind direkt und vier Monate nach dem Erhalt der Schulung signifikant besser darin legitime von betrügerischen Beispielen zu unterscheiden.

Die Erkennungsrate für alle legitimen und betrügerischen Beispielen, war über alle Zeitpunkte hinweg signifikant unterschiedlich, F(2,18) = 9.589, $p = 0.002$, eta^2[g] = 0.283. Vor der Schulung (Pre) ist die Sensitivität $d' = 0.76$, nach der Schulung (Post)$d' = 1.46$ und vier Monate danach (Retention) $d' = 1.26$. Anschließende Paarvergleiche (Bonferrini korrigiert) der einzelnen Zeitpunkte ergaben: Die Erkennungsrate zum Post Zeitpunkt war signifikant höher als zum Pre Zeitpunkt ($p < 0.001$), dagegen bestand kein signifikanter Unterschied zwischen Retention und Post ($p = 0.675$) oder Pre ($p = 0.079$). Somit lässt sich die Hypothese H-Komplett nicht bestätigen.

Vergleich	N	Differenz	korrigiertes p
Pre - Post	19	-6.26	<0.001*
Pre - Retention	19	-2.42	0.079
Post - Retention	19	1.26	0.675

Table 1: ANOVA mit Messwiederholung für alle Daten über alle drei Messzeiptunke. Der P-Wert wurde mit Bonferroni korrigiert. Stern (*) bedeutet ein signifikanter Unterschied.

4.4 Einzelne Phishing Tricks

Es gibt sehr unterschiedliche Entwicklungen der Erkennungsraten (Anzahl der richtigen Antworten / Anzahl der Beispiele in Prozent) abhängig vom Phishing Trick (siehe Table 2). Zunächst gehen wir auf die betrügerischen Beispiele ein. Am besten schneidet das Beispiele (2) einer zufälligen URL mit einem Buttom in der E-Mail (100%, 100% und 100%) ab. Ähnlich gut haben zwei weitere Beispiele funktioniert: Das Beispiel (2a) mit dem zu erwartenden Wer-Bereich im Pfad (95%, 100% und 100%) sowie (1) die *unplausible Nachricht* mit einem falschen Absender (100%, 100% und 95%). Vier Beispiele haben sich in unterschiedlichen Maß über die drei Zeitpunkte hinweg verbessert: (2a) Subdomain (47%, 84% und 95%), (2b) Substitute (79%, 89% und 100%), (2) Typo (74%, 89% und 84%) und (3) Anhang PDF.EXE (85%, 100% und 100%). Bei allen diesen Beispielen ist eine positive Entwicklung gegenüber vor der Schulung festzustellen und nur bei einem Beispiel ein leichter Abstieg zu Retention. Vier Beispiele zeigen schwankende Ergebnisse d.h. sie verbessern sich zunächst (Post) und fallen aber wieder ab (wenn auch nicht auf das Niveau von davor): (1) Nonsense (84%, 95% und 68%), Erweiterung (2c)(63%, 95% und 84%), (Spezial) Missmatch (89%, 100% und 89%) und (3) Anhang EXE (67%,100% und 80%). Davon ist lediglich Nonsense mit 68% für Retention niedriger als der Ausgangswert. Bei den betrügerischen Beispielen gibt es zwei, die entweder schlechter werden oder in einem nicht zufriedenstellenden Bereich sich verbessern: (2b) Swap (32%, 63% und 47%) und (3) Anhang DOC (100%, 62% und 50%). Swap endet also zum Retention Zeitpunkt im Rate-Bereich (um die 50% bei zwei Antwortmöglichkeiten) und Anhang Doc verschlechtert sich sogar hin zum Rate-Bereich.

