Studie zur Artefaktauswahl zur Messung von IT-Security-Awareness

Tim Jüliger
2651712

Universität Bonn
Erstgutachter: Prof. Dr. Michael Meier
Zweitgutachter: Jun.-Prof. Dr.-Ing. Delphine Reinhardt

Inhaltsverzeichnis

1 Einleitung

Die Digitalisierung erreicht zunehmend jegliche Facetten der Welt. Nicht nur berufliche Aspekte und Handel, sondern auch sämtliche Aspekte des Alltags, wie zum Beispiel Kommunikation, Finanzdienstleistungen und Einkäufe, werden vermehrt digital und über das Internet durchgeführt. Dies eröffnet jedoch auch neue Möglichkeiten für Cyberkriminalität.
Zwar ist es sehr kompliziert die Schäden von Angriffen zu messen, doch machen viele Quellen deutlich, dass diese Schäden Jahr für Jahr steigen und Cyberkriminalität ein zunehmendes Problem darstellt. Dennoch können durch den korrekten Umgang mit IT-Infrastruktur und Erkennung der Anzeichen von IT-sicherheitsrelevanten Vorfällen die Schäden von erfolgreichen Angriffen vermindert und die Erfolgsrate von Angriffen reduziert werden.
Der Risikofaktor Mensch ist laut einigen Quellen einer der wichtigsten Faktoren im Umgang mit der IT-Sicherheit von Unternehmen. Grund hierfür ist, dass viele Angriffe erst durch menschliches Fehlverhalten ermöglicht werden. Hierzu zählen beispielsweise unverschlüsselt gespeicherte Daten, unbedachte Weitergabe sensibler Daten oder öffentlich zugängliche Wechseldatenträger. [1] Einige Quellen sprechen sogar davon, dass 70% aller Sicherheitsprobleme in Unternehmen durch menschliche Fehler verursacht werden. [2]
Bisher steht die individuelle IT-Security-Awareness der Nutzer bei der Prävention von Angriffen im Hintergrund. Zwar existieren Schulungen zur IT-Sicherheit, die auch erfolgreich durchgeführt werden und sich positiv auswirken, jedoch gehen diese nicht gezielt auf Schwachstellen der Nutzer ein, sondern übermitteln jedem Teilnehmer die gleichen Inhalte. Das vorherige Verständnis der IT-Sicherheit wird also vor den Schulungen nicht geprüft und ob eine tatsächliche Verbesserung des Verständnisses stattfindet, kann so nicht bestimmt werden.
Um gezielt Nutzer im korrekten Umgang zu schulen und Defizite in der Kompetenz bezüglich der IT-Sicherheit zu erkennen, benötigt man Möglichkeiten, die IT-Security-Awareness des Anwenders zu messen. Nur so können Schulungen gezielt und individuell Schwachstellen beseitigen und potentielle Schäden minimiert werden. Diesbezüglich fehlt allerdings noch Grundlagenforschung.
Die Leitfrage, deren Antwort bislang unklar ist, lautet: „Lässt sich IT-Security-Awareness messen?“ Um sich einer Antwort anzunähern, setzt sich diese Arbeit mit der Frage auseinander, wie Nutzer auf Artefakte von IT-sicherheitsrelevanten Vorfällen reagieren und was diese Reaktionen über das IT-Sicherheitsbewusstsein von Nutzern aussagen. Hierfür werden zwei Studien durchgeführt. Zunächst sucht die erste Studie nach häufigen Reaktionen auf eine Vielzahl von Artefakten von Angriffen. Die Umfrage richtet sich daher an die allgemeine Bevölkerung.
Die Folgestudie richtet sich ausschließlich an Experten aus der IT-Sicherheit und verwandten Bereichen. Diese Studie soll eine Bewertungsgrundlage schaffen, indem die Teilnehmer gebeten werden, die häufigen Reaktionen der ersten Studie zu bewerten und Schätzungen abzugeben, wie diese Reaktionen für eine Messung von IT-Security-Awareness zu bewerten sind.

2 Grundlagen

Zu Beginn müssen einige Grundlagen dargelegt werden. Diese stammen zum größten Teil aus dem ITS.APT-Projekt, welches ebenfalls den thematischen Rahmen für diese Arbeit definiert.
Zunächst werden Artefakte genauer erläutert und IT-Security-Awareness genau definiert. Dann werden einige allgemeine Grundlagen zum Entwurf von Studien und dazu relevanten Aspekten beschrieben. Darauf folgen einige verwandte Arbeiten und zuletzt einige Erklärungen zum ITS.APT-Projekt selber.

2.1 Artefakte

Artefakte sind der Hauptfokus dieser Arbeit. Sie werden im ITS.APT-Projekt wie folgt definiert:

Definition 1 *An object, item, event or effect that is artificially constructed, observable and part of or caused by an attack is called artifact. [3]*

Ein Artefakt ist also laut Definition ein künstlich erzeugter Vorfall oder ein künstlich erzeugtes Element eines Systems, das direkt aus einem Angriff folgt. Diese Vorfälle beziehungsweise diese Elemente müssen beobachtbar sein. Dies ermöglicht Nutzern auf ein Artefakt zu reagieren. Wichtig ist ebenfalls folgende Differenzierung:

Definition 2 *A 1^{st}-order artifact is persistent with regard to its environment variability. 2^{nd}-order artifacts are visual representations of one (or more) 1^{st}-order artifact(s).*

Diese Definition unterscheidet Artefakte erster und zweiter Ordnung. Artefakte erster Ordnung sind unabhängig in Bezug zu ihrer Umgebung. Es handelt sich dabei meistens um die rein technische Darstellung eines Elements, also zum Beispiel einem abgelaufenen Zertifikat. Artefakte zweiter Ordnung beschreiben die visuelle Aufbereitung von einem oder mehreren Artefakten erster Ordnung. Diese Aufbereitung wird meist von Anwendungen oder dem Betriebssystem erzeugt. Bei dem gegebenen Beispiel eines abgelaufenen Zertifikats wäre beispielsweise die angezeigte Warnung eines Webbrowsers das entsprechende Artefakt zweiter Ordnung.
Abbildung 1 veranschaulicht Artefakte verschiedener Ordnungen. Wie man erkennen kann, sieht der Nutzer das Artefakt erster Ordnung nur durch die visuelle Aufbereitung dieses Artefakts durch sein System. Er reagiert also nicht direkt auf ein Artefakt erster Ordnung, sondern stets auf das dazugehörige Artefakt zweiter Ordnung.
Nahezu alle Angriffe erzeugen Artefakte und die Nutzer werden gezwungen auf diese Artefakte mit speziellen Handlungen zu reagieren. In diesem Kontext ist auch das Ausbleiben einer solchen Handlung beziehungsweise das bewusste Ignorieren eines Artefakts eine solche Handlung. Diese Reaktionen können offensichtlich beobachtet und bewertet werden. Jedoch fehlt eine einheitliche Methode, um

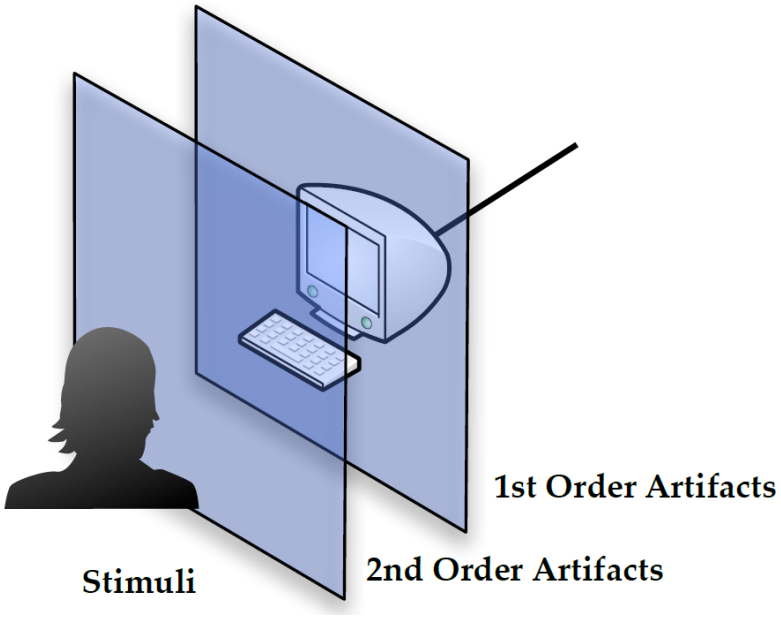

Abbildung 1. Eine Veranschaulichung der Artefaktordnungen. [3]

diese Reaktionen auf verschiedene Artefakte zu bewerten und so eine möglichst vergleichbare Einschätzung der IT-Security-Awareness zu ermöglichen. Um solch eine Methode zu erstellen müssen möglichst viele Reaktionen von Experten bewertet werden, was das Ziel dieser Studie darstellt.

2.2 IT-Security-Awareness

Es existieren einige verschiedene Definitionen für IT-Security-Awareness. Einige Definitionen beziehen sich hierbei auf den passiven Umgang und das erhöhte Interesse an verschiedenen Themen der IT-Sicherheit. Andere Definitionen beziehen sich auf alle Nutzer innerhalb eines Unternehmens und definieren IT-Security-Awareness über deren Verständnis für die Bedeutung von IT-Sicherheit, die persönliche Verantwortung an der Sicherheit des Unternehmens und an diese Anforderungen angepasstes Verhalten. [4]
Beide Ansätze besitzen gute und wichtige Aspekte, jedoch wird für diese Arbeit die Definition des ITS.APT-Projektes verwendet. Diese lautet wie folgt:

Definition 3 *IT security awareness is situation awareness limited to elements directly or indirectly related to IT security.*

Hierbei beschreibt Situationsbewusstsein das Wahrnehmen von Elementen eines Systems in einem Zeitrahmen, das Verstehen ihrer Bedeutung im Bezug auf die

gegebene Situation und die Projektion des zukünftigen Status des Systems. IT-Security-Awareness ist also ein auf IT-sicherheitsrelevante Elemente beschränktes Situationsbewusstsein und beschreibt das Wahrnehmen von Schutzmaßnahmen und Gefahren eines Systems, ein Verständnis ihrer Bedeutung für einen sicheren Zustand des Systems und das Wissen über die Folgen der wahrgenommenen Elemente. Wahrnehmung, Verständnis und Projektion beschreiben drei Stufen von IT-Security-Awareness, wobei das Erreichen aller Stufen ein hohes Maß an IT-Security-Awareness beschreibt und das Fehlen von einer oder mehrerer Stufen ein niedriges Maß darstellt. [3]
Das Verwenden dieser Definition hat einige Vorteile gegenüber den anderen Definitionen. Einerseits ist diese Definition sehr eindeutig. Es existieren keine großen Interpretationsmöglichkeiten, wie es bei der Definition über das Verständnis für die Bedeutung von IT-Sicherheit der Fall ist. Bei dieser kann diskutiert werden, wie genau besagtes Verständnis zu verstehen ist. Durch die Aufteilung in drei Stufen erlaubt die gewählte Definition auch ein gewisses Maß an Messbarkeit . Eine klare Möglichkeit, das Interesse von Menschen an einem Thema zu messen ist jedoch ein eher komplexes Problem. Auch eine Gewichtung der verschiedenen Aspekte der anderen Definitionen könnte ein Problem darstellen. Ebenfalls ein entscheidendes Argument für diese Definition ist die Konsistenz der Definitionen dieser Arbeit und des ITS.APT-Projekts.

