Advances in Intelligent Systems and Computing 593

Denise Nicholson *Editor*

Advances in Human Factors in Cybersecurity

Proceedings of the AHFE 2017 International Conference on Human Factors in Cybersecurity, July 17–21, 2017, The Westin Bonaventure Hotel, Los Angeles, California, USA

Springer

Advances in Intelligent Systems and Computing

Volume 593

About this Series

The series “Advances in Intelligent Systems and Computing” contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing.

The publications within “Advances in Intelligent Systems and Computing” are primarily textbooks and proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

More information about this series at http://www.springer.com/series/11156

Denise Nicholson
Editor

Advances in Human Factors in Cybersecurity

Proceedings of the AHFE 2017 International Conference on Human Factors in Cybersecurity, July 17–21, 2017, The Westin Bonaventure Hotel, Los Angeles, California, USA

Springer

Editor
Denise Nicholson
Soar Technology, Inc.
Belle Isle, FL
USA

ISSN 2194-5357 ISSN 2194-5365 (electronic)
Advances in Intelligent Systems and Computing
ISBN 978-3-319-60584-5 ISBN 978-3-319-60585-2 (eBook)
DOI 10.1007/978-3-319-60585-2

Library of Congress Control Number: 2017943021

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Advances in Human Factors and Ergonomics 2017

AHFE 2017 Series Editors

Tareq Z. Ahram, Florida, USA
Waldemar Karwowski, Florida, USA

8th International Conference on Applied Human Factors and Ergonomics and the Affiliated Conferences

Proceedings of the AHFE 2017 International Conference on Human Factors in Cybersecurity, July 17−21, 2017, The Westin Bonaventure Hotel, Los Angeles, California, USA

Advances in Affective and Pleasurable Design	*WonJoon Chung and Cliff (Sungsoo) Shin*
Advances in Neuroergonomics and Cognitive Engineering	*Carryl Baldwin*
Advances in Design for Inclusion	*Giuseppe Di Bucchianico and Pete Kercher*
Advances in Ergonomics in Design	*Francisco Rebelo and Marcelo Soares*
Advances in Human Error, Reliability, Resilience, and Performance	*Ronald L. Boring*
Advances in Human Factors and Ergonomics in Healthcare and Medical Devices	*Vincent G. Duffy and Nancy Lightner*
Advances in Human Factors in Simulation and Modeling	*Daniel N. Cassenti*
Advances in Human Factors and System Interactions	*Isabel L. Nunes*
Advances in Human Factors in Cybersecurity	*Denise Nicholson*
Advances in Human Factors, Business Management and Leadership	*Jussi Kantola, Tibor Barath and Salman Nazir*
Advances in Human Factors in Robots and Unmanned Systems	*Jessie Chen*
Advances in Human Factors in Training, Education, and Learning Sciences	*Terence Andre*
Advances in Human Aspects of Transportation	*Neville A. Stanton*

(continued)

(continued)

Advances in Human Factors, Software, and Systems Engineering	*Tareq Z. Ahram and Waldemar Karwowski*
Advances in Human Factors in Energy: Oil, Gas, Nuclear and Electric Power Industries	*Paul Fechtelkotter and Michael Legatt*
Advances in Human Factors, Sustainable Urban Planning and Infrastructure	*Jerzy Charytonowicz*
Advances in the Human Side of Service Engineering	*Louis E. Freund and Wojciech Cellary*
Advances in Physical Ergonomics and Human Factors	*Ravindra Goonetilleke and Waldemar Karwowski*
Advances in Human Factors in Sports, Injury Prevention and Outdoor Recreation	*Tareq Z. Ahram*
Advances in Safety Management and Human Factors	*Pedro Arezes*
Advances in Social & Occupational Ergonomics	*Richard Goossens*
Advances in Ergonomics of Manufacturing: Managing the Enterprise of the Future	*Stefan Trzcielinski*
Advances in Usability and User Experience	*Tareq Ahram and Christianne Falcão*
Advances in Human Factors in Wearable Technologies and Game Design	*Tareq Ahram and Christianne Falcão*
Advances in Communication of Design	*Amic G. Ho*
Advances in Cross-Cultural Decision Making	*Mark Hoffman*

Preface

Our daily life, economic vitality, and national security depend on a stable, safe, and resilient cyberspace. We rely on this vast array of networks to communicate and travel, power our homes, run our economy, and provide government services. Yet, cyberintrusions and attacks have increased dramatically over the last decade, exposing sensitive personal and business information, disrupting critical operations, and imposing high costs on the economy. The human factor at the core of cybersecurity provides greater insight into this issue and highlights human error and awareness as key factors, in addition to technical lapses, as the areas of greatest concern. This book focuses on the social, economic, and behavioral aspects of cyberspace, which are largely missing from the general discourse on cybersecurity. The human element at the core of cybersecurity is what makes cyberspace the complex, adaptive system that it is. An inclusive, multidisciplinary, holistic approach that combines the technical and behavioral element is needed to enhance cybersecurity. Human factors also pervade the top cyberthreats. Personnel management and cyberawareness are essential for achieving holistic cybersecurity.

This book will be of special value to a large variety of professionals, researchers, and students focusing on the human aspect of cyberspace, and for the effective evaluation of security measures, interfaces, user-centered design, and design for special populations, particularly the elderly. We hope this book is informative, but even more that it is thought provoking. We hope it inspires, leading the reader to contemplate other questions, applications, and potential solutions in creating safe and secure designs for all.

A total of six sections presented in this book:

I. Cybersecurity Tools and Analytics
II. Cybersecurity Interface and Metrics
III. Human Factors in Cyber-Warfare
IV. Human Dimension and Visualization for Cybersecurity
V. Cybersecurity Training and Education
VI. Privacy and Cultural Factors in Cybersecurity

Each section contains research paper that has been reviewed by members of the International Editorial Board. Our sincere thanks and appreciation to the board members as listed below:

Grit Denker, USA
Ritu Chadha, USA
Frank Greitzer, USA
Jim Jones, USA
Anne Tall, USA
Mike Ter Louw, USA
Elizabeth Whitaker, USA
Hanan A. Alnizami, USA

July 2017 Denise Nicholson

Contents

Cybersecurity Tools and Analytics

Cybersecurity Interface and Metrics

Human Factors in Cyber-Warfare

Human Dimension and Visualization for Cybersecurity

Cybersecurity Training and Education

Privacy and Cultural Factors in Cybersecurity

Cybersecurity Tools and Analytics

Cybersecurity Management Through Logging Analytics

Michael Muggler, Rekha Eshwarappa, and Ebru Celikel Cankaya(✉)

Department of Computer Science, University of Texas at Dallas, Richardson, TX 75080, USA
{mxm121531,rbe140030,exc067000}@utdallas.edu

Abstract. To make cybersecurity efforts proactive rather than solely reactive, this work proposes using machine learning to process large network related data: We collect various performance metrics in a network and use machine learning techniques to identify anomalous behavior. We introduce the novel idea of using weighted trust to prevent corruption of classifiers. Our design combines all aspects of a log management system into one distributed application for a data center to effectively offer logging, aggregation, monitoring and intelligence services. For this, we employ a three-component log management system: (1) to actively extract metrics from machines, (2) to aggregate and analyze extracted metrics to detect anomalous behavior, and (3) to allow reviewing collected metrics and to report on anomalous behavior observed. Our system runs at network and application layers and is concerned with risk mitigation and assessment. Several machine learning techniques are compared w.r.t. their classification, as well as detection performances.

Keywords: Log management system · Cybersecurity · Anomaly detection

1 Introduction

Cybersecurity efforts can greatly benefit from machine learning as machines can classify log data in near real time, and learn as attackers become more sophisticated. Moreover, machine learning algorithms improve the decisions made based on log data.

Large organizations have very large clusters of machines that support business critical operations. These machines are networked and in many cases provide services exposed over the public internet. These machines are running FOSS and or COTS software that can become vulnerable due to variety of reasons. With well-resourced, sophisticated and motivated adversaries it becomes a management problem to effectively defend data centers. It is increasingly important for data center operators to properly monitor, detect and mitigate against attacks as fast as possible.

Traditional cybersecurity in an organization relies on human operators to collect and interpret log data. While automated rule-based IDS exist to help in this effort, attackers constantly learn how to thwart those systems. Machine learning mechanisms can be retrained when attackers change their modus operandi. Cybersecurity efforts can then greatly benefit from machine learning. Machines can classify log data in near real time and learn as attackers become more sophisticated.

D. Nicholson (ed.), *Advances in Human Factors in Cybersecurity*,
Advances in Intelligent Systems and Computing 593, DOI 10.1007/978-3-319-60585-2_1

Automated solutions to provide cyber-defense against computer networks can be classified into two approaches: proactive and reactive. The most commonly employed approach is reactive. In this approach it is not until a vulnerability is exploited that administrators react to mitigate it [1]. It is the costly havoc directly resulting from the exploit, though, that prompts cybersecurity professionals to act more proactively. The proactive approach uses countermeasures running in real time. These countermeasures have the goal of preventing exploits before happening [1]. IDS, firewalls, network attached antivirus and antimalware blades, proxies and network quarantine servers are examples of proactive measures [3]. IDS and SIEM systems collect data from multiple sources combined with rule-based filtering techniques to distinguish malicious and benign activities [4]. These rules are defined by a policy, which should be the center of focus because as attackers learn, the policies must also change to reflect the sophistication of the adversary.

Machine learning allows us to consider a new form of proactive defense. Using machine learning, we can combine traditional rule-based IDS with an intelligent agent (IA). These agents monitor the system, a network of machines in real time. Using prior knowledge of what suspicious behavior looks like they can then identify potentially malicious activity and make decisions to prevent it. Unlike humans, these agents are not adversely affected by the volume of data produced by a large system.

2 Background and Related Work

A log is a record of the events occurring within a system, which is a cyber-physical entity having multiple components as workstations, servers, network devices, doors, badge readers, and employees. Entities in the system produce data in the form of text files called logs. Logs are composed of log entries that contain information pertaining to what, when and why an event has happened. Logs are generated by many software such as antimalware and antivirus, firewalls and intrusion prevention systems, operating systems on servers, workstations, network devices, and applications. In a large system of multiple components, multiple log entries for the same event maybe stored [2].

Logs are not only useful for security purposes, but also serve for modern information-driver businesses via many functions such as troubleshooting, performance, optimization and auditing which renders the collection and storage of logs so important. Companies typically follow logging procedures as defined by one or more of the governing methods as FISMA, HIPAA, SOX, GLBA and PCI DSS [4]. The volume and velocity of logging information has increased greatly as organizations use more computing resources. This has created the need for a clearly defined process for generating, transmitting, storing, analyzing and disposing of this data. Log management systems are essential to ensuring that log data are stored in sufficient detail for an appropriate period of time [1].

Log analytics is a critical task in log management systems. Analytics allow organizations to identify security incidents, fraudulent activity and operational problems. The better the analytics the quicker issues can be resolved and prevented in the future. Reactive security systems often use routine or scheduled log analysis. Whereas the proactive approach uses a real-time approach where log events are classified in real time and shown to administrators automatically. Two fundamental problems exist for log management systems to resolve: First is a balancing act, i.e. balancing the limited computing resources available for processing and storage with the large and continuous volume of logging data. Second is the high number of log sources, inconsistent log content and formats, inaccurate time stamps. Log management systems need to securely collect, normalize, sanitize, and analyze log data in an efficient manner.

Organizations must establish policies and procedures for log management. At minimum they must define what will be collected, how it will be used and how it will have to be destroyed. They should prioritize the importance of logging. Without metrics important decisions cannot be made. Metrics allow the establishment of specific norms that allow an organization to identify whether these norms are followed, as leaving them are key indicators of an issue.

Organizations should also create and maintain a logging management infrastructure that meets its needs. For example, machines not taking part in the logging process are perfect vectors for adversaries. Organizations should provide staff with the proper tools and support them in their responsibilities. Logging events that go unnoticed could be those that are indicative of an attack.

Organizations should establish a standard for log management operational practices, which means they should monitor the status of all log sources, archiving, and maintenance of the logging infrastructure. Maintenance activities include checking for upgrades and patches, and acquiring, testing and deploying new or upgraded components in a timely manner. It is also important to synchronize logging host actions to ensure that a causal path, that is a trace from the action and the events that transpired in the system from a specific action are recorded. Further reconfiguration may be required when policy, technology or other factors change.

Organizations should establish how log data should be used. This includes analytics. Documenting and reporting anomalies are the most important deliverable that results from a logging infrastructure. Business planning steps should include how this will be accomplished, in what timeframe and who will be responsible for mitigating the identified anomalies.

The most valuable logging data comes from IDSs [5]. IDS monitors events occurring in a network of computer systems. Using specific techniques, they analyze network data (such as payload, protocol, and devices involved in the communication) for signs of possible attacks. When IDS identifies an attack, it flags and reports it. The IDS determines if an attack has occurred by looking for violations against policies. These policies are defined such that IDS can interpret. It is an intrusion prevention system, which is not typically part of IDS that prevents the attack after it is identified. A system that accomplishes both tasks is called an IDPS.

In general, IDS operates for anomaly or signature detection. In anomaly detection an IDS will determine whether a deviation from the established normal has occurred. In signature detection the IDS uses patterns defined by rules for well-known attacks to identify intrusions. Anomaly detection can be improved by applying machine learning techniques such as neural networks, genetic algorithms and support vector machines.

Applying ensemble techniques like bagging and boosting allow classifiers to minimize false positives/negatives.

What a rules-based IDS accomplishes can be described as a pattern recognition task where raw data represented as a vector is either processed unlabeled (for unsupervised learning), or labeled (for supervised learning). In this method we maintain which data points are considered normal operation. Deviation from these norms increases the probability that a specific vector represents malicious activity. We can use supervised learning by applying class labels to some data points along with ensemble techniques to create a highly accurate semi-supervised classification algorithm.

Single classifiers such as Naïve-Bayes, k-nearest neighbor, support vector machines and neural networks are typically used in IDSs [4]. It is a new idea to combine multiple classifiers. We can also employ decision trees to allow changing IDS policies as new attacks are discovered.

An intrusion prevention system is composed of several components as sensors to collect information, agents to make decisions based on the collected information, and management consoles to allow administrators to oversee and validate the operation of the IDS/IDPS. Though adversaries target each component, agents are the most vulnerable because they make the decision to block an attack. The agent is what employs the classifiers to make decisions based on the collected log data.

There is a tradeoff between the source of data and the use of it: Though it is very important to collect data from multiple sources for best performance, logs coming from heterogeneous sources can result in issues. Moreover, a single log source can generate multiple logs, e.g. an application storing authentication information has one log for network activity and another log for successful authentication requests. Inconsistent timestamps and logging content must be normalized and sanitized in such logs.

Each log source records certain piece of information in its log entries, while filtering unnecessary information. However, what one source deems unnecessary may be useful to the learning algorithms later. Hence, it is very important to capture as much data as possible at the source before filtering by the constraints on the agent learning algorithms. Another complication occurs when each source represents common values inconsistently, such as different date formats, protocols, and inadequate network information.

Another issue with multi-source logging is inconsistent timestamps. Each log source, or host, will reference its own internal clock for each log event. If a host's clock is inaccurate, the timestamps in its logs will also be inaccurate. This can make analysis of logs more difficult, particularly when logs from multiple hosts are analyzed. Following is an example of timestamp drift: Consider two events A and B. If there exists a causal path from A to B, i.e. logical timestamp of A is less than that of B then we conclude that A happened before B. However due to clock drift we get the following relation: $A \rightarrow B$, $L(A) \leq L(B)$, but clock drift $D(L(A)) \geq D(L(B))$, thus $L(A) + D(L(A)) \geq L(B) + D(L(B))$, thus $B \rightarrow A$. This contradiction needs to be resolved.

The final and most important issue with multi-source logging is the trustworthiness of the source: if a host is compromised then the log data it provides may also be compromised. We use weight to determine the level of trust for each source, where a lower level of trust does not affect the result of classification but a higher level of trust does. In an enterprise network, this is implemented by assigning lowest level of trust to the endpoints, while assigning higher level of trust to the internal network devices, which should be harder to access. Typically, this is a weighted sum, whose weights are determined by depth in the network, time since last update, running average of vulnerabilities, and average severity of vulnerabilities for the log source.

Log management is the key component of IDS. For IDS/IDPS to be effective, it must have all necessary log data so that it can make decisions. The log management infrastructure has three tiers. The first tier is the log generation. Hosts run the logging client applications, which make log data available to servers in the second tier. Ideally, machines can join/leave the logging infrastructure with minimal effect to the second tier. Log generation is accomplished by parsing extracted data from core resources. Core resources can be metrics and/or files on an operating system (OS). Typically, all applications running on an OS should provide log data. This can be accomplished on a Linux by using strace, perf, ltrace, ptrace and proc/pseudo file system [6].

The strace utility allows the caller to inspect system calls a program makes, catch all inter-process signals and inspect child processes spawned by a program. The utility allows the caller to filter system calls by types. Further, it allows inspection of the arguments passed to these system calls. The perf utility allows the caller to profile how many CPU instructions are executed and the number of cache misses. Programs have a specific modus operandi described by the syscalls and perf metrics. If a program is compromised, these values will be different.

The ltrace utility allows the caller to see what library calls were made by a process. This is very similar to strace including most of the same bugs strace has. Another utility is ptrace, which allows caller to check each thread. This utility has the ability to attach to a single thread in a multithreaded process and can determine the specific corrupt thread in the event of a multithreaded application compromise. The proc/pseudo file system contains important metrics that allows the logging infrastructure to inspect each process running on the system, look at environment variables on a per-application basis, watch what files are opened/read/written, and monitor memory usage for each process.

In the log generation, first tier, the application running on the host will typically perform parsing, sanitizing, aggregation and projection. In the projection step, the application will decide what data to expose to the second tier. Parsing can be using a utility, reading a log file written by an application or checking a metric using an API. Sanitization is done by filtering invalid or unexpected values and noise. This is dependent on the core resource used. Aggregation is an important phase used to extract the most important data and save resources. It is accomplished by filtering, which is the suppression of log entries from second tier by grouping these entries logically. Typically, filtering should not adversely affect the generation of log data that is needed by second tier. Aggregation also involves generating data counts such as sum, min/max, and average values. It is critical to perform aggregation soon to assure effective use of resources. Further log reduction through these methods makes event correlation much easier. The first tier should maintain an order; however, timestamps should be assigned on receiving of the data at the second tier.

The second tier performs log analysis and storage. This tier typically operates as a network application on a server. One or many servers can work together to handle large volumes of log data which is typically stored and queried on a single database. One or more servers receive log data from the first tier. This data is received either in real-time or in scheduled batches depending on the amount of log data, and how constrained network resources are.

The second tier also performs log normalization in order to save resources. Log normalization is the process of converting each field in a log event to a specific data representation and category. It may happen by merging or decomposing some fields. For example, a timestamp from one log may contain more significant digits than what is required by the logging system. Further normalization may add other features to the log entry such as a digital signature (via MD5 or SHA1) to ensure integrity.

The third and final tier is the log monitoring and analytics layer, where management console and report generation are performed. This layer contains agents in the form of one or many classifiers to analyze log data and yield result, which could be either clear or an indicator of one or more attacks. In the event of a probable attack, the prevention system that resides in this layer makes changes to the network environment to prevent the attack. This tier also assigns a priority to each message based on source and type. The third tier is critically important because it is where we find correlations between events across multiple sources and extract causal paths. Log viewing is another feature provided by this layer. Log entry displays should be in a human-readable format that is easy to search and find causal paths. Over time old log data must be securely destroyed, where all entries from a log that is no longer needed are removed. The reason for removal is obvious: existing data may allow an attacker to identify what information the training algorithms in the agent ingest to make decisions. This information may allow attackers to construct specific work-arounds against the data. Disposal is clearly defined by NIST standards for data destruction.

3 Implementation and Results

We aim at bringing all aspects of a log management system with features that are typically offered in an IDPS into one distributed application for a data center. Our design promises to effectively offer logging, aggregation, monitoring and intelligence services against adversaries that are getting increasingly aggressive, creative and careful. We collect various performance metrics from machines in the network and use machine learning to identify anomalous behavior, which adversaries cannot hide. We train classifiers to detect patterns of this behavior that correspond to known attacks.

Our log management system architecture is illustrated in Fig. 1 below:

We use Python platform for implementation because of its ubiquitous support on popular Linux distributions. Our application uses minimal dependencies and requires only the standard library. Three components of our system are logging deamon (for data collection), data aggregation (for data analytics), and web application (for display) that communicate over a RESTful HTTP API. Each component is loosely coupled, where low level component initiates communication with high level component. Login deamon reads application logs, runs utility commands, and extracts data as a file with a timestamp. The data aggregator stores received logs on a database in a round-robin fashion. It uses programmable thresholds to generate alerts. It also uses machine learning

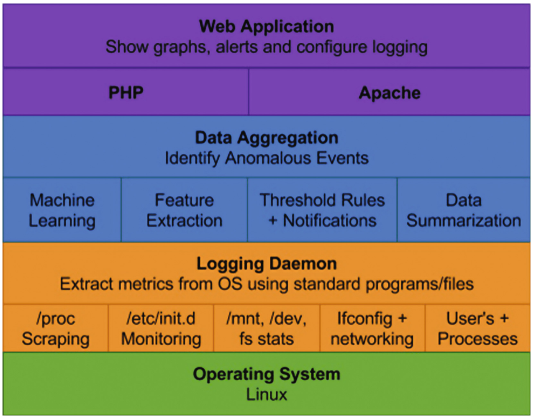


Fig. 1. Our log management system architecture.

techniques to determine the probability that monitored system is compromised. The web application provides statistical data in the form of graphs and displays alerts. The user can adjust the thresholds and rules for alerts, configure machine learning and the data extracted by the logging deamon.

The first component will actively extract metrics from machines in a system. Metrics are small bits of information that describe the state of a machine. The second component is responsible for aggregating and analyzing the extracted metrics to detect anomalous behavior. This is the core of the agent, which uses machine learning with feedback implying that the system gets smarter over time. The third component allows administrators and management to view collected metrics and reports on any anomalous behavior observed.

3.1 Component I: Logging and Trace Daemon

Insufficient logging is the root cause of cybersecurity incidents going unnoticed. As an example, the top 10 cybersecurity incidents caused by database compromise could have been avoided with proper activity logging [7]. This component runs on all monitored machines within an organization. Its purpose is to resolve insufficient log data problem. The tasks supported by Component I are displayed in Table 1:

As seen in Table 1, the logging performed by component I is designed to look for *indicators of compromise (IOC)*. IOC are derived from several sources within the machine as determined by the OS. This component runs as a daemon, which is a background process. It uses Linux commands and built-in Linux pseudo file system to monitor processes. In Linux I/O devices, processes and memory are all accessible via file system. This daemon runs several tasks on a schedule of 1, 5, 15, 30 and 60 s. Typically, processor intensive tasks are executed on a longer interval.

Table 1. The tasks supported by component I: logging and trace daemon.

Task name	Description	Schedule	IOC
Filesystem memory usage (fsusage)	Retrieves size, amount used and available memory of each filesystem, and mountpoint information	5 m	Total storage (Int) Total used (Int) Total available (Int) Percent used (Int)
Service status (initdstat)	For each service listing in/etc./init.d it will check the status by calling the '*status*' argument against the scripts	15 m	Service count (Int) Running count (Int) Issue count (Int)
Load averages (loadavg)	Retrieves the average load over one, five and fifteen minutes by looking at the/proc/loadavg pseudo-file	1 m	Avg 1 min (Int) Avg 5 min (Int) Avg 15 min (Int)
Memory information (meminfo)	Retrieves total amount of physical RAM used, memory allocated in the virtual address space, and RAM used as cache memory all in KB	1 m	Total Memory (Int) Free Memory (Int) Percent Used (Float) Virtual Memory Cached Memory
Network interface status (netint)	Retrieves the number of packets sent/received, errors sent/received and interfaces up/down	1 m	Errors in (Int) Errors out (Int) Packets in (Int) Packets out (Int) Int up (Int) Int down (Int)
Users online (online)	Retrieves a list of usernames and IP addresses of the currently logged in users	1 m	Users (Dict) Count online (Int)

This daemon looks for the log collection service, which is component II. It will attempt to connect to this service to deposit the collected log data. If it cannot connect, it will log that it failed to connect and will continue to ensure fault tolerance. As also addressed in future work section, we plan to use service discovery to allow this component to discover a log collection service even if one goes down.

The trace daemon is the most innovative feature of our design: It adapts Linux tools strace, ltrace and ptrace to collect information about how a program behaves, such as system calls, dynamic library calls, and memory and register values of the program. These data are collected in streams. We then generate a power set of the streams and compare it against a trained model in Weka to determine if it is malicious or not.

Well-resourced attackers follow this recipe for an attack:

1. Reconnaissance: Identify the target and vulnerabilities in the target's network.
2. Entry: Use one or more exploits to gain access into the network.
3. Movement: Laterally move across the network to gain a wider attack surface.
4. Achieve: Disrupt, extract and manipulate.

Monitoring and logging becomes important at each of these steps. During reconnaissance attackers use port scanning and fingerprinting tools. These tools are detected by increased CPU activity and packet errors in the network interfaces, which are detected by this tool. During the entry phase, the attacker will run one or more commands to install a backdoor, rootkit and execute commands. These commands will result in increased CPU, users changing online, and new syscalls being executed which can all be detected. During movement, other computers may experience new CPU usages, which are also detected.

The types of logs traditionally collected by organizations exist in three categories: System logs which are endpoints, authentication and applications. Typical network logs are email, firewall, VPN systems, and networked devices. Application logs are one of proxies, DNS, DHCP, FTP, SQL Server, antivirus, or web applications. A real time solution that automatically isolates the key events is accomplished by these services.

3.2 Component II: Aggregation and Analytics Engine

The purpose of this utility is to collect the logs from the logging and monitoring agents and make sense of the data. The idea is to understand how cybersecurity incidents transpire and detect them as early as possible. It is important to understand that regardless of how mundane or trivial an action performed on a machine is, whenever combined with other actions, it can be potentially devastating. The goal is to look at all events across machines and applications, then piece together a timeline and check the events against a trained classifier to determine its potency. That is if the series of actions represent an incident that needs to be further investigated.

Configuration is an important part of security. During an organizations cybersecurity planning session, they must define what they consider important indicators of compromise. That is what sequence of events represent normal operations. This is what is used to train the classifier. The organization identifies where bad behavior may occur. It also introduces scenarios or use cases to define what good (normal) and bad (anomalous) behavior should look like. Then logs that are required to fully cover or shatter the behavior class need to be identified. Finally, the events or line items provided in the logs are defined so that they can be extracted. These activities are listed in Fig. 2 below.

The analytics engine uses machine learning. This offers the greatest flexibility in the rapidly changing cybersecurity environment [8]. Classifiers are able to be automatically retrained over time, especially when adversaries change their tactics. A variety of security applications already employee this technique. For example, spam detection and filtering schemes all rely on a Naïve-Bayes classifier. However, when discussing the use of machine learning it is important to evaluate the security of the classifiers.

Assume the attacker has gained control of a machine. We define the level of control that attacker has as a function of what commands the attacker can access. Consider a set

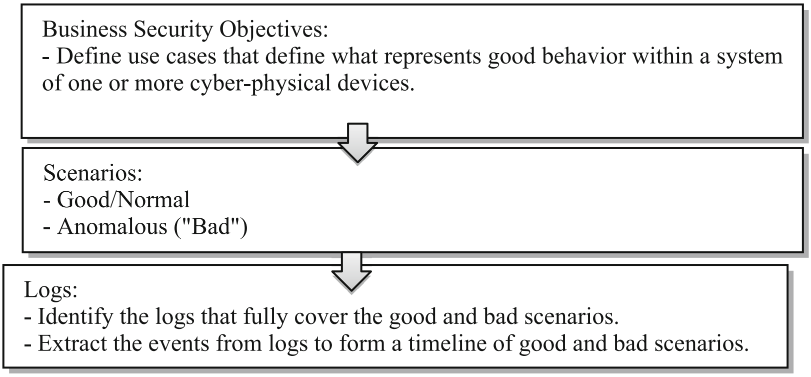

Business Security Objectives:
- Define use cases that define what represents good behavior within a system of one or more cyber-physical devices.
Scenarios:
- Good/Normal
- Anomalous ("Bad")
Logs:
- Identify the logs that fully cover the good and bad scenarios.
- Extract the events from logs to form a timeline of good and bad scenarios.

Fig. 2. The training process of the aggregation and analytics engine.

of all possible commands a Linux OS can accept through shell. Then a subset of all commands would become accessible by the compromised user account. This subset is equal to the full set if the compromised user is root. In any case, if the attacker has access to a subset of at least one command then he can saw the logging data provided to the analytics engine.

Potentially, an attacker can add new commands that will prompt events to be logged and these events will be input to the training phase of a classifier. Therefore, we must consider security implications of training a classifier: can the attacker degrade performance of the classifier? Here degradation means the attacker can trick the classifier into yielding false positives/negatives. This way attacker can effectively leverage the knowledge of the machine learning mechanism that renders it useless to system administrators. We consider this event to be a form of DoS, where service of proper classification is denied.

If it is possible for an adversary to manipulate the classifier, then we must find ways to defend against it. The goal is to prevent an adversary from leveraging knowledge of the classifiers, mitigate any potential impact from manipulating the training process, and prevent the exploit of specific properties or assumption of the learning process. Leveraging knowledge does not necessarily imply that we obscure which classifiers are used. This is also true when we consider how to prevent the exploitation of classifier properties.

We propose a trivial solution that uses a technique used by ensemble methods in machine learning: When we retrain the classifier, we retain old classifiers training data [9]. Consider a timeline where we represent each time classifier is trained. We can look at any point in time and split the timeline into two as before attacker control and after attacker control. After an attacker gains control of a system, we can assume the retrained classifier may be broken in some way. However, the older classifier was still trained on good data, i.e. data that has not been compromised yet.

As specified in Fig. 3, our design employs 3 types of training. We use good (untainted) training data as test data for the new classifier. Everything that the new classifier has, which is potentially tainted for having ingested attacker-controlled data, is used to train a new classifier. This new classifier is assigned a greater weight than the bad classifier. We still use bad classifier because it is not entirely useless, on the contrary more useful than the classifier that existed before retraining. In fact, to mitigate an attack we want to use a classifier for as long as it is potentially useful.

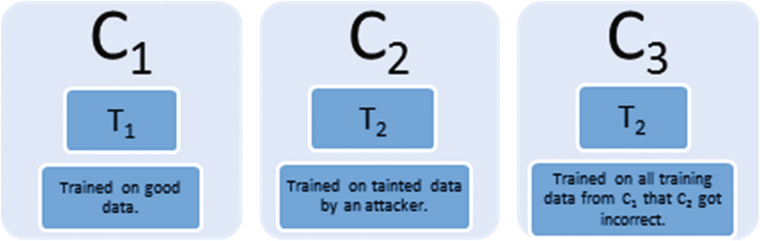

C_1
T_1
Trained on good data.
C_2
T_2
Trained on tainted data by an attacker.
C_3
T_2
Trained on all training data from C_1 that C_2 got incorrect.

Fig. 3. Different types of training.

A machine learning system is comprised of several components:

1. A mechanism that ingest training data and produces a hypothesis function F.
   (a) The training data is extracted from log events that have been classified.
   (b) Typically, classification is done using humans or other means.
   (c) Labeling, which is outside the scope of this paper.
2. Select one or more high performing classifiers using k-fold cross validation.
3. Use binary adaptive boosting on the top n high performing classifiers.
4. Construct a final classifier based on adaptive boosting and retrain according to historical classifier performance.

$$F_T(x{:}Event\ Entry\ Vector) := \frac{Wrong\left(F_{C_2}, T^1\right)}{|T_1| + |T_2|} F_{C_2}(x) + \mu F_{C_3}(x) \tag{1}$$

Where

$$0 \le \mu \le \frac{Wrong\left(F_{C_2}, T^1\right)}{|T_1| + |T_2|} F_{C_2} \tag{2}$$

is some discriminate factor.

$$F_{C_i}(x) = \begin{cases} -1, & if\ classifier\ C_i = -1 \\ 1, & otherwise \end{cases} \tag{3}$$

A vector represents a series of events that occurred in a system being monitored. Adaptive boosting allows us to generate two or more hypothesis functions with varying degrees of success and focus. Focus is the primary concern. This allows us to select multiple types of classifiers that are good at classifying certain system events. Boosting, an ensemble technique, allows us to lower the bound on the error, i.e. to minimize the number of false positives. By further applying the technique described above, we also minimize the number of false negatives. A system needs to be wary of attacks against a machine learning system, such as forceful intrusion attacks that cause false positive and negatives [10].

3.3 Component III: Web-Based Management Console

The goal of component III is to provide graphical representations of the collected data, allow data-scientists and security analyst to monitor system, and generate management reports. It uses Node.js, Python, and Monogdb to store the data. Mongodb provides capped collections with is akin to a round-robin database. We use Angular JS to provide interface and charting capabilities.

3.4 The Testing Environment

As seen in Fig. 4, we use a virtual network of four virtual machines for testing purposes: the first two vms (vm0, vm1) run the logging daemon, the third vm (logger) runs the application, analytics, and MongoDB processes. The last runs Kali Linux for attacking vm0 and 1.

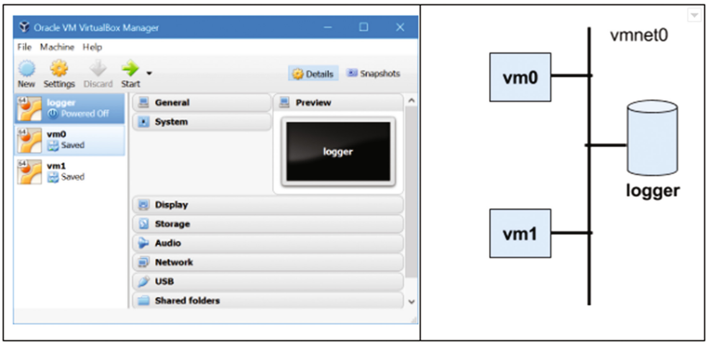


Fig. 4. Ubuntu 64-bit installations on VirtualBox and the network diagram of virtual machines.