Bei den legitimen Beispielen gibt es vier Trends. Im Folgenden sollen die Namen der korrespondierenden Phishing Trick Beispiele genannt werden, um die Zuordnung zu erleichtern. Hier soll aber festgehalten werden, dass diese Beispiele wie in Kapitel 3.3 zu sehen keine betrügerischen Tricks enthalten. Drei Beispiele bleiben relativ gleich: (2a) Subdomain (84%, 100% und 84%), (2b) Substitute (95%, 100% und 84%) und (2c) Erweiterung (89%, 100% und 89%). Bei allen drei Beispielen werden zum Post-Zeitpunkt die best möglichen Werte erreicht und danach findet wieder ein leichter Abfall statt. Das Substitute Beispiel fällt dabei auf einen leicht niedrigeren Wert als beim Pre-Zeitpunkt zurück. Ein Beispiel verbessert sich hin zum Post und Retention Zeitpunkt, aber verbleibt im Rate-Bereich: (3) Anhang EXE (46%, 67% und 56%). Bei sechs Beispielen stellen wir eine Verbesserung vom Pre zu den anderen beiden Zeitpunkten fest: (1) Absender (74%, 95% und 89%), (2) Random URL (42%, 58% und 68%), (2a) Path (79%, 95% und 89%), (2b) Typo (39%,), (Spezial) Missmatch (74%, 84% und 89%) und (3) Anhang PDF.EXE (0%, 85% und 80%).

	Pre			Post			Retention		
	P	L	G	P	L	G	P	L	G
(1) Absender	1,00	0,74	0,87	1,00	0,95	0,97	0,95	0,89	0,92
(1) Nonsense	0,84	0,84	0,84	0,95	0,89	0,92	0,68	0,89	0,79
(2) Random URL	1,00	0,42	0,71	1,00	0,58	0,79	1,00	0,68	0,84
(2a) Subdomain	0,47	0,84	0,62	0,84	1,00	0,92	0,95	0,84	0,89
(2a) Path	0,95	0,79	0,87	1,00	0,95	0,97	1,00	0,89	0,95
(2b) Substitute	0,79	0,95	0,87	0,89	1,00	0,95	1,00	0,84	0,92
(2b) Typo	0,74	0,39	0,56	0,89	0,53	0,71	0,84	0,74	0,79
(2b) Swap	0,32	0,95	0,63	0,63	1,00	0,82	0,47	0,95	0,71
(2c) Erweiterung	0,63	0,89	0,76	0,95	1,00	0,97	0,84	0,89	0,87
(Spezial) Missmatch	0,89	0,74	0,82	1,00	0,84	0,92	0,89	0,89	0,89
(3) Anhang PDF.EXE	0,85	0,00	0,42	1,00	0,85	0,92	1,00	0,80	0,90
(3) Anhang DOC	1,00	0,54	0,77	0,62	0,50	0,56	0,50	0,56	0,53
(3) Anhang EXE	0,67	0,46	0,56	1,00	0,67	0,83	0,80	0,56	0,68
Gesamt	0,78	0,66	0,72	0,91	0,83	0,87	0,84	0,80	0,82

Table 2: Durchschnittliche Erkennungsrate von betrügerischen und legitimen Beispielen für alle drei Zeitpunkte (0.0 = von keiner Person erkannt, 1.0 von allen Personen erkannt). P = Phis, L = Legitim, G = Gesamt

5 DISKUSSION

Anders als in Reinheimer et al. [19], musste die Hypothese bezogen auf die vier Monate verworfen werden. Eine mögliche Erklärung für dieses Phänomen, könnte die geringe Anzahl an Teilnehmern darstellen. Da wir eine große Effektstärke mit eta^2[g] = 0.283 haben

und der eigentliche P-Wert knapp an der Grenze liegt, gehen wir davon aus, dass mit einer größeren Teilnehmerzahl ein signifikantes Ergebnis erzielt worden wäre. Mit der erzielten Effektstärke und einer Power von 0,95 würden 37 Teilnehmer benötigt um einen signifikanten Unterschied zwischen Pre und Retention zu finden (G*Power Analyse). Zudem schnitten die Teilnehmer/Innen bereits in dem Pre-Quiz gut ab, was einen signifikanten Unterschied über einen längeren Zeitraum erschwert nachzuweisen.