2.3 Grundlagen des Studien-Designs

Das Ziel eines guten Studien Designs ist es eine möglichst hohe Rücklaufquote und zeitgleich eine möglichst hohe Anzahl von Teilnehmern zu erreichen, sowie möglichst aussagekräftige, repräsentative Ergebnisse. Hierzu gibt es einige wichtige, grundlegende Aspekte. Um möglichst viele Personen zur Teilnahme zu bewegen ist eine gute und überzeugende Einladungs-Mail unentbehrlich. Diese sollte soweit möglichst eine direkte Ansprache besitzen und dem Teilnehmer erklären, warum diese Umfrage durchgeführt wird. Ebenfalls sollte eine direkte Bitte an den Teilnehmer enthalten sein an der Umfrage teilzunehmen und es sollte ein Grund angegeben werden, weshalb er von der Umfrage profitiert. Abgeschlossen wird die Einladung mit einer kurzen Danksagung. Ebenfalls wichtig ist das Verwenden von aktiven Ausdrücken, das Vermeiden von Füllwörtern und zu langen und komplexen Sätzen. Natürlich sollte bei der Wortwahl auch bedacht werden, welche Personengruppen angesprochen werden. [5]
Zu Beginn der eigentlichen Umfrage sollten zunächst kurze, interessante Fragen gestellt werden. Dennoch sollten besonders wichtige Fragen bereits möglichst früh in der Umfrage platziert und auf nicht essenzielle Fragen verzichtet werden. Fragen sollten möglichst einfach und klar formuliert und Mehrdeutigkeiten verhindert werden. Auch Abkürzungen und für die Teilnehmer eventuell unbekannte Fachausdrücke sollten nicht vorkommen. Stattdessen sollten nicht wertende, komplette Sätze verwendet werden. Die Umfrage sollte in Themenblöcke unterteilt werden und innerhalb der Blöcke möglichst nur eine Sorte von Fragen vorkommen. Zudem sollten Erklärungen gegeben werden, wie auf die Fragen zu

antworten ist. Den letzten Block sollten die demographischen Fragen bilden [6]. Selbstverständlich sollte die Umfrage vor ihrer Verbreitung ausgiebig getestet werden, um Fehler auszuschließen. [7]
Je nach gewünschten Daten können verschiedene Fragetypen verwendet werden. Open-ended questions erlauben den Teilnehmer eigene Antworten zu formulieren und einzugeben. Dieser Fragetyp bietet die Möglichkeit, eigene Gedanken und Ideen einzubringen, erhöht jedoch auch den Aufwand der Umfrage. Die erhaltenen Daten sind von einer höheren Varianz, jedoch schwerer zu analysieren. [7]
Eine weitere Sorte an Fragen sind close-ended questions. Bei diesen Fragen werden Antwortmöglichkeiten vorgegeben und Teilnehmer können eine oder mehrere Antworten auswählen. Diese Fragen sind einfacher zu analysieren, da die gesammelten Daten nicht so differenziert sind, jedoch müssen hier alle relevanten Antwortmöglichkeiten vorher bekannt sein. Außerdem müssen sich die einzelnen Antwortmöglichkeiten gegenseitig ausschließen, sollte nur eine von ihnen wählbar sein. Zwar kann man diesen Fragetyp mit einer Eingabemöglichkeit für die Teilnehmer erweitern, jedoch sollte dann vorab überlegt werden, wie diesen Daten verwendet werden, da diese oft aufgrund der Problematik der Analyse im Nachhinein nicht verwendet werden. [7]
Der letzte für diese Arbeit relevante Fragentyp sind Bewertungsfragen. Bei diesen handelt es sich um close-ended questions, bei denen vorgegebene Elemente auf einer numerischen Skala bewertet werden sollen. Besonders wichtig ist hier die Wahl einer entweder geraden oder ungeraden Skala. Eine ungerade Skala bietet einen mittleren Wert, wobei gerade Skalen dies nicht tun und es so ermöglichen zu sehen, in welche Richtung die Teilnehmer bei der Bewertung tendieren. [7]
Die Verwendung einer Fortschrittsanzeige erhöht zwar die Menge vollständig abgeschlossener Datensätze, jedoch nicht mit statistischer Signifikanz [8]. Für die Befragungsdauer einer Studie sind 6 Tage ausreichend. Nach dieser Zeit werden nahezu keine neuen mehr Datensätze ausgefüllt werden. Ein längerer Umfragezeitraum bietet sich allerdings an, wenn über längere Zeit neue Einladungen verschickt werden oder wenn eine Erinnerungsmail versendet werden soll. Diese sollte dann nach 6 Tagen eingeplant und versandt werden. [9]
Die Studie zur Befragungsdauer bezieht sich jedoch nur auf Einladungen über E-Mail. Es lässt sich vermuten, dass sich die Verbreitung über soziale Netzwerke ähnlich verhält, jedoch auf Grund der möglichen Weiterleitung der Einladungen länger noch neue Teilnehmer erreicht.
Bei der Feldzeit von Studien sollte ebenfalls darauf geachtet werden, dass die Stichproben nicht verzerrt werden. Dies kann durch eine zu kurze Befragungsdauer geschehen, da Personen mit unregelmäßigem und seltenen Internetgebrauch eine geringere Auswahlwahrscheinlichkeit haben. Auch die nicht gleichmäßige Abbildung der Wochentage kann zu einer Verzerrung der Stichprobe führen. [10]
Ein weiteres wichtiges Element des Studien-Designs ist die Validität. Validität bezieht sich auf das Vermeiden von systematischen Fehlern bei der Erstellung von Umfragen. Man unterscheidet zwischen interner und externer Validität. Interne Validität bezeichnet die Resistenz der Ergebnisse gegenüber äußeren, nicht eingerechneter Einflüsse. Eine Studie besitzt also eine hohe interne Validität, wenn

ein Ergebnis direkt aus einem gegebenen Einfluss folgt. Man kann diese durch Kontrollgruppen und Kontrollfragen erreichen. Externe Validität beschreibt die Möglichkeit, die Ergebnisse einer Studie zu verallgemeinern und die Ergebnisse auf andere Gruppen zu übertragen. Um externe Validität zu erreichen muss interne Validität gewährleistet sein. Ein zu vermeidender Fehler, der externe Validität verhindert, ist eine nicht repräsentative Stichprobe. [11]

2.4 Related Work

IT-Security-Awareness ist in der Forschung ein Feld mit in problematischem Maß fehlenden, gemeinsamen Grundlagen. Das sind die Ergebnisse einer von Tsohou et al. an der University of the Aegean durchgeführten Meta-Studie von 48 verschiedenen Publikationen, die sich primär oder sekundär mit dem Thema IT-Security-Awareness auseinandersetzen. Laut dieser Studie fehlen zur effektiven Forschung einerseits einheitliche Definitionen der relevanten Elemente, sowie klare Ziele, die mittels IT-Security-Awareness erreicht werden sollen. Auch die verwendeten Methodiken einiger Studien werden als ungeeignet eingeordnet. Dies soll zukünftige Forschung zu konkreten Methoden der IT-Security-Awareness anregen. Zusätzlich fehlt es vielen Studien an theoretische Grundlagen. [12]
Die Autoren der Studie schlussfolgern, dass die fehlenden Klarheiten in Bezug auf bedenkliche Sachverhalte der IT-Sicherheit bei vielen Betroffenen (Forscher, Manager, Fachmänner) eventuell zu Frustrationen mit Anstrengungen zu Steigerung von IT-Security-Awareness führen könnten. Dies könnte einen Grund für die anhaltende Problematik von IT-Sicherheit in Unternehmen darstellen. Mögliche zukünftige Forschung zur Erläuterung unklarer Problematiken in organisatorischen Situationen könnte positiv dazu beitragen, diese Frustrationen abzubauen und so eine Problemlösung ermöglichen. [12]
Eine der wenigen Arbeiten, die sich direkt mit der Problematik des Messens des IT-Sicherheitsverständnisses von Nutzern auseinandersetzt, ist der Vortrag „De-humanizing Human Vulnerability Assessment“ von Laura Bell. Die Autorin argumentiert gegen das alleinige Verbessern von Systemen und für das Testen von Nutzern, auch wenn dies eventuell Bedenken auslöst, da eine Bewertung der Nutzer immer auch zu Konflikten führen kann. [13]
Das entwickelte System soll in drei Schritten systematisch Schwachstellen aufdecken. Zunächst werden die Verbindungen der Nutzer untereinander erschlossen. Dies soll personalisierte Angriffe ermöglichen und aufzeigen, welche Risiken durch etwaige Verbindungen verstärkt werden könnten. Die zweite Phase testet die Nutzer. Hierfür werden personalisierte Angriffe mittels verschiedener Vektoren durchgeführt. Diese beinhalten zum Beispiel E-Mails, soziale Netzwerke, SMS und Wechseldatenträger. Dies sind zwar hauptsächlich Methoden des Social Engineering, jedoch werden diese hier mit dem Ziel verwendet IT-Security-Awareness zu messen. Im letzten Schritt werden die gesammelten Daten analysiert, und Veränderungen über Zeit überwacht. So sollen die Trainingserfolge gemessen werden, und Schwachstellen gefunden und beseitigt werden. [13]
Leider existiert zu dem Vortrag keine vollständige Ausarbeitung, sodass Details

nicht verglichen werden können. Das im Vortrag entwickelte Konzept wird jedoch in der Wirtschaft verwendet. [14]

2.5 ITS.APT

Das vom Bundesministerium für Bildung und Forschung geförderte Verbundsprojekt „IT-Security Awareness Penetration Testing“ bildet den Rahmen für diese Arbeit. Bei diesem Projekt arbeiten die AG IT-Sicherheit der Rheinischen Friedrich-Wilhelms-Universität Bonn, die Abteilung „Fachgebiet Allgemeine Psychologie: Kognition“ der Universität Duisburg-Essen, die Enno Rey Netzwerke GmbH, das Universitätsklinikum Schleswig-Holstein, Lübeck, das Institut für Informations-, Telekommunikations- und Medienrecht der westfälischen Wilhelms-Universität Münster und das Unabhängige Landeszentrum für Datenschutz Schleswig-Holstein, Kiel zusammen. Das Projekt befasst sich mit der Problematik der Messung der IT-Sicherheit bei Betreibern kritischer Infrastrukturen mittels klassischem Penetration Testing, da dieses zwar die Infrastruktur testet, jedoch den Faktor Mensch nicht einbezieht. [15]
Ziel des Projektes ist die Erweiterung des klassischen Penetration Testing um eben diesen Faktor Mensch. Besonders die Frage, ob und wie das IT-Sicherheitsbewusstsein von Nutzern für die Sicherheit eine Rolle spielt, konnte bisher nicht eindeutig bewiesen und gemessen werden und steht im Fokus des Projekts. [15]
Um diese Fragen zu beantworten, wird, neben einigen studentischen Rahmenarbeiten, eine umfassende Feldstudie an dem Universitätsklinikum Schleswig-Holstein durchgeführt. Nach einer sorgfältigen Evaluation soll ein Tool entwickelt werden, mittels dessen kosteneffizient die Messung des kollektiven IT-Sicherheitsbewusstseins eines Unternehmens gemessen werden kann. Dies kann ebenfalls wieder als solider Ansatzpunkt für zukünftige Forschung verwendet werden, zum Beispiel zur Messung der tatsächlichen Auswirkungen von Schulungen auf nicht nur das IT-Sicherheitsbewusstsein der Teilnehmer, sondern auch deren Effekt auf die IT-Sicherheit der Infrastruktur. [15]
Problematisch sind hierbei Aspekte des Datenschutzes und des Arbeitsrechts, da die Feldstudie in einem realistischen Umfeld, nämlich dem des Universitätsklinikums Schleswig-Holstein an dessen Mitarbeitern durchgeführt werden soll. Aus diesem Grund muss eine enge Zusammenarbeit von IT-Sicherheitsexperten, Juristen und Psychologen stattfinden. [15]