4 Conclusion and Future Work

We introduce the novel idea of using weighted trust to prevent corruption to classifiers while processing host data in a networking environment. Machine learning techniques used in this paper are based on the well-respected tool Weka. Our initial implementation yields promising results by allowing as to build the infrastructure of a log management system.

In the future, we would like to accomplish the following: Add more classifiers to Weka to create a comprehensive classifier frame, and improve training, testing and classification pipeline as it is currently a manual process. We also plan collecting more data from the logging system (component I) and better integrate the scheduled and continuous tasks (tracing), add better charts, searching, event linking and report generation. We also plan to retrain our stream processing model to see whether it can improve anomaly detection performance.

References

1. Steinberger, R.: Proactive vs. Reactive Security. http://www.crime-research.org/library/Richard.html
2. Application Security, Deconstructed and Demystified, Infosec Institute (2011). http://resourcesinfosecinstitute.com/application-security-deconstructed/
3. Kent, K., Souppaya, M.: Guide to Computer Security Log Management. NIST Special Publication 800-92 (2006)
4. Tsai, C.F., Hsu, Y.F., Lin, C.Y., Lin, W.Y.: Intrusion detection by machine learning: a review. Expert Syst. Appl. **36**(10), 11994–12000 (2009)
5. Scarfone, K., Mell, P.: Guide to Intrusion Detection and Prevention Systems. National Institute of Standards and Technology, Gaithersburg. Special Publication 800-94 (2007)
6. Lee, W., Stolfo, S.J.: Learning patterns from unix process execution traces for intrusion detection. AAAI technical report WS-97-07, pp. 50–56 (1997)
7. Ramanan, S.: What are the top 10 cybersecurity breaches of 2015? https://www.quora.com/What-are-the-top-10-Cyber-security-breaches-of-2015
8. Huang, L., Joseph, A.D., Nelson, B., Rubinstein, B., Tygar J.D.: Adversarial Machine Learning. In: AISec 2011, pp. 43–58 (2011)
9. Blum, A.L., Langley, P.: Selection of relevant features and examples in machine learning. Artif. Intell. **97**, 245–271 (1997)
10. Barreno, M., Nelson, B., Sears, R., Joseph, A.D., Tygar, J.D.: Can machine learning be secure? In: ASIACCS 2006, pp. 16–25 (2006)

Adaptive Weak Secrets for Authenticated Key Exchange

Phillip H. Griffin(✉)

Griffin Information Security, 1625 Glenwood Avenue, Raleigh, NC 27608, USA
phil@phillipgriffin.com

Abstract. This paper describes biometric-based cryptographic techniques that use weak secrets to provide strong, multi-factor and mutual authentication, and establish secure channels for subsequent communications. These techniques rely on lightweight cryptographic algorithms for confidential information exchange. Lightweight algorithms are suitable for use in resource constrained environments such as the Internet of Things where implementations require efficient execution, limited access to memory and small code size. Password Authenticated Key Exchange, and Biometric Authenticated Key Exchange protocols based on user knowledge extracted from biometric sensor data, both rely on weak secrets. These secrets are shared between a client and an access controlled server, and used as inputs to Diffie-Hellman key establishment schemes. Diffie-Hellman provides forward secrecy, prevents user credentials from being exposed during identity authentication attempts, and thwarts man-in-the-middle and phishing attacks. This paper describes the operation of these protocols using an adaptive knowledge substitution process that frequently modifies the weak secrets used for protocol operation without requiring disruptive user password changes. The password substitution strings used to implement this process can be far longer and more complex than the weak secrets people can easily memorize. The process described in this paper allows people with diverse abilities to use simple, easily recalled, quickly entered passwords and still benefit from the strength of long, complex strings when operating cryptographic protocols.

Keywords: Authentication · Biometrics · Key exchange · Password · Security

1 Introduction

Information and Communications Technologies (ICT) provide a variety of services and make available opportunities that can enrich peoples lives and benefit our society as a whole. Recent research reveals that ICT-connected devices constitute the "technology with the greatest impact in promoting the inclusion of persons with disabilities" [1]. The growing ubiquity of smart phones and networked devices in the Internet of Things (IOT) heralds "a new age not only of information sharing in general", but an era of new opportunities to provide services to "disabled and non-disabled communities alike" [1].

In a world of over a "billion persons living with disabilities" [1], it is import that ICT applications and services are universally accessible. Access control systems that follow Universal Access (UA) design guidance can help remove barriers to ICT access

D. Nicholson (ed.), *Advances in Human Factors in Cybersecurity*,
Advances in Intelligent Systems and Computing 593, DOI 10.1007/978-3-319-60585-2_2

and reduce the exclusion of the elderly and infirm [1]. Following UA principles can help all people enjoy the benefits of securely "accessing, participating and being fully-included in social, economic and political activities" [1].

Universal access is a methodology that incorporates human factors into user interface design in an effort to provide "the utility of modern information technology to as broad a range of individuals as possible" [2]. Considering the vast differences between individuals, who may be young, elderly, healthy, infirm, disabled or not disabled, provision of a single, monolithic access control interface is not likely to achieve the goals of UA. Serving the needs of a diverse population requires offering people choices in the ways they gain access to information and communications systems. There is greater potential for integrating "security and usability effectively" in access control systems based on "biometrics than with other authentication methods" [2]. This makes biometric technologies a "natural choice for implementing authentication in UA systems" [2].

People have diverse abilities that may impede or prohibit their use of a particular access control method or interface. Individuals afflicted with "dyslexia can have problems in remembering the digits in the correct order", or have trouble spelling or reading [3]. This can make password-based access control using a keyboard device difficult. Users with degenerative arthritis, "limited use of arms or hands", and those with a "cognitive impairment will find most biometric systems much easier to use and provide them a greater level of security" [3]. Since every individual is not capable of using every type of computer input device or every biometric technology type, authentication systems with user interface designs that offer users a variety of choice alternatives will be capable of offering access to greater numbers of users.

Identity authentication is a critical security control for managing the risk of unauthorized access to information and communications technology (ICT) systems. The cost of deploying credentials that enable strong user authentication can be prohibitive. User convenience can also be an issue and creating effective, inclusive design can be a challenge. Offering authentication methods that include passwords and biometrics or that combine the two can lead to low-cost, secure solutions that are convenient and easy to use by persons with diverse abilities.

2 Biometric-Based Cryptographic Techniques

Weak secrets are those "that can be easily memorized" by a user and that are often "chosen from a relatively small set of possibilities" [4]. Passwords, passphrases and Personal Identification Numbers (PIN) are examples of weak secrets. They are easily recalled by users, typically short in length, and are composed from a limited set of characters. Weak secrets are commonly used in access control systems today, and serve as a *something-you-know* identity authentication factor.

Weak secrets also play a role in authenticated key exchange (AKE) protocols, where they function as shared secret inputs to a Diffie-Hellman key exchange process. Password Authenticated Key Exchange (PAKE) is a protocol that allows two remote parties "to establish a secure communication channel" between them "without relying on any external trusted parties" [5]. Establishment of the secure channel is based "on a shared low-entropy password", a weak secret known to both parties. This shared secret is used in the PAKE protocol to provide implicit identity authentication [5].

In a Password Authenticated Key Exchange (PAKE) protocol the confidentiality of user authentication credentials is protected by encryption from identity theft, man-in-the-middle (MITM), and phishing attacks during transfer [6]. PAKE has been suggested as a way to remediate these attacks in the Transport Layer Security (TLS) protocol by inserting PAKE following the TLS handshake [7]. This approach still relies on digital certificates, which can be cost prohibitive in some applications. In practice, neither digital certificates nor TLS are needed by PAKE for access control systems to achieve mutual and multifactor authentication.

Biometric Authenticated Key Exchange (BAKE) is a "biometrics-based protocol for authenticated key exchange" [8] that relies on PAKE. The BAKE protocol extracts "knowledge shared by communicating parties" needed to operate a PAKE protocol "from data collected by biometric sensors" [8]. Once extracted, this user knowledge is input to a PAKE protocol to derive an encryption key. This key is used to protect a user biometric sample, a *something-you-are* identity authentication factor, during a user authentication attempt. By including a biometric sample in the user credentials protected during transfer by PAKE and BAKE, both protocols can achieve 2-factor user authentication.

Telebiometric Aauthentication Objects (TAO) are "tagged physical objects" that have been associated with a user by a relying party. This association allows TAO to be used as a *something-you-have* identity authentication factor. These objects are "functionally coupled with biometric sensors and connected to a telecommunications network" [9]. TAO combine telecommunications networks with biometric sensors to enable identity authentication and user identification services. These 'smart objects' enable IoT access controls that offer "strong, low cost mutual and multi-factor authentication" that are frequently readily available (i.e., smart phones) and can be easy for many people to use [6].

During the user authentication phase of a PAKE or BAKE protocol, TAO can be included in the user credentials to provide an additional identity authentication factor. By combining biometric authentication with registered TAO during operation of an AKE protocol, 2- and 3-factor user authentication can be achieved. User credential transfer and subsequent information exchange needed to achieve mutual authentication require that all data transfers be protected by strong encryption.

3 Internet of Things Security Limitations

Building a world of universal healthcare, ambient assisted living, and IoT-based services for reliable delivery to remote environments requires secure, universal access to ITC resources. As the 5th generation (5G) of mobile and wireless networks replace existing infrastructure, "future networks are likely to benefit from high reliability and security, very high speeds and increased reach and mobility" [10]. Though coming improvements in network security are helpful, implementers still need to ensure "data protection and privacy" of stored user data [10]. They must also protect the authenticity and confidentiality of sensitive user authentication credentials and the end-to-end "secure, reliable and consistent exchange of data between devices, applications and platforms" [10].

As the expanding IoT ages, it will contain ever growing numbers and types of devices, "information technology systems and software applications" that once deployed, must maintain their ability to continue "to communicate, exchange data, and use the information that has been exchanged" [11]. Effective encryption solutions are needed that can perform well on both small IoT devices, and on larger platforms in the data centers they access. These solutions must be capable of being implemented, not only on high speed networks and resource rich servers, but on the small computing devices that will still be common on the IoT for many years to come.

The need to secure devices in the IoT has fueled research and development of a family of lightweight cryptography solutions, "cryptographic primitives, schemes and protocols tailored to extremely constrained environments" [12]. The term 'lightweight' should not be viewed negatively. The term does not imply that lightweight cryptography is 'weak', but that it offers efficiencies in its "execution time, runtime memory (i.e. RAM) requirements, and binary code size" [12]. Lightweight algorithms can provide "the cryptographic strength needed to protect sensitive user credentials during identity authentication, and during subsequent communications" [13].

Both PAKE and BAKE rely on Diffie-Hellman key exchange for cryptographic key establishment. The user establishes a key to protect their credentials when attempting access, and the accessed server establishes the same key to perform mutual authentication and protect client-server information exchange during transfer. Once a key is available, a symmetric key algorithm is used to protect the confidentiality of user credentials during an authentication attempt and subsequent communications.

User credentials may include a biometric sample collected from the user to provide a *something-you-are* identity authentication factor. Credentials may also include one or more physical objects associated with the user biometric reference template and known to the server [9]. These authentication objects may be tagged objects that have been pre-registered with the server for use as *something-you-have* authentication factors [9]. When these objects are coupled with telecommunications-enabled biometric sensors, they can be "used for mutual and multifactor authentication in access control systems" [13].

4 Lightweight Cryptographic Algorithms

The Advanced Encryption System (AES) algorithm is considered "an excellent and preferred choice" for "almost all block cipher applications" [14]. However, the AES algorithm is "not suitable for extremely constrained environments such as RFID tags and sensor networks" [14]. These environments are common in the IoT, where applications may require "security and hardware efficiency" [14], but are constrained by limited power, communications bandwidth, or processing capabilities.

The ISO/IEC 29192 lightweight cryptography standard specifies symmetric key-based cryptographic primitives for block cipher, stream cipher, hash function, and Message Authentication Code (MAC) algorithms. Part 2 of the series will soon define four symmetric block ciphers, the PRESENT, CLEFIA, SIMON, and SPECK algorithms described in the following tables. The PRESENT and CLEFIA lightweight algorithms first appeared in the current version of the standard, the 2012 edition. Both algorithms had been introduced some five years earlier, PRESENT at the CHES 2007 Workshop on Cryptographic Hardware and Embedded Systems [14] and CLEFIA at FSE 2007, the Fast Software Encryption Workshop [15].

As shown in Table 1., the PRESENT cipher has “a block size of 64 bits and a key size of 80 or 128 bits” [16]. PRESENT requires 32 processing rounds, with each of the rounds consisting of a “sequence of simple transformations” [16]. Each processing round introduces a new round key, with the last round key used for final processing. The creators of PRESENT considered hardware efficiency in their design resulting in an implementation that required only 1580 GE to encrypt a 64-bit block using an 80 bit key [14].

Table 1. PRESENT algorithm characteristics

PRESENT-128 and PRESENT-80			
Block size (bits)	Key length (bits)	Number of rounds	Round keys
64	128	31	32
64	80	31	32

As shown in Table 2., the CLEFIA cipher has “a block size of 128 bits and a key size of 128, 192 or 256 bits” [16]. The number of processing rounds and the number of round keys needed varies by key length. Longer keys have greater processing requirements. CLEFIA has a structure “based on a generalized Feistel network” [13] and that is used “data processing part and the key schedule” [16].

Table 2. CLEFIA algorithm characteristics

CLEFIA			
Block size (bits)	Key length (bits)	Number of rounds	Round keys
128	256	26	52
128	192	22	44
128	128	18	36

Work began in 2015 to add SIMON and SPECK block cipher families to a revision of the ISO/IEC 29192-2 standard. Approval of this revision is expected in 2017, but the revised standard has yet to be published. SIMON and SPECK are relatively recent block cipher proposals created by “researchers from the National Security Agency (NSA)” of the United States [17].

Both algorithms offer “efficient and secure” encryption that provide a means of achieving solutions that are “low-cost and easy to implement and deploy on multiple platforms” [17]. These algorithms target a range of platforms and applications, from “mobile devices, through RFID tags to electronic locks” [17]. Their cryptographic strength and efficiency makes them “appealing for use in IoT applications” [13].

SIMON and SPECK both offer “very competitive performance, small memory footprint” that beats “most existing lightweight ciphers in terms of efficiency and compactness" [17]. Both block cipher algorithms are based on "very simple and elegant" designs built on the Addition/Rotation/XOR (ARX) philosophy [17]. The class of ARX algorithms rely on a set of "simple arithmetic operations: modular addition, bitwise rotation (and bitwise shift) and exclusive-OR" [18].

Table 3. SIMON, and SPECK algorithm characteristics

SIMON and SPECK	
Block size (bits)	Key length (bits)
128	256
128	192
128	128
96	144
96	96
64	128
64	96
48	96

As shown above in Table 3, the range of key sizes to be standardized for SIMON and SPECK span those supported by both their PRESENT and CLEFIA predecessors. Both algorithms offer cryptographic strength sufficient to protect user credentials and subsequent information exchange in the operation of the BAKE and PAKE protocols. They make flexible IoT implementation designs to manage security risks possible, offering "great performance on hardware and software platforms" [19] The SIMON block cipher is "designed towards hardware applications and SPECK for software applications" [19].

5 Adaptive Password Substitution Strings

When a user first establishes an account on a multi-user computer system, they are assigned a system-unique identifier. This account name or user identifier (user ID) is presented by the user along with identity authentication credentials during subsequent login events. Information management and security information used to control user access may be associated with a user account name and stored by the system.

One or more user identity authenticators, such as a password or biometric reference value will also be stored and associated with the user ID. To establish a biometric authenticator, then user must enroll in a biometric system to create a biometric reference template for each biometric type being enrolled. Biometric reference templates are used by the access control system to match user biometric samples during authentication attempts subsequent to enrollment.

When Telebiometric Authentication Objects (TAO) are used to authenticate a user in an access control system, an identifier of each user possession must be associated with the user ID or biometric reference template of the user. When a BAKE or PAKE protocol is used for multifactor user authentication, user selected knowledge information must also be bound to the user ID and known to the server before being used for identity authentication and to operate a BAKE or PAKE protocol.

User knowledge information known to a server and associated with a user account can be used as a *something-you-know* authentication factor. This knowledge can be presented to the system in many ways and formats, ranging from a simple password entered through a keyboard device, to a PIN entered using a smart phone touch screen, to human speech recorded by a microphone, to "observations of a sequence of gestures collected by an image-based biometric authentication system" [13]. For use as an input to an AKE protocol, each type of knowledge presentation must be presented to the protocol in a character string format. For example, the words of a human speaker can be extracted from a voice biometric sensor using speech recognition techniques and formed into a password string.

It is usual to consider "gestures based on American Sign Language (ASL) hand signs" [13] as single character values that collected together may be short, and easy for the user to recall and present. This can lead to AKE inputs that may be easily guessed by an attacker, or to system-forced frequent changes to user passwords. Such changes may be disruptive to users and lead to behaviors that thwart security goals.

User-memorized passwords can be associated with complex password 'substitution strings' selected by a server. The password and substitution strings can be securely stored on the server and preloaded on a user controlled device at the time a password is selected by the user and registered to the system. This password to substitution string mapping is illustrated in the first two columns of Table 4.

Table 4. Password substitution strings before and after mutual authentication acknowlegement

<table>
<tr><th colspan="2">User-memorized password</th><th>1st Substitution string value</th><th>2nd Substitution string value</th></tr>
<tr><td> </td><td>A</td><td>N|f4&64ejotU$5$E</td><td>PoQd,8H'*6Z0v|oH</td></tr>
<tr><td> </td><td>H</td><td>7#ktM0tzcbvz/+ uN</td><td>+Qm\2XE&nw]vgGy|</td></tr>
<tr><td> </td><td>F</td><td>B[p8Gu56Wg54TjQj</td><td>F_lH.(uU67Jgq2 ~ O</td></tr>
<tr><td> </td><td>E</td><td>/7|-:?%Xc|X$Tsv/</td><td>;}-c%y.,rS[Pm:h:</td></tr>
</table>

In Table 4, the user presents the password ‘AHFE’ to the access control system using ASL hand signs [20]. Prior to operation of a PAKE protocol, each letter is mapped to its associated substitution value to become the effective password string, “N|f4&64ejotU$5$E7#ktM0tzcbvz/+uNB[p8Gu56Wg54TjQj/7|-:?%Xc|X$Tsv/”. This derived value is used as the password input to the PAKE protocol, which uses the Diffie-Hellman protocol to create an encryption key. This key is used with a cipher to protect user credentials sent to the server, along with an unprotected user ID, during the authentication process.

On receipt of the encrypted user message, the server uses the plaintext user ID to located the password substitution strings of the user. The server uses these string to form the effective password needed to derive the same symmetric encryption key used to encrypt the message, then decrypts the ciphertext. Once the user has been authenticated, the server responds to assure the user of its identity.

During this final mutual authentication step of the PAKE protocol, all information exchange are encrypted using the shared secret key. During this protected communication, the server can create and load a new set of password substitution strings on the user device and on the server. On both devices, the current password substitution strings and the new strings are maintained until the user responds to the server, indicating the new strings have been received.

The server may then update their copy of the password substitution strings to a new set of values to be used during the next user authentication attempt, as shown in the third column of Table 4. The user can also update their local copy of the new strings without any changes to their actual password value, ‘AHFE’. Both user and server may maintain the replaced strings to mitigate the risk of substitution string update errors.

In this way the user and the server can dynamically adapt to new effective password values without disruptive changes to the familiar password memorized by the user. This adaptive processes can be performed as frequently as each user access, and effectively provides the user with an automated, one-time-password capability. This reduces the likelihood from forced user password changes of “access to an account by an attacker who has captured the account’s password” and who can guess the new password chosen by the user as a replacement based on their prior selections [21].

6 Conclusion

Biometric Authenticated Key Exchange (BAKE) and its underlying Password AKE protocol rely on weak secrets that can be used to provide strong, affordable mutual and multifactor authentication. Both protocols protect user credentials during identity authentication, enable forward secrecy, and are resistant to man-in-the-middle and spoofing attacks. They can leverage lightweight block ciphers to secure communications when using AES is not practical. BAKE and PAKE do not require users to manage digital certificates or to rely on the existence of a functioning public key infrastructure. When offered as choice alternatives, these security techniques provide support for universal user access.

The lightweight block ciphers defined in ISO/IEC 29192-2 are designed for use in resource constrained environments, such as those found in the IoT. Lightweight cryptography is not weak, but uses fewer resources than algorithms commonly found in desktop and data center environments. The algorithms can protect user credentials during identity authentication attempts, and they can provide confidentiality services during subsequent communications.

Once BAKE and PAKE have established a secure channel for communications, user password substitution strings can be securely refreshed. These user password proxies can be changed as frequently as needed without changing the underlying user password. This process ensures that complex, frequently changing secrets that are far too difficult for a user to memorize are used as inputs to BAKE and PAKE, while ensuring user convenience is maintained. User can avoid frequent password changes, choose easily recalled and easily entered passwords, and still enjoy the security benefits of password complexity and frequent password changes.

References

1. ICT Consultation: The ICT opportunity for a disability-inclusive development framework (2013). http://www.itu.int/accessibility. Accessed 25 Feb 2017
2. Mayron, L.M., Hausawi, Y., Bahr, G.S.: Secure, usable biometric authentication systems. In: International Conference on Universal Access in Human-Computer Interaction, pp. 195–204. Springer, Heidelberg, July 2013. https://www.researchgate.net/profile/Gisela_Bahr/publication/. Accessed 22 Feb 2017
3. Center for excellence in universal design: cardholder authentication (2013). http://universaldesign.ie/Technology-ICT/Irish-National-IT-Accessibility-Guidelines/Smart-Cards/Making-Smart-Card-Services-Accessible/Cardholder-Authentication/. Accessed 25 Feb 2017
4. International Organization for Standardization/ International Electrotechnical Commission: ISO/IEC 11770-4
5. Hao, F., Shahandashti, S.F.: The SPEKE protocol revisited. In: Chen, L., Mitchell, C. (eds.) Security Standardisation Research: First International Conference, SSR 2014, pp. 26–38, London, UK, 16–17 December 2014. https://eprint.iacr.org/2014/585.pdf. Accessed 23 Feb 2017
6. Griffin, P.H.: Biometric-based cybersecurity techniques. In: Advances in Human Factors in Cybersecurity, pp. 43–53. Springer, Switzerland (2016)
7. Griffin, P.H.: Transport layer secured password-authenticated key exchange. Inf. Syst. Secur. Assoc. (ISSA) J. **13**(6) (2015)
8. Griffin, P.H.: Biometric knowledge extraction for multi-factor authentication and key exchange. Procedia Comput. Sci. **61**, 66–71 (2015). Complex Adaptive Systems Proceedings, Elsevier B.V.
9. Griffin, P.H.: Telebiometric authentication objects. Procedia Comput. Sci. **36**, 393–400 (2014). Complex Adaptive Systems Proceedings, Elsevier B.V.
10. International Telecommunications Union (ITU) Broadband Commission for Sustainable Development: Digital Health: A Call for Government Leadership and Cooperation between ICT and Health (2017). Accessed 28 Feb 2017. http://www.broadbandcommission.org/Documents/publications/WorkingGroupHealthReport-2017.pdf

11. World Health Organization, Atlas of eHealth Country Profiles 2015: The use of eHealth in support of universal health coverage. http://www.who.int/goe/publications/atlas_2015/en/. Accessed 28 Feb 2017
12. Dinu, D., Le Corre, Y., Khovratovich, D., Perrin, L., Großschädl, J., Biryukov, A.: Triathlon of lightweight block ciphers for the internet of things. IACR Cryptology ePrint Archive, p. 209 (2015)
13. Griffin, P.: Secure authentication on the internet of things. In: IEEE SoutheastCon, April, 2017
14. Bogdanov, A. et al.: PRESENT: an ultra-lightweight block cipher. In: Paillier P., Verbauwhede I. (eds.) Cryptographic Hardware and Embedded Systems - CHES 2007. Lecture Notes in Computer Science, vol. 4727. Springer, Heidelberg (2007). https://link.springer.com/chapter/10.1007/978-3-540-74735-2_31. Accessed 22 Jan 2017
15. Shirai T., Shibutani K., Akishita T., Moriai S., Iwata T.: The 128-bit blockcipher CLEFIA. In: Biryukov A. (ed.) Fast Software Encryption, FSE 2007. Lecture Notes in Computer Science, vol. 4593. Springer, Heidelberg (2007). https://link.springer.com/chapter/10.1007/978-3-540-74619-5_12. Accessed 18 Jan 2017
16. International Organization for Standardization (ISO)/International Electrotechnical Commission (IEC): ISO/IEC 29192-2 Information technology – Security techniques – Lightweight cryptography – Part 2: Block ciphers (2012)
17. Biryukov, A., Roy, A., Velichkov, V.: Differential analysis of block ciphers SIMON and SPECK. In: Fast Software Encryption, pp. 546–570. Springer, Heidelberg (2014)
18. Biryukov, A., Velichkov, V., Le Corre, Y.: Automatic search for the best trails in arx: application to block cipher speck. In: Fast Software Encryption–FSE (2016)
19. Bhasin, S., Graba, T., Danger, J., Najm, Z.: A look into SIMON from a side-channel perspective. In: 2014 IEEE International Symposium on Hardware-Oriented Security and Trust (HOST), pp. 56–59. IEEE (2014)
20. Vicars, W.: American Sign Language (ASL) (2011). http://www.lifeprint.com. Accessed 14 Jan 2017
21. Zhang, Y., Monrose, F., Reiter, M.K.: The security of modern password expiration: an algorithmic framework and empirical analysis. In: Proceedings of the 17th ACM Conference on Computer and Communications Security, pp. 176–186. ACM (2010)

Internet of Things and Distributed Denial of Service Mitigation

Mohammed AlSaudi Ali(✉), Dyaa Motawa, and Fahad Al-Harby

The College of Computer and Information Security,
Naif Arab University for Security Sciences, Riyadh, Saudi Arabia
msaudi@arrowad.sa, dnmotawa@moj.gov.sa,
fmalharby@nauss.edu.sa

Abstract. Concerns about security on the Internet of Things (IoT) cover data privacy and integrity, access control and availability. IoT abuse in distributed denial of service (DDoS) attacks is a major issue, as the limited computing, communications, and power resources of typical IoT devices are prioritised in implementing functionality rather than security features. Incidents involving attacks have been reported, but without clear characterisation and evaluation of threats and impacts. The main purpose of this work is to mitigate DDoS attacks against the IoT, by studying new technologies and identifying possible vulnerabilities and potential malicious uses, and building protections against them. The simulation results show that the proposed scheme is effective in mitigating DDoS attacks on IoT.

Keywords: Internet of Things (IoT) · Mitigation · Distribution denial of service (DDoS) attack · Botnets

1 Introduction

Poor security on many IoT devices makes them soft targets and often victims may not even know they have been infected. The Internet of Things (IoT) has become an issue due to the increasing number of DDoS attacks. In many cases, people and companies use IoT devices, such as digital cameras and DVR players, but do not know about the dangers of exploited DDoS attacks. With the rapid increase in the use of IoT devices, the potential power of a DDoS attack when using IoT devices could increase dramatically. Research by Peraković et al. analysed the impact of the IoT on the volume of DDoS attacks. The paper stated that the rise in the number of IoT devices and the low level of protection implemented offers the possibility to create a botnet network that is able to generate significantly greater amounts of attack traffic. The analysis also showed a correlation between the increase in IoT devices and the increase in DDoS attacks [1].

Recent Gartner research estimated there to be more than 2.9 billion connected IoT devices in consumer smart home environments in 2015. These connected devices could provide a much larger surface for attackers to target home networks. The use of weak passwords is a security issue that has repeatedly been seen in IoT devices. These devices may not have a keyboard, so configuration has to be done remotely. Unfortunately, not all vendors force the user to change the devices' default passwords and

D. Nicholson (ed.), *Advances in Human Factors in Cybersecurity*,
Advances in Intelligent Systems and Computing 593, DOI 10.1007/978-3-319-60585-2_3

many have unnecessary restrictions which make the implementation of long, complex passwords impossible.

1.1 Objective of the Study

Consideration of how to create countermeasures and mitigate against DDoS attack, especially with a huge number of IoT devices, is the objective of this study.

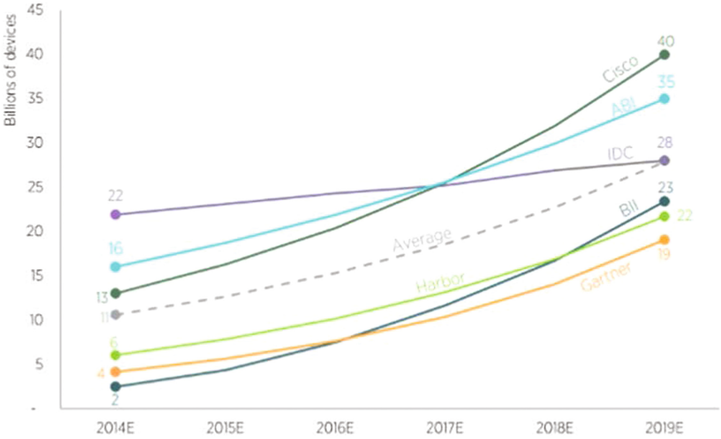
2 Background

In this section, we review the related works carried out in this domain. Information security has been a field of increasing importance in the information field. DDoS attacks have been shown to pose a threat to web services for some time and this has become a more serious threat recently due to advancements in internet technology and IoT.

2.1 Internet of Things (IoT) [2]

Today, we are experiencing rapid innovation and a vast array of developments in technology and subsequent increases in human interaction with machines through the Internet of Things (also called machine to machine or objects). Internet of Things describes the connection of devices to the internet using embedded software and sensors to communicate, collect and exchange data with one another. With IoT, the world is wide open, as is the IP of each device used to access it.

The reality of the IoT market around the world is undoubtedly emerging. Vendors are evolving their solutions in a supply-driven market that is on the edge of becoming a demand-driven market (Fig. 1).



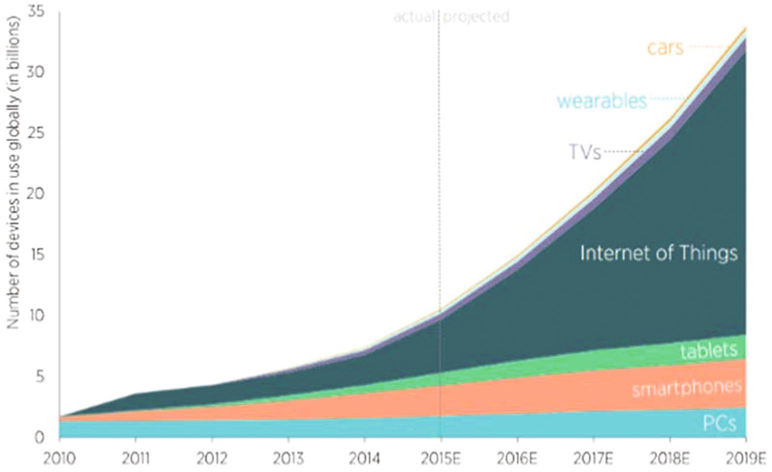
Source: John Greenough, "The Internet of Things is Rising: How the IoT Market Will Grow Across Sectors," *Business Insider Intelligence*, October 8, 2014. Produced by Adam Thierer and Andrea Castillo, Mercatus Center at George Mason University, 2015.

Fig. 1. Internet of Things (IoT) grow market

- Cisco Project believes that 40 billion intelligent things will be connected and communicating by 2019.
- ABI Research estimates that more than 35 billion networked devices will be in use by 2019.
- International Data Corporation (IDC) notes that around 28 billion networked devices were in use by 2012 and that 212 billion devices will be connectable by 2020, 15% (around 31.8 billion) of which will be installed and operational by the end of 2020.
- Gartner anticipates that 19 billion IoT devices will be in operation by 2019 and 25 billion devices will be online by 2020.
- Harbor Projects estimates that 21.7 billion IoT devices will be connected and in use by 2019.
- Machine Research reports that roughly 7.2 billion "machine to machine connected consumer electronic devices" will be in global use by 2023.
- Business Insider Intelligence (BII) estimates that there will be a total of 23.4 billion IoT devices connected by 2019 and that adoption will be driven by enterprise and manufacturing sectors.

2.2 Distributed Denial of Service Attack (DDoS)

A DDoS is a type of DoS attack that uses a large number of computers, infected by a worm or a Trojan horse, to launch simultaneous attacks on a single target over a very short period. This causes the system to slow or shut down, thereby denying users the ability to use it. This is done by sending a large amount of requests simultaneously from the attacker's host; this is called flooding and is done in order to prevent services from being provided to legitimate users (Fig. 2).



Source: John Greenough, "The Internet of Everything 2015," *Business Insider Intelligence*. Produced by Adam Thierer and Andrea Castillo, Mercatus Center at George Mason University, 2015.

Fig. 2. The internet of everything: devices in use globally

What Does the Attacker Want? There are several reasons why an attacker would like to cause DDoS. It could be a group of people who would like to bring down a specific webpage or website in order to keep it isolated from business. Thus the company might lose all its online transactions and end up failing. Rivalry in business could be one main factor for these attacks. Another reason which increases the frequency of these attacks is the simplicity involved. A beginner could perform this type of attack effortlessly without having much technical expertise. Attackers often post their attacking tools and scripts online to aid others who like to carry out similar operations. There are websites and forums that give out tools along with instruction manuals to make it easier for anyone to carry out such attacks. People who carry out such attacks without having actual knowledge about them are called ‘Script Kiddies’.

There could also be other reasons for attacks, for example, a group might not like the content published on a specific website and would like to bring it down.

Types of DDoS Attacks? There are two types of DDoS attacks:

(a) Network-centric attacks: overload a service by using up bandwidth.
(b) Application-layer attacks: overload a service or database with application calls.

2.3 Information Security (Countermeasures Against DDoS)

The DDoS attack has opened up an important conversation about internet security and volatility. Not only has it highlighted vulnerabilities in the security of IoT devices that need to be addressed, but it has also sparked further dialogue in the internet infrastructure community about the future of the internet. As we have in the past, we look forward to contributing to that dialogue [3].