Entsprechend ist es wichtig, dass in diesem Bereich weitere Studien durchgeführt werden. Daher sollte in Zukunft der Zeitabschnitt zwischen drei und sechs Monate genauer betrachtet und analysiert werden, da das Ergebnis dieser Studie (anders als in [19]) bereits mögliche Probleme ab dem vierten Monat aufweist.

Nun zu den besonderen Ergebnissen der Beispiele: Das Beispiel mit einem Word Dokument wurde vor der Schulung besser erkannt. Mögliche Erklärungen hierfür sind: Im Rahmen der Schulung wurde thematisiert, dass nach dem Ausschalten von Makros keine Gefahr mehr besteht. Womöglich wurde nach der Schulung diese Annahme getroffen und deshalb Word als ungefährlich angenommen. Umgekehrt wurden die beiden legitimen Beispiele mit PDF Anhängen häufig als Phish eingestuft. Hier könnte eine Ursache sein, dass in der Schulung davor gewarnt wird, dass zunächst einmal jedes Dateiformat Schadsoftware enthalten könne. Unsere Ergebnisse zeigen, dass zusätzlich zum bereits mehrfach erforschten Thema der Phishing Tricks mit Links, die Anhänge im gleichen Umfang untersucht werden sollten. Hier sind die Tricks viel schwieriger in einer Studie abzubilden, weil es weniger Eindeutigkeit (außerhalb der Verwendung von exe-Dateien oder gänzlich unbekannten Dateiformaten) für die Beurteilung gibt. Dennoch ist es wichtig die entsprechenden Stellen in der Schulung kritisch zu überprüfen und hier ggf. Qualitatives Feedback zu sammeln.

Das Phish-Beispiel 'Swap' (papyal.de statt paypal.de) wird im Vergleich zu den anderen Beispielen seltener als Phish erkannt. Dies liegt vermutlich an dem Beispiel selbst. Hierbei wird die Eigenschaft des menschlichen Gehirns ausgenutzt, dass Wörter im Ganzen gelesen werden. Wenn wir ein Wort nicht mit voller Konzentration Buchstabe für Buchstabe lesen, in dem ein Buchstabendreher enthalten ist, wird dieser häufig überlesen. Erschwerend kommt hinzu, dass die Darstellung der URL bei gängigen Anwendungen (wie denen, die wir in der Studie genutzt haben) sub-optimal ist, weil die Anzeige klein ist und die Buchstaben nah beieinander stehen. Diesem Problem könnte mit mehr Beispielen in der Schulung und womöglich früherer Auffrischung solcher Tricks begegnet werden.

Das betrügerische Beispiel 'Nonsense' wurde durch die Teilnehmer/Innen eher richtig bewertet, jedoch ist der Erkennungswert bei dem Retention-Quiz niedriger als beim Pre-Quiz. Im Beispiel wird nach persönlichen Daten gefragt, die der Organisation bekannt sein sollten. Vielleicht dachten einige Teilnehmer/Innen - entgegen der klaren Hinweise in der Schulung -, dass diese Mail berechtigt sei, da in manchen Situationen eine solche Nachricht gegebenenfalls realistisch erscheint. Dies sollte in Zukunft auch gegen Ende der Schulung erneut betont werden.

Die Studie weist einige Limitationen auf, die im Folgenden diskutiert werden: In einer neuen Studie könnte mit einer größeren Stichprobe im Rahmen der Schulung noch weitere Gruppen für eine noch breit gefächerte Retention Messung erhoben werden. Insgesamt wurde im Bezug auf die Anzahl der Beispiele ein Kompromiss versucht zu finden. Einerseits sollten die verschiedenen Phishing Tricks so ausführlich wie möglich abgebildet werden. Ziel war es eine möglichst gute Aussagekraft über diese verschiedenen Tricks und deren Erkennungsrate zu haben. Andererseits sollten die Teilnehmer/Innen nicht mit einer Masse an Beispielen überbelastet werden. Damit keine Übermüdungseffekte dazu führen, dass die Erkennungsrate durch diese Effekte negativ beeinflusst wird. Deshalb wurde sich dazu entschieden, dass nicht immer jedes Beispiel in jedem Fragebogen beantwortet werden muss.