3 Studiendesign und Studienplattform

Im Verlauf dieser Arbeit werden zwei Studien durchgeführt. In der Vorstudie sollen zunächst möglichst viele Artefakte mit dazugehörigen Reaktionen gefunden werden, indem eine möglichst große Anzahl an normalen Nutzern digitaler Systeme befragt wird. Zusätzlich werden noch einige grundsätzliche Angaben zum IT-Sicherheitsbewusstsein gesammelt, sodass am Ende, zusätzlich zu dem eigentlichen Ziel einer Methode zur Messung von IT-Security-Awareness, auch einige

Rückschlüsse auf das Sicherheitsbewusstsein der Teilnehmer möglich sind, indem die Angaben mit den Einschätzungen der zweiten Studie verglichen werden. Diese zweite und eigentliche Hauptstudie soll Bewertungen des Artefaktpools durch Experten der IT-Sicherheit einholen. Da die Menge an Experten vor allem im deutschen Sprachraum sehr beschränkt ist, wird diese Studie zweisprachig auf deutsch und englisch durchgeführt.

3.1 SoSci Survey

Als Studienplattform wird SoSci Survey gewählt. Dieses Softwarepaket ermöglicht professionelle Umfragen, die jedoch mit sehr wenig Einarbeitung intuitiv erstellt werden können. Es existieren mehr als 30 Fragetypen, von denen Gebrauch gemacht werden kann. Auch selbstdefinierte Fragetypen sind theoretisch möglich. Zusätzlich lassen sich Filter erstellen, die bei gewissen Angaben im Fragebogen unterschiedliche Folgefragen erlauben. Auch die zufällige Auswahl mehrerer Fragebögen ist möglich, wenn zum Beispiel aus Zeitgründen nicht alle Teilnehmer alle Fragen beantworten können. Bilder, Diagramme, Audiodateien und Filme lassen sich ebenfalls problemlos einbinden. Ein weiterer Vorteil von SoSci Survey ist, dass die gesammelten Daten am Ende in einem auswählbaren, weiter benutzbaren Datenformat vorliegen. Zur Auswahl stehen zum Beispiel SPSS, GNU R und eine allgemeine Darstellung mittels CSV-Dateien (Character Separated Values). [16]
Des Weiteren erlaubt SoSci Survey problemlose Pretests, in denen man einen Studienentwurf an Tester versenden kann. Ebenfalls nützlich ist die Möglichkeit, in Echtzeit die Anzahl der Teilnehmer zu sehen und schon während der Umfrage die Daten einzusehen, die bereits eingetragen worden sind. Zusätzlich bieten die Betreiber von SoSci Survey das kostenfreie Hosten von wissenschaftlichen Umfragen ohne kommerziellen Hintergrund an. Umfragen mittels SoSci Survey sind auf bis zu 5000 Teilnehmer ausgelegt. Allerdings dient es nur zur Erhebung von Daten und nicht deren Auswertung. Zur Auswertung wird also ein eigenes Programm benötigt. [16]
Ein weiterer Vorteil von SoSci Survey sind Aspekte des Datenschutzes. Die von SoSci Survey verwendeten Server nutzen eine SSL-Verschlüsselung, die verhindert, dass Dritte die gesendeten Daten mitlesen können. Des Weiteren werden keine IP-Adressen gespeichert, was ein späteres Zusammenführen der Datensätze mit deren Autor verhindert. Für das Sammeln von E-Mail-Adressen existiert eine getrennte Speicherung. Dies gewährleistet eine vollständige Anonymität der eingegebenen Daten. Außerdem werden für die Fragebögen keine Cookies verwendet. Bei Cookies handelt es sich um Datenfragmente, die eine Identifizierung von Nutzern ermöglichen. Diese können ein Datenschutz-Problem verursachen. [16]

3.2 Auswertungswerkzeug

Die Studien werden mittels SPSS ausgewertet. Die von IBM entwickelte SPSS Statistics Software ermöglicht eine Übersicht über alle erhobenen Rohdaten, als auch die statistische Analyse dieser und das Erstellen von Diagrammen. Die Software wird verwendet, da dieses Dateiformat von SoSci Survey unterstützt wird und mit nur wenig Einarbeitungszeit effektiv genutzt werden kann. Ein weiteres Argument für die Nutzung von SPSS ist, dass so alle Funktionen in einer Umgebung verfügbar sind, und nicht für einzelne Teile, also die statistische Analyse sowie die visuelle Aufbereitung der Ergebnisse, unterschiedliche Programme benötigt werden. [17]

3.3 Vorstudie

In den folgenden Abschnitten werden Details der Vorstudie erläutert.

Ziel der Vorstudie Das Ziel der Vorstudie ist das Erzeugen eines möglichst großen Pools an Artefakten der IT-Sicherheit sowie an tatsächliche Reaktionen auf diese Artefakte. Zusätzlich sollen Daten erhoben werden, die es ermöglichen, die Ergebnisse der Hauptstudie anzuwenden und Aussagen über die IT-Security-Awareness der Teilnehmer der Vorstudie zu tätigen. Zwar wird diese Einschätzung nicht repräsentativ sein, lässt allerdings erste Rückschlüsse auf mögliche Fehler der zu erarbeitenden Methode zur Messung von IT-Security-Awareness zu.

Aufbau der Vorstudie Die Vorstudie unterteilt sich in 3 inhaltliche und einen demographischen Abschnit. Der erste Abschnitt besteht aus Fragen, die zu gegebenen Artefakten Reaktionen abfragen. Neben einer Auswahl an möglichen Reaktionen können Teilnehmer der Umfrage auch eigene Reaktionen angeben. Dieser Abschnitt dient nicht nur dem Sammeln von Reaktionen, sondern soll den Teilnehmern auch einige Beispiele für Artefakte liefern. Diese Beispiele sollen den Teilnehmern helfen, den zweiten Teil der Studie zu beantworten. Auf eine formale Definition von Artefakten wird bei der Vorstudie verzichtet, da die Umfrage auch ohne detailliertes Wissen über Artefakte möglich ist und die erhöhte Komplexität einige Teilnehmer abbrechen lassen könnte.
Der zweite Abschnitt besteht ausschließlich aus Fragen mit Freitextantworten. Hier sollen die Teilnehmer zu den Bereichen Browser, E-Mail, Betriebssystem, andere Programme und Sonstiges eigene Artefakte und ihre Reaktionen auf diese angeben. Dieser Teil der Umfrage dient direkt dem Zweck, einen möglichst großen Artefakt- und Reaktionspool aufzubauen.
Der dritte und letzte inhaltliche Abschnitt besteht aus einigen Kontrollfragen. Diese Fragen bestehen aus kurzen Situationsbeschreibungen, bei denen die Teilnehmer in die Rolle eines Nutzers treten sollen und ihre Reaktionen auf indirekt

beschriebene Artefakte erläutern. Bei diesen Artefakten handelt es sich um die gleichen Artefakte wie im ersten Abschnitt. Der Zweck von Kontrollfragen liegt darin, die Konsistenz der Antworten zu überprüfen und so weitere Rückschlüsse zu ermöglichen. Es werden keine Kontrollfragen zum zweiten Abschnitt gestellt, da in diesem Abschnitt keine neuen Erkenntnisse gewonnen werden können und die Dauer der Umfrage sonst zu groß wird, was die Abbrecherquote zu stark erhöhen könnte.
Der letzte Abschnitt besteht aus demographischen Fragen und Fragen zum Nutzerverhalten der Teilnehmer. Die Fragen zur Demographie sollen Rückschlüsse auf Unterschiede zwischen verschiedenen Teilnehmergruppen ermöglichen. Gefragt wird hier nach Alter, Geschlecht, dem formalen Bildungsgrad und der beruflichen Beschäftigung. Fragen zum Nutzerverhalten zielen hingegen darauf, Rückschlüsse auf das Sicherheitsverständnis der Teilnehmer zu erlauben. Hierbei werden die Verwendung von Virenschutzprogrammen, die wöchentliche Zeit am Computer, die Dauer, wie viele Jahre lang das Internet bereits verwendet wird, das bevorzugte Betriebssystem, das verwendete Virenschutzprogramm, das verwendete E-Mail-Programm, präventive Maßnahmen, Unterschiede im Umgang mit IT-Sicherheit zwischen mobilen Geräten und Heim-PCs und eine eigene Einschätzung des IT-Security-Verständnisses abgefragt. Diese Fragen zum Nutzerverhalten stehen allerdings ganz zu Beginn der Umfrage, um auch Abbrecher der Umfrage besser einschätzen zu können.

Abbildung 2 zeigt einen Ausschnitt der vierten Seite der Vorstudie. Im oberen Teil sieht man das Logo der Universität Bonn, welche verwendet wurde, um Teilnehmern der Umfrage einen offiziellen Eindruck zu vermitteln und so die Rücklaufquote zu erhöhen. Ebenfalls im oberen Teil zu sehen ist der Fortschrittsbalken. Der hier angezeigte Prozentsatz bezieht sich jedoch nicht auf die Anzahl der ausgefüllten Fragen, sondern lediglich dem Prozentsatz der bearbeiteten Seiten. Im unteren Teil der Abbildung sieht man zwei der verwendeten Fragen. Alle Fragen folgen dem hier sichtbaren Schema. Nach dem Fragetext folgen Satzfragmente, die mögliche Reaktionen darstellen. Nach einer Reihe dieser Antwortmöglichkeiten folgt noch ein Feld zur Eingabe von Freitextantworten, in dem eigene Reaktionen angegeben werden können. Diese Seite enthält 9 derartige Fragen. Die verwendeten Fragetypen im ersten und dritten inhaltlichen Abschnitt sind close-ended questions, wobei diese um die Möglichkeit zur Eingabe eines weiteren Elements erweitert sind. Im zweiten inhaltlichen Abschnitt werden open-ended questions verwendet, da diese zur Eingabe von neuen Artefakten und dazugehörigen Reaktionen notwendig sind.

Teilnehmer und Verbreitung der Vorstudie Die Umfrage richtet sich an alle Nutzer digitaler Systeme. Es gibt hierbei keine besonderen Einschränkungen zur Auswahl der Teilnehmer, jedoch können nur Personen teilnehmen, die Zugriff auf einen Computer und das Internet besitzen, da die Umfrage Online durchgeführt wird.