IoT Malware – Common Traits [4]. While IoT malware is becoming more sophisticated, the fact that it is being used mostly for DDoS attacks allows us to distinguish several common traits that are seen within the variety of existing malware families. As far as malware distribution goes, attackers take a straightforward approach. While some malware variants need to be manually installed on the device, the most common method consists of a scan for random IP addresses with open Telnet or SSH ports, followed by a brute-force attempt to login with commonly used credentials.

IoT Malware Families [4]. Below are the most recognisable and prevalent malware families targeting embedded devices (Table 1).

Table 1. IoT malware families

Linux.Dofloo (aka AES.DDoS, Mr. Black)	Linux.Xorddos (aka XOR.DDoS)
Linux.Pinscan/Linux.Pinscan.B (aka PNScan)	Linux.Darlloz (aka Zollard)
Linux.Kaiten/Linux.Kaiten.B (aka Tsunami)	Linux.Ballpit (aka LizardStresser)
Linux.Routrem (KTN-Remastered, KTN-RM)	Linux.Moose
Linux.Routrem (aka Remainten)	Linux.Wifatch (aka Ifwatch)
Linux.Gafgyt (aka GayFgt, Bashlite)	Linux.LuaBot

Vulnerable Devices [4]. Most IoT malware targets non-PC embedded devices. Many are Internet-accessible but, because of their operating system and processing power limitations, they may not include any advanced security features. Embedded devices are often designed to be plugged in and forgotten after a very basic setup process. Many don't get any firmware updates or owners fail to apply them and the devices tend to only be replaced when they've reached the end of their lifecycle. As a result, any compromise or infection of such devices may go unnoticed by the owner and this presents a unique lure for the remote attackers.

DDoS Source Countries 2016. Analysis of a Symantec honeypot that collects IoT malware samples found that the highest number of IoT attacks originated in China, accounting for 34% of attacks seen in 2016. Further, 28% of attacks stemmed from the US, followed by Russia 9%, Germany 6%, the Netherlands 5%, and Ukraine 5%. Vietnam, the UK, France, and South Korea rounded out the top ten [4] (Fig. 3).

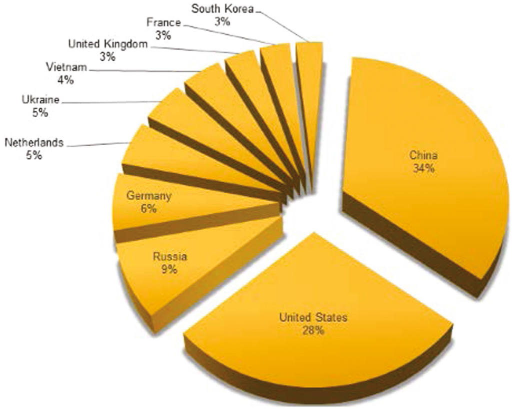


Fig. 3. Top ten attack origins on monitored IoT honeypot in 2016

This quarter marks a full year with China as the top source country for DDoS attacks with just under 30% of attack traffic this quarter, as shown in Fig. 4. Importantly, the proportion of traffic from China has been reduced from 56%, which has had a significant effect on the overall attack count, and has led to the 8% drop in attacks seen quarter 3.

Cross-Platform Malware [4]. It is quite simple for the attackers to cross-compile their malware for a variety of architectures. While the most common targets are the ×86, ARM, MIPS, and MIPSEL platforms, attackers continue to expand the number of potential targets and have also been creating variants for PowerPC, SuperH and SPARC architectures. By doing so, the list of potentially vulnerable devices increases, with more web servers, routers, modems, NAS devices, CCTV systems, ICS systems, and other devices added to the list of potential targets.

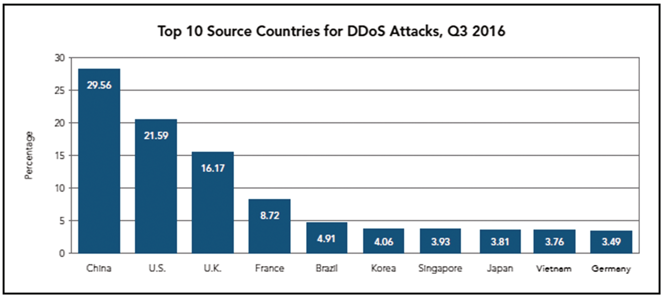


Fig. 4. Top 10 source countries for DDoS attacks, Q3 2016

Top Passwords (*Password Dictionary*) [4]. Attacks on Symantec's honeypot also revealed the most common passwords IoT malware used to attempt to log into devices. Not surprisingly, the combination of 'root' and 'admin' leads the chart, indicating that default passwords are frequently never changed. The default credentials (user name: *admin* and password: *admin*) also feature highly. As reported in May 2016, an old vulnerability in routers allowed the worms targeting embedded devices to spread across thousands of networks' routers running outdate firmware. It looks like the attackers behind IoT malware still count on the presence of unpatched routers in the wild. Further down the charts we see the default credential combination for the Raspberry Pi devices (user name: pi and password: raspberry), which indicates a growing trend of attackers specifically targeting this platform.

Top 10 Brute-force (*Dictionary*) default usernames and passwords used against IoT devices (Table 2).

Table 2. Top 10 default usernames and passwords used against IoT devices

#	Username	Password	#	Username	Password
1	root	admin	6	DUP admin	password
2	admin	root	7	test	1234
3	DUP root	123456	8	oracle	test
4	ubnt	12345	9	postgres	qwerty
5	access	ubnt	10	pi	raspberry

3 Methods and Results

3.1 "New Technologies Pose New Threats" [5]

Technology has changed our lives for the better; there is no doubt about it. However, it has also introduced various risks into them. In fact, this is one of the most interesting things about technology: its effect depends on the people behind it. Sadly, alongside inspiring figures that move technology and the world forward, there is always a group abusing it for the worst. Some corporations are constantly studying new technologies and identifying possible vulnerabilities and potential malicious uses, and building protections against them. The mission is to stay one step ahead of malware developers.

By using *IBM SPSS* to output some results from survey analysis, we focused on the important questions after checking the reliability analysis by using *Cronbac Alpha* as:

a. Cronbach's Alpha (*Reliability Analysis*)

Factor is **0.704** and it's greater than **0.60** it means have a good reliability (Table 3).

Table 3. Cronbach's alpha (reliabilities analysis)

Case processing summary			
		N	%
Cases	Valid	6	10.3
	Excluded[a]	52	89.7
	Total	58	100.0

Reliability statistics	
Cronbach's alpha	N of items
.704	11

[a]Listwise deletion based on all variables in the procedure.

b. By using frequencies for part (have Smart Building Devices) we consider the below table (Table 4).

Table 4. Frequencies for part (have Smart Building Devices)

Have smart devices					
		Frequency	Percent	Valid percent	Cumulative percent
Valid	(لا) No	34	58.6	63.0	63.0
	(نعم) Yes	20	34.5	37.0	100.0
	Total	54	93.1	100.0	
Missing	System	4	6.9		
Total		58	100.0		

c. By using Frequencies for part (BIoT Connected to the Internet) we consider the below table (Table 5).

Table 5. Frequencies for part (BIoT connected to the internet)

BIoT connected web					
		Frequency	Percent	Valid percent	Cumulative percent
Valid	(لا) No	2	3.4	10.5	10.5
	(نعم) Yes	17	29.3	89.5	100.0
	Total	19	32.8	100.0	
Missing	System	39	67.2		
Total		58	100.0		

d. By using Frequencies for part (Configuration devices by yourself) we consider the below table (Table 6).

Table 6. Frequencies for part (configuration devices by yourself)

BIoT config by yourself					
		Frequency	Percent	Valid percent	Cumulative percent
Valid	(لا) No	15	25.9	75.0	75.0
	(نعم) Yes	5	8.6	25.0	100.0
	Total	20	34.5	100.0	
Missing	System	38	65.5		
Total		58	100.0		

e. By using Frequencies for part (Change the default username and password) we consider the below table (Table 7).

Table 7. Frequencies for part (change default username & password)

Default password changed					
		Frequency	Percent	Valid percent	Cumulative percent
Valid	(لا) No	28	48.3	52.8	52.8
	(نعم) Yes	25	43.1	47.2	100.0
	Total	53	91.4	100.0	
Missing	System	5	8.6		
Total		58	100.0		

f. By using Frequencies for part (Tools and means of protection) we consider the below table (Table 8).

Table 8. Frequencies for part (tools and means of protection)

Have a protection					
		Frequency	Percent	Valid percent	Cumulative percent
Valid	(لا) No	22	37.9	41.5	41.5
	(نعم) Yes	18	31.0	34.0	75.5
	(لا أعلم) I do not know	13	22.4	24.5	100.0
	Total	53	91.4	100.0	
Missing	System	5	8.6		
Total		58	100.0		

3.2 Software Used

The use of the following software helped us to obtain the results: Tableau 9.3, SurveyMonkey, Microsoft Forms, and IBM SPSS Statistics.

4 Recommendations

Firstly, the FBI [6] suggests precautionary measures to mitigate a range of potential DDoS threats and IoT compromises, to include, but not be limited to the following:

- Have a DDoS mitigation strategy ready ahead of time and keep logs of any potential attacks.
- Implement an incident response plan that includes DDoS mitigation and practice this plan before an actual incident occurs. This plan may involve external organisations such as an Internet service provider, technology companies that offer DDoS mitigation services, and law enforcement. Ensure that the plan includes the appropriate contacts within these external organisations. Test activate an incident response team and third party contacts.
- Implement a data back-up and recovery plan to maintain copies of sensitive or proprietary data in a separate and secure location. Back-up copies of sensitive data should not be readily accessible from local networks.
- Review reliance on easily identified Internet connections for critical operations, particularly those shared with public facing web servers.
- Ensure upstream firewalls are in place to block incoming UDP packets.
- Change default credentials on all IoT devices.
- Ensure that software or firmware updates are applied as soon as the device manufacturer releases them.

When researching the subject of anti-DDoS, we found one example from a famous telecommunication company in Saudi Arabia have this technique already. This is Mobily Co. and it mentions on their site that "In today's technology-run business environment, a Distributed Denial of Service (DDoS) attack is one of the most crippling threats to companies. These attacks are increasingly targeted at specific businesses and government agencies. With multiple variations, such attacks can cause downtime, drive up bandwidth costs, result in customer churn, and can eventually lead to severe financial losses. The Mobily anti-DDoS offering is a part of Mobily Managed Security Services. It provides a cloud-based DDoS detection and mitigation service. You can implement a robust anti-DDoS service without investing in expensive hardware or professional services. Mobily anti-DDoS proactively monitors organisational traffic patterns from within Mobily Data Centres. The in-country traffic monitoring and analyses survey the national traffic within the Kingdom's borders.

4.1 Research Recommendations

Based on the results of the case study and analysis of the survey, we are highlighting the mitigation options against DDoS attacks that exploit IoT devices, i.e. outlining countermeasures. There are three categories situated around responsibilities of three groups:

Manufacturers should be making some modifications to firmware to countermeasure against attacks, e.g.

a. They should be improving the firmware of these devices by making it mandatory to change the factory default username and password after the first login, and highlighting that this must be not equal to the factory default username and password. This will mitigate many DDoS attacks to exploit smart devices.
b. They should make the Remote Access option ***Opt-in***, which means this option is unchecked by default, so the user must check it to activate it.
c. They should add a Firewall to the device and make it ***Opt-out***, which means this option is checked by default, so the user must uncheck it to disable it.
d. There should be collaboration with companies that focus on protection of information security, to produce ***Anti-Attack***, which could be riveted with Firmware in their devices.

Vendors and ISPs should implement mitigation systems and increase awareness of countermeasures to attacks, such as:

a. They should be implementing and providing a cloud-based DDoS detection and mitigation service to prevent this kind of attacks, e.g. using Anti-DDoS Service.
b. They should be spreading the culture of protection and information security against these kinds of attacks, by increasing awareness of the risks that we will face if we leave devices with factory default settings.
c. Technicians must change the factory default user name and password, and tell end users how to change those settings when necessary.

End users should be aware of the countermeasures necessary to prevent attacks, e.g.: Users need to have knowledge about the importance of using security software; and users must change the factory default username and password, or change it after the initial configuration by others.

5 Conclusion

The presented results show that DDoS is a powerful technique that enhances attack capabilities. IoT infrastructures offer a huge attack surface in terms of DDoS that has not yet been widely explored in attacks. Given the saturation identified in tests, it is expected that IoT devices, when used as reflectors, will be hit at least as severely as victims. Fortunately, current best practices for prevention are available and can be used to mitigate some of the attacks. In short, the threat is real, but there are ways to deal with it; however, it requires efforts in management and enhancement of IoT software.

References

1. Radware Ltd.: DDoS Handbook. Radware, Mahwa (2015)
2. Castillo, A., Thierer, A.: Projecting the growth and economic impact of the internet of things. Mercatus Center at George Mason University, Washington (2016)
3. Dyn Analysis Summary of Friday October 21 DDoS Attack. http://dyn.com/blog/dyn-analysis-summary-of-friday-october-21-attack

4. Symantec Security Response. https://www.symantec.com/connect/blogs/iot-devices-being-increasingly-used-ddos-attacks
5. Check Point Software Technologies Ltd. Akamai's [state of the internet]/security. http://blog.checkpoint.com/2016/04/12/new-technologies-pose-new-threats/
6. FBI: Distributed denial of service attack against domain name service host highlights vulnerability of "Internet of Things" devices. In: Private Industry Notification 161026-001 FBI Cyber Division, New York (2016)

Eye Tracking Graphical Passwords

Martin Mihajlov[1(✉)] and Borka Jerman-Blazic[2]

[1] Ss. Cyril and Methodius University, Goce Delcev 9b, 1000 Skopje, Macedonia
martin@eccf.ukim.edu.mk
[2] Jozef Stefan Institute, Jamova 16, 1000 Ljubljana, Slovenia
borka@e5.ijs.si

Abstract. In this paper, we investigate the cognitive process behind graphical password selection by using eye-tracking. The goal of the study is to discover how users perceive and react to graphical authentication during graphical password selection, which is valuable for improving the design concepts in novel authentication mechanisms. As a result, we present the initial results of the study noting cognitive differences based on gender, and we define user profiles for enrolment and authentication processes.

Keywords: Eye tracking · Graphical authentication · Graphical passwords

1 Introduction

In the past decade, graphical authentication has been proposed as a viable solution to usable security issues in authentication. The main idea behind this concept lies in dual-code theory, which suggests that the different subsystems for processing verbal and pictorial information help enhance memory when text and pictures can be related to each other [1]. Hence, the foundation for graphical authentication relies on the fact that humans have a vast memory for images, which, as shown by Bower et al. [2], is not affected by the person's cognitive abilities. Furthermore, Shepard [3] has shown the existence of improved performance in both recall and recognition for pictorial over verbal representations. Consequently, graphical authentication systems have been designed to authenticate the user based on image input with the aid of a mouse [4, 5], stylus [6, 7] or touch screen [8].

The study of graphical password selection is highly dependent on the properties of the authentication mechanism itself and the selected methodology for analysis. In order to extend the research of graphical password selection, in our study we used eye-tracking as an evaluation methodology. Our goal was to trace how participants processed graphical authentication during enrollment and authentication sessions.

The graphical authentication mechanisms used in this study, ImagePass, uses single-object images as units of the authentication key. In order to enter the graphical password the user has to click on a series of images presented in a 4 × 4 grid, which contains both valid and decoy authentication images. When both the sequence and the clicked images are correct, the user is granted permission to access the system [9]. An extensive analysis of the graphical password properties has shown that there are gender differences between

D. Nicholson (ed.), *Advances in Human Factors in Cybersecurity*,
Advances in Intelligent Systems and Computing 593, DOI 10.1007/978-3-319-60585-2_4

male and female users in the type of images selected for the authentication key [10]. Hence, in this study we use eye tracking methodology in order to better understand the cognitive process of selecting images in graphical authentication. This is underlined by two theoretical assumptions, which relate eye-movement with cognitive processing. The first assumption, immediacy, relates that information processing takes place at the same time that the information is encountered, regardless of the level on its occurrence. The second, eye-mind assumption, suggests that observed visual information is directly processed due to the fact that human gaze is closely linked to the focus of attention [11].

While it has been suggested as a technology to be used in the process of authentication [12, 13], eye tracking research in the field of graphical authentication is scarce. Two studies have been performed on the Passpoints system where the authors investigate whether eye fixations can predict the location points for graphical passwords [14, 15]. For analyzing gender differences in graphical password selection, eye tracking has only been used provisionally used as a method for analysis [16]. When it comes to processing images, recently, eye tracking has been as a methodology to investigate the process of learning from an illustrated science text by examining the effects of using a concrete or abstract picture to illustrate the text [17]. As for gender-based perception, eye tracking was used to quantify gender differences in visual landmark utilization during navigation [18].

2 Methodology

2.1 Participants and Equipment

Data from at least 30 participants is required in order to have valid eye tracking results [19]. Therefore, for the purpose of the experiment 33 participants, 18 male and 15 female were recruited through direct contact and advertisements. Demographically, the ages of the participants ranged from 19 to 32 with a mean age of 27.2 years. All of the participants had 20/20 vision and had either a medium or a high web experience. Regarding education, 54% had some college education, while 46% were college graduates or higher. None of the participants had any previous experience with graphical authentication or eye tracking.

All of the sessions in this experiment were conducted at the Laboratory for Open and Network Systems at the Jožef Stefan Institute in Ljubljana, Slovenia. Eye movements were collected with the Tobii T60 eye tracker sampling eye position at a rate of 60 Hz, with a fixation time of 200 ms, latency of 33 ms and drift of 0.1°. To display stimulus pages and to define regions of interest, the Tobii Studio 1.5.4 software was used. The content was viewed on a 17″ TFT monitor a personal computer running Windows 7. The eye tracker was successfully calibrated for all participants, thus complete eye movement data was recorded for all subjects.

2.2 Experiment Design

The ImagePass graphical authentication mechanism was deployed on a remote web server from March to June 2015. The experiment consisted of one enrollment session and one closing session taking place in a laboratory, and up to five authentication sessions performed over different time intervals. The laboratory sessions were eye tracked in a controlled environment, while the intermittent sessions were performed at the participants' convenience.

To avoid analyzing authentication as a primary task, all of the participants were misdirected, and informed that the purpose of the experiment is an analysis of their Internet search-behavior patterns. For this reason, an additional module was developed and attached to the ImagePass system with superficial functionalities intended to simulate search behavior analysis. Predefined search tasks were loaded in the module in form of a question that could be answered with a multiple-choice response. For each session four different search tasks were defined where the participant had to use a specific search service or visit a particular site, find the requested information and answer the multiple choice question.

In the first eye tracking session data was collected from the all the Enrollment screens, while in the closing eye tracking session data was collected only for the Authentication screen. The screens were categorized in areas of interest (AOIs) with the most observed variable being fixation time on a specific AOI. The same AOIs were defined for the Login and the Choose Username screens. The area around the username textfield was defined as Username, and the top-right corner of the screen was defined as Create Account. The Select Password screen had the most complex AOI. Initially the AOI were divided as Info, the text area in the top left side of the screen, Selected Password, the area that shows the clicked images, Confirm, the area on the lefts side of the screen around the function buttons and Image Selection the area around the authentication imageset. However, in order to analyze the graphical password selection process in more detail, the Image Selection AOI was further subdivided into five rows and six columns. Finally, the Authentication screen was similarly divided into: Info, Selected Password, Confirm and Authentication Grid.

2.3 Procedure

The eye tracked enrollment behavior of the participants was analyzed through individual 20–25 min sessions taking place in a period over 2 weeks. During each session, the participant was introduced to the nature of the experiment before signing a consent form. In addition, the participant filled out an interest questionnaire to determine their demographic characteristics. In order to familiarize the participant with the eye tracking hardware a practice session was set-up in which the eye tracker was calibrated to the participant's eye movements and a photographic test image was shown to the participant for two seconds. In order to get results that are more relevant and treat authentication as a secondary task, the participants were intentionally misdirected that the purpose of the experiment is to analyze search behavior through an online application, which tracks and logs their search activities.

During the enrollment eye tracking sessions, the participants had to perform two tasks: enroll to the ImagePass system, and Complete Search Tasks. The enrollment task was subdivided into three subtasks with each subtask taking place on a different screen interface: create account, create username and choose graphical password. The second task was to complete four search tasks where the participants had to find some specific information on the World Wide Web through a particular search service and answer a multiple choice question regarding that information. Over the following period of 4 weeks, the participants were asked to login to the system remotely on a weekly basis and complete a new set of search-based tasks. For the final session, the participants were invited back to the laboratory to essentially eye-track their behavior after continuous use of the system.

To analyze the eye tracking data the individual users' viewing behaviors were analyzed through gaze replays and by observing the respective gaze plots. In addition, the fixation times on predefined areas of interest were also evaluated in order to determine specific user scan path profiles and gender specific behavior. The data for all search tasks in all sessions was discarded as irrelevant.

2.4 Research Questions and Hypotheses

A preliminary eye tracking study on a previous prototype produced initial observations that pointed towards potential differences between male and female participants when using graphical authentication (Mihajlov et al. 2013). To explore the potentially cognitive nature of this finding on users' perception of the system this follow up study analyzes the patterns of users based on gender. In addition, as stated previously, graphical passwords have a more cognitive nature than other general online-tasks. The selection process in graphical authentication, especially when single everyday objects are used as parts of the authentication key, will have a personal component. This individuality would be based on the users' everyday habits, familiarity and interactions with the selected objects. Consequently, it would be expected that the users' scan path patterns would not follow the general scanpath theory. Formally, the hypotheses are stated as follows:

- *H1:* There is an observable difference in perception of the graphical authentication system between male and female users.
- *H2:* There is a difference in general scan path patterns between ImagePass and expected general scan patterns for web-based applications.

3 Results and Discussion

To evaluate the authentication process, all of the eye tracking sessions were analyzed separately. The potential differences between male and female participants were noted as observations, as using statistical tests to evaluate the differences in their performance would yield a sample too small for relevant results.

While creating the username, the tested participants spent more time on the Username AOI (M = 4.62, SD = 4.07) than the Create User AOI (M = 2.67, SD = 1.47). This difference is observably larger for female users (M = 5.74, SD = 4.11, and M = 1.66, SD = 1.52) than male users (M = 3.70, SD = 4.48, and M = 3.47, SD = 1.18), which, nevertheless, is within the expected range.

The analysis of the data for the Graphical Password Selection screen yielded results that are more complex. On average participants spent 35.56 s (SD = 4.67) on selecting their graphical password, without any noticeable differences between male and female participants. Male participants paid more attention to the Selected Password AOI (M = 3.84, SD = 1.93) and Info AOI (M = 4.95, SD = 4.77), than female participants (M = 2.99, SD = 3.10, and M = 2.30, SD = 1.11). On the other hand, female participants spent more time observing the Image Selection AOI (M = 27.61, SD = 2.78), than male participants (M = 22.59, SD = 3.44). The details of the descriptive statistics are given in Table 1.

Table 1. Descriptive statistics for fixation lengths on AOIs for select password screen

		Not on AOI	Row 1	Row 2	Row 3	Row 4	Row 5	Selected password	Info	Confirm
Male	M	1.45	6.62	5.29	3.96	3.24	3.48	3.84	4.95	2.50
	SD	0.71	3.51	2.40	3.87	3.67	3.77	1.93	4.77	1.00
Female	M	1.24	8.12	5.05	5.35	5.05	4.04	2.99	2.30	2.03
	SD	0.65	9.10	1.34	2.2	1.07	0.20	3.10	1.11	2.87

The Image Selection AOI was subdivided into five rows in order to analyze the data more precisely. There is a decrease of attention as the user looks further down in this AOI, with a higher drop between the first and second row and lower drops for the following rows. This would imply that users are more likely to select images from the higher than the lower rows.

When the Image Selection AOI is subdivided into five columns, the results are slightly different. There is a slight up-down variation as the participant observes the first three columns, before the attention start decreasing for the last three columns. When this attention is subdivided by gender groups, the results show a different pattern for the first three columns. Female participants pay more attention to the first column and then their attention drops for the remaining columns. However, male participants have an increase in attention from column 1 to column 3 with a sharp drop for the remaining three columns.

Generally, the first fixation for all of the participants was within the Info AOI. They would then proceed with selecting the graphical password by fixating on the Selection Grid AOI. While selecting the graphical password, the participants would notice the clicked images appearing in the Selected Password AOI. They would also notice the New Images button however, the button remained unclicked for all the testing sessions. Finally, the focus would fall within the Confirm AOI as they click the Select button to finish the selection process. For the selection of the graphical password itself, the following two general user profiles can be determined:

'Get Me Out of Here' Profile. In this profile the user views graphical, and probably all authentication mechanisms as a potential nuisance. The graphical password selection grid gets a perfunctory glance, and the graphical password itself is selected quickly, usually by either repetitively clicking one image, or alternating the clicks between two images. The location of the selected images in the graphical password selection grid is within the first three rows and the first three columns with the other images being almost completely ignored. This profile is more likely to belong to a male rather than a female participant (Fig. 1).

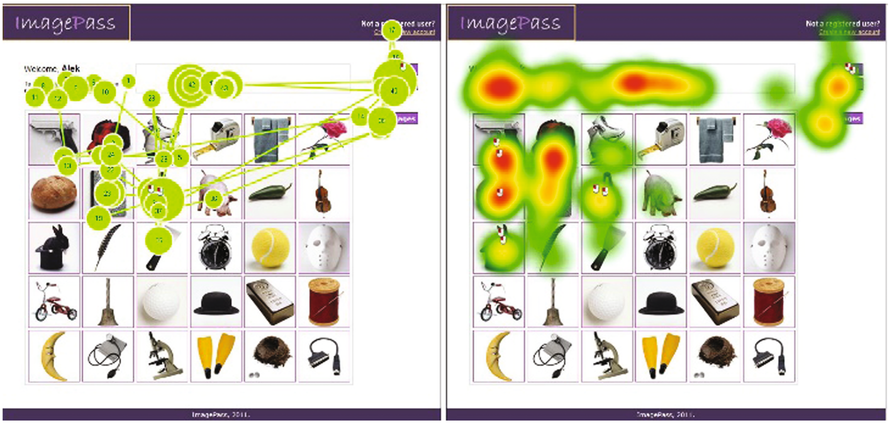


Fig. 1. Sample gaze plot and aggregated heat map for the GMOH profile.

'Let Me Think' Profile. In this profile, the user indulges into graphical password selection more carefully. The images in the graphical password selection grid are viewed longer and more attentively. The selection of the images constituting the graphical password is more varied including more rows and columns from the selection grid, although there is still a preference for higher rows, of left columns. There is either no repetition of images or one repetition of an image in a few cases in the selected authentication key. This profile is more likely to belong to a female rather than a male participant.

As mentioned previously, the authentication process was eye tracked 1 month after the enrollment to the ImagePass system. Most of the participants followed the protocol of accessing the system on a weekly basis to perform Internet search task and answer the related questions. A few participants did not adhere to the procedure and had authenticated either once or zero times to the system and had difficulties remembering the graphical password. There were no observable differences between male and female participants. Based on the data analysis the following two user profiles can be observed during authentication:

'No Problem' Profile. In this profile, the user has no difficulty in recognizing the graphical password and uses that graphical password to authenticate to the system. As the page loads, they immediately focus on the Authentication Grid AOI and quickly scan the presented images until they focus on the first image from their authentication imageset. Depending on the grid positions of the first image, the user recognizes and clicks the image within 10 to 20 fixations from the initial page load. When the initial image is in the lower part of the identification grid the user temporarily fixates on subsequent images belonging to the authentication imageset before focusing and clicking on the first image. Occasionally, the user fixates on the Selected Password AOI to check how many images have been clicked. With a sporadic glance to the Info AOI, the user finalizes the authentication process by focusing on the Confirm AOI and clicking the Login button (Fig. 2).

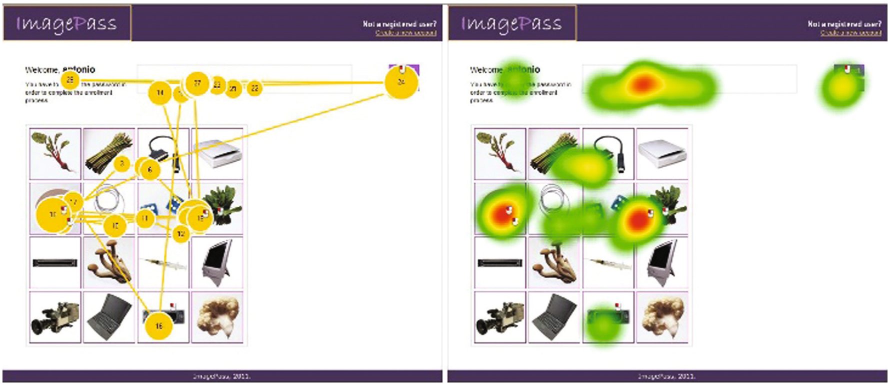


Fig. 2. Sample gaze plot and aggregated heat map for the NP profile.

'What Was It' Profile. In this profile, the user does not recognize the graphical password immediately and in this case, the observed behavior is expectantly more erratic. The user focuses on the Authentication Grid AOI and carefully scrutinizes all the images in order to recognize their corresponding authentication set. After the wrong graphical password is entered, the user starts shifting the focus more chaotically, between the Authentication Grid AOI, the Selected Password AOI and the Info AOI. In the meantime, a variation of the images in the imageset is clicked. In the end, either the user recognizes the graphical password after few attempts or in the two instances where the graphical password was not correctly recognized the user opted for the "Create a new account" option, as password recovery was not available.

4 Conclusion

In this brief paper, we reported on an eye-tracking study, which accesses the cognitive process during graphical password selection. Generally, the results of the experiment reinforce both proposed hypothesis. The differences in perception of the ImagePass graphical authentication mechanism are observable only during enrollment and are not evident during continuous authentication to the system, which is in support of hypothesis 1. In addition, the existence of different profiles during enrollment and authentication is in support of hypothesis 2, as scanpath theory requires the presence of repetitive patterns, which are dissimilar between persons for a specific stimulus and between stimuli for a specific person.

In future research we would like to analyze how a mobile environment influences users who authenticate with graphical authentication mechanisms. We would also like to devise setups that would allow for a mobile eye tracking study in order to analyze the graphical password selection process in more detail.

References

1. Paivio, A.: Mental representations: a dual-coding approach. Oxford University Press, New York (1986)
2. Bower, G.H., Karlin, M.B., Dueck, A.: Comprehension and memory for pictures. Mem. Cogn. **3**(2), 216–220 (1975)
3. Shepard, R.M.: Recognition memory for words, sentences and pictures. J. Verbal Learn. Verbal Behav. **6**, 156–163 (1967)
4. Wiedenbeck, S., Waters, J., Birget, J.C., Brodski, A., Memon, N.: PassPoints: design and longitudinal evaluation of a graphical password system. Int. J. Hum. Comput. Stud. **63**, 102–127 (2005)
5. Chiasson, S., van Oorschot, P.C., Biddle, R.: Graphical password authentication using cued click points. In: European Symposium on Research in Computer Security, pp. 359–374 (2007)
6. Blonder, G.E.: Graphical password. U.S. Patent 5559961. Lucent Technologies, Inc., New Jersey (1995)
7. Jermyn, I., Mayer, A., Monrose, F., Reiter, M., Rubin, A.: The design and analysis of graphical passwords. In: Proceedings of the 8th USENIX Security Symposium (1999)
8. Real User Corporation: Two Factor authentication, graphical passwords – passfaces (2012). http://www.passfaces.com
9. Mihajlov, M., Jerman-Blazič, B., Ilievski, M.: ImagePass-designing graphical authentication for security. In: 7th International Conference on Next Generation Web Services Practices (NWeSP), pp. 262–267. IEEE (2011)
10. Mihajlov, M., Jerman-Blažič, B., Shuleska, A.C.: Why that picture? Discovering pass-word properties in recognition-based graphical authentication. Int. J. Hum. Comput. Interact. **32**(12), 975–988 (2016)
11. Just, M.A., Carpenter, P.A.: A theory of reading: from eye fixations to comprehension. Psychol. Rev. **87**, 329–354 (1980)
12. Hoanca, B., Mock, K.: Secure graphical password system for high traffic public areas. In: Proceedings of the 2006 Symposium on Eye Tracking Research and Applications, p. 35. ACM (2006)
13. Kinnunen, T., Sedlak, F., Bednarik, R.: Towards task-independent person authentication using eye movement signals. In: 2010 Symposium on Eye-Tracking Research and Applications, pp. 187–190 (2010)
14. LeBlanc, D., Chiasson, S., Forget, A., Biddle. R.: Can eye gaze predict graphical passwords? In: 4th ACM Symposium on Usable Privacy and Security (SOUPS), Pittsburgh, USA (2008)
15. LeBlanc, D., Forget, A., Biddle. R.: Guessing click-based graphical passwords by eye tracking. In: IEEE Privacy, Security, Trust (PST), Ottawa, Canada (2010)
16. Mihajlov, M., Trpkova, M., Arsenovski, S.: Eye tracking recognition-based graphical authentication. In: 7th International Conference on Application of Information and Communication Technologies, pp. 1–5. IEEE (2013)
17. Mason, L., Pluchino, P., Tornatora, M.C., Ariasi, N.: An eye-tracking study of learning from science text with concrete and abstract illustrations. J. Exp. Educ. **81**(3), 356–384 (2013)
18. Andersen, N.E., Dahmani, L., Konishi, K., Bohbot, V.D.: Eye tracking, strategies, and sex differences in virtual navigation. Neurobiol. Learn. Mem. **97**(1), 81–89 (2012)
19. Eraslan, S., Yesilada, Y., Harper, S.: Eye tracking scanpath analysis on web pages: how many users? In: 9th Biennial ACM Symposium on Eye Tracking Research and Applications, pp. 103–110. ACM (2016)

Understanding and Discovering SQL Injection Vulnerabilities

Abdullaziz A. Sarhan(✉), Shehab A. Farhan, and Fahad M. Al-Harby

Naif Arab University for Security Sciences, Riyadh, Saudi Arabia
{aalwany,shomidi}@moe.gov.sa, fmalharby@nauss.edu.sa

Abstract. The Internet has become very important today and a large part of everyday life, so it is vital to focus on security for web applications and mobile services, so as to protect electronic commerce, electronic government, social media and all electronic services that transfer information through it. News reports of attacks on services are frequent. Hackers use vulnerabilities in software or hardware to destroy services, and one of the common vulnerabilities is SQL injection. This vulnerability comes down to poor coding practices of junior programmers writing SQL dynamics at the back end. This paper creates a case study that considers two scenarios using ASP.NET 2015 and SQL Server 2014. In the first scenario, we check whether SQL injection exists or not, then make an SQL injection from the front end and add it to the SQL statement that exists at the back end. Then we hack the website. In the second scenario, we attempt to create a solution to protect this website. The research paper confirms that SQL injection already exists in ASP.NET 2015 (web form) and SQL Server 2014.