Ein weiterer kritischer Punkt ist der Zugriff auf das Material nach der Schulung. Es ist nicht auszuschließen, dass die Teilnehmer/Innen diese Materialien für eine spätere Verwendung gespeichert oder ausgedruckt haben. In diesem Kontext erschien es uns allerdings wichtiger, dass die Teilnehmer/Innen weiterhin Zugriff auf Schulungsmaterialien haben, sodass diese Einschränkung in Kauf genommen wurde. Um diesen Einfluss quantifizierbar zu machen, könnte in einer weiteren Studie untersucht werden, ob der Zugriff auf die Materialien einen Einfluss auf den Effekt über einen längeren Zeitraum hat.

6 FAZIT

Wie die Ergebnisse zeigen, hilft die interaktive Präsenzschulung den Teilnehmer/Innen dabei, legitime von betrügerischen Nachrichten zu unterscheiden. Auch über einen längeren Zeitraum bleibt dieser Effekt für die meisten Beispiele erhalten. Dennoch ist der Gesamt-Effekt bereits nach vier Monaten nicht mehr signifikant. Um empfehlen zu können, was ein guter Zeitpunkt für die Auffrischung ist, sind weitere Studien zu unterschiedlichen Zeitpunkten notwendig.

Die Ergebnisse für die einzelnen Phishing Tricks zeigen, welche Tricks insbesondere nach vier Monaten besonders Probleme machen. Diese Erkenntnis wird genutzt, um die Unterlagen für die Schulung weiter zu verbessern.

7 DANKSAGUNG

Die Forschung wurde unterstützt durch die Helmholtz Gemeinschaft (HGF), Unterthema Engineering Secure Systems (ESS) am Karlrsuher Institut für Technologie (KIT). Ebenso danken wir Herr Rühl und dem Polizeipräsidium Südhessen für die Kooperation.

REFERENCES

[1] M Butavicius, K Parsons, M Pattinson, A McCormac, D Calic, and M Lillie. Understanding Susceptibility to Phishing Emails: Assessing the Impact of Individual Differences and Culture. *International Symposium on Human Aspects of Information Security & Assurance (HAISA)*, 2017.

[2] Marcus Butavicius, Kathryn Parsons, Malcolm Pattinson, and Agata McCormac. Breaching the human firewall: Social engineering in phishing and spear-phishing emails. *Australasian Conference on Information Systems*, 2016.

[3] Casey Canfield, Alex Davis, Baruch Fischhoff, Alain Forget, Sarah Pearman, and Jeremy Thomas. Replication: Challenges in using data logs to validate phishing detection ability metrics. *Symposium on Usable Privacy and Security (SOUPS)*, 2017.

[4] Casey Canfield, Baruch Fischhoff, and Alex Davis. Using Signal Detection Theory to Measure Phishing Detection Ability and Behavior. In *Symposium on Usable Privacy and Security (SOUPS)*, 2015.

[5] Casey Canfield, Baruch Fischhoff, and Alex Davis. Quantifying Phishing Susceptibility for Detection and Behavior Decisions. *Human Factors: The Journal of Human Factors and Ergonomics Society*, 58:1158–1172, 2016.

[6] Casey Inez Canfield and Baruch Fischhoff. Setting Priorities in Behavioral Interventions: An Application to Reducing Phishing Risk. *Risk Analysis*, 38:826–838, 2018.

[7] Gamze Canova, Melanie Volkamer, Clemens Bergmann, and Benjamin Reinheimer. Nophish app evaluation: lab and retention study. In *NDSS workshop on usable security*, 2015.

[8] Iain Embrey and Kim Kaivanto. Many Phish in the C: A Coexisting-Choice-Criteria Model of Security Behavior. *arxiv*, 2018.

[9] Timothy Kelley and Bennett I Bertenthal. Real-world decision making: Logging into secure vs. insecure websites. *Proceedings of the USEC'16*, 2016.