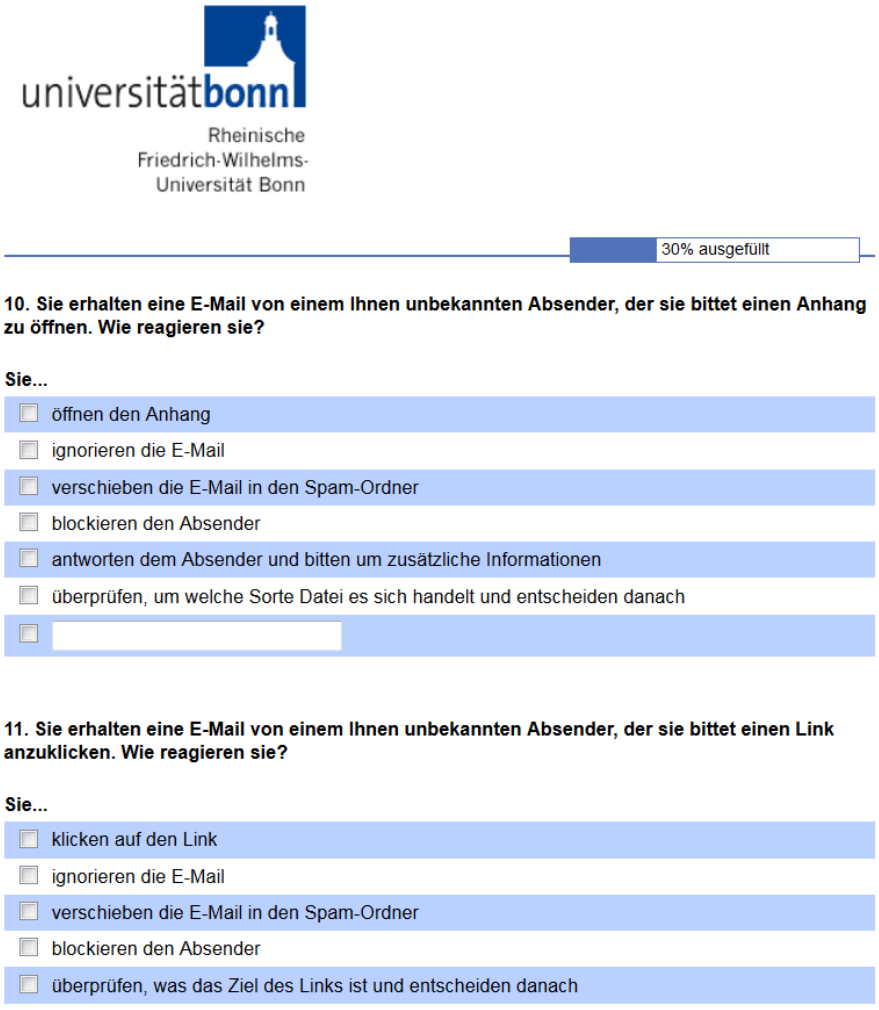

Abbildung 2. Ein beispielhafter Ausschnitt einer Umfrageseite der Vorstudie

Die Umfrage wird sowohl über Soziale Medien, als auch über E-Mail-Verteiler verbreitet. Eine Auswahl der angeschriebenen Fachschaften befindet sich in Abbildung 3. Eine vollständige Auflistung ist jedoch nicht möglich, da nicht alle angeschriebenen Fachschaften erfasst wurden. Da allerdings die mehrheit der angeschriebenen Verteiler hierbei entweder Fachschaftslisten oder Informatikgruppen sind, wird ein starker Teilnehmerbias erwartet. So wird im Vergleich zur allgemeinen Bevölkerung ein größerer Anteil an Akademikern und sehr computeraffinen Teilnehmern erwartet. Dies bewirkt , dass die Ergebnisse in Bezug auf die finale Bewertung des IT-Sicherheitsbewusstseins nicht repräsentativ sind. Ob dies ein Nachteil für das Ziel des Erstellens eines möglichst großen Artefaktpools ist, oder doch ein Vorteil lässt sich jedoch nicht mit Sicherheit sagen.

Fachschaftslisten	
Aachen	**Bonn**
Matematik Physik Informatik	Geschichte
Chemie	Geographie
Biowissenschaften	Geodäsie und Geoinformation
Architektur	GeKoSka
Bauingenieurwesen	Ethnologie/Altamerikanistik
Maschinenbau	ELW
Rohstoffe und Entsorgungstechnik	Informatik
Materialwissenschaft und Werkstofftechnik	Chemie
Geowissenschaften und Ressourcenmanagement	Biologie
Geographie und Wirtschaftsgeographie	Archäologie
Elektrotechnik und Informationstechnik	English and American Studies
Philosophie	Agrarwissenschaften
Lehramt	
Kommunikationswissenschaft	
Wirtschaftswissenschaften	
Medizin	
Zahnmedizin	
Logopädie	

Abbildung 3. Eine Liste mit für die Vorstudie angeschriebenen Fachschaften

Auswertung der Vorstudie Die Auswertung der Vorstudie unterteilt sich in zwei Bereiche, nämlich das Sammeln von Artefakten und Reaktionen aus den Freitextantworten und die Analyse der angegebenen Reaktionen im Hinblick auf ihre Aussage über das IT-Sicherheitsbewusstsein des Teilnehmers. Für den ersten Bereich wird lediglich eine übersichtliche, tabellarische Darstellung der Datensätze mittels SPSS verwendet. Für den zweiten Bereich müssen die Da-

tensätze mit den Ergebnissen der Hauptstudie verknüpft werden. Dafür werden die Datensätze in sinnvolle Sets eingeteilt, die in der betrachteten Variable, wie zum Beispiel dem Alter oder dem angegebenen Verständnis von IT-Sicherheit, übereinstimmen. Diese Sets werden dann mit den Ergebnissen bewertet.

Erwartete Ergebnisse der Vorstudie Die erwarteten Ergebnisse der Vorstudie unterteilen sich in zwei Gruppen. Die erste Gruppe beinhaltet die gesuchten Artefakte und die dazugehörigen Reaktionen. Diese Gruppe ist für den Rest der Arbeit wichtiger, jedoch außerhalb der zweiten Studie relativ uninformativ. Die zweite Gruppe sind die erhobenen Daten bezüglich der tatsächlichen Reaktionen der Nutzer auf eine Auswahl von Artefakten, die vor der ersten Studie das Resultat einiger Überlegungen waren und als Beispiel verwendet werden.
Einige interessante Ergebnisse aus dieser zweiten Gruppe sind die Beantwortungen einiger vorher erarbeiteten Hypothesen. Diese Hypothesen sind in der Folgenden Liste zu sehen.

1. Junge Erwachsene reagieren am besten auf Artefakte
2. Personen die wenig Zeit am PC verbringen ignorieren am häufigsten Artefakte
3. Männer und Frauen reagieren verschieden auf Artefakte
4. Höhere Bildung verbessert den Umgang mit Artefakten
5. Phishing-Mails werden zu sehr großem Teil gelöscht/ignoriert
6. Eine Verlangsamung des PCs wird meist ignoriert
7. Personen die ein gutes Verständnis von IT-Sicherheit angeben benutzen nicht Internet Explorer
8. Mac-Nutzer geben ein geringeres Verständnis von IT-Sicherheit an
9. Personen die wenig Zeit am PC verbringen, geben ein geringeres Verständnis von IT-Sicherheit an

Erwartet wird jedoch auch, dass die gefundenen Daten einen starken Bias besitzen und die Stichprobe nicht groß genug ausfällt, um Aussagen über die allgemeine Bevölkerung zu treffen. Dennoch sollten die Daten einen gewissen Einblick ermöglichen. Wie groß und bedeutend dieser ist, hängt allerdings stark von der erfolgreichen Durchführung der Studie und deren Erfolg ab.

3.4 Hauptstudie

In den folgenden Abschnitten werden Details der Hauptstudie erläutert.

Ziel der Hauptstudie Das Ziel der Hauptstudie sind die Bewertungen der Reaktionen auf die einzelnen Artefakte aus einem möglichst großen Artefaktpool. Diese Bewertungen sollen es ermöglichen, mittels der angegebenen Reaktionen von Nutzern Aussagen über deren IT-Sicherheitsbewusstsein zu treffen. Zusätzlich zu der direkten Bewertung der einzelnen Reaktionen, soll auch herausgefunden werden, ob die angegebenen Artefakte tatsächlich Aussagen über

das IT-Sicherheitsbewusstsein zulassen. Beispielsweise könnten zu komplexe Artefakte für die Bewertung ungeeignet sein, wenn selbst sehr sicherheitsbewusste Nutzer nicht wissen, was eine gute Reaktion auf dieses Artefakt ist.
Ebenfalls sollen in der Vorstudie nicht gefundene Reaktionen gesammelt werden. Eine vollständige Bewertung dieser ist zwar nicht möglich, da diese nicht nachträglich in die Umfrage eingefügt werden können, jedoch können auch zu diesen neuen Elementen eine gewisse Tendenz ihres Einflusses auf die IT-Security-Awareness angegeben werden.

Aufbau der Hauptstudie Die Hauptstudie besteht aus 3 Teilen. Nach einer kurzen Begrüßung werden zunächst Artefakte genau definiert. Im Gegensatz zur Vorstudie ist eine solche Definition für die Hauptstudie von großer Bedeutung, da ein genaues Verständnis von Artefakten zur korrekten Einschätzung benötigt wird. In diesem Abschnitt werden keine Fragen gestellt.
Der zweite und größte Teil der Umfrage besteht aus den Bewertungen der Reaktionen auf einzelne Artefakte, die in der Vorstudie gesammelt wurden, sowie Einschätzungen der Aussagekraft der Artefakte selber. Die Bewertungsfragen bestehen immer aus dem Artefakt, das im Fragetext genannt und erläutert wird, sowie den dazugehörigen Reaktionen. Jede Reaktion kann mittels eines Schiebereglers mit Werten von 1 bis 20 bewertet werden. Diese Abstufung wurde gewählt, da so zwar genügend kleine Abstufungen möglich sind, jedoch nicht nachträglich eine neue Skala entworfen werden muss, weil die Bewertungen zu kleinstufig sind. Außerdem hat diese Skala eine gerade Anzahl an Stufen und erlaubt so die Tendenz des Nutzers zu erkennen. Dieser Abschnitt umfasst 6 der 10 Seiten der Hauptstudie.
Der dritte und letzte Teil der Umfrage umfasst einige demographische Fragen. Hier werden Geschlecht, Alter, Erfahrungen im Bereich IT-Sicherheit und die Zeit, die in der IT-Sicherheit oder verwandten Gebieten gearbeitet worden ist, abgefragt.
Um den Zeitaufwand pro Teilnehmer zu reduzieren und so eine möglichst hohe Rücklaufquote zu erzielen, wird die Umfrage in zwei gleich große Umfragen aufgeteilt. Hierfür bleiben der erste und dritte Abschnitt der Umfragen identisch, jedoch bekommt jede Umfrage nur die Hälfte der zu bewertenden Artefakte, nämlich nur 25 Stück. Die Zuordnung der einzelnen Teilnehmer zu der jeweiligen Umfrage wird durch SoSci-Survey durchgeführt, wobei die Teilnehmer per Zufall auf die Umfragen aufgeteilt werden.

Teilnehmer und Verbreitung der Hauptstudie Die Hauptstudie richtet sich an Experten der IT-Sicherheit, sowohl aus der Wirtschaft als auch aus der Forschung. Da es sich hierbei um eine sehr beschränkte Menge an Personen handelt und auch wirklich nur diese Personen befragt werden sollen, wird die Umfrage hauptsächlich durch persönliche E-Mails verbreitet. Einige Experten verbreiten die Studie an ihre Abteilungen und Angestellten.
Es werden über 100 Experten angeschrieben. Bei diesen handelt es sich hauptsächlich um Professoren und deren Mitarbeiter und Teilnehmer von verschiede-

nen Tagungen der IT-Sicherheit. Einige Experten werden über weitere Onlinesuchen gefunden.