Keywords: Attack · Web application · SQL · Injection · Vulnerabilities

1 Introduction

In today's information age and with the perpetration of the internet of things (IoT), all trading transactions, education, economics, marketing and other services use electronic services and observe immense transfers of information via the Internet. These transfers are exposed to threats by cyber criminals because the Internet has hardware and software vulnerabilities. One common vulnerability is SQL injection attack, which occurs due to poor coding practices and lack of security awareness in some organizations. Many studies have considered this vulnerability and provided solutions to protect services, but when undertaking research on the topic through a case study, we found it still exists in ASP.NET 2015 (web form) and SQL Server 2014. We therefore provide a possible solution and some recommendations to help avoid SQL injection vulnerability.

1.1 Web Application

A web application or "web app" is a software program that runs on a web server. Unlike traditional desktop applications, which are launched by the operating system, web apps must be accessed through a web browser, as they are applications on the web [1]. A web

D. Nicholson (ed.), *Advances in Human Factors in Cybersecurity*,
Advances in Intelligent Systems and Computing 593, DOI 10.1007/978-3-319-60585-2_5

application therefore must be programmed in a language that is understandable by a web browser. Web browsers understand a finite amount of languages which means that web applications must be programmed in one of them to be understood. The following is a list of dominant languages that web applications can be programmed in: HTML, DHTML, XHTML, XML, Flash, JavaScript, Java, PHP, ASP, ActiveX and AJAX (a combination of JavaScript and XML).

1.2 Web Application Structure

Web applications now come in several forms, including two-tier and three-tier versions. This refers to the number of levels of the application. The three-tiered approach is most common at present and represents presentation, application and storage. The presentation layer is the web browser, while the application layer resides on the server and includes the files in the particular programming languages and/or some sort of server technology that helps translate information to the other layers (Ruby on Rails, PHP, etc.). The storage layer is generally a database that stores the information that is passed to the other layers. The application layer is basically the brains of the web application and allows the other two layers to interact in a more user-friendly way by supplying both with the required information. With the advent of Web 2.0, web applications have become abundant. Web 2.0 created the ability to share information, collaborate across multiple computers and operate across multiple operating systems, with a user interface that can be edited by the user. If you're doing more than just reading content on a site, if you're interacting with other users and/or editing the colours, layouts and options of your web interface, you're most likely using a web application [2]. Based on this, a web application can be defined as a software program based on HTML, JavaScript and CSS which is executed in a web browser and accessed locally or on the web.

2 Background

2.1 Web Application Security

Web application security is the process of securing confidential data stored online from unauthorized access and modification. This is accomplished by enforcing stringent policy measures. Security threats can compromise the data stored by an organization, as hackers with malicious intentions try to gain access to sensitive information. The aim of web application security is to identify the following: critical assets of the organization, genuine users who may access the data, the level of access provided to each user, various vulnerabilities that may exist in the application, data criticality and risk analysis on data exposure, appropriate remediation measures.

2.2 Software Development Security Problems

Some software development problems that result in software that is difficult or impossible to deploy in a secure fashion have been identified as "deadly sins in software security". These twenty problem areas are in software development (also called software engineering). The problem areas are described in the following sections.

2.3 SQL Injection

A SQL injection attack consists of insertion or "injection" of a SQL query via the input data from the client to the application [3]. A successful SQL injection exploit can read sensitive data from the database, modify database data (Insert/Update/Delete), execute administration operations on the database (such as shutdown the DBMS), recover the content of a given file present on the DBMS file system and, in some cases, issue commands to the operating system. SQL injection attacks are a type of injection attack, where by SQL commands are injected into the data-plane, in order to effect the execution of predefined SQL commands [2]. SQL injection is the vulnerability that results when an attacker has the ability to influence the Structured Query Language (SQL) queries that an application passes to the backend database [4].

2.4 History of SQL Injection

SQL injection as an attack method was first publicised as a side note to a comprehensive Microsoft web services exploitation article. The article first appeared in the fifty-fourth article of Phrack, a digital periodical that covers hacking topics. Titled "NT Web Technology Vulnerabilities" the article was written by Rainforest Puppy of the Wire Trip security group, and discussed Microsoft SQL and ASP injection exploits. Rainforest Puppy approached the injection technique as a side note to more serious vulnerabilities. The Phrack article served as a starting point for SQL injection research. An official advisory concerning the ability to batch commands was posted by the Allaire group (now part of Macromedia) several months later. The Allaire advisory, advisory number ASB99-04, generalised attack methods against ColdFusion, ASP, Sybase SQL, and Microsoft SQL Applications. SQL injection was still a very new concept, and Allaire's advisory assumed that only a very narrow group of systems was vulnerable, due to variables being encapsulated with quotes. Microsoft countered with the argument that the issues identified by Rainforest Puppy were not vulnerabilities, but they were features. The response prompted Rainforest Puppy to aggressively pursue his research of SQL injection techniques. The next exploitation and subsequent publication by Rainforest Puppy proved Microsoft's claim to be untrue. "How I hacked Packet Storm—a look at hacking www threads via SQL" by Rainforest Puppy, demonstrated how to directly circumvent some of the barriers to SQL injection that were assumed by the Allaire advisory. This article was the first to introduce a directed, successful attack using SQL injection and proved how easy it was to actively circumvent implied security features. The structure of the database was enumerated through random input and close analysis of the error reports generated by feeding the database the random data. Through his analysis of the database structure and his understanding of SQL query syntax, Rainforest Puppy defeated the limits of quotation by "breaking out", using additional sets of quotes to bypass SQL format constraints and injecting his own commands into the database [5].

2.5 How SQL Injection Works

Prospective customers, employees and business partners may all have the right to store or retrieve information from a database online. A website may allow any site visitor to submit and retrieve data. Legitimate access for visitors could include a site search, sign up forms, contact forms, logon forms and all of these provide a window into a database. The various points of access are quite possibly incorporated in "off-the-shelf" applications or may be custom applications set up just for a particular business's site. The forms and their supporting code likely come from many sources, have been acquired at different times and possibly installed by different people.

SQL injection is the use of publicly available fields to gain entry to a private database. This is done by entering SQL commands into form fields, instead of the expected data. Improperly coded forms will allow a hacker to use them as an entry point to a database, at which point the data in the database may become visible, as could access to other databases on the same server, or possibly even other servers in the network. Website features such as contact forms, logon pages, support requests, search functions, feedback fields, shopping carts and even the functions that deliver dynamic web page content, are all susceptible to SQL injection attacks, because the fields presented for visitor use must allow at least some SQL commands to pass through directly to the database [6]. In order to run malicious SQL queries against a database server, an attacker must first find an input within the web application that is included inside an SQL query [4].

In order for an SQL injection attack to take place, the vulnerable website needs to directly include user input within an SQL statement. An attacker can then insert a payload that will be included as part of the SQL query and run against the database server (Fig. 1).

2.6 Protecting Applications from SQL Injection Attacks

It is possible to protect an application from an SQL injection attack by using parameters (stored procedures) on the database SQL server and by using parameters in the backend, but we do not advise that. If we try to enter by using SQL injection, as shown below, we can't do the log in procedure in the database.

```
Select * from secUsers where secUsers.isActive = 1 and
secUsers.userName = @userName and secUsers.userPass = @pass-
word
```

2.7 Case Study

The case study contains two scenarios: in the first scenario we check whether SQL injection exists or not, while the second scenario creates solutions to protect the application from attacks by cyber criminals. In this case, we used ASP.NET2015 (web form) and SQL Server 2014.

Case Study Requirements. Windows 7 or higher, .Net framework 4.6.1, ASP.NET 2015 (web form), SQL Server 2014, good specification for hardware.

First Scenario. We hack the application by using SQL injection weakness. To do we follow these steps:

1. In ASP.NET from file menu chose new project and select ASP.NET web application.
2. Type project name
3. Select web form and click ok
4. Add new page (login.aspx)
5. Add two controls, textbox type (txtUserName, txtPassword)
6. Add submit button (btnLogin)
7. We type the code in btnLogin_click () event
8. Add another page its name is main page.
9. Open SQL server then create the table, e.g.
10. Create username = SQL injection test and userPass = 123
11. Run the application and enter correct username and password.
12. Press login button. Then we access the application to main page.
13. We try login to the application using the wrong password or user name.

When we press the login button, we cannot access the application. We try login to the application using the wrong password or user name, and to do that we inject the SQL dynamic statement. When we press the login button, we can access the application using SQL injection vulnerability from the back end to the front end of the website, by adding

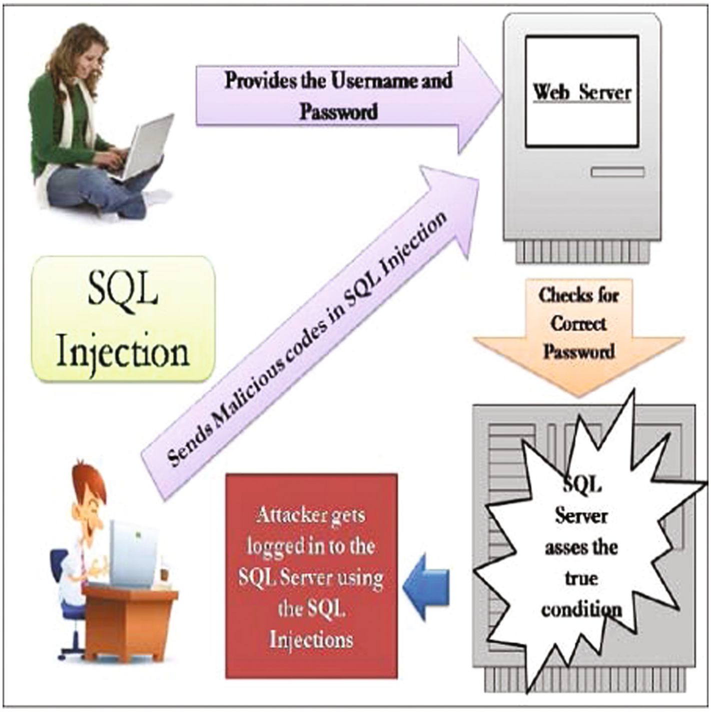

Provides the Username and Password
Web Server
SQL Injection
Sends Malicious codes in SQL Injection
Checks for Correct Password
SQL Server asses the true condition
Attacker gets logged in to the SQL Server using the SQL Injections

Fig. 1. How SQL injection works [4]

the injection (or ‘1’ = 1) to the condition in the SQL dynamic statement. We can now access the application even if the password is not correct.

Second Scenario. In the second scenario, we create a solution to protect the application from attacks by cyber criminals. We use parameterised rather than SQL dynamic statement. In this scenario, we used a stored procedure to retrieve data or validate usernames and passwords. Also, it is possible to use the parameterised technique at the back end, but we do not advise this solution, as a professional programmer should avoid hard code.

When running the application again, typing the wrong username/password, adding this injection (or ‘1’=’1) to the password and pressing the login button, the results are as below. Although we applied the same scenario as before, in this scenario we are unable to access the application. The reason being that we addressed SQL injection vulnerability by using a parameterised technique to retrieve data and validate the username and password.

3 Conclusion

Based on the results of the case study applied in the research paper, SQL injection attacks still exists in ASP.NET2015 when typing poor coding in the back end, without a parameterized technique, as demonstrated in the first scenario. It is possible to protect the web application from SQL injection attacks by using parameterized technique with store procedure, as shown in the second scenario. Through this research paper, we learned the following lessons: training should be given to programmers to avoid SQL injection attacks; knowledge gained should then be transferred between the team; and there should be a commitment within the organization’s policy for the team leader to undertake peer reviews of coding. Finally, it is important to avoid concatenations in SQL statements when using the stored procedure.

Information is the most important asset and achieving security of data stored on the web has to be the utmost priority in this competitive world market. Attacks exploiting security vulnerabilities occur in databases of applications through the injection of code. Vulnerabilities are becoming opportunities for attackers to gain access and manipulate system resources. The lack of a good mechanism for accessing the application at design level is exposed. In the end, a multi-layer web security strategy is the best solution, drawing on the strengths of all relevant technologies. Considering the riskiness of the SQL injection threat, an Adaptive Database Firewall should be a distinguished element in every solution.

3.1 Lessons Learned

(a) SQL injection is the most common website vulnerability on the Internet. It takes advantage of non-validated input vulnerabilities to pass SQL commands through a web application for execution by a backend database.

(b) Threats of SQL injection include authentication bypass, information disclosure and data integrity and availability compromise.

(c) SQL injection can be categorized as error-based SQL injection or blind SQL injection.
(d) Database administrators and web application developers need to follow a methodological approach to detect SQL injection vulnerabilities in web infrastructures.
(e) Pen testers and attackers need to follow a comprehensive SQL injection methodology and use automated tools such as SQL Map for successful injection attacks.
(f) Major SQL injection countermeasures involve input data validation, error message customization, database access privilege management and a database firewall, which monitors the networks between the application servers and databases [7].

References

1. Sharpened Productions Web Application (2016). http://techterms.com/definition/web_application
2. Investintech.com Inc.: What is a Web Application (2016). http://www.investintech.com/content/webapplication/
3. Whitman, M., Mattord, H.J.: Principles of Information Security, 4th edn. Cengage, Boston (2012)
4. Clark, J., Alvarez, R.M., Hartley, D., Hemler, J., Kornbrust, A., Meer, H., O'Leary-Steele, G., Revelli, A., Slaviero, M., Stuttard, D.: SQL Injection Attacks and Defense United States of America (2009)
5. Rolston, B.: Attack Methodology Analysis: SQL Injection Attacks. US-CERT Control Systems Security Center, Idaho Falls (2005)
6. Shegokar, A.M., Manjaramkar, A.K.: A survey on SQL injection attack, detection and prevention techniques. Int. J. Comput. Sci. Inf. Technol. **5**(2), 2553–2555 (2014). ISSN 0975-9646,2014/05
7. Alrajhi, M., Alothman, M., Aldosari, A., Othman, A.: Understanding and Discovering SQL Injection Vulnerabilities and Countermeasures. NAUSS (2016)

Grid Framework to Address Password Memorability Issues and Offline Password Attacks

Paul Biocco and Mohd Anwar(✉)

North Carolina A&T State University, Greensboro, NC 27411, USA
manwar@ncat.edu

Abstract. Passwords today are the most widely used form of authentication, yet have significant issues in regards to security due to human memorability limitations. Inability to remember strong passwords causes users generally to only satisfy the bare minimum requirements during an enrollment process. Users having weak passwords are vulnerable to offline password attacks, where an adversary iteratively guesses the victim's password and tests for correctness. In this paper, we introduce a new password scheme, Grid framework, that takes advantage of current encryption technologies and reduces the user's effort to create a strong password. The Grid Framework scheme translates an easy-to-remember sequence on a grid into a complex password consisting of randomly selected uppercase, lowercase, numeric, and special symbols with a minimum length of eighteen characters that the user is not required to memorize. The Grid Framework results in a system that increases memorability for secure authentication.

Keywords: Authentication · Memorability · Passwords

1 Introduction

User authentication is required for secure access to private data that pertains to a user, manage their own information, or a combination of both. Typically, authentication is done through either something the user has, is, or knows. One of the most common forms of authentication that is based on a user's knowledge is the password. There has been a significant amount of research from delving into securing passwords to simply creating systems that would not require a password. Despite the effort in finding stronger alternative forms of authentication, passwords are still the most widely used on the Web.

Unfortunately, there is a major memory versus security issue in regards to the password. While the password is great at protecting information when it is hard to guess, people sacrifice security to favor memorability, resulting in passwords that are easy to compromise via offline attacks [1] such as dictionary attacks [2], brute force attacks [3], and rainbow attacks [4]. Users tend to favor the use of dictionary words in their password, which make them more susceptible to dictionary attacks, especially in comparison to the passwords that are randomly generated. However, users struggle to remember a long (e.g., an eight-character) password of random alphanumeric characters, forcing them to write their password down where it may be lost or vulnerable to theft. This paper attempts to address the question, "how can we create strong memorable passwords?".

D. Nicholson (ed.), *Advances in Human Factors in Cybersecurity*,
Advances in Intelligent Systems and Computing 593, DOI 10.1007/978-3-319-60585-2_6

In response to this question, we propose the Grid framework. The Grid framework, at its core, is a method to condense passwords from random characters to clicks on a colored Grid, effectively replacing standard user interface (UI) with a username and an interactive grid instead of a password. Its main function is to create an invisible, random password that is used to authenticate the user into the system. Each standard password is at least eighteen characters long, including both alphanumeric characters and symbols, by creating a sequence of characters on a grid. Because of both the random nature of the password generation and the sheer length of the generated password, we believe that combining the Grid framework combined with current password encryption methods will yield in an authentication system more secure than the user-generated passwords. We also believe that the simplified nature of the Grid framework is more memorable than that of most randomly generated passwords with eight alphanumeric characters.

The paper is organized as such: Sect. 2 covers related works. Section 3 presents the overall Grid framework schematic. Section 4 presents the memorability aspect of the Grid framework. Section 5 describes the Grid framework's advantage in dealing with offline attacks. Section 6 discusses future works and we conclude in Sect. 7.

2 Related Work

We surveyed authentication techniques, starting with the password. According to an experiment done in the paper written by Yan and Blackwell [5], self-created passwords are easier to remember than randomly generated passwords via blindly choosing letters on a grid. However, these same user-created passwords were more often cracked with offline password attacks such as dictionary attacks, permutation attacks, and user information attacks. The randomly generated passwords resulted in significantly stronger passwords resistant to those same attacks, but were twice as hard to remember in the user's opinions, and took over five times longer to fully remember their password than users who chose their own password.

The most common technique for making users create more secure passwords involve restricting passwords with constraints. For example, most websites have a minimum length requirement for their passwords. Some websites require both uppercase and numeric characters when creating a password. However, Adams, Sasse, and Lunt observed that adding more stringent requirements resulted in overall weaker passwords [6]. Information revolving around personal information such as names and birthdays was often used in passwords, making them susceptible to personal information attacks. Users were also more likely to repeat passwords across multiple platforms if the password they used was complex.

Mnemonics are one of few ways to create secure passwords. In a blog post by Bruce Schneier, an eminent cybersecurity scholar, discusses a password creation technique using a full sentence [7]. Using the sentence "When I was seven, my sister threw my stuffed rabbit in the toilet", he creates the password "WIw7, mstmsritt". By only using the case, punctuation, and first letters of the sentence, a password that satisfies all typical password constraints.

The research done on graphical passwords supports intuitiveness of our approach. Sonia Chiasson et al. created a visual authentication system with a highly positive enrollment rate and successful authentication rate [8]. The users, on average, highly rated the easiness to create a graphical password. Interestingly, when asked if they prefer text-based passwords to graphical passwords, their average response was 4.9 on a 10-point scale, showing that graphical passwords are at least as much desired as textual passwords.

3 Grid Framework Approach

Overall there are four parts to the Grid framework: username, seed number, representative characters, and the user's password. The username is what the user's alias will be for the web application, often associated with an email. The seed number is a string that will randomly produce the representative characters, making sure that each Grid's character content matches no other user. Representative characters are the set of characters invisible to the user that are associated with the user's clicks on the Grid. Each tile (one square on the grid) represents the three characters, concatenated and stored to the password field when clicking on the tile. The user's password is generated through the combination of the representative characters, concatenated in the order each tile was clicked.

The Grid framework's design is broken down into two perspectives: the user view and the system view – the system is defined as the server side of the Grid's framework. The rest of this section is divided into three stages: enrollment, system interaction, and storage. These stages will dictate the format for the rest of this paper.

3.1 Enrollment

As described in Fig. 1, the enrollment process starts with the new user creating a username that must not be taken by an existing user. This username is then associated with a unique seed value and temporarily stored; both values are deleted if the enrollment process is canceled at any time. The seed, which is unique per user, is used to generate the set of representative characters associated with the user's "Grid". These character values are used to generate a set of muted colors, to be rendered on the tiles of a six-by-four grid. The user then will create their sequence, minimum length of six, on the grid and repeat it to confirm the sequence. The sequence is converted to its character equivalent string using the representative characters to create a password. Finally, the password along with the username is hashed and stored in the system's database.

The enrollment through Grid is similar to that of a traditional texted-based password enrollment. In both forms of authentication, the user creates a username, a sequence (of clicks in Grid, similar to characters in text-based passwords), and then is required to repeat the sequence. This similarity makes Grid-based authentication easy to use, and less likely to get rejected due to unfamiliarity [9].

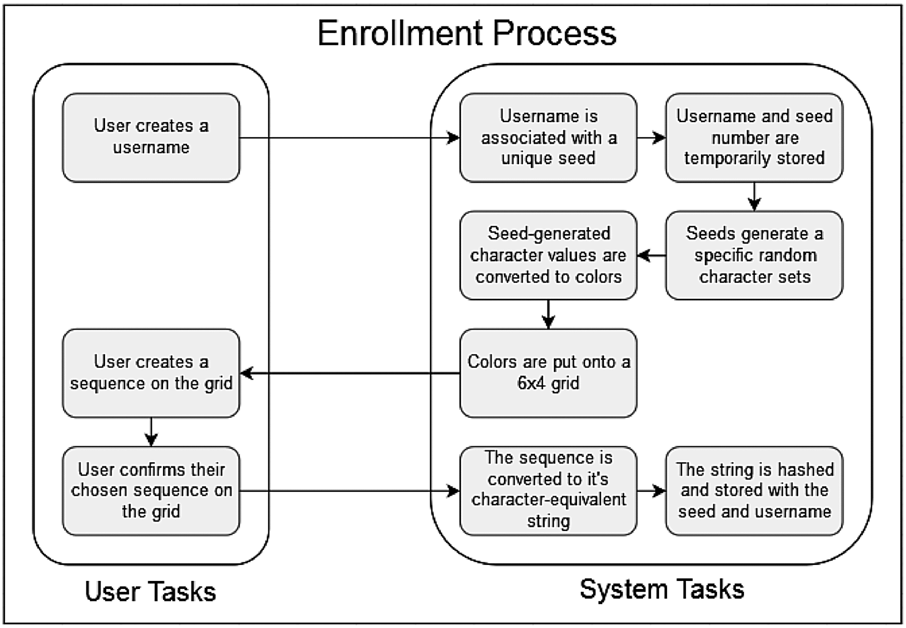

Enrollment Process
User creates a username
Username is associated with a unique seed
Username and seed number are temporarily stored
Seeds generate a specific random character sets
Seed-generated character values are converted to colors
Colors are put onto a 6x4 grid
User creates a sequence on the grid
User confirms their chosen sequence on the grid
The sequence is converted to it's character-equivalent string
The string is hashed and stored with the seed and username
User Tasks
System Tasks

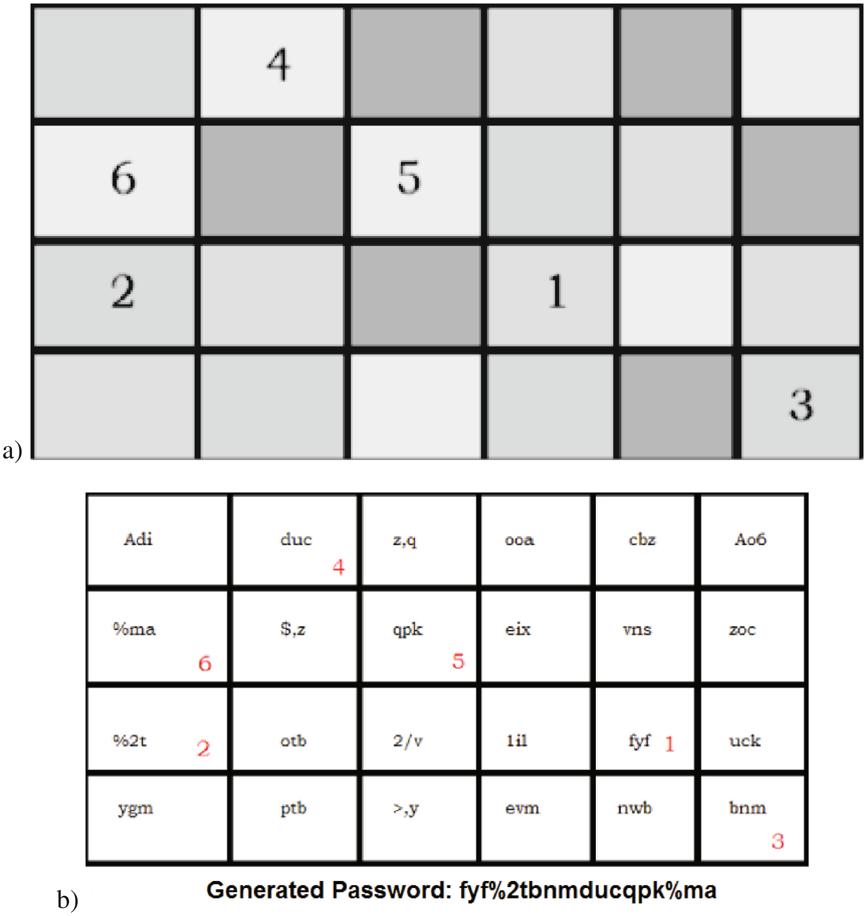
Fig. 1. The enrollment process of the Grid framework is described in terms of user tasks and system tasks.

Grid authentication has additional requirements beyond storage, including seed generation, image mapping, and interface loading. We have not yet considered how much processing power Grid would require if employed with a significant user base.

3.2 General Interaction

From the user's perspective (shown in Fig. 2a), system interactions are minimal. Users only see the Grid with multi-colored tiles (Sequence numbers not shown to the user). By clicking on a tile, users can input their chosen sequence to generate their password and authenticate. Putting in an incorrect sequence results in a temporary account lockout, rejecting login attempts for a designated timeframe. Using colors may cause predictability issues to arise, similar to other image-based authentication techniques [10]. Significantly contrasting colors may cause consistent attention to be drawn to specific tiles. To combat this, the framework will group color schemes together, only allowing similar shades and hues to be picked for the same grid, defusing potential points of interest [11].

From the system's perspective, depicted in Fig. 2b, each tile on the grid has three characters. Clicking on a tile will generate three characters, appended together to create a password used for authentication. The representative character set, when stored, is a 72-character string – each tile contains three characters which are evenly divided among all 24 tile in the grid. The password for authentication created through the user's clicks is significantly longer and more secure than an average user-generated password [5]. In the example



a)

	4				
6		5			
2			1		
					3

b)

Adi	duc 4	z,q	ooa	cbz	Ao6
%ma 6	$,z	qpk 5	eix	vns	zoc
%2t 2	otb	2/v	1il	fyf 1	uck
ygm	ptb	>,y	evm	nwb	bnm 3

Generated Password: fyf%2tbnmducqpk%ma

Fig. 2. (a) A standard Grid from the user's perspective with an example sequence. The numbers represent the order of mouse click interactions by the user (i.e., 1 = first click, 2 = second click, etc.). (b) A Grid from the system's view, showing the representative characters and the generated password based on the displayed sequence.

provided, with six clicks on the grid results in an eighteen-character password. For the random generation of characters inside each grid, all typical characters on a keyboard, 10 digits, 26 lowercase letters, 26 uppercase letters, and 33 symbols, are used.

The Grid framework's design allows for minor alterations on the system side. Special characters implemented into the Grid's framework beyond the basic 33 symbols can be used to further strengthen the Grid framework. Using alt-codes characters such as ◙, ¼, and «, adding more length to each grid's representative character sets, and creating a rejection system of patterns deemed too predictable (such as clicking the same position six times) are all programmatically feasible design alterations possible for the Grid framework. Other possible customizations include letting the user pick their own colors for their authentication grid after enrolling, or creating a color blind option.

3.3 Storage

Since the main information that the Grid framework uses is text, everything can be encrypted or hashed using an algorithm of choice, determined by the system's engineer. Like passwords, secure storage requires hashing or encryption to avoid storing raw information in a database.

The password, seed, and username will be secured using one-way hashing functions. The username and click sequence resulting in the password is always to be input by the user, shown in Fig. 3, and the seed is stored only to make sure no duplicate seeds are created. The system generated character sequence (corresponding to mouse clicks on the Grid), however, must be encrypted as it goes in both from server side to client side and vice versa.

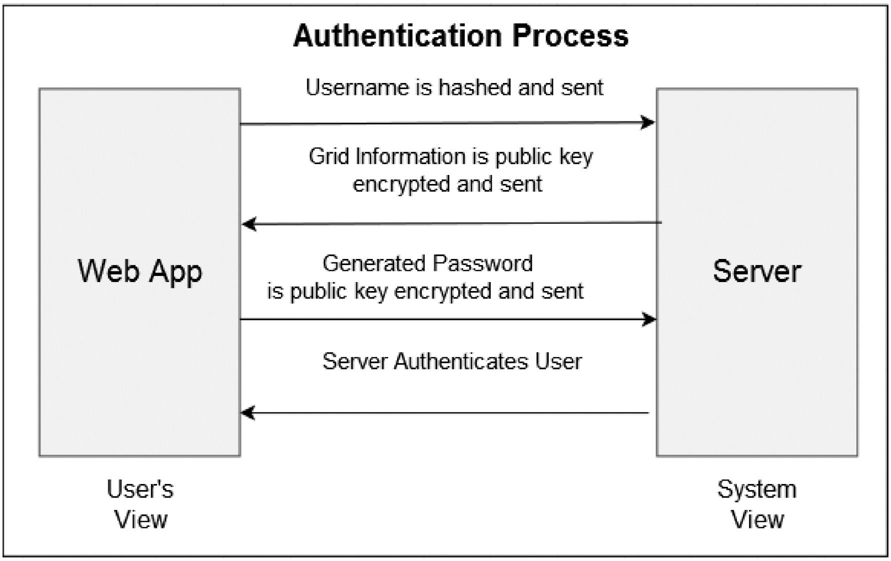

Authentication Process
Username is hashed and sent
Grid Information is public key encrypted and sent
Web App
Generated Password is public key encrypted and sent
Server
Server Authenticates User
User's View
System View

Fig. 3. Overall authentication process, showing the interactions between the server and web application.

All information exchanges should use public key encryption such as RSA for secure transmission of the password and username from the interface to the authentication server. Since web applications are sometimes accessed using public WiFi, users will not always have a secure channel, rendering them as an insecure option [12, 13].

4 Grid Memorability vs. Password Memorability

Passwords generated from the Grid framework are stronger than user-generated passwords [5]. Despite this, Grid's memorability is not a problem. By representing multiple characters per tile, we can reduce a password of eighteen characters long down to six positions on a grid. This will make it easier to remember a Grid's sequence instead of a randomly generated password. Weak passwords are still more memorable than a Grid's sequence as they often use words, names, and other already-familiar bits of information that are easier to recall [14].

Passwords that are hard to guess are hard to remember simply from their random nature [5]. Mnemonics are used to create more secure passwords by using specific feature recognitions [7]. For example, a secure password can be created by taking the first character in each word of a sentence. The sentence "Today I went to the super and bought some fish", results in the password "TIwttsabsf", which looks more random than an average user-generated.

Mnemonics are not strictly verbal, and can also be auditory or visual and can be created to remember a Grid's sequence [15]. As shown in Fig. 4, a Grid's sequence can be set to a pattern. In the figure, each tile selected after the first selection is a "knight's move" in chess away, always the exact same distance away from the previous location. Other visual mnemonics would also include drawing a picture on the Grid, such as a check mark, a smile, or a lightning bolt.

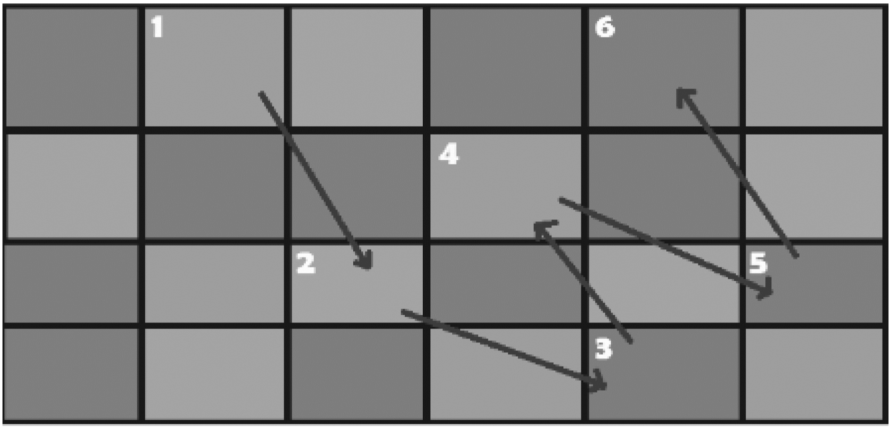


Fig. 4. Showing a Grid with a sequence following a set pattern, only using "knights" moves from chess.

Feature analysis is another important aspect in grid's memorability. Again referring to Fig. 4, we can note this sequence has many notable features that are often recognized. The first position in the sequence is in the top row and second column and is on a tile of specific color. Naturally, the brain takes hold of such details in visual pattern recognition [16]. Users utilize spatial feature recognition as well. Taking advantage of the brain's natural tactile encoding, a user will feel general characteristics of their own Grid sequence [17]. In Fig. 4, when clicking on tiles 2 through 6, a zigzagging motion is made with the mouse, a characteristic that is subconsciously notable to users [17]. Features similar to these zigzags are not limited to only pattern based sequences. Even randomly created sequences will have more subtle tactile features, an advantage over text-based password authentication for memorability.

5 Grid and Offline Password Attacks

Offline password attacks such as brute force attacks, rainbow tables, dictionary attacks, and hybrid attacks are ineffective against strong and long passwords [18]. Naturally, Grid framework's authentication generates a random minimum eighteen-character long password to be hashed. Breaking passwords of this size would be too resource intensive [19] to be viable with today's technology.