[10] Ponnurangam Kumaraguru, Justin Cranshaw, Alessandro Acquisti, Lorrie Cranor, Jason Hong, Mary Ann Blair, and Theodore Pham. School of phish: a real-world evaluation of anti-phishing training. In *Symposium on Usable Privacy and Security (SOUPS)*, pages 1–12, 2009.

[11] Alexandra Kunz, Melanie Volkamer, Simon Stockhardt, Sven Palberg, Tessa Lottermann, and Eric Piegert. Nophish: evaluation of a web application that teaches people being aware of phishing attacks. *Informatik 2016*, 2016.

[12] Elmer Lastdrager, Inés Carvajal Gallardo, Pieter Hartel, and Marianne Junger. How effective is anti-phishing training for children? In *Thirteenth Symposium on Usable Privacy and Security ({SOUPS} 2017)*, pages 229–239, 2017.

[13] Gang Liu, Guang Xiang, Bryan A Pendleton, Jason I Hong, and Wenyin Liu. Smartening the crowds: computational techniques for improving human verification to fight phishing scams. In *Symposium on Usable Privacy and Security (SOUPS)*, pages 1–13, 2011.

[14] Jaclyn Martin. *Something Looks Phishy Here: Applications of Signal Detection Theory to Cyber-Security Behaviors in the Workplace*. PhD thesis, University of South Florida, 2017.

[15] Jaclyn Martin, Chad Dubé, and Michael D Coovert. Signal Detection Theory (SDT) Is Effective for Modeling User Behavior Toward Phishing and Spear-Phishing Attacks. *Human Factors: The Journal of Human Factors and Ergonomics Society*, 60:1179–1191, 2018.

[16] Christopher B Mayhorn and Patrick G Nyeste. Training users to counteract phishing. *Work (Reading, Mass.)*, 41 Suppl 1:3549–52, 2012.

[17] María M Moreno-Fernández, Fernando Blanco, Pablo Garaizar, and Helena Matute. Fishing for phishers. Improving Internet users' sensitivity to visual deception cues to prevent electronic fraud. *Computers in Human Behavior*, 69:421–436, 2017.

[18] Stephan Neumann, Benjamin Reinheimer, and Melanie Volkamer. Don't be deceived: the message might be fake. In *International Conference on Trust and Privacy in Digital Business*, pages 199–214. Springer, 2017.

[19] Benjamin Reinheimer, Lukas Aldag, Peter Mayer, Mattia Mossano, Reyhan Duezguen, Bettina Lofthouse, Tatiana von Landesberger, and Melanie Volkamer. An investigation of phishing awareness and education over time: When and how to best remind users. In *Symposium on Usable Privacy and Security (SOUPS)*, pages 259–284, 2020.

[20] Steve Sheng, Bryant Magnien, Ponnurangam Kumaraguru, Alessandro Acquisti, Lorrie Faith Cranor, Jason Hong, and Elizabeth Nunge. Anti-phishing phil: the design and evaluation of a game that teaches people not to fall for phish. In *Symposium on Usable Privacy and Security (SOUPS)*, pages 88–99, 2007.

[21] Harold Stanislaw and Natasha Todorov. Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, 31(1):137–149, 1999.

[22] Simon Stockhardt, Benjamin Reinheimer, Melanie Volkamer, Peter Mayer, Alexandra Kunz, Philipp Rack, and Daniel Lehmann. Teaching phishing-security: which way is best? In *IFIP International Conference on ICT Systems Security and Privacy Protection*, pages 135–149. Springer, 2016.

[23] Christopher Thompson, Martin Shelton, Emily Stark, Maximilian Walker, Emily Schechter, and Adrienne Porter Felt. The web's identity crisis: understanding the effectiveness of website identity indicators. In *28th USENIX Security Symposium (USENIX Security 19)*, pages 1715–1732, 2019.

[24] Kai Florian Tschakert and Sudsanguan Ngamsuriyaroj. Effectiveness of and user preferences for security awareness training methodologies. *Heliyon*, 5(6):e02010, 2019.

[25] Tianjian Zhang. Knowledge expiration in security awareness training. *Annual ADFSL Conference on Digital Forensics, Security and Law*, 2, 2018.