Auswertung der Hauptstudie Die Auswertung der Studie erfolgt mittels SPSS. Die Rohdaten werden so tabellarisch dargestellt und anschließend visuell aufbereitet. Dafür werden aus den Bewertungen der Reaktionen auf die einzelnen Artefakte Boxplots erstellt. Boxplots werden verwendet, da sich mittels diesen komplexe statistische Werte einfach und auf einen Blick erkennbar darstellen lassen.
Ein weiterer Vorteil von Boxplots besteht darin, dass diese alle Daten erkennbar machen. Sowohl Median, als auch oberes und unteres Quartil und sogar Ausreißer sind erkennbar.
Problematisch ist lediglich die Auswertung der Einschätzungen bezüglich der Aussagekraft der Artefakte selbst. Diese lassen sich nicht einfach visuell aufbereiten und müssen deshalb in Textform ausgedrückt werden.

Erwartete Ergebnisse der Hauptstudie Auch für die Hauptstudie werden einige Hypothesen erarbeitet. Bei diesen handelt es sich hauptsächlich um Einschätzungen von Reaktionen und die erwartete Bewertbarkeit von Artefakten. Eine Auflistung der Artefakte kann in der folgenden Liste eingesehen werden.

1. Internet Explorer Nutzer sind weniger IT-Security-Aware
2. Das Verwenden eines Virenscanners ist grundsätzlich eine gute Reaktion
3. Das Ignorieren von (möglichen) Artefakten ist nie gut
4. Den Verbindungsaufbau nach einer Zertifikationswarnung abzubrechen ist die richtige Entscheidung
5. Nach einem Blue-Screen den Virenscanner zum überprüfen zu nutzen ist eine korrekte Entscheidung
6. Sicherheitsprogramme nach der Aufforderung eines unbekannten Programms zu deaktivieren ist schlecht
7. Passwörter in Klartext zu versenden weil der IT-Support dies sagt ist falsch
8. „Freezed“ ein PC ist das Abstellen des Stroms keine optimale Reaktion
9. Bei einer überhitzenden Grafikkarte (Bitcoin-Mining) dem PC durch Ausschalten eine Pause zu geben ist nicht optimal

Für die Bewertbarkeit von Artefakten existiert nur eine Hypothese. Erwartet wird, dass lediglich einfach bemerkbare und technisch nicht zu anspruchsvolle Artefakte für die Bewertung von IT-Security-Awareness verwendet werden können. Ist ein Artefakt schwer zu bemerken, so spricht eine falsche Reaktion vermutlich eher nicht für ein geringes Sicherheitsbewusstsein. Auch technisch zu anspruchsvolle Artefakte sollten für die Bewertung eher nicht geeignet sein, da das für eine korrekte Reaktion benötigte Wissen zu spezifisch ist und nicht von einem normalen Nutzer zu erwarten ist.

4 Durchführung der Studien

Der folgende Abschnitt beschreibt die tatsächliche Durchführung der Studien.

4.1 Durchführung der Vorstudie

Fragebogen	Datensätze abgeschlossen / gesamt / Klicks		
Soziodemografie (Demonstr sociodemographics	136	216	431
Gesamt	**136**	**216**	**431**

Einzelstatistik zu Ausstiegsseiten
Bitte oben den entsprechenden Fragebogen anklicken
Soziodemografie (Demonstration)

Letzte bearbeitete Seite	Datensätze abgeschlossen / gesamt / kumulativ		
Seite **10**	136	136	136
Seite **9**	0	1	137
Seite **7**	0	8	145
Seite **6**	0	7	152
Seite **5**	0	42	194
Seite **4**	0	1	195
Seite **3**	0	18	213
Seite **2**	0	3	216
Gesamt	**136**	**216**	

Insgesamt wurden 431 Aufrufe (Klicks) für diesen Fragebogen aufgezeichnet (einschließlich versehentlicher doppelter Klicks, Aufrufe durch Suchmaschinen, ...).

Abbildung 4. Eine genaue Auflistung der abgeschlossenen Datensätze der Vorstudie

Der Befragungszeitraum der Vorstudie ist von 24.06.2016 bis einschließlich 13.07.2016, also 2 Wochen und 6 Tage. Wie in Abbildung 4 zu sehen wurden 216 Fragebögen ausgefüllt, wobei 136 Datensätze vollständig ausgefüllt worden sind. Insgesamt wurde der Fragebogen 431-mal angeklickt. Abbildung 5 zeigt die Menge an Teilnehmer (orange) und Klicks (grau) pro Tag im Befragungszeitraum. Obwohl die Umfrage bereits am 24.06. gestartet wurde, sind die ersten Datensätze erst am 27.06. An diesem Datum begann die tatsächliche Verteilung des Links zum Fragebogen. Grund dafür ist, dass die Umfrage erst sehr spät startet und keine Teilnehmer verloren gehen sollen, die eventuell über das Wochenende verreist sind. Deshalb werden die Einladungen montagmorgens versendet.
Ebenfalls auffällig ist die hohe Anzahl an Abbrechern auf Seite 5, wo 42 Teilnehmer beziehungsweise 52.5% aller Abbrecher die Umfrage beendet haben. Auf

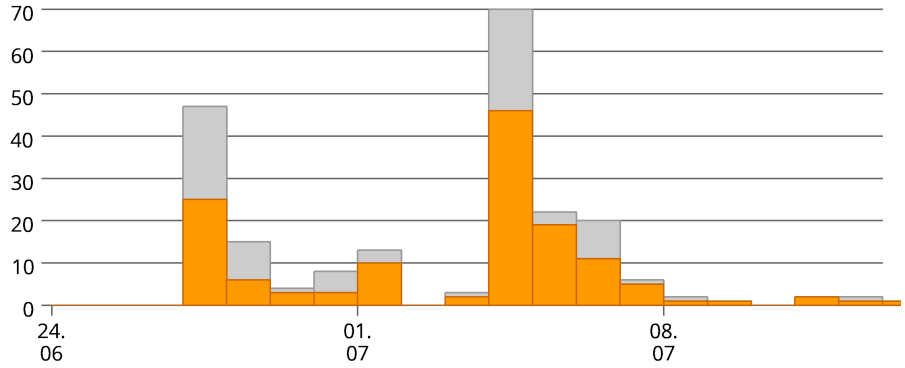

Abbildung 5. Eine zeitliche Darstellung der Teilnehmerzahlen der Vorstudie

dieser Seite befindet sich eine kurze, sehr oberflächliche Erklärung von Artefakten und eine kurze Erläuterung des folgenden Teils, nämlich der Sammlung von Artefakten und dazugehörigen Reaktionen.
Über den Grund für diese Erhöhung an Abbrüchen kann nur gemutmaßt werden. Eine Möglichkeit ist, dass die Erklärung zu Artefakten viele Teilnehmer zum Abbrechen bewegt hat, da diese nicht in ausreichenden Details erklärt hat, was Artefakte sind und die Teilnehmer sich zu unsicher waren, um fortzufahren. Dies würde zwar der anfänglichen Theorie widersprechen, dass eine genaue Erklärung für die Vorstudie nicht nötig ist, kann allerdings nicht ausgeschlossen werden. Eine weitere Möglichkeit besteht darin, dass die Teilnehmer die benötigte Zeit unterschätzt haben und auf Grund der Anweisungen für den nächsten Abschnitt die Umfrage beendet haben.

4.2 Durchführung der Hauptstudie

Der Befragungszeitraum der Hauptstudie ist von 30.08.2016 bis einschließlich 18.09.2016, also 2 Wochen und 6 Tage. Wie in Abbildung 6 deutlich zu sehen, werden zwei Mal Einladungen zur Umfrage versendet. Die erste Gruppe von Einladungen werden am 2.09.2016 versendet, die zweite Gruppe am 11.09.2016. Wie in Abbildung 7 zu sehen, sind 29 Datensätze vollständig abgeschlossen worden und insgesamt sind 41 Datensätze bearbeitet worden. Diese verteilen sich auf Fragebogen qnr04 (Questionaire Number 4) mit 14 vollständigen und insgesamt 19 teilweise ausgefüllten Datensätzen, sowie auf qnr03 (Questionaire Number 3) mit 15 vollständigen und insgesamt 22 bearbeiteten Datensätzen. Die Bezeichnungen qnr03 und qnr04 entstammen hierbei der Entwicklung der Umfragebögen, wobei qnr01 und qnr02 lediglich zum Testen benutzt wurden. Insgesamt wurde die Umfrage 101-mal angeklickt. Insgesamt wurde die deutsche Version der Umfrage 27-mal bearbeitet, während die englische Version nur 14 mal bearbeitet worden ist.

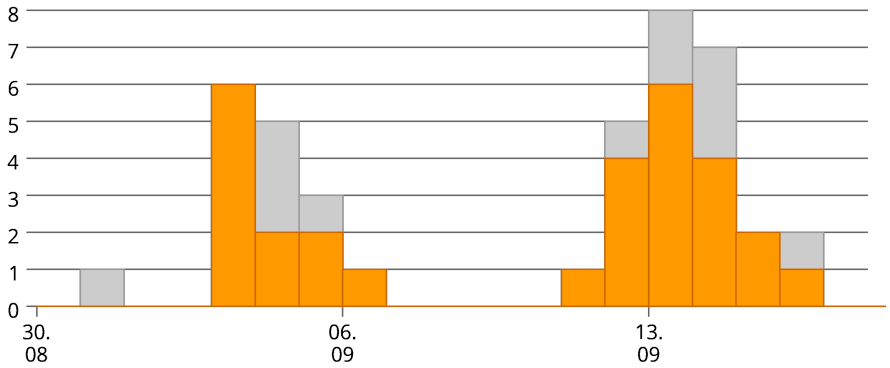

Abbildung 6. Eine zeitliche Darstellung der Teilnehmerzahlen der Hauptstudie

Fragebogen	Datensätze abgeschlossen / gesamt / Klicks		
Fragebogen qnr04	14	19	44
Fragebogen qnr3	15	22	57
Gesamt	**29**	**41**	**101**

Einzelstatistik zu Ausstiegsseiten

Bitte oben den entsprechenden Fragebogen anklicken

Fragebogen

Letzte bearbeitete Seite	Datensätze abgeschlossen / gesamt / kumulativ		
Seite **10**	14	14	14
Seite **8**	0	1	15
Seite **7**	0	1	16
Seite **5**	0	1	17
Seite **4**	0	1	18
Seite **3**	0	1	19
Gesamt	**14**	**19**	

Insgesamt wurden 44 Aufrufe (Klicks) für diesen Fragebogen aufgezeichnet (einschließlich versehentlicher doppelter Klicks, Aufrufe durch Suchmaschinen, ...).

Letzte bearbeitete Seite	Datensätze abgeschlossen / gesamt / kumulativ		
Seite **10**	15	15	15
Seite **9**	0	1	16
Seite **5**	0	3	19
Seite **3**	0	3	22
Gesamt	**15**	**22**	

Insgesamt wurden 57 Aufrufe (Klicks) für diesen Fragebogen aufgezeichnet (einschließlich versehentlicher doppelter Klicks, Aufrufe durch Suchmaschinen, ...).