Brute force attacks trying to crack a password generated from a sequence of size six will have to go through over 397 decillion combination (95^{18}), selecting one character out of 95 characters located on the keyboard, eighteen times in a row. Both dictionary attacks and hybrid attacks will simply not work due to the random nature of these generated passwords. Hybrid and dictionary attacks are only useful when targeting passwords with dictionary words and names, atypical of passwords randomly generated [20]. Rainbow tables, too, would struggle to break passwords of such a large size. Computationally intensive, cracking with a rainbow table would need 397 decillion hashes to be computed and recorded for strictly six mouse inputs.

If the password was ever cracked, there would be a secondary step that would be required before the attacker could authenticate as the user. The attacker would need to break down the password into its parts, then associate each subpart with the correct tile on the grid. Since these characters cannot be seen from the user end, the attacker would have to decrypt the 72-character long representative character content string. Without this, hacking the stored password would not be sufficient to decipher the input sequence to be entered through the grid.

Additionally, it is notable that the Grid framework's authentication process is immune to keylogging malware [21]. Only the username would ever be revealed via keylogging. Inputting the sequence would require a mouse or touchpad. The Grid framework would have the same security as a virtual keyboard, effectively shutting out keylogging malware.

6 Future Works

The research scope for Grid framework is quite broad. In order to prove that Grid is a viable replacement of the current username and password in memorability, a Grid framework, web-based or offline, must be created and tested, then compared with the baseline password schemes. Secondly, the Grid framework needs to be tested for enrollment and authentication times to compare with standard enrollment and authentication speeds of a username and password. If the Grid framework proves to be too unwieldy to authenticate into, adjustments in the framework's design may be required to match the speeds of enrollment. Finally, a full proof-of-concept must be completed. A basic demonstration of the Grid framework needs to be launched and penetration tested for any additional vulnerabilities that might have been unforeseen. Attempting to find vulnerability on the Grid framework from the user interface will also be an additional mandatory step.

7 Conclusion

We attempt to create an authentication system that will address the major concerns of password memorability and offline attacks. Caused by the inability to remember randomized passwords, users generally tend to pick significantly weaker passwords, often easy to break via offline password attacks such as dictionary attacks. This paper discusses the Grid framework, created to make strong but easy to remember password. Clicking on tiles would append together a password too strong to be cracked via offline password attacks while simultaneously reducing the amount of effort required to memorize the Grid's sequence. The sequence would be more difficult to remember than poor and average passwords, but easier than strong passwords. Additionally, the password that the sequence creates will make standard offline attacks computationally infeasible.

The Grid framework utilizes only current technologies that exist. Using RSA public key encryption, the username, and the contents of the grid can be given to the user without the possibility of interception. Overall, the system has clear advantages over the standard password to withstand offline attacks from the large password sizes and advantages over the standard password in memorability from the replacement of random actions to positions on a grid.

Acknowledgments. This work was supported in part by the U.S. National Science Foundation under Grant SES-1318501.

References

1. Dhamija, R., Perrig, A.: Deja Vu-a user study: using images for authentication. In: USENIX Security Symposium, 14 August 2000, vol. 9, p. 4 (2000)
2. Jablon, D.P.: Extended password key exchange protocols immune to dictionary attacks (1997)
3. Pliam, J.O.: On the incomparability of entropy and marginal guesswork in brute-force attacks. In: International Conference on Cryptology in India, pp. 67–79. Springer, Heidelberg (2000)
4. Avoine, G., Bourgeois, A., Carpent, X.: Fingerprint tables: a generalization of rainbow table (2013)
5. Yan, J., Blackwell, A.: Password memorability and security: empirical results (2004)
6. Adams, A., Sasse, S.A., Lunt, P.: Making passwords: secure and usable (1997)
7. Schneier, B.: Choosing a secure password (2014). https://www.schneier.com/blog/archives/2014/03/choosing_secure_1.html
8. Chiasson, S., Forget, A., Biddle, R., von Orschot, P.C.: Influencing users towards better passwords: persuasive cued click-points (2008)
9. Marques, J.M., Yzerbyt, V.Y., Leyens, J.P.: The black sheep effect: extremity of judgments towards ingroup members as a function of group identification. Eur. J. Soc. Psychol. **18**(1), 1–6 (1988)
10. Renaud, K., De Angeli, A.: My password is here! An investigation into visuo-spatial authentication mechanisms. Interact. Comput. **16**(6), 1017–1041 (2004)
11. Baik, M., Suk, H.J., Lee, J., Choi, K.: Investigation of eye-catching colors using eye tracking. In: IS&T/SPIE Electronic Imaging, 14 March 2013, p. 86510W. International Society for Optics and Photonics (2013)

12. Cheng, N., Wang, X.O., Cheng, W., Mohapatra, P., Seneviratne, A.: Characterizing privacy leakage of public WiFi networks for users on travel. In: 2013 Proceedings IEEE INFOCOM, 14 April 2013, pp. 2769–2777. IEEE (2013)
13. Boyko, V., MacKenzie, P., Patel, S.: Provably secure password-authenticated key exchange using Diffie-Hellman. In: International Conference on the Theory and Applications of Cryptographic Techniques, 14 May 2000, pp. 156–171. Springer, Heidelberg (2000)
14. Orgill, G.L., Romney, G.W., Bailey, M.G., Orgill, P.M.: The urgency for effective user privacy-education to counter social engineering attacks on secure computer systems. In: Proceedings of the 5th Conference on Information Technology Education, 28 October 2004, pp. 177–181. ACM (2004)
15. Schmalzl, L., Nickels, L.: Treatment of irregular word spelling in acquired dysgraphia: selective benefit from visual mnemonics. Neuropsychological Rehabil. **16**(1), 1–37 (2006)
16. Dill, M., Wolf, R., Heisenberg, M.: Visual pattern recognition in Drosophila involves retinotopic matching. Nature **365**(6448), 751–753 (1993)
17. Courtney, S.M., Ungerleider, L.G., Keil, K., Haxby, J.V.: Object and spatial visual working memory activate separate neural systems in human cortex. Cereb. Cortex **6**(1), 39–49 (1996)
18. Shay, R., Komanduri, S., Durity, A.L., Huh, P.S., Mazurek, M.L., Segreti, S.M., Ur, B., Bauer, L., Christin, N., Cranor, L.F.: Can long passwords be secure and usable? In: Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems, 26 April 2014, pp. 2927–2936. ACM (2014)
19. Cheon, J.H.: Security analysis of the strong Diffie-Hellman problem. In: Annual International Conference on the Theory and Applications of Cryptographic Techniques, 28 May 2006, pp. 1–11. Springer, Heidelberg (2006)
20. Tasevski, P.: Password attacks and generation strategies. Tartu University, Faculty of Mathematics and Computer Sciences, 21 May 2011
21. Vishnani, K., Pais, A.R., Mohandas, R.: An in-depth analysis of the epitome of online stealth: keyloggers; and their countermeasures. In: International Conference on Advances in Computing and Communications, 22 July 2011, vol. 22, pp. 10–19. Springer, Heidelberg (2011)

Cryptanalysis and Improvement of an Advanced Anonymous and Biometrics-Based Multi-server Authentication Scheme Using Smart Cards

Chunyi Quan[1], Hakjun Lee[1], Dongwoo Kang[1], Jiye Kim[2], Seokhyang Cho[3], and Dongho Won[1](✉)

[1] Information Security Group, Sungkyunkwan University, Suwon, South Korea
{sikwon,hjlee,dwkang,dhwon}@security.re.kr

[2] Department of Mobile Internet, Daelim University College, Anyang, South Korea
jykim.isg@gmail.com

[3] Department of Information and Communication, Pyeongtaek University, Pyeongtaek, South Korea
cshlch@ptu.ac.kr

Abstract. In conventional single-server environment, a user must register to every server if he/she wants to access numerous network services. It is exceedingly hard for users to generate different robust passwords and remember them with corresponding identities. To solve this problem, many multi-server authentication schemes have been proposed in recent years. In 2017, Chang et al. improved Chuang and Chen's scheme, arguing that their scheme provides higher security and practicability. However, we demonstrate that Chang et al.'s scheme is still vulnerable to outsider attack and session key derived attack. In addition, we also find that both malicious user and server can carry out user impersonation attack in their scheme. In this paper, we propose a new biometrics-based authentication scheme that is suitable for use in multi-server environment. Finally, we show that the proposed scheme improves on the level of security in comparison with related schemes.

Keywords: Authentication · Multi-server · Biometrics · Smart card

1 Introduction

In 1981, Lamport [1] proposed the first remote password authentication scheme under insecure network. However, his scheme is proved to be insecure against guessing attacks. Therefore, smart card based scheme were considered as a solution and came into sight. By utilizing smart cards, instead of keeping a verification table, participants are allowed to store secret information into a smart card which improves security to a new level. After that, other novel schemes [2, 3] which adopt biometrics were introduced for further enhancement. However, all aforementioned schemes [1–3] are designed for single-server environment which makes users extremely inconvenient to

D. Nicholson (ed.), *Advances in Human Factors in Cybersecurity*,
Advances in Intelligent Systems and Computing 593, DOI 10.1007/978-3-319-60585-2_7

access resource from servers because they must register to each server separately. To solve this problem, a new authentication structure for multi-server environment was introduced and several related schemes have been proposed [4–8].

In 2010, Yang and Yang [4] proposed a biometric password-based multi-server authentication scheme using smart card which enables users to register for only once and then be qualified to access all servers. Unfortunately, their scheme costs vast computational resource due to the heavy use of modular exponentiation operations. In the same year, Yoon and Yoo [5] proposed an improved scheme based on elliptic curve cryptosystem. He [6] demonstrated that their scheme cannot resist privileged-insider attack, masquerade attack and stolen smart card attack. In 2014, Chuang and Chen [7] presented a scheme under the assumption that all servers are trusted and achieves both high efficiency and security. However, Chang et al. [8] proved that Chuang and Chen's scheme is insecure against stolen smart card attack, forgery attack and has privacy preservation issue. Furthermore, Chang et al. indicated that in traditional biometric-based scheme the authentication may fails due to the slight difference between imprinted biometrics and original ones. Therefore, they adopted functions defined in Dodis et al.'s work [9] and proposed an enhanced scheme, claiming that their scheme satisfies all desirable security requirements. In this paper, after careful analysis, we find that Chang et al.'s scheme is vulnerable to outsider attack and session key derived attack. In addition, both malicious user and server can carry out user impersonation attack in their scheme. To resolve these vulnerabilities, we propose a new biometric-based authentication scheme that is suitable for multi-server environment. In particular, the comparison on security level between our scheme and other related schemes [2–5, 8] implies that our scheme can defend against a number of attacks including the ones of Chang et al.'s scheme.

The rest of the paper is organized as follows: In Sect. 2, we introduce basic concepts of secure sketch presented by Dodis et al. In Sects. 3 and 4, we review and cryptanalyze Chang et al.'s scheme. Section 5 describes the proposed scheme. Sections 6 and 7 gives a detailed security and performance analysis where our scheme is compared with related schemes, respectively. Finally, in Sect. 7, we conclude this paper.

2 Secure Sketch

The major problem of biometrics-based authentication scheme is that the imprinted biometric can slightly differentiate with the original template since some noise are unavoidably introduced into the reproducing process. To rectify this weakness, Chang et al. [8] adopted Dodis et al.'s function [9] which is defined that a $(\mathcal{M},\ m, m', t)$ secure sketch is a randomized map $SS : \mathcal{M} \rightarrow \{0,1\}^*$ in which m is min-entropy, m' is the lower bound of average m and t refers to the number of tolerated errors.

For distance function dis and vectors $w, w' \in \mathcal{M}$, a deterministic recovery function $Rec(w', SS(w)) = w$ exists which allows to recover w from its sketch $SS(w)$ and w' that is close to w as long as $dis(w, w') < t$ is satisfied. According to this definition, for any given binary $[n, k, 2t+1]$ error correcting code E, we set randomized map SS as a $(\mathcal{M},\ m, m+k-n, t)$-secure sketch and $SS(W;X) = W \oplus E(X)$, where n is string length, k indicates the dimension of codeword, W is uniform and X is a random parameter. There is a decoding function D can correct t errors maximum that $dis(W, W') < t$. D works as $D(W', S(W;X)) = X$. Lastly, we can set the recovery function $Rec(W', S(W;X)) = SS(W;X) \oplus E(D(W' \oplus SS(W;X))) = W$.

3 Review of Chang et al.'s Scheme

In this section, we briefly review the advanced anonymous and biometrics-based multi-server authentication scheme of Chang et al. [8]. Their scheme consists of following phases: server registration, user registration, login, authentication and password change. The notations used in this paper are described in Table 1.

Table 1. Notations

Notations	Description
U_i, S_j, SC_i	User, server and user's smart card
RC	Registration center
ID_i, SID_j	Identity of U_i and S_j
PW_i, BIO_i	Password and biometrics of U_i
x, y	The secret key and number of RC
$E(\cdot), D(\cdot)$	The encoding and decoding function based on Dodis et al.'s paper [9]
$h(\cdot)$	A secure hash function

3.1 Server Registration Phase

S_j sends a registration request to RC via a secure channel. RC accepts S_j and computes $k_1 = h\left(SID_j \parallel h(y)\right)$ and $k_2 = h(x \parallel y)$. Finally, RC sends k_1 and k_2 back to S_j.

3.2 User Registration Phase

1. U_i freely chooses his/her identity ID_i, password PW_i, and imprints his/her personal biometric information BIO_i into a special device. U_i randomly generates a number r_i that is only retained by himself/herself and computes $\alpha_i = BIO_i \oplus E(r_i)$, $V_i = h(PW_i) \oplus \alpha_i$ and $R_i = h(PW_i \oplus r_i)$. Afterwards, U_i transmits $\{ID_i, V_i, R_i\}$ to RC via a secure channel.
2. After receiving the registration request message from U_i, RC calculates $A_i = h(ID_i \parallel x)$, $B_i = h(ID_i \parallel R_i)$, $C_i = h^2(R_i) \oplus h(y)$, $D_i = h(R_i) \oplus A_i \oplus h(x \parallel y)$ and $E_i = h(A_i \parallel h(x \parallel y)) \oplus h(R_i)$.
3. Lastly, RC stores $\{V_i, B_i, C_i, D_i, E_i, h(.)\}$ into SC_i and sends it to U_i.

3.3 Login Phase

1. U_i inserts his/her SC_i into a card reader, inputs his/her ID_i^* and PW_i^*, imprints personal biometric information BIO_i^* via a special device.

2. SC_i employs inputted information to compute $R_i^* = h\left(PW_i^* \oplus D\left(V_i \oplus h\left(PW_i^*\right) \oplus BIO_i^*\right)\right)$ and verifies whether $h\left(ID_i^* \parallel R_i^*\right)$ equals to B_i. SC_i only proceeds to the next step when they are equal.
3. SC_i generates a random nonce n_i and computes $h(y) = C_i \oplus h^2\left(R_i^*\right)$, $M_1 = h\left(SID_j \parallel h(y)\right) \oplus n_i$, $CID_i = D_i \oplus h\left(R_i^*\right) \oplus h(n_i)$, $G_i = E_i \oplus h\left(R_i^*\right)$ and $CHECK_1 = h\left(h\left(SID_j \parallel h(y)\right) \parallel n_i \parallel G_i\right)$.
4. SC_i sends login request message $\{M_1, CID_i, CHECK_1\}$ to S_j.

3.4 Authentication Phase

1. Upon receiving the login request message from U_i, S_j first employs its secret k_1 to compute random nonce $n_i = M_1 \oplus k_1$ to check its freshness. If n_i is fresh, S_j subsequently computes $A_i = CID_i \oplus h(n_i) \oplus k_2$ and verifies whether $h(k_1 \parallel n_i \parallel h(A_i \parallel k_2))$ equals to $CHECK_1$. If it holds, S_j considers U_i as valid user.
2. S_j generates a random number n_j and computes $M_2 = n_j \oplus n_i \oplus k_1$, $SK = h(h(A_i \parallel k_2) \parallel n_i \parallel n_j)$ and $CHECK_2 = h(SK)$, followed by sending a response message $\{M_2, CHECK_2\}$ to U_i via a public channel.
3. SC_i retrieves random nonce n_j by computing $n_j = M_2 \oplus h\left(SID_j \parallel h(y)\right) \oplus n_i$ and checks its freshness. If n_j is fresh, SC_i then computes $SK = h\left(G_i \parallel n_i \parallel n_j\right)$ and checks if $h(SK)$ equals to $CHECK_2$. If the verification succeeds, SC_i computes $CHECK_3 = h\left(SK \parallel n_j\right)$ and sends it to S_j via a public channel.
4. After receiving $CHECK_3$ from U_i, S_j verifies whether $h\left(SK \parallel n_j\right)$ equals to $CHECK_3$ to reconfirm the authenticity of U_i. Then, U_i and S_j can start to communicate with the other party using the shared session key.

3.5 Password Change Phase

1. U_i inserts his/her SC_i into a card reader and inputs ID_i, PW_i and BIO_i.
2. SC_i computes $\alpha_i = V_i \oplus h(PW_i)$, $r_i = D(BIO_i \oplus \alpha_i)$ and $R_i = h(PW_i \oplus r_i)$, and verifies the condition $h(id_i \parallel R_i) = ?B_i$. If the it holds, SC_i asks U_i to submit a new password, otherwise password change request can be dropped.
3. U_i submits a new password PW_i^{new} and then SC_i employs it to compute $V_i^{new} = V_i \oplus h(PW_i) \oplus h\left(PW_i^{new}\right)$, $R_i^{new} = h\left(PW_i^{new} \oplus r_i\right)$, $B_i^{new} = h\left(ID_i \parallel R_i^{new}\right)$, $C_i^{new} = C_i \oplus h^2(R_i) \oplus h^2\left(R_i^{new}\right)$, $D_i^{new} = D_i \oplus h(R_i) \oplus h\left(R_i^{new}\right)$ and $E_i^{new} = E_i \oplus h(R_i) \oplus h\left(R_i^{new}\right)$. Finally, SC_i replaces V_i, B_i, C_i, D_i and E_i with V_i^{new}, B_i^{new}, C_i^{new}, D_i^{new} and E_i^{new}.

4 Cryptanalysis of Chang et al.'s Scheme

In this section, we cryptanalyze Chang et al.'s scheme [8] and demonstrate that their scheme possesses some security vulnerabilities. According to the threat model described in [10–12], an adversary can eavesdrop, modify and intercept any message in the public channel, and that an adversary can extract all information stored in the smart card by carrying out power analysis [11]. Under these two assumptions, the scheme has the following security problems and the descriptions are given below.

4.1 Outsider Attack

A malicious server $\mathcal{A}$ is aware of secrets k_1 and k_2 that are authenticated from RC and can retrieve A_i and n_i after receiving login request message $\{M_1, CID_i, CHECK_1\}$ from U_i during the authentication phase. If $\mathcal{A}$ steals SC_i which belong to the user he/she is communicating with and extracts parameters $\{C_i, D_i\}$ from it, he/she can compute $h(R_i) = D_i \oplus A_i \oplus k_2$ and then obtains the encrypted secret number of RC by calculating $h(y) = C_i \oplus h^2(R_i)$, which is the same for each user. Therefore, $\mathcal{A}$ may be able to launch other attacks with the knowledge of RC's secret $h(y)$.

4.2 Session Key Derived Attack

Suppose a malicious server $\mathcal{A}$ obtains RC's secret $h(y)$ in the previous attack. He/she can easily compute the session key that is transmitted between any user and server. The attack proceeds as follows:

1. $\mathcal{A}$ eavesdrops login request message $\{M_1, CID_i, CHECK_1\}$ between U_i and S_j, and computes $n_i = h\big(SID_j \parallel h(y)\big) \oplus M_1$ and $A_i = CID_i \oplus h(n_i) \oplus k_2$.
2. Then, $\mathcal{A}$ eavesdrops S_j's response message $\{M_2, CHECK_2\}$, retrieves the nonce n_j by computing $n_j = M_2 \oplus h\big(SID_j \parallel h(y)\big) \oplus n_i$. Afterwards, $\mathcal{A}$ can obtain the session key by computing $SK = h\big(h(A_i \parallel k_2) \parallel n_i \parallel n_j\big)$.

4.3 User Impersonation Attack

Although Chang et al. [8] claim that their scheme can endure user impersonation attack, however after careful analysis we find that an adversary $\mathcal{A}$ can still impersonate as a legitimate user to cheat with S_j. Especially in Chang et al.'s scheme, $\mathcal{A}$ can either be a malicious server or user. Suppose $\mathcal{A}$ is a malicious server who obtains RC's secret $h(y)$ by means of the attack we described in Sect. 4.1. In addition, each server is allocated with same secret value k_2 from RC. He/she can perform this attack by follows:

1. $\mathcal{A}$ intercepts the login request message $\{M_1, CID_i, CHECK_1\}$ sent from legal U_i to S_j and computes $n_i = h\big(SID_j \parallel h(y)\big) \oplus M_1$ and $A_i = CID_i \oplus h(n_i) \oplus k_2$.

2. $\mathcal{A}$ generates a random number n_i^*, then computes $M_1^* = h\big(SID_j \parallel h(y)\big) \oplus n_i^*$, $CID_i^* = A_i \oplus k_2 \oplus h\big(n_i^*\big)$ and $CHECK_1^* = h\big(h\big(SID_j \parallel h(y)\big) \parallel n_i^* \parallel h(A_i \parallel K_2)\big)$ and sends the forged login request message $\big\{M_1^*, CID_i^*, CHECK_1^*\big\}$ to S_j.
3. S_j retrieves $n_i^* = M_1^* \oplus k_1$ using the request message. Since n_i^* is chosen within valid time interval, S_j proceeds to compute $A_i = CID_i \oplus h\big(n_i^*\big) \oplus k_2$ and verify the condition $h(k_1 \parallel n_i \parallel h(A_i \parallel k_2)) = ?CHECK_1$. Obviously, the condition holds, therefore S_j authenticates $\mathcal{A}$ as legal user and computes $M_2 = n_j \oplus n_i^* \oplus k_1$, $SK = h\big(h(A_i \parallel k_2) \parallel n_i^* \parallel n_j\big)$ and $CHECK_2 = h(SK)$, where n_j is the random number generated by S_j. Finally, S_j reply $\mathcal{A}$ with $\{M_2, CHECK_2\}$.
4. After receiving the response message, $\mathcal{A}$ retrieves $n_j = m_2 \oplus h\big(SID_j \parallel h(y)\big) \oplus n_i^*$, $SK = h\big(A_i \oplus k_2 \parallel n_i^* \parallel n_j\big)$ and computes $CHECK_3 = h\big(SK \parallel n_j\big)$. Afterwards, $\mathcal{A}$ sends mutual authentication message $CHECK_3$ to S_j.
5. Upon receiving the authentication message from $\mathcal{A}$, S_j continues to proceed the scheme. Lastly, S_j is mistakenly convinced that $\mathcal{A}$ is a legitimate user and agrees on the session key SK with him/her.

If $\mathcal{A}$ is a malicious user, he/she still can launch this attack by follows:

1. $\mathcal{A}$ obtains RC's secret $h(y)$ by calculating $h(y) = C_a \oplus h^2(R_a)$, where C_a is stored in $\mathcal{A}$'s smart card and R_a can be recovered from $R_a = h(PW_a \oplus D(V_a \oplus h(PW_a) \oplus BIO_a))$. by using his/her PW_a and BIO_a.
2. $\mathcal{A}$ intercepts the login request message $\{M_1, CID_i, CHECK_1\}$ sent from U_i to S_j and computes $n_i = h\big(SID_j \parallel h(y)\big) \oplus M_1$ and $A_i \oplus k_2 = CID_i \oplus h(n_i)$.
3. $\mathcal{A}$ steals SC_i and extracts $\{V_i, B_i, C_i, D_i, E_i, h(.)\}$ from it by using power analysis. Then, $\mathcal{A}$ calculates $h(R_i) = D_i \oplus A_i \oplus k_2$ and $G_i = E_i \oplus h(R_i)$.
4. $\mathcal{A}$ computes $M_1^* = h\big(SID_j \parallel h(y)\big) \oplus n_i^*$, $CID_i^* = A_i \oplus k_2 \oplus h\big(n_i^*\big)$ and $CHECK_1^* = h\big(h\big(SID_j \parallel h(y)\big) \parallel n_i^* \parallel h(A_i \parallel k_2)\big)$, where random number n_i^* is chosen by $\mathcal{A}$ freely. Then $\mathcal{A}$ forges login request message $\big\{M_1^*, CID_i^*, CHECK_1^*\big\}$ and sends it to S_j.
5. Upon receiving the message from $\mathcal{A}$ who manages to impersonate as legal user U_i, the message can successfully pass S_j's verification.
6. Perform steps 3 to 5 in aforementioned attack that $\mathcal{A}$ is a malicious server. Finally, S_j authenticates $\mathcal{A}$ and shares the same session key with him/her.

5 The Proposed Scheme

This section proposes an improved biometrics-based authentication scheme that is suitable for use in multi-server environment. The proposed scheme comprises three participants: user (U_i), server (S_j), registration center (RC), and five phases: server registration, user registration, login, authentication, and password change.

5.1 Server Registration Phase

The server registration phase of proposed scheme is same as Chang et al.'s scheme [8].

5.2 User Registration Phase

1. U_i conducts in the same method as described in step 1 in Sect. 3.2.
2. Upon receiving the registration request message from U_i, RC computes $A_i = h(ID_i \| x)$, $B_i = h(ID_i \| R_i)$, $C_i = h(R_i) \oplus h(y)$, $D_i = A_i \oplus h(x \| y)$ and $E_i = h(A_i \| h(x \| y)) \oplus h(R_i \| h(y))$.
3. RC issues SC_i which contains $\{V_i, B_i, C_i, D_i, E_i, h(.)\}$ and sends it to U_i.

5.3 Login Phase

1. U_i inserts SC_i into a card reader, inputs ID_i^*, PW_i^* and BIO_i^*. SC_i first computes $R_i^* = h\left(PW_i^* \oplus D\left(V_i \oplus h\left(PW_i^*\right) \oplus BIO_i^*\right)\right)$ and verifies whether $h\left(ID_i^* \| R_i^*\right)$ equals to B_i. If it generates negative result, this phase can be terminated.
2. SC_i generates a random nonce n_i and computes $h(y) = C_i \oplus h\left(R_i^*\right)$, $M_1 = h\left(SID_j \| h(y)\right) \oplus n_i$, $CID_i = D_i \oplus h(n_i)$, $G_i = E_i \oplus h\left(R_i^* \| h(y)\right)$ and $CHECK_1 = h\left(h\left(SID_j \| h(y)\right) \| n_i \| G_i\right)$.
3. SC_i sends the request message $\{M_1, CID_i, CHECK_1\}$ to S_j.

5.4 Authentication Phase

1. S_j first checks the validity of the request message by verifying the freshness of random nonce $n_i = M_1 \oplus k_1$. If it holds, S_j computes $A_i = CID_i \oplus h(n_i)$ and verifies whether $h(k_1 \| n_i \| h(A_i \| k_2))$ equals to $CHECK_1$. If the condition holds, S_j authenticates U_i. Otherwise, the session is aborted.
2. S_j further generates a random number n_j and computes $M_2 = n_j \oplus n_i \oplus k_1$, $SK = h\left(h(A_i \| k_2) \| n_i \| n_j\right)$ and $CHECK_2 = h(SK)$. Then, S_j sends the response message $\{M_2, CHECK_2\}$ to U_i.
3. The rest of the authentication phase is same as Chang et al.'s scheme.

5.5 Password Change Phase

1. U_i inserts his/her SC_i into a card reader, then keys his/her ID_i and PW_i, and imprints personal biometric information BIO_i via a special device.
2. SC_i retrieves $\alpha_i = V_i \oplus h(PW_i)$, $r_i = D(BIO_i \oplus \alpha_i)$ and $R_i = h(PW_i \oplus r_i)$, and verifies the whether $h(id_i \| R_i)$ is equal to B_i. If it holds, U_i is allowed to type a new password, otherwise this phase can be aborted.
3. U_i types a new password PW_i^{new}. SC_i calculates $h(y) = C_i \oplus h(R_i)$, $V_i^{new} = V_i \oplus h(PW_i) \oplus h\left(PW_i^{new}\right)$, $R_i^{new} = h\left(PW_i^{new} \oplus r_i\right)$, $B_i^{new} = h\left(ID_i \| R_i^{new}\right)$, $C_i^{new} = C_i \oplus h(R_i) \oplus h\left(R_i^{new}\right)$ and $E_i^{new} = E_i \oplus h(R_i \| h(y)) \oplus h\left(R_i^{new} \| h(y)\right)$. Lastly, SC_i replaces V_i, B_i, C_i and E_i with V_i^{new}, B_i^{new}, C_i^{new} and E_i^{new}.

6 Cryptanalysis of Proposed Scheme

In this section, we cryptanalyze the proposed scheme and examines its security against various attacks. As described in Sect. 5, to achieve least increase on computational cost, our scheme modifies little in user registration phase and login phase based on Chang et al.'s scheme [8] and provides higher security. Therefore, all security features mentioned in [8] are also met in our scheme. In addition, we comparatively give an analysis between our scheme and previous schemes [2–5, 8], which is illustrated in Table 2.

Table 2. Comparison on security level between proposed scheme and related schemes

Features	Ours	[8]	[3]	[5]	[4]	[2]
Outsider attack	Yes	No	Yes	Yes	Yes	Yes
Session key derived attack	Yes	No	Yes	Yes	Yes	Yes
User impersonation attack	Yes	No	No	No	Yes	No
Off-line password guessing attack	Yes	Yes	Yes	Yes	Yes	No
Server spoofing attack	Yes	Yes	No	Yes	No	No
Stolen smart card attack	Yes	Yes	No	No	Yes	No

Yes: The scheme can resist the attack. No: The scheme cannot resist the attack

6.1 Resistance to Outsider Attack

Assume an adversary $\mathcal{A}$ is a malicious server who is aware of $k_1 = h\left(SID_j \parallel h(y)\right)$ and $k_2 = h(x \parallel y)$, however he/she cannot obtain $h(y)$ by computing $h(y) = h(R_i) \oplus C_i$, where C_i is stored in SC_i. Only possessing correct ID_i, PW_i and BIO_i can retrieve random number r_i and further compute R_i. The possibility that $\mathcal{A}$ obtains ID_i and PW_i simultaneously is extremely small, and BIO_i cannot be forged or obtained since it is imprinted by U_i via a special device. Furthermore, $h(R_i)$ is only applied to constitute C_i, which means $\mathcal{A}$ is not capable of obtaining it from operating with any other parameters. Therefore, our scheme prevents $\mathcal{A}$ from launching outsider attack.

6.2 Resistance to Session Key Derived Attack

The session key is computed as $SK = h\left(h(A_i \parallel k_2) \parallel n_i \parallel n_j\right)$, where $A_i = h(ID_i \parallel x)$, $k_2 = h(x \parallel y)$, random numbers n_i and n_j are generated by U_i and S_j, respectively. Assume an adversary $\mathcal{A}$ somehow obtains ID_i, he/she cannot compute SK without the knowledge of secrets x and y that are only known by RC. $\mathcal{A}$ cannot retrieve random numbers n_i and n_j neither, since they must be computed by using $h(y)$ and k_1, which indicates that only legal user and server can compute these two random nonces. Therefore, $\mathcal{A}$ cannot reveal session key SK by any means in the proposed scheme.

6.3 Resistance to User Impersonation Attack

Assume that an adversary $\mathcal{A}$ intercepts all messages $\{M_1, M_2, M_3, CID_i, CHECK_1, CHECK_2, CHECK_3\}$ between U_i and S_j through a public network, steals SC_i and extracts all information $\{V_i, B_i, C_i, D_i, E_i, h(.)\}$. However, $\mathcal{A}$ cannot forge login request message $\{M_1, CID_i, CHECK_1\}$, where $M_1 = h(SID_j \parallel h(y)) \oplus n_i$, $CID_i = D_i \oplus h(n_i) = A_i \oplus h(x \parallel y) \oplus h(n_i)$ and $CHECK_1 = h(h(SID_j \parallel h(y)) \parallel n_i \parallel G_i) = h(h(SID_j \parallel h(y)) \parallel n_i \parallel h(A_i \parallel h(x \parallel y)))$, because secrets x and y are only known to RC, n_i is a random nonce that is generated by U_i. Furthermore, $\mathcal{A}$ cannot generate $\{M_1, CID_i, CHECK_1\}$ without A_i, which can be exclusively obtained by S_j. If the adversary $\mathcal{A}$ is a malicious user or server, he/she is capable of retrieving some parameters within $\{n_i, h(SID_j \parallel h(y)), h(x \parallel y), h(y)\}$. However, as described in Subsects. 6.1 and 6.2, it is impossible for $\mathcal{A}$ to obtain all parameters that form a valid login request message $\{M_1, CID_i, CHECK_1\}$ to impersonate as a legitimate user. Hence, our scheme can resist user impersonation attack.

7 Performance Analysis

In this section, we compare our scheme with other related schemes [2–5, 8] on computational cost during login and authentication phase, which is illustrated in detail in Table 3. Notations used in this section are described as follows. T_h refers to the time to execute a one-way hash function for a single time. T_E and T_D are defined as the time taken to perform one encoding or decoding operation based on Dodis et al.'s definition [9]. T_{ecc} is the computation time that one elliptic curve operation requires. T_e indicates the computation time for one modular exponentiation operation. The computational parameter T_f indicates the computation time to execute fuzzy extractor for once. Although our scheme requires one more hash operation during login phase compared with Chang et al.'s scheme, however it consumes an extremely small amount of time. Considering the security enhancement of proposed scheme, the increased computation cost is worthy.