Abbildung 7. Eine genaue Auflistung der abgeschlossenen Datensätze der Hauptstudie

Ebenfalls interessant ist die Verteilung der Ausstiegsseiten. Diese ist bei der Hauptstudie sehr gleichmäßig verteilt, was dafür sprichen könnte, dass es keine Seite gab, auf der grobe Probleme existieren, die eine Mehrheit der Teilnehmer zum abbrechen bewegt hat.

5 Ergebnisse und Evaluation

Die folgenden Abschnitte beschreiben die Ergebnisse der beiden Studien, sowie eine allgemeine Evaluation der Studien.

5.1 Ergebnisse der Vorstudie

Die wichtigsten Ergebnisse der Vorstudie sind die gefundenen Artefakte. Eine vollständige Auflistung aller gefundener Artefakte befindet sich in Abbildung 8. Hierbei stammen 9 Artefakte (Phishing-Mail mit Anhang, Phishing-Mail mit Link, Verlangsamung des Systems, Freeze des Systems, kurzzeitiger Freeze des Systems, Aufforderung zum Deaktivieren des Virenscanners, Festplattengeräusche, Zertifikatswarnung und Spam vom eigenen Account) aus vor der Studie stattgefundenen Überlegungen. Im Verlauf der Vorstudie wurden also 41 neue Artefakte gefunden. Diese beinhalten jedoch auch einen Austausch mit Teilnehmern des ITS.APT-Projekts, in dessen Verlauf ebenfalls eine Liste mit Artefakten erstellt worden ist.
Interessant und in starken Kontrast zu der vor den Studien aufgestellten Hypothese sind die angegebenen Reaktionen auf Verlangsamungen des PCs. Die Hypothese war, dass dieses Artefakt meistens ignoriert wird. Die gesammelten Daten widersprechen dieser Hypothese jedoch stark, da lediglich 6 von 196 Personen angegeben haben, dies zu ignorieren. Dies bedeutet, das entweder die Hypothese falsch ist, oder die Angaben inkorrekt sind, möglicherweise da die Reaktion sehr falsch wirkt.
Bestätigt wurde hingegen die Hypothese, dass Nutzer des Internet Explorers im Schnitt ein geringeres Verständnis von IT-Sicherheit angeben. So liegt über alle Teilnehmer der Vorstudie der Median bei 15, bei Nutzern des Internet Explorers jedoch nur bei 8. Insgesamt haben nur 10 Teilnehmer angegeben, diesen Browser zu verwenden. Im Kontrast dazu, und zur aufgestellten Hypothese, geben Nutzer eines Macs ein sehr stark erhöhtes Verständnis von IT-Sicherheit an, mit einem Median von 19 bei 25 Nutzern dieses Betriebssystems. Zur Erinnerung, die Bewertungsskala geht von 1 (sehr gering) bis 20 (sehr hoch).
Entgegen der 3. Hypothese der Vorstudie gibt es keine grundsätzlichen Unterschiede zwischen den Reaktionen zwischen männlichen und weiblichen Teilnehmern. Lediglich bei Reaktionen, die genaueres Überprüfen der Artefakte beinhalten, lässt sich ein deutlich erhöhter Prozentsatz an männlichen Teilnehmern erkennen, die diese Reaktion angeben. Des ist beispielsweise der Fall beim Überprüfen eines Anhangs einer Phishing-E-Mail, bevor entschieden wird, wie mit der Datei umgegangen wird.

1. Verlangsmung des PCs
2. Festplattengeräusche
3. Hoher Leistungsverbrauch im Leerlauf
4. Abgelaufenes SSL-Zertifikat (kurz, lang, invalid)
5. Unbekannte Prozesse
6. Unauthorisierte Online-Abbuchungen
7. Nicht selbstverfasste E-Mails im "Gesendet"-Ordner
8. Fremde Curserbewegungen
9. Browserwarnung fehlendes https
10. Verschlüsselung der Festplatte
11. Phishing-Mails
12. Blue Screen
13. Änderungen an Einstellungen
14. Unerwünschte Toolbars
15. Spontanes Öffnen von Webseiten/Programmen
16. Unerwünschte Downloads
17. Aufflackern einer Console
18. Zugriffswarnung Webcam
19. Ungewöhnliche Darstellung einer Webseite
20. IFrames / Hidden Frames
21. Ungewollte Weiterleitung
22. XSS-Warnungen
23. s/mime ohne Verifizierung
24. Mails mit gefälschtem Absender
25. Unsichere Anhänge
26. Langsames Hochfahren
27. Unbekannte neue Ordnerstrukturen
28. Virenschutzwarnung
29. UAC Pop-Up
30. Unbekannte Icons in der Taskleiste
31. Login mit korrektem PW nicht möglich
32. IDS Warnung
33. Programm verlangt inoffizielle Updates
34. Internetzugriff von Programmen ohne Onlinefunktionalitäten
35. Versteckte Installationen von anderen Programmen während Installation
36. Fehlende Hardware
37. Anrufer des vermeintlichen Kundendienstes verlang PW
38. Fremde Hardware
39. Unbekannte Updates, die auf der Herstellerwebseite nicht existieren
40. Spam vom eigenen Account
41. Konstantes Laden einer Webseite
42. Unbekannte Änderungen an der eigenen Webseite
43. Lüftergeräusche
44. Gehackter Mailaccount eines Kollegen
45. Unbekannte Bounces
46. Warnungen von einem fremden Antivirus-Programm
47. Verzögerung beim Tippen
48. Geänderte Uhrzeit/Datum

Abbildung 8. Die Vollständige Auflistung der in der Vorstudie gefundenen Artefakte

5.2 Ergebnisse der Hauptstudie

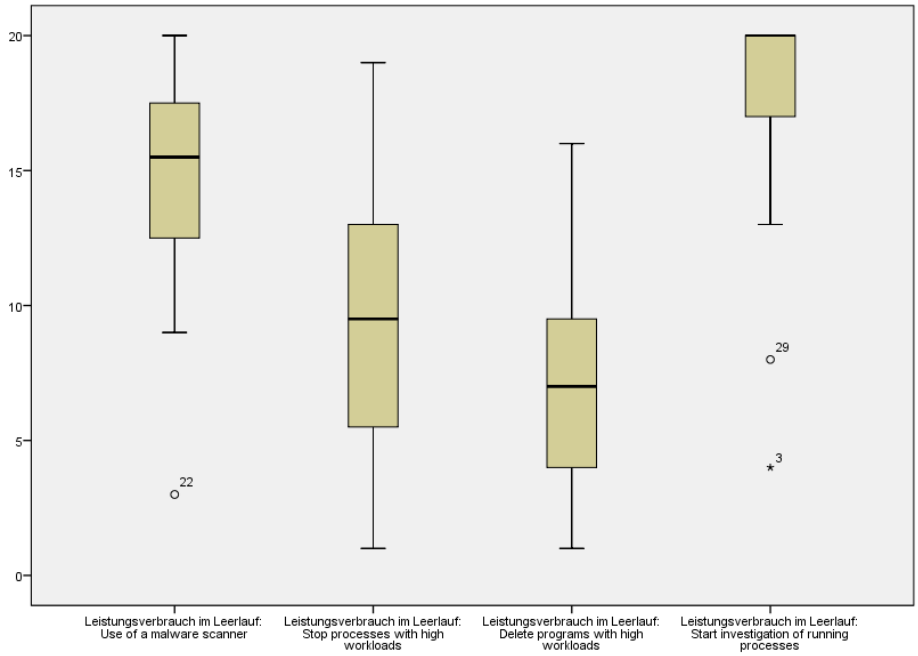

Abbildung 9. Die als Boxplots aufbereiteten Datensätze zum Artefakt „erhöhter Leistungsverbrauch im Leerlauf“ (n=16)

In diesem Abschnitt werden einige ausgewählte Ergebnisse der Hauptstudie vorgestellt, die aus bestimmten Gründen besonders interessant sind. Alle Diagramme, auch die hier behandelten, können im Anhang in Originalgröße eingesehen werden. Zunächst werden jedoch einige grundsätzliche Ergebnisse zusammengefasst. Wie erwartet und auch in den Hypothesen zur Hauptstudie beschrieben, ist das Ignorieren von Artfakten grundsätzlich sehr schlecht bewertet worden. Die einzigen Ausnahmen hierbei sind Artefakte, bei denen das Ignorieren jegliche Gefahr für den Nutzer verhindert, wie zum Beispiel bei Phishing E-Mails. Auch bei diesen ist das Löschen der E-Mail und das Blocken des Absenders jedoch besser bewertet, da dies auch zukünftige Angriffe verhindern kann. Im Kontrast zur aufgestellten Hypothese ist das Verwenden eines Virenscanner nach einem Blue Screen nur neutral bewertet worden.
Insgesamt wurden sehr vorsichtige Reaktionen, wie das Blockieren von Zugriffen aus möglicherweise schädlichen oder unbekannten Quellen eher positiv bewertet und jegliche Art des Zulassens von Artefakten negativ bewertet. Inkonsistent ist jedoch im Gegensatz dazu das Informieren von Tech Support und Vorgesetzten bewertet worden, was mal eher positiv und mal eher negativ bewertet worden

ist. Bestätigt wurde auch die Hypothese, dass das Abbrechen der Verbindung nach einer Zertifikatswarnung grundsätzlich eine korrekte Entscheidung ist.
Zur besseren Visualisierung wurden aus den Datensätzen Boxplots erstellt. Ein gutes Beispiel ist in Abbildung 9 zu sehen. Die Abbildung zeigt die Ergebnisse zum Artefakt „erhöhter Leistungsverbrauch im Leerlauf“ in Bezug auf die vier Reaktionen „Use of a malware scanner“, „Stop processes with high workloads“, „Delete programs with high workloads“ und „Start investigation of running processes“. Deutlich zu erkennen sind hier jeweils zwei positiv bewertete Reaktionen und zwei tendenziell eher negativ bewertete Reaktionen. Sowohl die Verwendung eines Virenschutzprogramms als auch das Untersuchen von laufenden Prozessen wurden sehr positiv bewertet. Bei Ersterem liegt der Median bei 15.5 und bei Zweiterem sogar bei 20, der höchsten möglichen Bewertung. Die anderen beiden Reaktionen, das Stoppen von Prozessen mit hohem Ressourcenverbrauch und das Löschen von Programmen mit hohem Ressourcenverbrauch wurden beide eher negativ bewertet. Der Median liegt bei dem Ersten bei 9.5 und bei dem Zweiten bei 7. Deutlich zu sehen ist jedoch auch, dass die Werte bei allen Reaktionen, besonders bei den beiden negativen, sehr weit gestreut sind. Besonders bei der zweiten Reaktion, „Stop processes with high workloads“ ist diese Streuung sehr stark. Hier liegt die Interquartildistanz, also die Differenz der Werte in den beiden Quartilen, bei 8 (40% der Gesamtlänge) und die Differenz zwischen Minimum und Maximum bei 18. Trotz den offensichtlichen Differenzen der Meinungen über die Reaktionen haben 14 Teilnehmer das Artefakt „Leistungsverbrauch im Leerlauf“ als aussagekräftig im Hinblick auf die IT-Security-Awareness des Nutzers bewertet und 4 Teilnehmer als nicht aussagekräftig.

Ein weiteres interessantes Beispiel ist das in Abbildung 10 dargestellte Artefakt „Login mit richtigem Passwort schlägt fehl“. Hier werden die Reaktionen „Contact the provider's tech support“, „Try variations of his password“, „Use of a malware scanner“ und „Reset the password“ bewertet. Hier wurden die erste und vierte Reaktion überwiegend positiv bewertet, die dritte Reaktion sehr neutral, wenn auch mit einer leicht negativen Tendenz, und die zweite Reaktion eher negativ. Besonders interessant ist hierbei die sehr hohe Varianz der Bewertungen. Diese liegt bei allen Reaktionen außer der ersten bei mehr als 40. Die hohe Streuung der Bewertungen lässt sich auch in der Abbildung sehen, da die einzelnen Boxplots nahezu die gesamte Skala abdecken und auch die Quartile sehr lang sind. Dies ist besonders sichtbar bei der vierten Reaktion, bei der das untere Quartil 9 Einheiten lang ist und die gesamte Interquartildistanz bei 13 liegt. Dies spricht für sehr große Meinungsunterschiede zwischen den einzelnen Teilnehmern. Das bedeutet, dass für eine sicherere Bewertung größere Stichproben hilfreich wären, besonders da die großen Unterschiede doch dafürsprechen, dass eventuell noch detailliertere Befragungen notwendig sein könnten um zu verstehen, wie diese zu Stande kommen. Trotz den sehr großen Meinungsdifferenzen haben 11 Personen das Artefakt als aussagekräftig über die IT-Security-Awareness eines Nutzers bewertet und nur 5 Personen haben es als nicht aussagekräftig bewertet. Die Diskrepanz zwischen der Stichprobengröße 14 und den

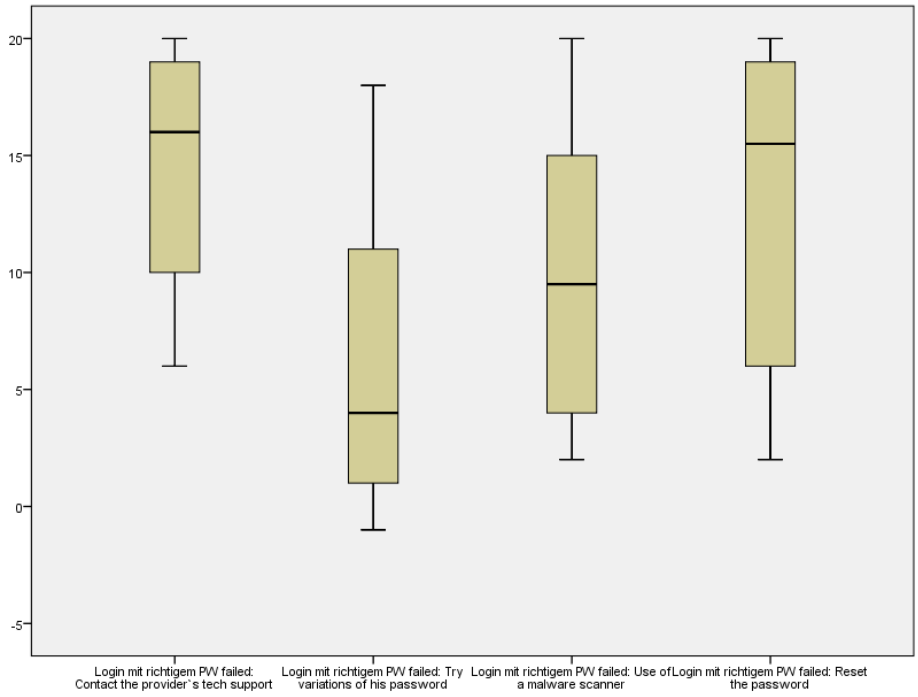

Abbildung 10. Die als Boxplots aufbereiteten Datensätze zum Artefakt „Login mit richtigem Passwort schlägt fehl“ (n=14)

abgegebenen Bewertungen 16 wird in der Evaluation diskutiert.

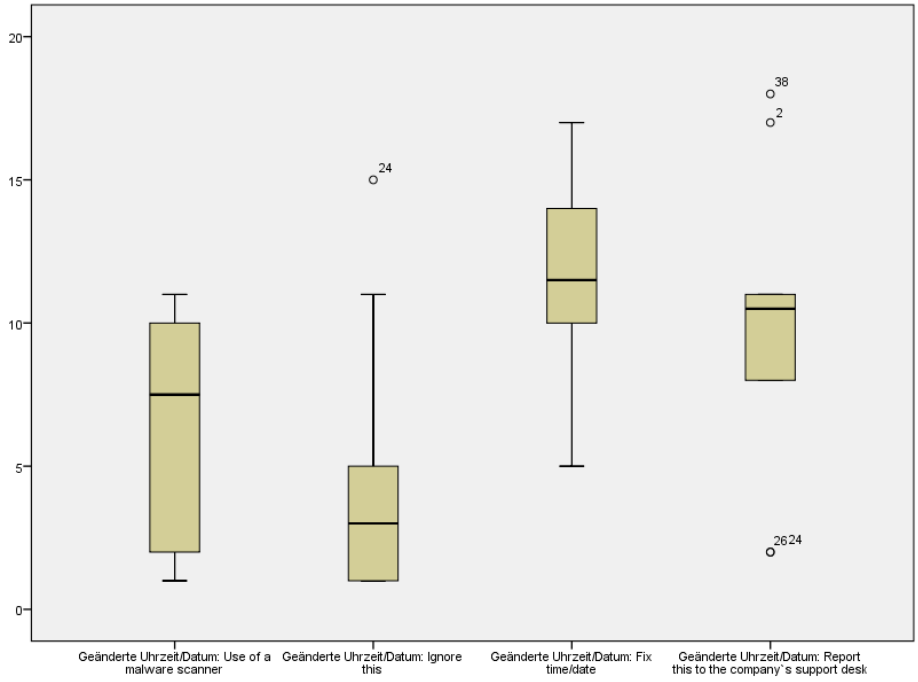

Abbildung 11. Die als Boxplots aufbereiteten Datensätze zum Artefakt „Geänderte Uhrzeit/Datum“ (n=10)

Ein relativ überraschendes Beispiel sind die Bewertungen zum Artefakt „Geänderte Uhrzeit/Datum“. Bei diesem Artefakt werden die Reaktionen „Use of a malware scanner“, „Ignore this [artifact]“, „Fix time/date“ und „Report this to the company’s support desk“ bewertet. Hier wurden die ersten beiden Reaktionen negativ bewertet und die letzten beiden liegen mittig auf der Bewertungsskala. Besonders interessant ist hierbei, dass das Verwenden eines Virenschutzprogramms tatsächlich konstant eher negativ bewertet wurde, obwohl dieses in 75% der Artefakte positiv oder neutral bewertet wurde. Unglücklicherweise kann über mögliche Erklärungen nur gemutmaßt werden. Auch die Suche nach Hilfe beim zuständigen Tech Support ist überraschend wenig positiv bewertet worden, obwohl das Bitten um Hilfe eigentlich bei Ungewissheit der Nutzer eher als korrektes Handeln erwartet worden ist. Auch hier fehlen genaue Erklärungen, wieso diese Bewertungen getroffen wurden.

5.3 Evaluation

Obwohl im Verlauf der Arbeit große Mengen an Daten gesammelt wurden, existieren einige Problematiken, sowohl mit der Methodik, als auch den verwendeten Auswertungswerkzeugen. Ein Beispiel, an dem Problematiken erläutert werden können, ist das Artefakt „Fehlende Hardware“, zu sehen in Abbildung 12. Besonders problematisch ist die Reaktion „Get new hardware at the company's support desk concealing the theft“. Offensichtlich ist diese Reaktion schlecht. Diese Meinung haben auch mehr als die Hälfte der Teilnehmer der Studie. Dennoch verteilen sich die restlichen 50% der Bewertungen über den gesamten Bewertungsbereich. Das Maximum liegt tatsächlich beim Maximum der Skala, nämlich 20. Dies sollte eigentlich nicht der Fall sein, da das Verheimlichen von Diebstahl definitiv nicht für die IT-Security-Awareness von Nutzern spricht. Diese Problematik könnte verschiedene Gründe haben. Einerseits könnte es für ein grundsätzliches Missverständnis der Bewertungsskala sprechen, also, dass die Skala verkehrt herum interpretiert worden ist. Andererseits könnte es ein Indiz dafür sein, dass die Fragen, beispielsweise aus Zeitgründen, nicht vollständig gelesen worden sind und so fehlerhaft bewertet worden sind. Diese Fehler können, besonders auf Grund der leider sehr geringen Stichprobengröße der Hauptstudie, die Ergebnisse in zu großem Maß verzerren. Zwar wirken sich solche Verzerrungen stärker auf den Durchschnitt als auf den hier verwendeten Median aus, aber dennoch werden hier Werte verzerrt, die andere Rückschlüsse verhindern könnten.

Eine weitere Problematik kann anhand eines weiteren Beispiels veranschaulicht werden. Abbildung 13 zeigt die Boxplots für das Artefakt „Konstantes Laden“. Dieses Artefakt bezieht sich auf das kontinuierliche Laden von Daten auf einer Webseite. 5 Teilnehmer haben angegeben, dieses Artefakt sei nützlich zur Bewertung von IT-Security-Awareness, 8 Teilnehmer haben dem widersprochen. Dennoch können hier zwei Problematiken aufgezeigt werden. Einerseits liegen alle Reaktionen in ihren Bewertungen sehr eng zusammen, was eine tatsächliche Aussage über das korrekte Vorgehen in Anbetracht des Artefakts unmöglich macht, andererseits sind alle Reaktionen eher negativ. Dies erzeugt das Problem, dass Nutzer bei diesem Artefakt im Rahmen einer Messung ihrer IT-Security-Awareness in allen Fällen eine negative Bewertung erhalten würden. Dies macht das Artefakt unbrauchbar zur Bewertung von IT-Security-Awareness, obwohl 38% der Teilnehmer dieser Aussage widersprechen.