Table 3. Comparison of computational cost in login and authentication phase between proposed scheme and related schemes

Phases	Ours	[8]	[3]	[5]	[4]	[2]
Login	$8T_h + 1T_D$	$7T_h + 1T_D$	$4T_h$	$2T_h + 1T_{ecc}$	$4T_h + 1T_e + 1T_f$	$2T_h$
Authentication	$10T_h$	$10T_h$	$13T_h$	$15T_h + 3T_{ecc}$	$4T_h + 4T_e$	$8T_h$
Total	$18T_h + 1T_D$	$17T_h + 1T_D$	$17T_h$	$17T_h + 4T_{ecc}$	$8T_h + 5T_e + 1T_f$	$10T_h$

8 Conclusions

In this paper, we analyze Chang et al.'s scheme and demonstrate that it possesses a number of security vulnerabilities including outsider attack, session key derived attack and user impersonation attack. To overcome these flaws, we propose an improved biometrics-based authentication scheme which retains the merits of Chang et al.'s scheme and also achieves a variety of security features. In addition, the cryptanalysis of this paper shows that our scheme rectifies weaknesses of Chang et al.'s scheme.

Acknowledgements. This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (No. R0126-15-1111, The Development of Risk-based Authentication Access Control Platform and Compliance Technique for Cloud Security).

References

1. Lamport, L.: Password authentication with insecure communication. Commun. ACM **24**(11), 770–772 (1981)
2. Das, A.K.: Analysis and improvement on an efficient biometric-based remote user authentication scheme using smart cards. IET Inf. Secur. **5**(3), 145–151 (2011)
3. Li, X., Niu, J.W., Ma, J., Wang, W.D., Liu, C.L.: Cryptanalysis and improvement of a biometrics-based remote user authentication scheme using smart cards. J. Netw. Comput. Appl. **34**(1), 73–79 (2011)
4. Yang, D., Yang, B.: A biometric password-based multi-server authentication scheme with smart card. In: 2010 International Conference on Computer Design and Applications (ICCDA), vol. 5, p. V5-554. IEEE (2010)
5. Yoon, E.J., Yoo, K.Y.: Robust biometrics-based multi-server authentication with key agreement scheme for smart cards on elliptic curve cryptosystem. J. Supercomput. **63**(1), 235–255 (2013)
6. He, D.: Security flaws in a biometrics-based multi-server authentication with key agreement scheme. IACR Cryptology ePrint Archive, 365 (2011)
7. Chuang, M.C., Chen, M.C.: An anonymous multi-server authenticated key agreement scheme based on trust computing using smart cards and biometrics. Expert Syst. Appl. **41**(4), 1411–1418 (2014)
8. Chang, C.C., Hsueh, W.Y., Cheng, T.F.: An advanced anonymous and biometrics-based multi-server authentication scheme using smart cards. Int. J. Netw. Secur. **18**(6), 1010–1021 (2016)
9. Dodis, Y., Reyzin, L, Smith, A.: Fuzzy extractors: how to generate strong keys from biometrics and other noisy data. In: International Conference on the Theory and Applications of Cryptographic Techniques, pp. 523–540. Springer, Heidelberg (2004)
10. Moon, J., Choi, Y., Jung, J., Won, D.: An improvement of robust biometrics-based authentication and key agreement scheme for multi-server environments using smart cards. PloS One **10**(12), e0145263.5 (2015)
11. Jung, J., Kang, D., Lee, D., Won, D.: An improved and secure anonymous biometric-based user authentication with key agreement scheme for the integrated EPR information system. PLoS One **12**(1), e0169414 (2017)
12. Kim, J., Lee, D., Jeon, W., Lee, Y., Won, D.: Security analysis and improvements of two-factor mutual authentication with key agreement in wireless sensor networks. Sensors **14**(4), 6443–6462 (2014)

Cryptanalysis of Chaos-Based 2-Party Key Agreement Protocol with Provable Security

Jongho Moon[1], Taeui Song[1], Donghoon Lee[1], Youngsook Lee[2], and Dongho Won[1](✉)

[1] Information Security Group, Sungkyunkwan University, Suwon, South Korea
{jhmoon, tusong, dhlee, dhwon}@security.re.kr
[2] Department of Cyber Security, Howon University, Gunsan, South Korea
ysooklee@howon.ac.kr

Abstract. In a public communication environment, a remote user authentication scheme for establishing a secure session between a user and a server is a very important factor. Authentication schemes, which originate from a password-based authentication scheme, apply some mathematical algorithms to securely share session keys between users and servers. In a remote user authentication scheme, safety is a very important factor, but it is also important to reduce computational cost. Therefore, even if a mathematical algorithm is applied, it is necessary to select an algorithm that consumes a small amount of computation. Recently, Luo et al. proposed a chaos-based two-party key exchange protocol and claimed that the proposed scheme solved the off-line password guessing attack and was safe from other common attacks. They used a Chebyshev chaotic maps. This algorithm is used in many authentication schemes because it consumes a small amount of computation. However, we find that Luo et al.'s scheme is still insecure. In this paper, we show the problems of Chebyshev chaotic maps and demonstrate how an attacker can attempt some attacks.

Keywords: Chaotic map · User authentication · Key agreement

1 Introduction

Password-based authentication schemes have been widely used for decades with the development of communication technologies. Since Lamport [1] proposed the first password-based authentication scheme with insecure communication in 1981, password-based authentication schemes have been extensively investigated. The problem with password-based authentication, however, is that the server must maintain a password table to verify the legitimacy of the login user.

Key agreement is as important as the user authentication for secure communication over a public channel between two or more participants. Most of recent key agreement schemes are originated from Deffie–Hellman (D–H) key agreement which was proposed by Diffie and Hellman [2]. However, the original Deffie-Hellman key agreement is vulnerable against man-in-the middle attack.

D. Nicholson (ed.), *Advances in Human Factors in Cybersecurity*,
Advances in Intelligent Systems and Computing 593, DOI 10.1007/978-3-319-60585-2_8

To overcome this problem, the researchers began to apply certain mathematical algorithms. One of them is chaos. Chaos is a type of deterministic random process generated by a nonlinear dynamic system that can be used to design digital chaos-based cryptosystems.

In 2007, Xiao et al. [3] proposed a chaos-based key agreement scheme based on utilizing chaotic public-key cryptosystem [4]. Comparing to the traditional schemes in the part of key agreement, it could reduce computation complexity. However, Guo and Zhang [5] pointed out that Xiao et al.'s [3] scheme could not resist the server spoofing and denial-of-service (DoS) attacks. Furthermore, in Guo and Zhang [5] proposed an enhanced scheme, and claimed that their scheme could resist the security flaws of Xiao et al.'s scheme. However, Liu et al. [6] demonstrated some crucial security flaws of Guo et al.'s scheme [5], and proposed an improved scheme. They use the timestamp value and hash function with the shared secret and some context information to bind multiple steps as organic whole, which makes every step message cannot be separately used to launch replay and DoS attacks. Unfortunately, Luo et al. [7] found that Liu et al.'s scheme is vulnerable to the off-line password guessing and server impersonation attacks, and proposed an improved scheme.

In this paper, we first introduce the Chebyshev chaotic maps [8]. Afterward, we review Luo et al.'s scheme and demonstrate that Luo et al.'s scheme is still insecure to the off-line password guessing and user impersonation attacks and has a problem about the session key security.

The rest of this paper is organized as follows. In Sect. 2, we briefly introduce the Chebyshev chaotic maps. In Sect. 3, we review Luo et al.'s scheme. In Sect. 4, we analyze Luo et al.'s scheme and show that their scheme is why vulnerable. Our conclusions are presented in Sect. 5.

2 Preliminaries

In this section, we briefly introduce the Chebyshev chaotic maps [8, 9] and threat assumptions.

2.1 Chebyshev Chaotic Maps

The Chebyshev polynomial $T_n(x)$ is an x polynomial of degree n.

Definition 1. Let n be an integer and x is a real number from the set $[-1, 1]$, so that the Chebyshev polynomial of degree n is defined as $T_n(x) = \cos(n \cdot arccos(x))$.

Definition 2. Given the two elements x and y $\in Z_p^*$, the Chaotic Maps Discrete Logarithm Problem (CMDLP) is whether the integer r can be found such that $y = T_r(x)$. The probability of $\mathcal{A}$ being able to solve the CMDLP is defined as $\Pr[\mathcal{A}(x, y) = r : r \in Z_p^*, y = T_r(x) \bmod p]$.

Definition 3. Given the three parameters x, $T_r(x)$ and $T_s(x)$, the Chaotic Maps Diffie-Hellman Problem (CMDHP) is whether $T_{rs}(x)$ can be computed such that $T_{rs}(x) = T_r(T_s(x)) = T_s(T_r(x))$.

2.2 Threat Assumptions

We introduce the Dolev-Yao [10] and some threat model [11, 12] to construct the threat assumptions which are describe as follows:

1. An adversary E can be either a user or server. Any registered user can act as an adversary.
2. An adversary E can eavesdrop every communication in a public channel, thereby capturing any message exchanged between a user and server.
3. An adversary E can alter, delete, or reroute the captured message.
4. Information can be extracted from the smart card by examining the power consumption [13] of the card.

3 Review of Luo et al.'s Scheme

In this section, we review Luo et al.'s authentication scheme. They assumed that the hash value h_{pw} is shared between user A and server B, where $h_{pw} = H(ID_A \| PW_A)$ and H(·) means the chaotic one-way hash function. For convenience, the notations are described in Table 1.

Table 1. Notations and descriptions.

Term	Description
A	The user
B	The server
ID_A	User's identity
ID_B	Server's identity
r, s	A random number
$\oplus$	Bitwise XOR operation
$\|$	Concatenation operation
$H(\cdot)$	Hash function
h_{pw}	The shares hash value
$K_{session}$	The session key

Figure 1 illustrates the authentication and key agreement phase of Luo et al.'s scheme. Details are as below.

1. $A \rightarrow B : \{X, ID_A\}$. User A chooses a random number r and computes $X = T_r(h_{pw})$ and then sends $\{X, ID_A\}$ to the server B.

2. $B \rightarrow A$: $\{Y, \beta, ID_B\}$. After receiving the login request message, the server B chooses a random number s, and computes $Y = T_s(h_{pw})$, $K_B = T_s(X)$ and $\beta = H(ID_A \,||\, X \,||\, ID_B \,||\, Y \,||\, K_B)$. Finally, the server B sends $\{Y, \beta, ID_B\}$ to the user A.
3. $A \rightarrow B : \{\alpha\}$. User A computes $K_A = T_r(Y)$ when receiving the response message, and then verifies the legitimacy of the equation, which is $\beta = H(ID_A \,||\, X \,||\, ID_B \,||\, Y \,||\, K_B)$. If this hold, the server B is authenticated successfully; otherwise, the user A terminates this session. Afterwards, the user A computes $\alpha = H(ID_A \,||\, X \,||\, ID_B \,||\, Y \,||\, K_A)$, and sends $\{\alpha\}$ to the server B.
4. The server B verifies whether $\alpha = H(ID_A \,||\, X \,||\, ID_B \,||\, Y \,||\, K_B)$ holds or not when receiving the message $\{\alpha\}$. If this hold, the user A is authenticated successfully; otherwise, the server B terminates this session.
5. The user A and server B compute the session key: $K_{session} = T_r(T_s(h_{pw})) = T_s(T_r(h_{pw})) = T_{rs}(h_{pw})$.

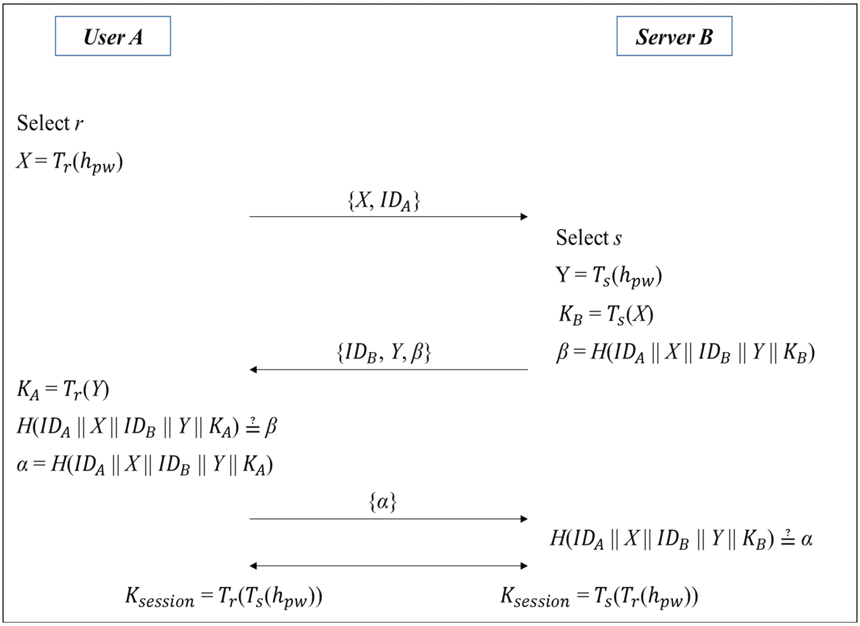

User A
Server B
Select r
$X = T_r(h_{pw})$
$\{X, ID_A\}$
Select s
$Y = T_s(h_{pw})$
$K_B = T_s(X)$
$\{ID_B, Y, \beta\}$
$\beta = H(ID_A \,||\, X \,||\, ID_B \,||\, Y \,||\, K_B)$
$K_A = T_r(Y)$
$H(ID_A \,||\, X \,||\, ID_B \,||\, Y \,||\, K_A) \overset{?}{=} \beta$
$\alpha = H(ID_A \,||\, X \,||\, ID_B \,||\, Y \,||\, K_A)$
$\{\alpha\}$
$H(ID_A \,||\, X \,||\, ID_B \,||\, Y \,||\, K_B) \overset{?}{=} \alpha$
$K_{session} = T_r(T_s(h_{pw}))$
$K_{session} = T_s(T_r(h_{pw}))$

Fig. 1. Luo et al.'s authentication and key agreement protocol.

4 Security Analysis of Luo et al.'s Scheme

Luo et al. claimed that their scheme is resistant to the off-line password guessing and impersonation attacks; however, we demonstrated that their scheme is still insecure to these attack types. We provide the details of these problems in the following subsections.

4.1 User Anonymity

In Luo et al.'s scheme, the user A sends the identity ID_A to the server B and the server B sends the identity ID_B to the user A. In the public communication channel, any adversary E can intercept or eavesdrop on the communicated message at any time. Therefore, their scheme does not provide user anonymity.

4.2 Off-line Password Guessing Attack

Suppose that an adversary E intercepts the communication messages $\{X, ID_A, Y, \beta, ID_B, \alpha\}$ between the user A and server B. The E can then obtain the password PW_A of the user A. The details are described as follows:

1. Adversary E guesses a password PW_A^*, and then he/she uses the published hash function to compute the $h_{pw}^* = H(ID_A \parallel PW_A^*)$.
2. Using the [4], the E computes $r' = \frac{\arccos(X) + 2k'\pi}{\arccos(h_{pw}^*)}, \forall k \in Z$.
3. By comparing the $T_{r'}(h_{pw}^*)$ with the X, if they are not equal, repeat seep 1 until they are equal.

4.3 Violation of the Session Key Security

Suppose that an adversary E intercepts the communication messages $\{X, ID_A, Y, \beta, ID_B, \alpha\}$ between the user A and server B, and obtains the password PW_A of the user A by the off-line password guessing attack. The E can then easily obtain the session key between the user A and server B. The details are described as follows:

1. Using the [4], the E computes $s' = \frac{\arccos(Y) + 2k'\pi}{\arccos(h_{pw})}, \forall k \in Z$.
2. The E can compute the shared session key $K_{session} = T_{s'}(X) = T_{s'}(T_r(h_{pw}))$.

4.4 User Impersonation Attack

Suppose that an adversary E intercepts the communication messages $\{X, ID_A, Y, \beta, ID_B, \alpha\}$ between the user A and server B, and obtains the password PW_A of the user A by the off-line password guessing attack. Then the E can easily impersonate the user A.

5 Conclusion

Recently various key-agreement schemes for secure communication have been proposed. In this paper, we have identified vulnerabilities in Luo et al.'s scheme in terms of off-line password guessing and user impersonation. We show how their scheme can suffer from these attacks. Finally, our further research direction ought to propose a secure user authentication and key agreement scheme which can solve these problems.

Acknowledgments. This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (No. R0126-15-1111, The Development of Risk-based Authentication·Access Control Platform and Compliance Technique for Cloud Security).

References

1. Lamport, L.: Password authentication with insecure communication. Commun. ACM **24**, 770–772 (1981)
2. Diffie, W., Hellman, M.E.: New directions in cryptography. IEEE Trans. Inf. Theor. **22**, 644–654 (1976)
3. Xiao, D., Liao, X., Deng, S.: A novel key agreement protocol based on chaotic maps. Inf. Sci. **177**, 1136–1142 (2007)
4. Bergamo, P., D'Arco, P., De Santis, A., Kocarev, L.: Security of public-key cryptosystems based on chebyshev polynomials. IEEE Trans. Circ. Syst. **52**, 1382–1393 (2005)
5. Guo, X., Zhang, J.: Secure group key agreement protocol based on chaotic hash. Inf. Sci. **180**, 4069–4074 (2010)
6. Liu, Y., Xue, K.: An improved secure and efficient password and chaos-based two-party key agreement protocol. Nonlinear Dyn. **84**, 549–557 (2016)
7. Lou, M., Zhang, Y., Khan, M.K., He, D.: An efficient chaos-based 2-party key agreement protocol with provable security. Int. J. Commun. Syst. 1–9 (2017)
8. Lee, C.C., Hsu, C.W.: A secure biometric-based remote user authentication with key agreement scheme using extended chaotic maps. Nonlinear Dyn. **71**, 201–211 (2013)
9. Moon, J., Choi, Y., Kim, J., Won, D.: An improvement of Robust and efficient biometrics based password authentication scheme for telecare medicine information systems using extended chaotic maps. J. Med. Syst. **40**, 1–11 (2016)
10. Dolev, D., Yao, A.: On the security of public key protocols. IEEE Trans. Inf. Theor. **29**, 198–208 (1983)
11. Moon, J., Choi, Y., Jung, J., Won, D.: An improvement of Robust biometrics-based authentication and key agreement scheme for multi-server environments using smart cards. PLoS ONE **10**, 1–15 (2015)
12. Choi, Y., Lee, D., Kim, J., Jung, J., Nam, J., Won, D.: Security enhanced user authentication protocol for wireless sensor networks using elliptic curves cryptography. Sensors **14**, 10081–10106 (2014)
13. Kocher, P., Jaffe, J., Jun, B., Rohatgi, P.: Introduction to differential power analysis. J. Cryptographic Eng. **1**, 1–23 (2011)

Cryptanalysis of Lightweight User Authentication Scheme Using Smartcard

Dongwoo Kang[1], Jaewook Jung[1], Hyungkyu Yang[2],
Younsung Choi[3], and Dongho Won[1(✉)]

[1] College of Information and Communication Engineering,
Sungkyunkwan University, 2066 Seobu-ro, Jangan-gu, Suwon-si,
Gyeongki-do 440-746, South Korea
{dwkang, jwjung, dhwon}@security.re.kr

[2] Department of Computer and Media Information,
Kangnam University, Seoul, South Korea
hkyang@kangnam.ac.kr

[3] Department of Cyber Security, Howon University, Impi-Myeon,
Gunsan-si, Jeonrabuk-Do 573-718, South Korea
yschoi@howon.ac.kr

Abstract. The mobile device market has grown rapidly, and as the internet becomes available wireless, it offers a variety of services to people such as browsing, file sharing, shopping anytime and anywhere. Contemporary, a smartcard comes to one of beneficial thing because of its convenience and lightweight. As smartcards become commercially available, on one side, smartcard based authentication scheme also actively researched. In 2016, Ahmed et al. proposed lightweight communication overhead authentication scheme with smartcard. Ahmed et al. argued that scheme they proposed was lightweight compared to the previously well-known other schemes, safe from multiple attacks, and satisfied multiple security features. However, we found that Ahmed et al.'s scheme also showed weaknesses and scheme's progress was incomplete. In this paper, we briefly introduce Ahmed et al.'s scheme and demonstrate that their scheme is still unstable to apply to user authentication environment using smartcard.

Keywords: Lightweight · User authentication · Smartcard · Network security · Security threat

1 Introduction

Since Lamport proposed the first password-based authentication scheme in 1981 [1]. Over the past few years, several studies have made on smart card-based user authentication [2–4]. However, the computer became more sophisticated and able to more calculations fasters, 3-factor user authentication schemes are under active research to safely protect user information [5–7]. The primary purpose of the authentication scheme using smart card is to verify and diagnose the reasonable user in a public channel environment where messages can be eavesdropped. Therefore, Smart card-based authentication scheme was proposed and developed continuously. Center of

D. Nicholson (ed.), *Advances in Human Factors in Cybersecurity*,
Advances in Intelligent Systems and Computing 593, DOI 10.1007/978-3-319-60585-2_9

development, the protection of the user's information likewise identity, password is also emerged as important problem even if the smartcard is stolen or lost and all information is leaked. Inevitably, there are some following security requirements when we proposed authentication scheme. Such as, security threat, anonymity [8, 9], mutual authentication [10, 11], and efficiency.

Ahmed et al. [12] in 2016 proposed that the new type of authentication scheme which use user's biometric information and preserve user anonymity and lightweight. Ahmed proved his scheme can resist various network security threats such as insider attack, replay attack, guessing attack, stolen-verifier attack, forgery attack, impersonate attack and so on. However, we discovered that it's scheme still unstable and inefficient. It cannot resist offline identity guessing attack and cannot provide session key confirmation property. Moreover, there is some risk of biometric information's recognition error.

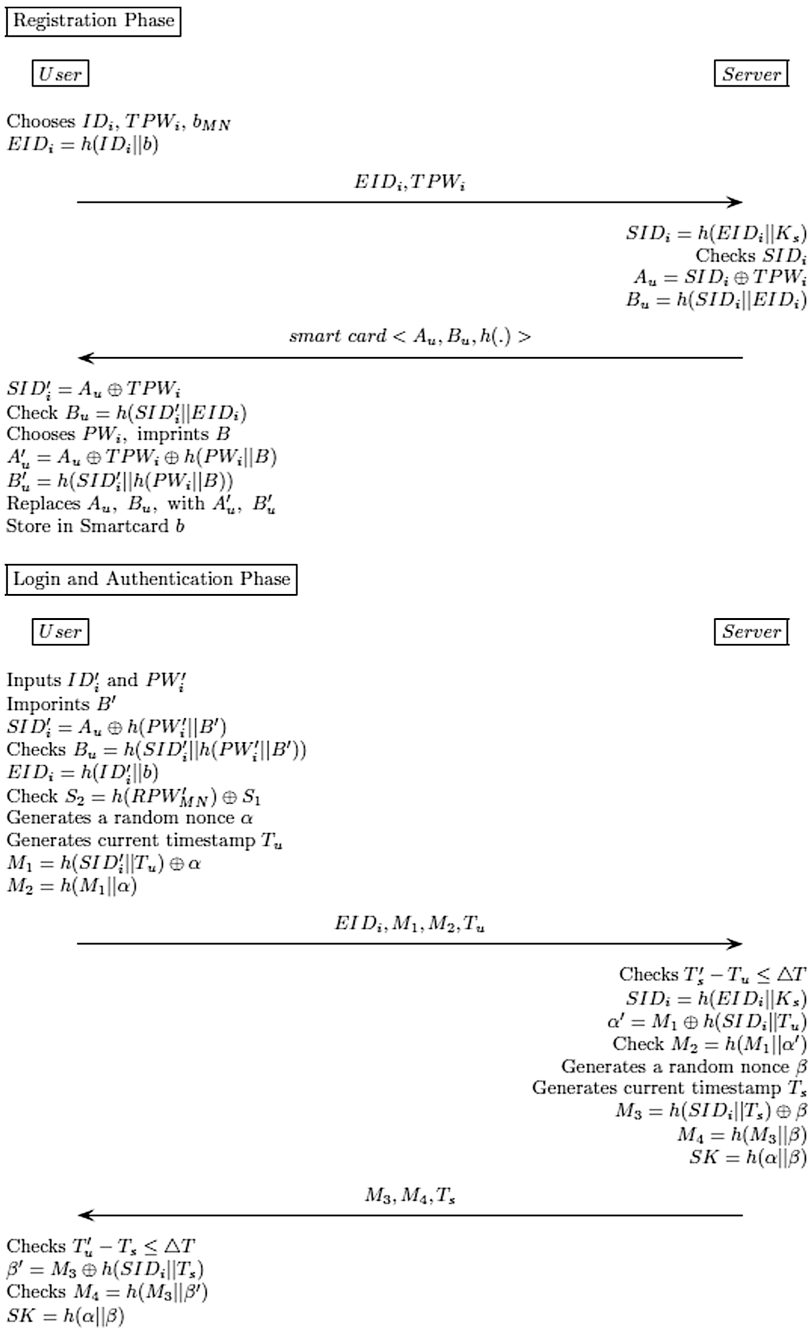
The rest of this paper is organized as follows: review Ahmed et al.'s user authentication scheme in Sect. 2. In Sect. 3, we point out security vulnerability and inefficiency. At the conclusion, we conclude this paper in Sect. 4.

2 Review in Ahmed et al.'s Scheme

This section reviews the lightweight user authentication scheme using smartcard proposed by Ahmed et al. In 2016, Ahmed et al.'s scheme consists of four phases: registration, login, authentication and password changing. The notations used in this paper are summed up as Table 1 and also details procedure of these phase are in Fig. 1.

Table 1. Notation used in this paper

Notations	Description
$\mathrm{U_i}$	A qualified user i
$\mathrm{ID_i}, PW_i$	User i's identity and password
TPW_i	User i's temporary password
B	User i's biometric information
S	A remote server
$\mathrm{K_s}$	Server secret key
T	Timestamp
ΔT	The maximum transmission delay
h(.)	A collision resistant one-way hash function
$\oplus$	The bitwise XOR operation
$\|\|$	String concatenation
$\dashrightarrow$	Secure channel
$\longrightarrow$	Public channel



Registration Phase

$User$ $Server$

Chooses ID_i, TPW_i, b_{MN}
$EID_i = h(ID_i||b)$

EID_i, TPW_i →

$SID_i = h(EID_i||K_s)$
Checks SID_i
$A_u = SID_i \oplus TPW_i$
$B_u = h(SID_i||EID_i)$

← $smart\ card < A_u, B_u, h(.) >$

$SID_i' = A_u \oplus TPW_i$
Check $B_u = h(SID_i'||EID_i)$
Chooses PW_i, imprints B
$A_u' = A_u \oplus TPW_i \oplus h(PW_i||B)$
$B_u' = h(SID_i'||h(PW_i||B))$
Replaces A_u, B_u, with A_u', B_u'
Store in Smartcard b

Login and Authentication Phase

$User$ $Server$

Inputs ID_i' and PW_i'
Imporints B'
$SID_i' = A_u \oplus h(PW_i'||B')$
Checks $B_u = h(SID_i'||h(PW_i'||B'))$
$EID_i = h(ID_i'||b)$
Check $S_2 = h(RPW_{MN}') \oplus S_1$
Generates a random nonce α
Generates current timestamp T_u
$M_1 = h(SID_i'||T_u) \oplus \alpha$
$M_2 = h(M_1||\alpha)$

EID_i, M_1, M_2, T_u →

Checks $T_s' - T_u \leq \triangle T$
$SID_i = h(EID_i||K_s)$
$\alpha' = M_1 \oplus h(SID_i||T_u)$
Check $M_2 = h(M_1||\alpha')$
Generates a random nonce β
Generates current timestamp T_s
$M_3 = h(SID_i||T_s) \oplus \beta$
$M_4 = h(M_3||\beta)$
$SK = h(\alpha||\beta)$

← M_3, M_4, T_s

Checks $T_u' - T_s \leq \triangle T$
$\beta' = M_3 \oplus h(SID_i||T_s)$
Checks $M_4 = h(M_3||\beta')$
$SK = h(\alpha||\beta)$

Fig. 1. Detail of Ahmed et al.'s Scheme

2.1 Registration Phase

Step 1. $\mathrm{U_i}$ chooses his/her identity $\mathrm{ID_i}$ and temporary password $\mathrm{TPW_i}$ and random number b. U_i computes $EID_i = h(ID_i||b)$. Then sends registration request message $<EID_i, TPW_u>$ to remote server S via secure channel.
Step 2. Remote server S that received registration request message, S computes $SID_i = h(EID_i||K_S)$. Then server checks SID_i and compared it server's verification table. Then, if it is already registered, rejects to prevent duplicated registration for same identity. Otherwise, S updates registered user list with SID_i. Then computes $A_u = SID_i \oplus TPW_i, B_u = h(SID_i \oplus EID_i)$. Then S issues smartcard and stored $\{A_u, B_u, h\}$ into smartcard and sends it to U_i via secure channel.
Step 3. U_i inserts a smartcard into a card reader and inputs his/her identity and temporary password ID'_i, TPW'_i, b' once again. Smartcard computes $EID'_i = h(ID'_i||b')$, $SID'_i = A_u \oplus TPW'_i, B'_u = h(SID'_i \oplus EID'_i)$ and compares B'_u with B_u which in smartcard. If not equal, smartcard terminates the registration session.
Step 4. U_i choose his/her own password PW_i and imprints biometric information such as fingerprint, iris. A smartcard computes $A'_u = A_u \oplus TPW_i \oplus h(PW_i||B), B'_u = h(SID_i||h(PW_i||B))$. Then replaces $A_u, B_u\ with\ A'_u, B'_u$ respectively and stores b into smartcard. Finally, $\{A'_u, B'_u, h, b\}$ stored in a smart card.

2.2 Login Phase

Step 1. $\mathrm{U_i}$ inserts his/her smartcard into card reader, and inputs $\mathrm{ID'_i}, \mathrm{PW}'_i, \mathrm{B}'$. Then, smartcard computes $\mathrm{SID'_i = A'_u \oplus h(PW'_i||b'), B'_u = h(SID'_i||h(PW'_i||B'))}$, and compares $\mathrm{B'_u}$ with $\mathrm{B_u}$ which in smartcard. If not equal, smartcard terminates the login session, and it holds, session proceed.
Step 2. Smartcard generates random nonce α, and current timestamp $\mathrm{T_u}$ and computes $\mathrm{EID_i = h(ID_i||b), M_1 = h(SID_i||T_u) \oplus \alpha, M_2 = h(M_1||\alpha)}$. Then smartcard sends login request message $\mathrm{<EID_i, M_1, M_2, T_u>}$ to server via public channel.

2.3 Authentication Phase

Step 1. Remote server S that received registration request message, S checks user's validity by $SID_i = h(EID_i||K_s)$ to check server's verification table and active user list. Also, S checks time delay by $T'_s - T_u \leq \Delta T$ which T'_s is current server's timestamp. If either condition is not satisfied, login request is rejected.
Step 2. Smartcard computes $\alpha' = M_1 \oplus h(SID_i||T_u), M'_2 = h(M_1||\alpha')$. Then, verify M'_2 equals to M_2, if different, session terminated.
Step 3. Server generates random nonce β, and current timestamp T_s and computes $M_3 = h(SID \parallel T_s), M_4 = h(M_3 \parallel \beta), SK = h(\alpha' \parallel \beta)$. Then, server sends mutual authentication message $<M_3, M_4, T_s>$ to user via public channel.
Step 4. User that received mutual authentication message, U_i checks time delay by $T'_u - T_s \leq \Delta T$ which T'_u is current user's timestamp. Then, calculates $\beta' = M_3 \oplus h(SID_i||T_s)$, $M'_4 = h(M_3||\beta')$. Then, verify M'_4 equals to M_4, if different, session terminated. Then, smartcard computes $SK = h(\alpha||\beta')$.

2.4 Password Change Phase

Step 1. User who wants to change his/her password inserts his/her smartcard into card reader, and inputs existing $\mathrm{ID}'_{\mathrm{i}}, \mathrm{PW}'_{\mathrm{i}}, \mathrm{B}'$. Then, smartcard computes $SID'_i = A'_u \oplus h(PW'_i \| b'), B'_u = h(SID'_i \| h(PW'_i \| B'))$, and compares B'_u with B_u which in smartcard. If not equal, smartcard terminates the password change phase, and it holds, phase proceeds.

Step 2. User inputs his/her new password PW_i^{new}, smartcard calculates $A_u^{new} = A_u \oplus h(PW_i \| B) \oplus h(PW_i^{new} \| B), B_u^{new} = h(SID_i \| h(PW_i^{new} \| B))$.

Step 3. Smartcard replaces A_u, B_u with A_u^{new}, B_u^{new} respectively. All these process without need to remote server's communication.

3 Cryptanalysis of Ahmed et al.'s Scheme

In this section, we point out security weakness of Ahmed et al.'s scheme. There are some adversary threat model are made analysis and design of the scheme.

(1) An adversary $\mathrm{U_a}$ can be either a user or remote server.

(2) An adversary $\mathrm{U_a}$ has total control over the public communication channel. Therefore, the adversary can intercept, insert, delete or modify any message transmitted via a public channel.

(3) An adversary $\mathrm{U_a}$ may steal legal user's smartcard and extract the information stored in it by means of analyzing the power consumption attack [13].

(4) An adversary $\mathrm{U_a}$ can easily guess low-entropy either password or identity in an offline guessing, but the guessing of two secret parameters is computationally infeasible in polynomial time [14].

3.1 Biometric Recognition Error

According the Ahmed et al.'s scheme, it applies a one-way hash function to biometric information. However, this hash function can be used to fingerprint of the data of an random size to fixed size. Unfortunately, biometric information has general restraint such as false acceptance and false rejection. This means that the output of the imprinted biometrics is not always same. Even though legal user inputs his/her own biometrics to the scanning device such as fingerprint sensors or iris recognition units, it is possible to different with initial biometrics information B. Therefore, even though person feels he/she inputs same biometrics information, the biometrics can generate different output. Furthermore, according to property of hash function, the large difference can be accrued. As a result, advanced techniques are needed to improve success rate of a legal user's verification such as fuzzy extractor or Bio hash function [15].

3.2 Lack of Explicit Session Key Confirmation Property

According to analysis provided in [16], an authenticated key scheme must have the explicit and implicit session key confirmation property. The implicit key confirmation property includes that the user assured that server could compute the session key. In addition, the explicit key confirmation property states that the user is assured that the server has actually computed the session key. Therefore, only the explicit key confirmation property provides the stronger assurances that and holds the same session key. A key agreement scheme that includes explicit key authentication is termed as authenticated key agreement with key confirmation (AKC) scheme. However, In the authentication phase of Ahmed et al.'s scheme, server computes $M_3 = h(SID_i||T_s) \oplus \beta, M_4 = h(M_3||\beta)$ and sends it to user. On receiving M_3 and M_4, user computes the session key as $SK = h(\alpha||\beta)$. However, the authentication message does not include the session key information. As a result, user cannot verify that session key is the correct session key which made by server. Therefore, the explicit session key confirmation property is not achieved in Ahmed et al.'s scheme.

3.3 Offline Identity Guessing Attack (User Anonymity)

Offline identity guessing attack is a threat in which an outsider eavesdrops the message and obtains smartcard's container to infer the identity of legal users. If an outsider attacker U_a steals the smart card and obtains parameters, $\langle A_u, B_u, b, h(.) \rangle$ and eavesdrops login request message $< EID_i >$. Then U_a can easily do offline identity guessing attack by following step.

Step 1. The attacker selects one of identity of nominee ID_i^*, calculate $h(ID_i^*||b)$ using hash function and b which contained in also smart-card.

Step 2. If 1)'s result is equal to $EID_i = h(ID_i||b)$, the attacker infers that ID_i^* is user U_i's identity ID_i.

Step 3. Otherwise, attacker selects another identity nominee and performs same steps, until he/she finds identity.

3.4 Inconvenience and User Verification Problem

In Ahmed et al.'s scheme, server verifies if user is already registered in registration phase by checking $SID_i = h(EID_i||x)$'s value. However, this verification method can cause duplicate registrations. Even if the same user attempts to re-register with the same identity, the value of b chosen randomly. Eventually, the value of the $EID_i = h(ID_i||b)$ is different even though same user. That is, it is meaningless to verify for already registered user by checking SID_i's value.

4 Conclusion

In 2016, Ahmed et al. proposed a new type of lightweight user authentication scheme that uses user's biometric information. Ahmed opinioned his scheme is resistance to famous attacks such as identity guessing attack, replay attack, insider attack, and provide user anonymity. Nevertheless, Ahmed et al.'s scheme is still unsecure and unstable. We showed to this paper, Ahmed et al.'s scheme cannot provide user anonymity and is risk of biometric recognition error due to hash function. Furthermore, user verification problem can be occurred because of inconvenience registration phase.

To conclude, our future research is to proposed more secure user authentication which preserving anonymity scheme, also it will be able to resist these threats.

Acknowledgments. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2010-0020210).

References

1. Lamport, L.: Password authentication with insecure communication. Commun. ACM **24**(11), 770–772 (1981)
2. Hwang, M.-S., Li, L.-H.: A new remote user authentication scheme using smart cards. IEEE Trans. Consum. Electron. **46**(1), 28–30 (2000)
3. Liao, Y.-P., Wang, S.-S.: A secure dynamic ID based remote user authentication scheme for multi-server environment. Comput. Stand. Interfaces **31**(1), 24–29 (2009)
4. Kang, D., et al.: Efficient and robust user authentication scheme that achieve user anonymity with a Markov chain. Secur. Commun. Netw. (2016)
5. Li, C.-T., Hwang, M.-S.: An efficient biometrics-based remote user authentication scheme using smart cards. J. Netw. Comput. Appl. **33**(1), 1–5 (2010)
6. Li, X., et al.: Cryptanalysis and improvement of a biometrics-based remote user authentication scheme using smart cards. J. Netw. Comput. Appl. **34**(1), 73–79 (2011)
7. Jung, J., et al.: An improved and secure anonymous biometric-based user authentication with key agreement scheme for the integrated EPR information system. PloS One **12**(1), e0169414 (2017)
8. Lee, H., et al.: Forward anonymity-preserving secure remote authentication scheme. KSII Trans. Internet Inf. Syst. **10**(3) (2016)
9. Chien, H.-Y., Chen, C.-H.: A remote authentication scheme preserving user anonymity. In: 19th International Conference on Advanced Information Networking and Applications, AINA 2005, vol. 2. IEEE (2005)
10. Yang, G., et al.: Two-factor mutual authentication based on smart cards and passwords. J. Comput. Syst. Sci. **74**(7), 1160–1172 (2008)
11. Kim, J., et al.: Security analysis and improvements of two-factor mutual authentication with key agreement in wireless sensor networks. Sensors **14**(4), 6443–6462 (2014)
12. Al Sahlani, A.Y.F., Lu, S.: Lightweight communication overhead authentication scheme using smart card. Indonesian J. Electr. Eng. Comput. Sci. **1**(3), 597–606 (2016)
13. Kocher, P., et al.: Introduction to differential power analysis. J. Cryptographic Eng. **1**(1), 5–27 (2011)
14. Amin, R., Biswas, G.P.: A secure light weight scheme for user authentication and key agreement in multi-gateway based wireless sensor networks. Ad Hoc Netw. **36**, 58–80 (2016)
15. Dodis, Y., Reyzin, L., Smith, A.: Fuzzy extractors: how to generate strong keys from biometrics and other noisy data. In: International Conference on the Theory and Applications of Cryptographic Techniques. Springer, Heidelberg (2004)
16. Blake-Wilson, S., Johnson, D., Menezes, A.: Key agreement protocols and their security analysis. In: IMA International Conference on Cryptography and Coding. Springer, Heidelberg (1997)

Cybersecurity Interface and Metrics

Modeling, Analysis and Control of Personal Data to Ensure Data Privacy – A Use Case Driven Approach

Christian Zinke[1(✉)], Jürgen Anke[2], Kyrill Meyer[1], and Johannes Schmidt[1]

[1] InfAI e.V, Hainstraße 11, 04109 Leipzig, Germany
{zinke,meyer,schmidt}@infai.org
[2] Hochschule für Telekommunikation Leipzig,
Gustav-Freytag-Str. 43-45, 04277 Leipzig, Germany
anke@hft-leipzig.de

Abstract. The compliance with data protection and privacy regulations such as the European General Data Protection Regulation (GDRP) is a challenging task for companies with complex IT landscapes. Current approaches lack of a technical integration with enterprise software systems and therefore require considerable manual effort to keep permissions and retention of data in line with data protection and privacy requirements. We propose an integrated information model to link data privacy requirements with software systems, modules and data to address this problem with the help of Information Lifecycle Management (ILM) functionality. The approach is illustrated with a use case of the compliant deletion of employee data upon fulfillment of the stated purpose.

Keywords: Data privacy · Information lifecycle management · Privacy model

1 Introduction

Data privacy is a term to describe regulatory, organizational and technical means to control and restrict the collection and dissemination of data by individuals and organizations. The general aim is to enable the self-determination of individuals with regard to their personal data, while at the same time allowing data-driven business processes and models in the information-driven economy. Data privacy breaches result in loss of trust[1] [1] and impose significant legal risks and financial costs for companies [2].

To face these risks and costs, data processing companies have to ensure normative conformity and technical feasibility (e.g. data security, protection as well as data privacy) at the same time. This is challenging as providing evidence to prove data privacy compliance for business applications is time-consuming, costly and rarely processed automatically. It is therefore not surprising that data privacy issues in IT demand processes are widely neglected and mostly processed as late as possible. With IT systems and data growing in complexity, the risks and costs of data privacy issues and possible breaches are rising. Thus, an integrated holistic approach is needed, combining data privacy issues and demand processes for given IT infrastructures. In this paper we

[1] Referring to Kearney's "crisis in trust" [1].

D. Nicholson (ed.), *Advances in Human Factors in Cybersecurity*,
Advances in Intelligent Systems and Computing 593, DOI 10.1007/978-3-319-60585-2_10

present such an integrated approach to support modeling, analysis and controlling of data privacy issues, thus providing compliance and ensuring technical feasibility.

2 Methodology

The methodology is use case driven and conducted within the design research approach. Therefore, a concrete problem will be described with the help of a suitable use case [3, 4]. The use case deals with a standard procedure of strict adherence to the retention period of personal data. Based on this, we analyze the requirements for a data privacy management approach, develop necessary design objects and a model, which is needed to provide a supporting tool addressing all relevant aspects of data privacy in that regard. On the technical side, the presented model will be linked to the information lifecycle management (ILM) approach in enterprise information systems. In our use case, we will demonstrate the linking to a SAP ERP system to ensure that personal data is deleted after a specific time period based on criteria modeled before. With that, the paper will demonstrate the potentials and limitations of the ILM concept for data privacy.

3 Use-Case

The use case is a simplified and generic case. Every company collects, processes and stores personal data from employees, customers and suppliers. Various laws and individual contracts regulate when and to what extends the usage of personal data is permitted. Important parts of these conditions or regulations are the purpose of data processing and the final retention period of the data in order to guarantee the compliance to the “data minimization” and “need-to-know” principles. After data collection, the company is able to access this data. If the purpose of data processing is fulfilled (e.g. an employee leaves the company), any further access to the respective data records has to be restricted, and the retention period begins. At the end of the retention period, the company has to ensure the final deletion of the data (Fig. 1).

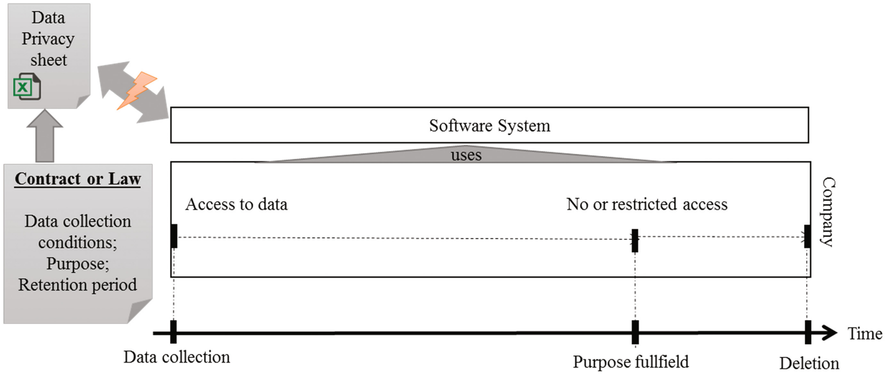


Fig. 1. Timeline of data handling without integrated privacy management

The presented case of a standard procedure of strict adherence to the retention period of personal data appears to be a challenging task for most companies. One of the reasons for this is that personal data privacy is often documented in general purpose office tools (e.g. as Word or Excel document), without any technical linkage to the real processed data, data categories or software systems. While privacy requirements can be handled manually is very small systems, the privacy of personal data cannot be manually controlled (analyzed or regulated) for a large and distributed IT. To alleviate this, data privacy requirements need to be described more formally in a model. The modeling approach needs to be linked both to the IT systems and to other existing mechanisms for handling data, such as the ILM approach. The paper aims to bridge the gap with the help of a privacy information model, which links data privacy rules and requirements to existing data sets and IT systems. It therefore allows analyzing, controlling and realizing the data privacy requirements for deletion.

4 Privacy Model

Based on the work of Anke et al. [5] and the presented use case, this section introduces at first the privacy design object. Afterwards, we review one existing approach that addresses the support of the deletion process – the ILM approach. The last section will introduce a linked model to ensure the compliance of data privacy requirements –in our case the retention period.

4.1 Privacy Design Objects

Within the uses case some basic objects which are needed to describe data privacy are indirectly introduced. In the following, we present these concepts which are part of the information model depicted in Fig. 2.

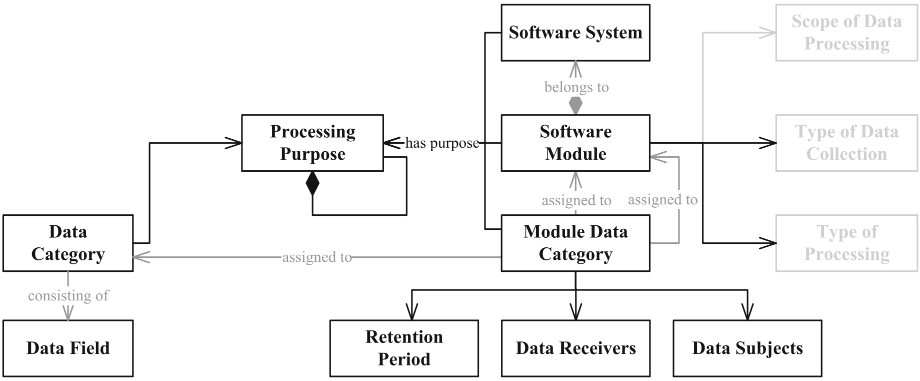


Fig. 2. Information model of software applications and data privacy concepts, see [5]

Personal data is stored in *Data Fields*, e.g. Name, Address, which can be grouped into *Data Categories*, e.g. Basic Data, Payroll Data etc. Modern integrated *Software Systems* such as SAP consist of multiple *Software Modules*, e.g. Organizational Management, Payroll Processing, Personnel Administration. Each of these modules requires certain data categories to fulfill its task. Multiple modules can use the same data category and vice versa.

As the data collection, processing and storage is facilitated by these modules, they determine the scope and type of data processing as well as the type of data collection, e.g. automated collection. As stated above, from a legal point of view it is required that processing of personal data is covered by a legitimate purpose. *Processing Purposes* can be hierarchical and also be assigned to the complete software system, a module or the module data. This allows a fine-grained modeling and control of relations between data, purposes and software applications.

Finally, the module data can also be assigned to *Data Subjects* (persons, the data is about), *Data Receivers* (persons that use the personal data to fulfill their tasks) and the associated retention period. It is important to note that data of a particular category can have different retention periods depending on the applicable combination of module and purpose. This leads to complex data management tasks in integrated systems that use a unified database for all modules.

4.2 Information Lifecycle Management

A popular definition for Information Lifecycle Management (ILM) is provided by SNIA Data Management Forum: “Information Lifecycle Management (ILM) is comprised of policies, processes, practices, and tools used to align the business value of information with the most appropriate and cost effective IT infrastructure from the time information is conceived through its final disposition” [6]. It is a strategic policy-based approach for IT systems in order to address the difficulties of enterprise data mobility and storage [7]. The ILM approach is used to organize data and storage classification as well as counting data access frequencies. Typically, a policy engine is applied to perform defined data flows [7]. Further, Short develops two ways to determine the information lifecycle. First, it can be defined by data access, from active data to data in archive. Second, it can be defined by data activities, such as collection, access and distribution, transformation, classification and archive [7]. Other authors like Tallon and Scannell [8] emphasize the stages of capturing, application and decline of data [8]. Thus, ILM focuses on policy-based data management, incl. storage, and is mostly about migration [9]. Retention and deletion are only minor aspects of ILM and will be applied through policies. The relation of deletion/retention of data and ILM can be conceptually modeled as shown in Fig. 3. Basically, ILM refers to data. Data will have a current status (e.g. locked), which is defined by policies implemented by rules. These rules also set the lifecycle stage of data, depending on different attributes, or meta-data, such as the usage frequency. The data is stored in a storage system, depending on lifecycle stage and data status. A(n information) lifecycle consists of lifecycle stages linked to the data.

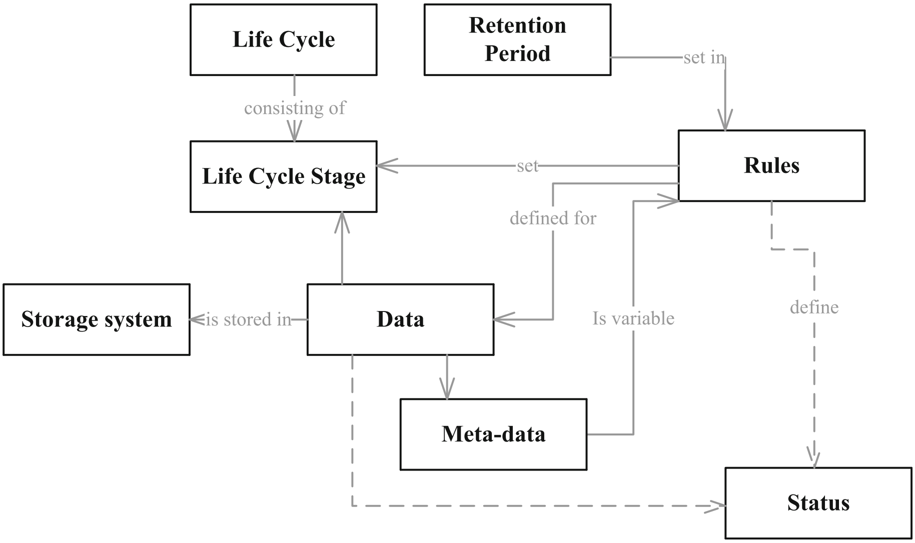


Fig. 3. Conceptual model of ILM

One of the most known implementation of ILM is the SAP NetWeaver Information Lifecycle Management. It contains an archive function and allows the definition of ILM rules, snapshots, the lockage of data as well as the deletion of data [10, 11]. Technically, the ILM controls the concrete storage and retention periods of data of a concrete IT system or module.

4.3 Linked Model

While ILM provides a lot of functions, from a data privacy perspective it only refers to retention periods. In order to ensure data privacy, all data privacy issues need to be modeled centrally for all software systems the company is responsible for. Thus, ILM helps to implement the specific data privacy requirement of consistent compliance of retention periods. Thus, it is not surprising, that the central linking concept for both introduced conceptual models is the retention period, which will be set by rules of the ILM. Other data such as modeled software systems and data categories help to identify the specific system and it's used ILM, which is conceptually shown in Fig. 4.

The application of the integrated model to the use case mentioned above is shown in Fig. 5. At first, the life cycle of data handling has to be extended data privacy model, where the framework for compliant data handling needs is defined. These modeled requirements are stored in the privacy model which is machine-readable. Based on that it is therefore possible to analyze and control the rules of the ILM which in turn manages data in specific software systems to keep them compliant with the data privacy requirements.

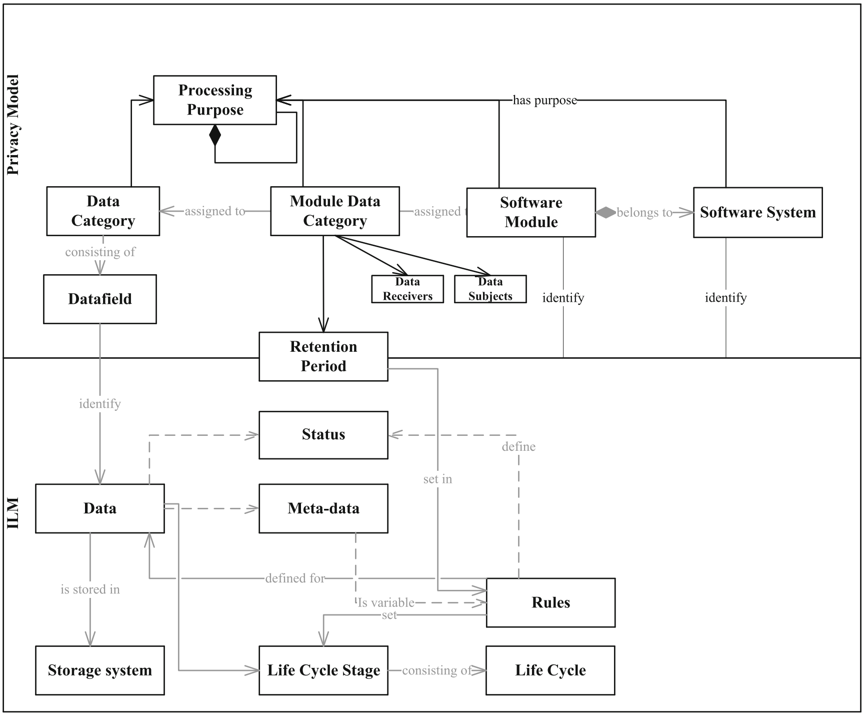


Fig. 4. Integrated conceptual model – ILM and privacy model

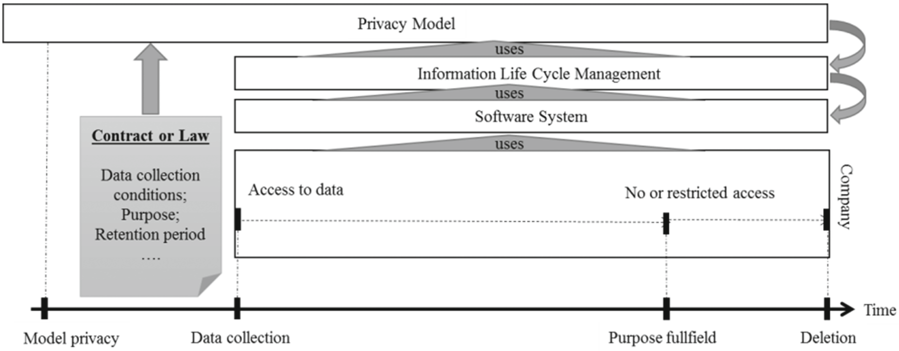
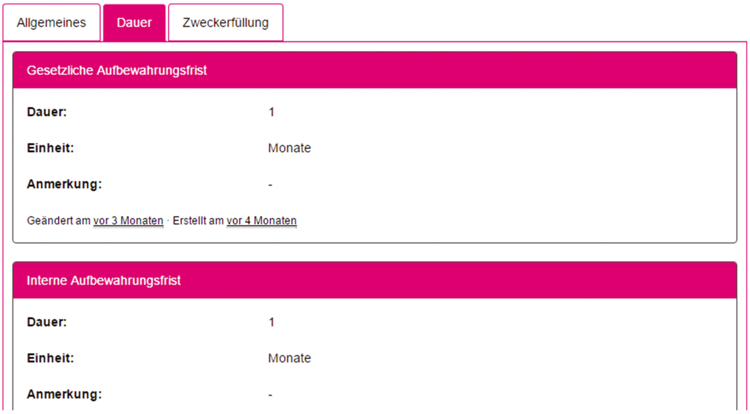


Fig. 5. Timeline of data handling with integrated privacy

To validate our model, we implemented the presented use-case into a SAP HCM as a proof of concept. We decided to demonstrate it for the Infotype “Address” (PA0006). According to the procedure introduced in Fig. 5, we firstly modeled the SAP system and the data privacy requirements (see Fig. 6).



Gesetzliche Aufbewahrungsfrist

Dauer:	1
Einheit:	Monate
Anmerkung:	-

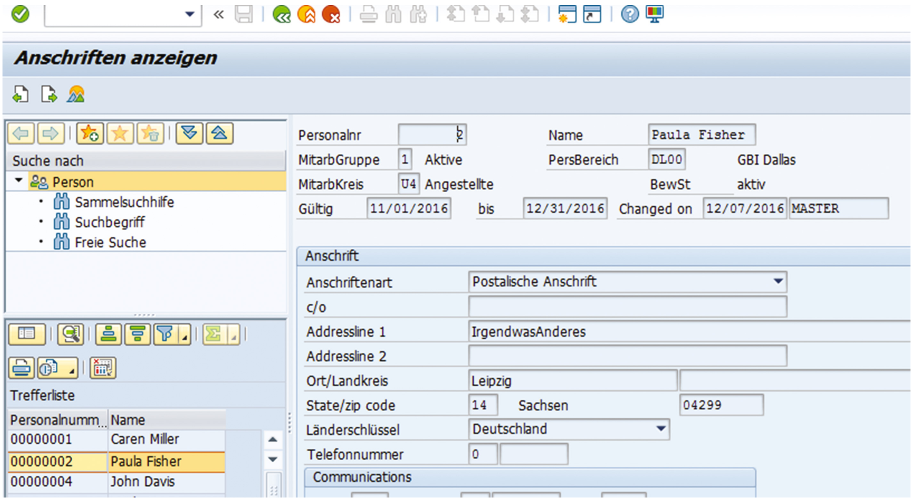
Geändert am vor 3 Monaten · Erstellt am vor 4 Monaten

Interne Aufbewahrungsfrist

Dauer:	1
Einheit:	Monate
Anmerkung:	-

Fig. 6. Screenshot of privacy modeling tool developed by InfAI

In the presented case, we set the retention period on one month. The used privacy tool developed by the InfAI is based on the introduced model (Sect. 4.1), as presented in Anke et al. [5]. Secondly, we added sample address data to the SAP HCM system (see Fig. 7).



Anschriften anzeigen

Personalnr	2	Name	Paula Fisher
MitarbGruppe	1 Aktive	PersBereich	DL00 GBI Dallas
MitarbKreis	U4 Angestellte	BewSt	aktiv
Gültig	11/01/2016 bis 12/31/2016	Changed on	12/07/2016 MASTER

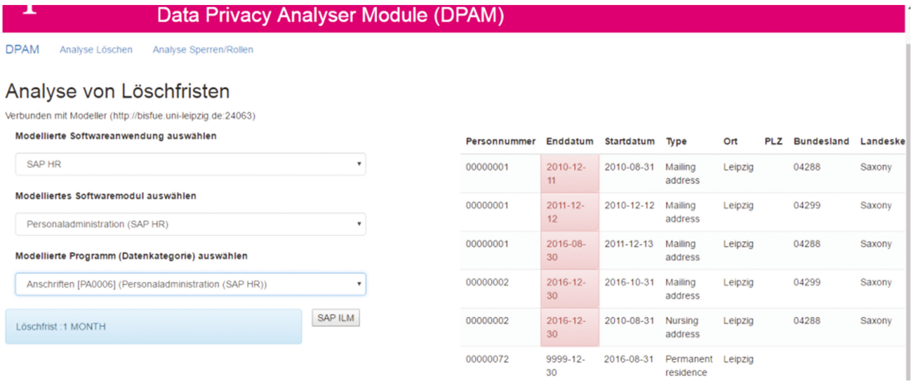
Anschrift

Anschriftenart	Postalische Anschrift
c/o	
Addressline 1	IrgendwasAnderes
Addressline 2	
Ort/Landkreis	Leipzig
State/zip code	14 Sachsen 04299
Länderschlüssel	Deutschland
Telefonnummer	0

Personalnumm...	Name
00000001	Caren Miller
00000002	Paula Fisher
00000004	John Davis

Fig. 7. Screenshot of on SAP address example

Thirdly, we developed a data privacy analyzer module to analyze the SAP system with the model. The module uses the SAP BAPI functions to read data sets from SAP HCM. As can be seen in Fig. 8, we choose the infotype to be analyzed from the model. Afterwards, data from the SAP system were analyzed to highlight data records that are not conform to the set retention period.



Data Privacy Analyser Module (DPAM)

Analyse von Löschfristen

Verbunden mit Modeller (http://bisfue.uni-leipzig.de:24063)

Modellierte Softwareanwendung auswählen: SAP HR

Modelliertes Softwaremodul auswählen: Personaladministration (SAP HR)

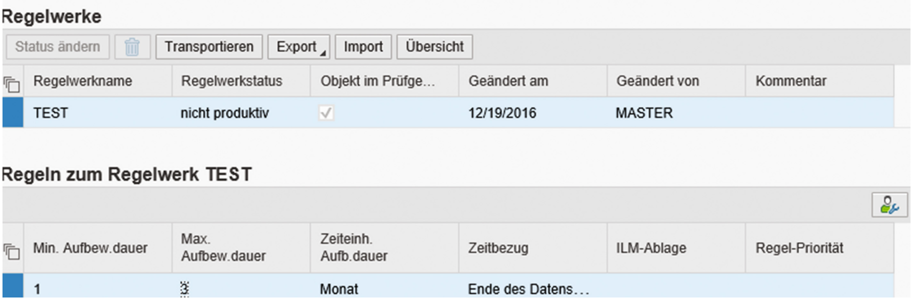
Modellierte Programm (Datenkategorie) auswählen: Anschriften [PA0006] (Personaladministration (SAP HR))

Löschfrist :1 MONTH

Personnummer	Enddatum	Startdatum	Type	Ort	PLZ	Bundesland	Landeske
00000001	2010-12-11	2010-08-31	Mailing address	Leipzig		04288	Saxony
00000001	2011-12-12	2010-12-12	Mailing address	Leipzig		04299	Saxony
00000001	2016-08-30	2011-12-13	Mailing address	Leipzig		04288	Saxony
00000002	2016-12-30	2016-10-31	Mailing address	Leipzig		04299	Saxony
00000002	2016-12-30	2010-08-31	Nursing address	Leipzig		04288	Saxony
00000072	9999-12-30	2016-08-31	Permanent residence	Leipzig			

Fig. 8. Screenshot of on data privacy analyzer

At this stage, it can be seen that the tested infotype PA0006 is not compliant with the privacy model. In order to rectify this, a complex procedure to change or imply the new rule into SAP is provided by the data privacy analyzer module. At least the new ILM rule has to be set productive within the SAP ILM toolset (Fig. 9.)



Regelwerke

Regelwerkname	Regelwerkstatus	Objekt im Prüfge...	Geändert am	Geändert von	Kommentar
TEST	nicht produktiv	✓	12/19/2016	MASTER	

Regeln zum Regelwerk TEST

Min. Aufbew.dauer	Max. Aufbew.dauer	Zeiteinh. Aufb.dauer	Zeitbezug	ILM-Ablage	Regel-Priorität
1	3	Monat	Ende des Datens...		

Fig. 9. Screenshot of SAP ILM

5 Summary and Future Work

In this paper, we have presented a model to link data privacy requirements with software systems and ILM to ensure data privacy compliance. While our work is still in an early phase, it shows a promising avenue to achieve a closed loop for automated enforcement of data privacy requirements in complex IT landscapes, even in times of distributed software systems, cloud computing and big data.

For this approach to work, we require the existence of an ILM system. ILM systems are designed to support the required functions. However, even if SAP supports ILM, many other enterprise software systems lack of an ILM functionality or active interfaces for managing data deletion from the outside. Nevertheless, the conceptual linkage helps to better understand the relation between data privacy requirement of retention periods and software systems with ILM.

The demonstrated data privacy requirement of retention period is only one example for implementing compliant data privacy requirements. While a lot of work in data protection and security, e.g. [12–14], anonymity, e.g. [15–17] or identification and authentication, e.g. [18–20] has been done, data privacy from legal viewpoint of software systems is still lacking in sufficient integrated mechanisms to model, analyze and control complex IT infrastructures, business processes and cross-company information flows.

References

1. Kearney, A.T.: rethinking personal data: a new lens for strengthening trust. In: World Economic Forum (2014). Accessed Nov 2014
2. Loomans, D., Matz, M., Wiedemann, M.: Praxisleitfaden zur Implementierung eines Datenschutzmanagementsystems: Ein risikobasierter Ansatz für alle Unternehmensgrößen. Springer, Heidelberg (2014)
3. Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S.: A design science research methodology for information systems research. J. Manage. Inf. Syst. (2007). doi:10.2753/MIS0742-1222240302
4. March, S.T., Smith, G.F.: Design and natural science research on information technology. Decis. Support Syst. (1995). doi:10.1016/0167-9236(94)00041-2
5. Anke, J., Berning, W., Schmidt, J., Zinke, C.: IT-gestützte Methodik zum Management von Datenschutzanforderungen. HMD (2017). doi:10.1365/s40702-016-0283-0
6. SNIA Data Management Forum: ILM Definition and Scope. An ILM Framework. http://www.snia-dmf.org/library/DMF-ILM-Vision2.4.pdf 20 July (2004). Accessed 2 Mar 2017
7. Short, J.E.: Information Lifecycle Management Concepts, Practices, and Value (2007)
8. Tallon, P.P., Scannell, R.: Information life cycle management. Commun. ACM **50**(11), 65–69 (2007)
9. Beigi, M., Devarakonda, M., Jain, R., Kaplan, M., Pease, D., Rubas, J., Sharma, U., Verma, A.: Policy-based information lifecycle management in a large-scale file system. In: Sixth IEEE International Workshop on Policies for Distributed Systems and Networks (POLICY 2005), Stockholm, Sweden, 6–8 June 2005, pp. 139–148 (2005). doi:10.1109/POLICY.2005.26
10. SAP: ILM-Erweiterungen in der Datenarchivierung. https://help.sap.com/saphelp_crm70/helpdata/de/c2/23e47063a74341a7641993dd03df94/content.htm. Accessed 2 Mar 2017
11. SAP: SAP NetWeaver Information Lifecycle Management. https://help.sap.com/saphelp_nw70ehp1/helpdata/de/7f/e188e04fdd462e8ec330bb80efc389/frameset.htm. Accessed 2 Mar 2017
12. Haynes, D., Bawden, D., Robinson, L.: A regulatory model for personal data on social networking services in the UK. Int. J. Inf. Manage. **36**(6), 872–882 (2016)
13. Gable, J.: Principles for protecting information privacy. Inf. Manage. J. **48**(5), 38–42 (2014)
14. Al-Jaberi, M.F., Zainal, A.: Data integrity and privacy model in cloud computing. In: Biometrics and Security Technologies (2014)
15. Loukides, G., Gkoulalas-Divanis, A., Shao, J.: Efficient and flexible anonymization of transaction data. Knowl. Inf. Syst. (2013). doi:10.1007/s10115-012-0544-3
16. Damiani, M.L.: Location privacy models in mobile applications: conceptual view and research directions. Geoinformatica (2014). doi:10.1007/s10707-014-0205-7

17. Domingo-Ferrer, J., Sánchez, D.: Database anonymization: privacy models, data utility, and microaggregation-based inter-model connections. Synth. Lect. Inf. Secur. Priv. Trust **8**(1), 1–136 (2016)
18. Kardas, S., Celik, S., Bingol, M.A., Kiraz, M.S., Demirci, H., Levi, A.: k-strong privacy for radio frequency identification authentication protocols based on physically unclonable functions. Wireless Commun. Mob. Comput. (2015). doi:10.1002/wcm.2482
19. Alavi, S.M., Baghery, K., Abdolmaleki, B., Aref, M.R.: Traceability analysis of recent RFID authentication protocols. Wireless Pers. Commun. (2015). doi:10.1007/s11277-015-2469-0
20. Hermans, J., Peeters, R., Preneel, B.: Proper RFID privacy: model and protocols. IEEE Trans. Mob. Comput. (2014). doi:10.1109/TMC.2014.2314127

Exploring the Discoverability of Personal Data Used for Authentication

Kirsten E. Richards(✉) and Anthony F. Norcio

University of Maryland, 1000 Hilltop Circle Baltimore, Baltimore, MD 21250, USA
kirsten5@umbc.edu

Abstract. The antinomic proposition of usable system authentication, an easily remembered and usable scheme for the proper user which is simultaneously unknown and unusable to any other entity, historically proves to be an elusive goal. While alternative propositions for authentication protocols are numerous, lacking in literature is foundational work directly relating potential authenticators with the discoverability of personal data online. This work presents a brief but foundational analysis of authentication and the connection between the authentication protocols and the inevitability of the introduction of personal data to the protocol to improve usability, particularly with regard to password based authentication. We investigate the discoverability, particularly whether another human, unacquainted with a specific individual, is able to purposefully find particular personal data commonly used in authentication protocols. In the study, five participants were asked to search for specific personal data regarding a sixth participant. Analysis of the results reveals consistent patterns in the personal data discovered by users. Analysis of discovered data lays a foundation for the improvement of current authentication systems as well as providing a proof of concept for the methodology and application recommendations to guide the creation of password alternatives with a goal towards the creation of usable, secure authentication systems.