Ein weiteres Problem liegt im Umgang mit SPSS. Bei Diagrammen über mehrere Reaktionen, so wie den vergleichenden Diagrammen mit den Boxplots aller Reaktionen zu einem Artefakt, werden einzelne Datensätze entfernt, wenn der Teilnehmer bei dem Artefakt eine der Reaktionen nicht bewertet hat. Dies reduziert die bereits geringe Stichprobengröße weiter. Dennoch sind diese Diagramme zu gut zum Vergleichen und um einen Überblick zu bekommen, als dass auf diese verzichtet werden könnte.
Eine weitere Problematik liegt im Umgang mit zusätzlich erhobenen Daten, wie zum Beispiel fehlende Reaktionen zu Artefakten. Diese lassen sich nur sehr

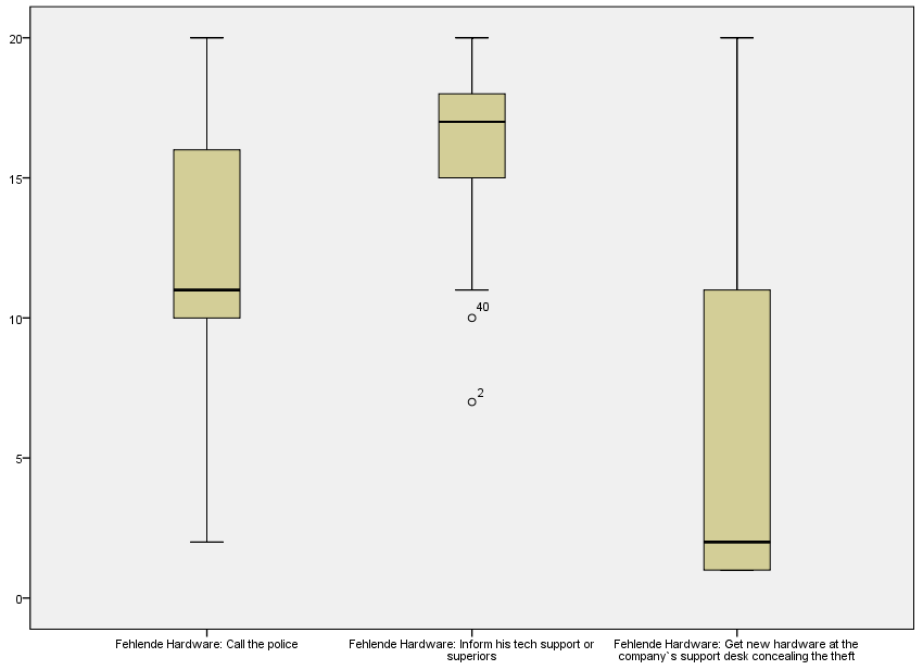

Abbildung 12. Die als Boxplots aufbereiteten Datensätze zum Artefakt „Fehlende Hardware“ (n=13)

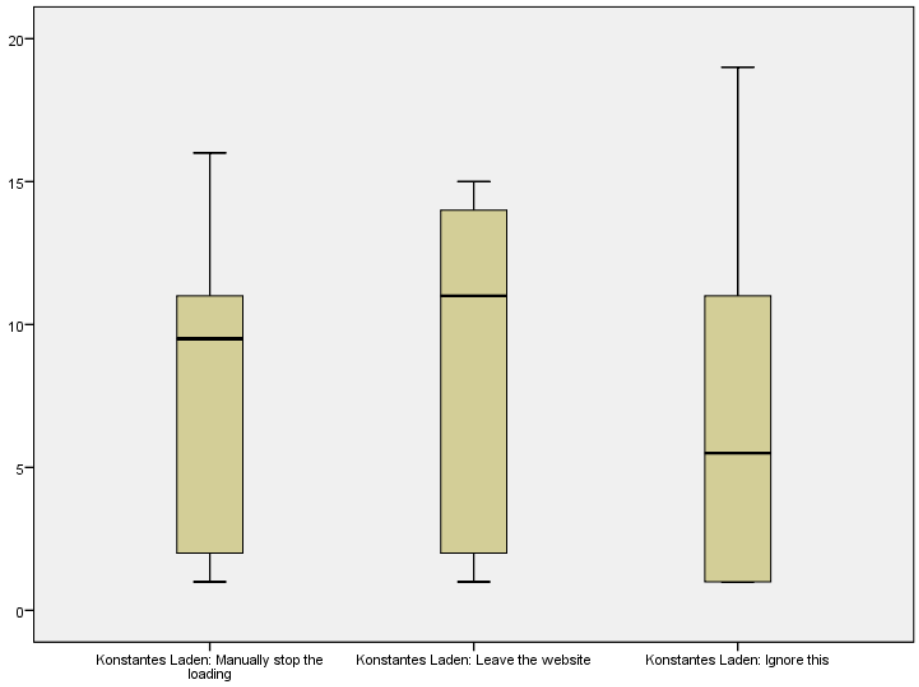

Abbildung 13. Die als Boxplots aufbereiteten Datensätze zum Artefakt „Konstantes Laden“ (n=10)

schwer auswerten und visuell aufbereiten, und sind somit nahezu unbrauchbar für mögliche, auf dieser Arbeit aufbauende Projekte. Die einzige Möglichkeit, um diese Daten in Zukunft zu verwenden, ist für das Erstellen einer zweiten Studie dieser Art, in der diese für eine neue, vielleicht umfassendere Bewertung verwendet werden.

6 Zusammenfassung

Zusammenfassend lässt sich sagen, dass IT-Security-Awareness noch reichlich Ansatzpunkte für Forschung besitzt, wie klare, nachvollziehbare Definitionen, ein klares Verständnis über die Folgen von verbesserter IT-Security-Awareness und darüber, wie man IT-Security-Awareness messen und steigern kann. Es handelt sich hierbei um eines der wichtigsten Themengebiete zur Entwicklung von effektiven Methoden um IT-Sicherheit in Unternehmen zu unterstützen, da nicht nur die technische Seite der IT-Sicherheit zum Schutz von sensiblen Inhalten und Infrastrukturen entscheidend ist, sondern auch eben jene Seite, die sich mit den Nutzern eben jener Inhalte und Infrastrukturen auseinandersetzt. Dieser Aspekt der IT-Sicherheit ist auch in Zeiten der stark voranschreitenden Digitalisierung und der zunehmenden Anzahl von an digitale Systeme gebundenen Existenzen noch nicht ausreichend gut erforscht und verstanden.
Im Verlauf dieser Arbeit sind zwei Studien entworfen und durchgeführt worden, deren endgültiges Ziel die Unterstützung von Forschung zum besseren Verständnis von IT-Security-Awareness ist. Insbesondere das Ziel der Entwicklung einer Möglichkeit zur Messung von IT-Security-Awareness soll mit den gesammelten Daten unterstützt werden. Diese wird im Rahmen des ITS.APT-Projekts umgesetzt.
Die Studien haben eine große Menge an Artefakten gefunden und Reaktionen auf diese Artefakte von Experten der IT-Sicherheit bewerten lassen. Dabei sind viele nützliche Daten gefördert worden, auch wenn im späten Verlauf der Arbeit einige Probleme in Methodik und Durchführung aufgetreten sind. Dennoch sollten die gefundenen Resultate und Überlegungen für zukünftige Arbeiten von Nutzen sein können, beispielsweise zur Entwicklung einer Skala mittels der die IT-Security-Awarness von Nutzern genau definiert werden kann.
Abschließend lässt sich noch sagen, dass diese Ergebnisse zwar vor allem in einem auf Unternehmen fokussierten Kontext nutzbar sind, aber das dennoch auch private Nutzer von einem besseren Verständnis von Aspekten der IT-Sicherheit profitieren können. Diese befinden sich im Internet genau wie Unternehmen stets Gefahren ausgesetzt, die ein gewissen Risiko darstellen können. Auch hier kann der korrekte Umgang mit Inhalten und Funktionen grobe Gefahren mitigieren. Zwar ist vollständige Sicherheit nahezu unmöglich zu gewährleisten, aber ein besseres Verständnis kann es Angreifern wenigstens schwerer machen und so vielleicht nach und nach mit allgemein steigendem Verständnis die Menge an erfolgreichen Angriffen doch reduzieren.

7 Ausblick

Die großen Mengen an erhobenen Daten ermöglichen ausgiebige, weiterführende Forschungen. Vor allem werden diese im Rahmen des ITS.APT-Projekts weiterverwendet, um dabei zu unterstützen, ein Modell zur Quantifizierung und Messung von IT-Security-Awareness zu entwickeln. Eins der wichtigsten Ziele dabei ist, IT-Security-Awareness vergleichbar zu machen und eventuell persönliche Stärken und Schwächen in Teilgebieten messbar zu machen.
Weiter mögliche Folgestudien könnten auf den gefundenen Daten aufbauen und Erkenntnisse verstärken oder im Zweifelsfall korrigieren. Besonders wichtig wären hier ähnliche Studien, die die gefundenen Erweiterungen, also beispielsweise zusätzliche Reaktionen, integrieren und eine höhere Stichprobengröße erreichen. Besonders eine größere Menge an Teilnehmern wäre wichtig, damit die gefundenen Ergebnisse stabiler werden und ihre Aussagekraft gefestigt wird. Auch zusätzliche Detailfragen wären möglicherweise wichtig, um einige der gefundenen Daten besser zu verstehen.

Literatur

1. Norbert Pohlmann and Hartmut F Blumberg. *Der IT-Sicherheitsleitfaden*. mitp, 2004.
2. Die menschliche seite der it-sicherheit. `http://www.heise.de/microsites/das-insider-portal/sicherheit/infografik-die-menschliche-seite-der-it-sicherheit/150/475/1563/`. abgerufen am 07.09.2016.
3. Arnolf Sykosch, Matthias Wübbeling, Oliver Matula, Christian Kollee. State of the art in it security awareness testing, 2014.
4. Bilal Khan, Khaled S Alghathbar, Syed Irfan Nabi, and Muhammad Khurram Khan. Effectiveness of information security awareness methods based on psychological theories. *African Journal of Business Management*, 5(26):10862, 2011.
5. ≪einladungsmail≫ anleitung zum erstellen einer einladungsmail. `https://www.onlineumfragen.com/pic/downloads/onlineumfragen_quickstart_Einladungsmail.pdf`. abgerufen am 13.09.2016.
6. Michael T Roberson and Eric Sundstrom. Questionnaire design, return rates, and response favorableness in an employee attitude questionnaire. *Journal of Applied Psychology*, 75(3):354, 1990.
7. Ellen Taylor-Powell. Questionnaire design: Asking questions with a purpose. *University of Wisconsin Extension*, 1998.
8. Mick P Couper, Michael W Traugott, and Mark J Lamias. Web survey design and administration. *Public opinion quarterly*, 65(2):230–253, 2001.
9. Die richtige feldzeit bei online-befragungen. `https://www.tns-infratest.com/presse/pdf/autorenbeitraege/Artikel_PuA_6-2005_Helmut_Leopold.pdf`. abgerufen am 13.09.2016.
10. Standards zur qualitätssicherung für online-befragungen. `http://www.adm-ev.de/fileadmin/user_upload/PDFS/Onlinestandards_D.PDF`. abgerufen am 13.09.2016.

11. Melissa DA Carlson and R Sean Morrison. Study design, precision, and validity in observational studies. *Journal of palliative medicine*, 12(1):77–82, 2009.
12. Aggeliki Tsohou, Spyros Kokolakis, Maria Karyda, and Evangelos Kiountouzis. Investigating information security awareness: research and practice gaps. *Information Security Journal: A Global Perspective*, 17(5-6):207–227, 2008.
13. De-humanizing human vulnerability assessment. `https://infocon.org/cons/SyScan/SyScan%202015%20Singapore/SyScan%202015%20Singapore%20presentations/SyScan15%20Laura%20Bell%20-%20Caring%20Less%20-%20De-humanizing%20Human%20Vulnerability%20Assessment.pdf`. abgerufen am 05.07.2016.
14. Safestack - agile application security. `https://safestack.io/index.html`. abgerufen am 26.09.2016.
15. State of the art in it security awareness testing. `https://itsec.cs.uni-bonn.de/itsapt/`. abgerufen am 25.06.2016.
16. Soscy survey(ofb - der onlinefragebogen). `https://www.soscisurvey.de/`. abgerufen am 17.09.2016.
17. Ibm spss - ibm analytics. `http://www.ibm.com/analytics/us/en/technology/spss/#what-is-spss`. abgerufen am 17.09.2016.

Eigenständigkeitserklärung

Hiermit versichere ich, dass ich die Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie Zitate kenntlich gemacht habe.

Bonn, den

Tim Jüliger