Keywords: User behavior · Password authentication · Personal data availability · Secondary authentication

1 Introduction

Authentication protocols frequently suffer from a paradoxical requirement: the creation of a system both usable and unusable in the proper context, both unforgettable and enigmatic. This challenge is uniquely suited to the Human Computer Interaction (HCI) community as it is through interaction and interface design that system compromises occur. In essence, we believe that all security problems are user interface and interaction problems, and thus can and should be positively and proactively addressed by the HCI community.

Passwords, the most common form of authentication, are notoriously difficult for users to maintain securely. Humans often attempt to circumvent the cognitive difficulties of passwords though the inclusion of personal data in their password to improve

D. Nicholson (ed.), *Advances in Human Factors in Cybersecurity*,
Advances in Intelligent Systems and Computing 593, DOI 10.1007/978-3-319-60585-2_11

memorability [1, 2]. Other forms of authentication suffer from the potential for compromise through alternative forms of authentication. From a HCI perspective, the use of personal data in password creation is almost certainly inevitable. Passwords are also likely to continue in use [3].

This work explores the personal data used in authentication and provides a case study assessment of the discoverability of personal data used in password authentication and secondary authentication protocols. Understanding empirically which personal data are more and less vulnerable to search is significant to advising users in creating stronger passwords and designing more secure authenticators and secondary authentication questions.

2 Background

Authentication and secondary authentication protocols rely on the use of personal data either by design or indirectly as a result of the interaction between the design and human cognition [4–6]. The various authenticators rely on personal data in different ways, a personal data element is often present either directly or indirectly. First, we explore personal data in password authentication, followed by other forms of authentication.

2.1 Password Authentication

Research consistently reveals that passwords, the most common form of authentication in use, are not easily managed from a human cognitive perspective [1, 7–9]. Given the cognitive difficulty of designing and maintaining secure passwords, it seems nearly inevitable that knowledge based will introduce personal data, well known to the user, to mitigate the difficulty of remembering a myriad of passwords. This difficulty has been characterized as instructing users to, “Pick something you cannot remember, and do not write it down” [10]. The challenge of recruiting users to proactively contribute to security through good password practices is observable in research [11, 12] and seems almost inevitable given the poor usability of passwords from a cognitive perspective. To improve the cognitive usability of passwords, users incorporate memorable data such as a birthdate or personally significant name [8]. The use of personal data for authentication is not limited to passwords, but also affects forms of authentication, both those designed to improve security as well as protocols designed to improve usability.

2.2 Additional Authenticators

Other authenticators rely on personal data as well. Possession authentication relies upon the user’s custody of some type of token – be it a phone, software, or another device. While in some ways superior to passwords in regard to usability, possession authenticators also involve a wide variety of susceptibilities, not the least of which is careless misplacement of a physical token [7]. Graphical passwords, promising from a usability perspective, have not been studied extensively from a personalized attack perspective, but are vulnerable to attacks where users can describe the authenticators [13]. Existence authentication, based on personal qualities and characteristics, such as biometrics might, from a cursory examination, seem impervious to personal data attacks. However, it is important to note that varied forms of data used in biometric style authenticators such as fingerprints or facial characteristics used in facial recognition are not seen as inherently private by users and are therefore less likely to be purposefully protected [7, 14]. Finally, attacks on social forms of authentication can be perpetrated by gaining knowledge of the target's social network. Like biometrics, social relationships are not generally considered "secret" and are thus vulnerable to attack and are unlikely to be protected and may be guessed through available information [15].

Even if the primary authentication protocol is impervious to a personal data attack, the system may be vulnerable to personal data attacks. Authenticators may be compromised by social engineering attacks which are often reliant on knowledge of personal [16–18]. Secondary authenticators, which are often designed around personal data knowledge, also provide an opportunity to attach systems regardless of the reliability of the primary authenticator on personal data [19, 20]. While secondary authenticators have been explored in research with regard personal knowledge of acquaintances [5], they have not to our knowledge, been addressed with regard to the discoverability of personal data by strangers.

2.3 Personal Data Availability

The combination of the use of personal data for primary and secondary authentication becomes potentially problematic when considering the immense volume and diversity of personal data available online [21]. Some connections have been made between known personal information and secondary authentication [5]. The direct connection of authenticators to personal data available on the web remains under developed in research.

2.4 Motivation

Many forms of authentication suffer from vulnerability to personal data based attacks. Researchers are challenged to improve the usability and security of passwords with an acknowledgement of the probability of persistence [3]. The study of personal data in authentication has far reaching implications to meet the research challenge of both improving passwords and improving design of alternative authenticators. We describe a methodology to explore and understand personal data availability, particularly personal data online. Furthermore, this research seeks to understand the vulnerability of online data to correct identification by human information seekers. The existence of personal data on the web is undeniable, however, the vulnerability of this data to discovery by human users addresses both the component of interaction in posting and proliferating data as well as understanding the discovery or search process for data.

The resulting data may be used to inform improvements in passwords themselves, an under-developed area in research, and also as a foundation for understanding how data availability should impact the development of novel authentication approaches. By illuminating the relationship between personal data's discoverability to human users online and personal data used in authentication, this work addresses a significant knowledge gap both within the HCI community with extended applications in a wide variety of information systems fields.

3 Methodology

The primary purpose of the study described is a proof of concept to demonstrate the pertinence of the research model and to provide preliminary insight into data availability online. Two different groups of participants were recruited to a multiphase study. An information source participant (source participant) served as a "target" for data retrieval and provided basic demographic data for identification as well as a photograph for use by seeker participants. Seeker participants attempted to identify specific personal data deemed of interest because of the use of the data in authentication. Seeker participants were asked to evaluate their familiarity with online searching with a Likert scale. Seeker participants were instructed to attempt to identify specific information regarding the source participant and report their findings via a survey which was administered in both paper and web form. The survey also asked seeker participants to report their search time for a particular data point, the location of discovered data and their perception of the relative difficulty of discovering the data with a Likert scale. Following collection of the personal data from seeker participants, the source participant verified the accuracy of submitted results.

The research design is intended to reveal what data were most likely to be found, the accuracy of data discovered, where the data were located by seekers, and the relative difficulty of accurately finding information as understood through the time spent searching, the perception of difficulty was reported in a Likert scale and the accuracy of identified personal data.

The methodology design allows for examinations of two aspects of human computer interaction. First, the human factor of providing the personal data online by a variety of actors such as the source participants themselves, family and friends, government entities, businesses and employers as a wide variety of entities may be responsible for personal data proliferation on social media and in other venues [22]. Second, the ease and reliability of individuals attempting to locate personal data. In the experiment described, the seeker participants' counterparts outside of the model might include employers seeking information about employees or prospective employees, government agencies performing background checks, and, of course, nefarious actors attempting to use personal data to gain unauthorized access to systems.

Five seeker participants were recruited for the study to attempt to discover information about one source participant. Three seeker participants were provided with the source participant's photograph, name and City and State location data while two were provided with only the source participant's name and location. This design choice was intended to reveal the significance of a photograph in identifying a source participant for future research.

4 Analysis

The study results are revealing to the research questions in several areas. First, we generally discuss the personal data and the accuracy of personal data explored by seeker participants. The data are then analyzed with attention to the significance of the photograph and name and location provided to seeker participants compared to only the provision of the name. Finally, the influence of the source participant's familiarity with the search process is discussed.

4.1 Data Discovery

Five data points were explored by seeker participants. Table 1 describes the general accuracy of data discovered by seeker participants. The personal data points explored by seekers are listed, followed by the number of unique guesses summarized across the searchers, and the accuracy, or percentages of correct answers compared to total guesses.

Table 1. Accuracy of personal data obtained

Personal data	Guesses	Accuracy
Mother's maiden name	3	0%
Nickname	2	20%
Children's names	5	20%
Pet's names	0	0%
Middle name	2	50%

There is notable consistency of the actual difficulty of accurate discovery across the searchers. Searchers were much more likely to discover a middle name. The significance of this finding is discussed further in Results.

The data were also analyzed for successful results which take into account the accuracy of the seeker participants themselves. This is particularly important for data points where seekers could supply more than one potential answer, such as children's names. Table 2 reveals the success of the seeker participants. The percentage of seeker participants answering the question correctly is reported. Finally, the mean perceived difficulty refers to the perception of difficulty on a six point Likert scale ranging from 1 – Impossible to 6 – Very Easy.

Table 2. Accuracy of seeker and perceived difficulty

Personal data	Accurate seekers	Difficulty scale
Mother's maiden name	0%	3.40
Nickname	20%	1.40
Children's names	60%	2.20
Pet's names	0%	1.00
Middle name	60%	2.25

The perceptions of perceived difficulty varied widely varied among participants, however, perception of ease did not always reflect an accurate answer. For examples, although Mother's Maiden Name is perceived as easily attainable, it was not answered correctly. The perception of difficult reflects the actual difficulty as revealed by the percentage of correct answers in other data points, such as Pet's Names, where the question is both perceived as difficult and is actually difficult based on the accuracy of guesses.

The most successfully identified data were also the data most closely and directly linked to the individual. For example, "middle name", correctly identified by 60% of data seekers, was identified directly from Whitepages.com or PeopleSmart.com by all successful guesses. Children's names were correctly identified via Facebook. Data less directly connect to the source individual, such as mother's maiden name, was not identified correctly in any instance.

4.2 Familiarity

Four of the five seeker participants rated themselves as "Extremely familiar" with online searches for information. One participant considered themselves "slightly familiar" with online searches for information. The latter participant was unable to identify any information correctly regarding the source participant. Of the "Extremely familiar" seeker participants, one in four did not report any data correctly. The remaining three participants provided all of the correct data to the study.

4.3 Sources

All successful answers were located from three sources: Facebook and Whitepages.com and PeopleSmart.com. While participants reported seeking answers from a variety of sources including Spokeo, Instagram, or using search engines to perform web searches, these additional sources did not result in a successful guess.

4.4 Photograph

The presence of a photograph was a factor in the correct and accurate identification of personal data compared to individuals only supplied with a name and location. Three individuals were supplied with a photograph. Of the three, two identified the correct source at least several times as revealed by the correct identification and were most likely to identify correct information on Facebook, which is rich in photographic media. The third source participant provided with a picture, did not correctly identify any data, spent less than three seconds on the survey page with the photograph and relied primarily on sources which were not photo-rich such as whitepages.com and newspaper articles such as obituaries.

Two participants were not provided with a photograph. Of these two, one seeker participant was not able to identify any data correctly. The second participant identified some data accurately and relied primarily on business sources such as Whitepages and PeopleSource but did not report use of social media to obtain information.

5 Results

Human factors are well understood to be a significant factor in maintaining secure authentication systems. However, many actors contribute to the availability of personal data online, which is directly involved in maintaining secure authentication systems. Influences in the discoverability of personal data that can contribute to personal security risk are varied.

The results of this study suggest several important contributions to the continued design of secure and usable authentication systems. The first is the consideration of the availability of data. Certain data were much more difficult for seekers to correctly identify compared to other data points. When the use of personal data in authentication may not be immediately rectified in authentication and secondary authentication protocols, the personal data used can and should be intentionally obscure to search. Furthermore, continued attention should be directed to sources of information, tied to real identity and relationships, about individuals.

The availability of data to searcher can also influence the discoverability of other personal data points. In our study, this is illustrated in by the presence of a photograph of the source participant. All seeker participants were provided with the same name and location information regarding the source participant, however, the individuals receiving the photograph as well were more likely to identify data correctly. As authentication protocols develop, designers will need to consider the vulnerability of their specific authentication protocol to discoverable data, particularly when considering data which is not considered private by users.

Finally, the skills of seekers themselves likely influence their success as revealed by the self-reported level of familiarity with web searches for data, however, a more diverse population is required to explore this finding in more depth.

Security breaches due to problems with authentication protocols are a “weakest link” proposition [9]. Avoiding the use of easily discovered data in the implementation of standard authentication protocols as well as in the development of new protocols will positively impact the security of current systems as well as improving the security outlook of newly developed authentication protocols.

6 Future Research

Research currently underway examines personal data across a wider pool of source participants, representing a varied demographic in terms of age and gender to address human factors in personal data. A larger group of seeker participant will provide the opportunity for a broader examination of information seeking behaviors that contribute to understanding human factors in information retrieval as well as providing for statistical analysis of variables affecting the search.

More research is needed to examine other forms of authentication and evaluate the personal data used in their design or employed by users to improve memorability. More research is currently being conducted to discover which data points remain more obscure to a variety of searchers in addition to exploring additional data used in various forms of authentication. Alternative authenticators should be evaluated on their avoidance personal data or at the very least, strategic use of personal data. Furthermore, this research space affords an opportunity to examine information search with a novel application. Finally, technology support for web content and searches will change over time and thus research into personal data availability will be ongoing as necessitated by the development of information systems.

References

1. Vu, K.-P.L., et al.: Improving password security and memorability to protect personal and organizational information. Int. J. Hum. Comput. Stud. **65**(8), 744–757 (2007)
2. Adams, A., Sasse, M.: Users are not the enemy. Commun. ACM **49**(12), 41–46 (1999)
3. Herley, C., Van Oorschot, P.: A research agenda acknowledging the persistence of passwords. IEEE Secur. Priv. **10**(1), 28–36 (2012)
4. Furnell, S.: Authenticating ourselves: will we ever escape the password? Netw. Secur. **2005**(3), 8–13 (2005)
5. Schechter, S., Brush, A.J.B., Egelman, S.: Its no secret: measuring the reliability of authentication via ‘secret’ questions. In: Proceedings of the 2009 30th IEEE Symposium on Security and Privacy, pp. 375–390 (2009)
6. Duggan, G.B., Johnson, H., Grawemeyer, B.: Rational security: modelling everyday password use. Int. J. Hum. Comput. Stud. **70**(6), 415–431 (2012)
7. Bonneau, J., et al.: The quest to replace passwords: a framework for comparative evaluation of web authentication schemes. In: IEEE Symposium on Security and Privacy, pp. 553–567 (2012)
8. Brown, A.S., et al.: Generating and remembering passwords. Appl. Cogn. Psychol. **18**(6), 641–651 (2004)
9. Sasse, M., Brostoff, S., Weirich, D.: Transforming the ‘weakest link’ a human-computer interaction approach to usable and effective security. BT Technol. J. **19**(3), 122–131 (2001)
10. Bonneau, J., et al.: Passwords and the evolution of imperfect authentication. Commun. ACM **58**(7), 78–87 (2015)
11. Grawemeyer, B., Johnson, H.: Using and managing multiple passwords: a week to a view. Interact. Comput. **23**(3), 256–267 (2011)
12. Pavlou, P.A.: State of the information privacy literature: where are we now and where should we go? MIS Q. **35**(4), 977–988 (2011)
13. Biddle, R., Chiasson, S., Van Orschot, P.C.: Graphical passwords learning from the first twelve years. ACM Comput. Surv. **44**(4), 1–41 (2012)
14. O’Gorman, L.: Comparing passwords, tokens, and biometrics for user authentication. Proc. IEEE **91**(12), 2021–2040 (2003)
15. Polakis, I., et al.: All your face are belong to us: breaking Facebook’s social authentication. In: Proceedings of the 28th Annual Computer Security Applications Conference, pp. 399–408. ACM, Orlando (2012)
16. Besnard, D., Arief, B.: Computer security impaired by legitimate users. Comput. Secur. **23**, 253–264 (2004)
17. Rhee, H., Kim, C., Ryu, Y.U.: Self-efficacy in information security: its influence on end users’ information security practice behavior. Comput. Secur. **28**(8), 816–826 (2009)
18. Furnell, S., Zekri, L.: Replacing passwords: in search of the secret remedy. Netw. Secur. **2006**(1), 4–8 (2006)

19. Reeder, R., Schechter, S.: When the password doesn't work: secondary authentication for websites. IEEE Secur. Priv. Mag. **9**(2), 43 (2011)
20. Schechter, S., Egelman, S., Reeder, R.W.: It's not what you know, but who you know: a social approach to last-resort authentication. In: CHI Conference, pp. 1983–1992, April 2009
21. Acquisti, A., Gross, R.: Imagined communities: awareness, information sharing and privacy on the facebook. In: Privacy Enhancing Technologies, pp. 36–58. Springer, Heidelberg (2006)
22. Benson, V., Saridakis, G., Tennakoon, H.: Information disclosure of social media users: does control over personal information, user awareness and security notices matter? Inf. Technol. People **28**(3), 426–441 (2015)

Human Centric Security and Privacy for the IoT Using Formal Techniques

Florian Kammüller(✉)

Department of Computer Science, Middlesex University, London, UK
f.kammueller@mdx.ac.uk

Abstract. In this paper, we summarize a new approach to make security and privacy issues in the Internet of Things (IoT) more transparent for vulnerable users. As a pilot project, we investigate monitoring of Alzheimer's patients for a low-cost early warning system based on bio-markers supported with smart technologies. To provide trustworthy and secure IoT infrastructures, we employ formal methods and techniques that allow specification of IoT scenarios with human actors, refinement and analysis of attacks and generation of certified code for IoT component architectures.

Keywords: Human factors · Security and privacy in the IoT · Formal methods

1 Introduction

The Internet of Things (IoT) denotes the combination of physical objects with their virtual representation in the Internet. It consists not only of human participants but "Things" as well. The IoT has a great potential to provide novel services to humans in all parts of our society. Amongst the biggest problems for this technology to catch on in critical applications are security flaws, due to technical restrictions, immaturity of software applications, and mainly a lack of transparency. The main trigger for security problems is human behaviour, either unintentional or malicious.

In this paper, we give an overview of how we apply formal techniques to enhance security and privacy of human centric IoT systems. We focus on healthcare aiming to support low-cost Alzheimer's diagnosis. We outline the process we use in the CHIST-ERA project SUCCESS. In detail, we report on using interactive theorem proving with Isabelle. We use this proof assistant for the modeling and attack analysis of infrastructures with humans and for the formal definition cryptographic. We apply the Isabelle Insider framework for human centric infrastructure analysis and the inductive approach for security protocol verification to support the secure IoT system development in the early security requirement phase as well as the technical network security level.

D. Nicholson (ed.), *Advances in Human Factors in Cybersecurity*,
Advances in Intelligent Systems and Computing 593, DOI 10.1007/978-3-319-60585-2_12

2 Background

This section provides a short summary of the techniques used in the process of formal development that we use in SUCCESS before highlighting the contributions of the current paper.

2.1 Overview of SUCCESS Project

The core idea of our approach is to use formal methods and verification tools to provide more transparency of security risks for people in given IoT scenarios.

SUCCESS will validate the scientific and technological innovation through pilots, one of which will be in collaboration with a hospital and will allow all stakeholders (e.g. physicians, hospital technicians, patients and relatives) to enjoy a safer system capable to appropriately handle highly sensitive information on vulnerable people while making security and privacy risks understandable and secure solutions accessible.

This international collaboration is funded by the European programme CHIST-ERA [1]. We apply techniques from hardware and software, user behaviour and human-computer interaction to a research pilot from the healthcare sector on supporting IoT monitoring techniques that are human understandable and can be certified by automated techniques.

- specification and verification techniques for secure IoT components and their composition [2],
- verification methods and risk assessment techniques [3] for IoT scenarios with models of human behavior [4], social interactions and human-system interactions,
- implementation and modeling languages with algorithms for the certification of safety, availability, secrecy, and trustworthiness across from the model to the platform [5].

2.2 Contribution of this Paper and Overview

This paper summarizes how the requirements for the IoT healthcare system lead to a high level formal specification in the Isabelle insider framework [6] which also allows attack tree analysis [4]. This first phase of the SUCCESS approach is summarized in Sect. 3 illustrated on a simplified architecture and use cases for the pilot case study. As a result of this first phase, the input to the BIP-based analysis and component architecture design and certified code generation is provided. We omit details on this phase since it is not within the scope of the paper. The output of this process however is a Java Script smartphone app capable of synchronous communication within the phone and via Bluetooth with sensors in the environment of the phone in the patient's home. The communication of the smartphone app with data servers in hospitals and other institutions (like research centers) is asynchronous and channeled via the Internet. It cannot be part of the certified code generation in BIP (which is restricted to synchronous communication). Therefore, we show up in Sect. 4 what is the state of the art of technically realizing secure communication for privacy sensitive data using web services and data interchange formats. In Sect. 5, we illustrate how to formally verify this communication using the inductive approach of security protocol verification in Isabelle (which is compatible with the Isabelle Insider framework as has recently been shown [7]).

3 Healthcare Case Study in Isabelle Insider Framework

The case study we use as a running example in this paper is a simplified scenario from the context of the SUCCESS project for Security and Privacy of the IoT [1]. A central topic of this project for the pilot case study is to support security and privacy when using cost effective methods based on the IoT for monitoring patients for the diagnosis of Alzheimer's disease. As a starting point for the design, analysis, and construction, we currently develop a case study of a small device for the analysis of blood samples that can be directly connected to a mobile phone. The analysis of this device can then be communicated by a dedicated app on the smart phone that sends the data to a server in the hospital.

3.1 Healthcare Scenario

In this simplified scenario, there are the patient and the carer within a room together with the smart phone (see Fig. 1).

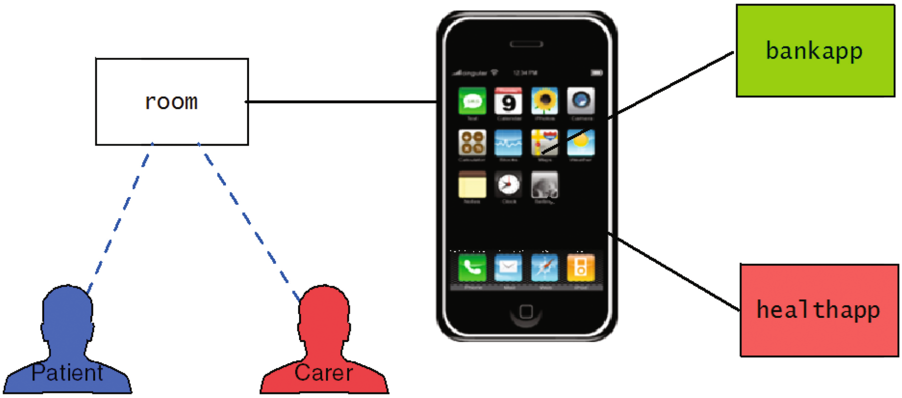


Fig. 1. Health care scenario: carer and patient in the room may use smartphone apps.

The carer has access to the phone to support the patient in handling the special diagnosis device, the smart phone, and the app. The insider threat scenario has a second banking app on the smart phone that needs the additional authentication of a "secret key": a small electronic device providing authentication codes for one time use common for private online banking.

Assuming that the carer finds this device in the room of the patient, he can steal this necessary credential and use it to get onto the banking app. Thereby he can get money from the patient's account without consent.

3.2 Isabelle Insider Framework Analysis

The Isabelle Insider framework enables formalization of the infrastructure as a graph of locations, like room or smartphone, in which human actors reside in locations and local policies are attached to them as well. The details of this modeling and analysis of the case study is given in [4]. As a brief illustration we give some excerpts here. The local policies are given by the following Isabelle definition and explained below.

```
local_policies G
(λ x. if x = room then {(λ y. True,{get, put, move}) }
      else (if x = sphone then {((λ y. has (y, ``PIN``)),
                   {put,get,eval,move}),(λy. True, {})}
            else (if x = healthapp then
                         {((λ y. (∃ n. (n @G sphone)∧
                           Actor n = y)),
                           {get,put,eval,move})}
                  else (if x = bankapp then
                          {((λy.(∃n.(n@G sphone)∧
                           Actor n=y ∧ has(y,``skey``))),
                           {get,put,eval,move})}
                        else {}))))
```

In this policy, any actor can move to the room and when in possession of the PIN can move onto the sphone and do all actions there. The following restrictions are placed on the two other locations.

- healthapp: to move onto the healthapp and perform any action at this location, an actor must be at the position sphone already;
- bankapp: to move onto the bankapp and perform any action at this location, an actor must be at the position sphone already and in possession of the skey.

3.3 Attack Tree Analysis

Attack Trees [8] are a graphical tree-based design language for the stepwise investigation and quantification of attacks. They have been integrated as an extension to the Isabelle Insider framework [6]. This integration extends the Insider model described in the previous section with a proof calculus and modelchecking semantics for attack trees. The extension allows stepwise refinement of attacks exhibiting possible attack paths. The refinement of attack trees is illustrated in Fig. 2 with the refined attack path highlighted.

The following refinement shows the logical expression of this attack refinement. It expresses that the carer can evaluate the money transfer on the bankapp by first stealing the skey, getting on the phone, on the bankapp and then evaluating.

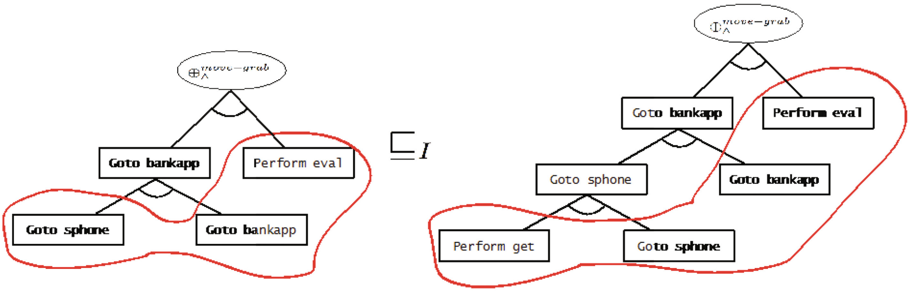


Fig. 2. Attack tree refinement enables stepwise attack path discovery.

$$\texttt{[Goto bankapp, Perform eval]}\ \oplus_{\wedge}^{\texttt{move}\blacksquare\texttt{grab}}$$

$$\sqsubseteq_{\texttt{hc\_scenario}}$$

$$\texttt{[Perform get, Goto sphone, Goto bankapp, Perform eval]}$$

$$\oplus_{\wedge}^{\texttt{move}\blacksquare\texttt{grab}}$$

The proof calculus uses the refinement to prove that the sequence of actions

$$\texttt{[Perform get, Goto sphone, Goto bankapp, Perform eval]}$$

represents an attack in the given infrastructure. The underlying semantics providing the notion of validity of an attack is based on the state transition relation defined in the modelchecking foundation (Kripke-structure over infrastructure states) we constructed in the Isabelle Insider framework.

The attack tree analysis enables formalizing the requirements and high level architecture of the pilot case study. The found attacks can be used to improve the security policies on the model to provide a security enhanced formal specification for the next phase of applying the BIP methodology to develop a component architecture for the target IoT infrastructure in which the security properties of the initial model are preserved and certified code for the components (sensors and smart phone) can be generated. We omit any details of this phase since they will be reported elsewhere. In addition, the attack trees and paths are naturally suited to visualize the security risks to users showing up potential attacks.

4 Security of Web Services for Mobile Devices

We now move to the level of the overall system architecture of SUCCESS in order to show up security and privacy risks of IoT devices connected to data servers via Internet and smart phone technology. In order to be compatible with existing standard technologies, the target code for the smartphone healthapp will be implemented in Java Script. This app represents the client side interface to the database servers in hospitals and other institutions, like research centers. Fortunately, the BIP methodology [2] is flexible enough to produce a Java Script app as certified target code for this component. However, BIP is designed for the formal development of synchronous systems. For the local scenario of sensors connected to a central hub like the smartphone either by physical link – like a blood sample sensor that can be connected via the micro usb or lightning port of the smartphone – or through close range networking protocols – like motion sensors communicating with the phone via Bluetooth [9], this is sufficient. Bluetooth is a packet-based protocol with a master-slave structure where all slaves share the master's clock, i.e., it is synchronous and thus amenable to the BIP code generation and certification process.

But the main data upload of the diagnosis data is to databases on external servers connected via Internet. This is asynchronous communication using web-services. The overall architecture is shown in Fig. 3 showing yet another Insider attack by the carer (discussed further below).

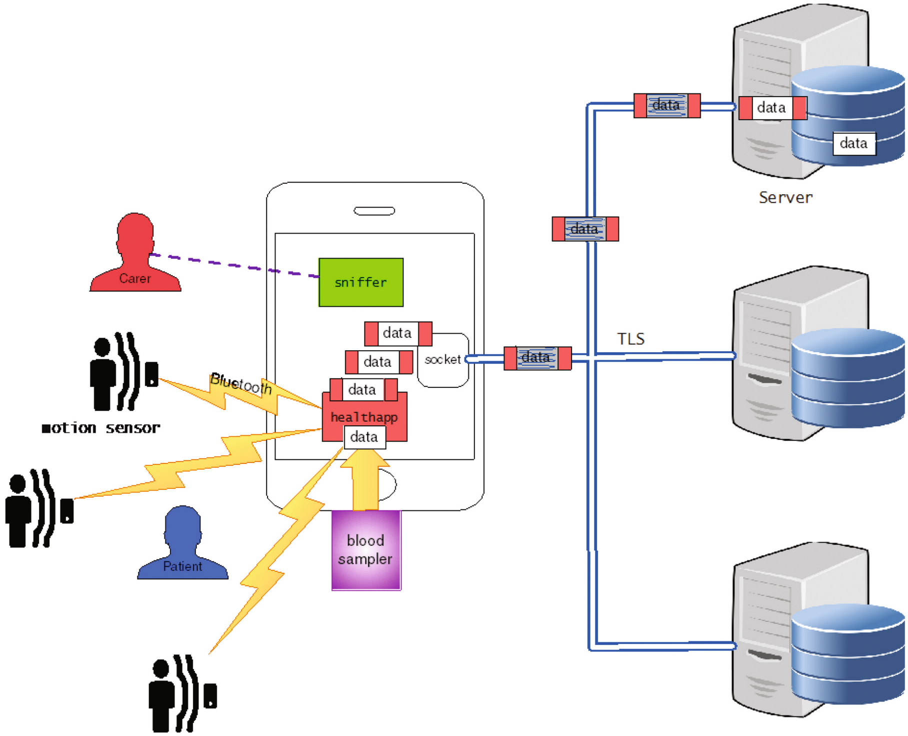


Fig. 3. Carer puts sniffer on smart phone eavesdropping on cleartext TCP packets.

Current standards of best practice for web services for mobile applications have settled on two combinations of technology (1) Java Script Object Notation (JSON) [10] over RESTful web services using http(s) or (2) eXtensible Markup Language XML over SOAP using Web Service Security (WSS) [11]. Solution (1) is more lightweight since the JSON data transfer standard is much less complex than XML. REST prescribes a standard format for web services that is also less complex than SOAP.

So from that perspective, it is a clear choice that in the context of mobile application the former is preferable to guarantee less resource consumption caused by an overhead of the SOAP/XML solution. The critical point is the consideration of security. While the combination of JSON over an https based RESTful web service is slick and appears sufficient it relies on the “s” in https, i.e. Transport Layer Security (TLS) (or Secure Socket Layer (SSL) how it was originally called and is still more widely known as).

TLS is a good standard solution providing point-to-point security between the http port or http proxy of the smart phone and its counterpart on the database servers. However, it does not provide end-to-end security. The difference is that in an end-to-end security connection the security protection would be between the healthapp and the database application on the server instead of in between the http socket of the smartphone usually on port 80 and the connected socket on the same port on the server as it is provided by a TLS connection. Do we need end-to-end security for SUCCESS?

Consider again Fig. 3: since the carer needs to have access to the smart phone to support the patient, he can still endanger privacy by the following attack. Suppose, we only use point-to-point security as given by TLS available on smart phones and servers by default. The carer can use his access to the smartphone to download a sniffer app from the app store, like Wireshark and thereby he can trace and intercept all message communication on the smartphone. This is again an insider attack since again the carer is the attacker. The CMU Insider Threat Guide provides the Insider Attack pattern of ambitious leader: if the carer would collaborate with an ambitious leader outside the home, he could install a specialized app on the phone that would forward intercepted packages from the healthapp to the server of the ambitious leader who could sell the data to interested parties or use it directly for blackmail.

Using the Isabelle Insider framework with its extension to attack trees [4, 6] this attack can be discovered and proved in the attack tree calculus.

$$\texttt{[Goto sphone, Perform put, Goto sniffer, Perform eval]} \oplus_{\wedge}^{\texttt{put-sniffer}}$$

It exposes an interesting challenge for the Isabelle Insider framework since an actor extends the infrastructure (and thus implicitly the local policies) by adding the new location sniffer.

5 Suggested Solution for Security of Web-Services for Mobile Devices

Practically, to introduce end-to-end security we could use JSON as data-interchange format and a RESTful web service with http and TLS. The standard use of http does not foresee the use of an individual protocol for the authentication and key establishment between the healthapp and the database servers, we can design a protocol on top of the transport layer security that enables our application to establish end-to-end rather than point-to-point security.

Designing your own protocol bears risks since security flaws can be introduced. However, in SUCCESS, we use formal methods and in particular Isabelle offers the inductive approach to security protocol verification, e.g. [12], which we use to verify our protocol.

5.1 Isabelle Inductive Approach

A TLS formalization exists already in Isabelle [12]. It uses the inductive approach to security protocol verification that is compatible with the Isabelle Insider framework as we have found in recent applications to an auction protocol [7]. So, for our purposes we can simply assume TLS to be available and base the formal modeling and verification of the current end-to-end extension for SUCCESS on it. Such formal representation of security protocols in the inductive approach and other formal approaches, provide abstract descriptions of protocols. These abstract descriptions usually express what is formalized in standards like in the case of TLS. They serve to prove security properties with mathematical rigour and machine support making the assumption that keys are not lost, and cryptographic algorithms are not broken. Concerning communication channels they use the common strong Dolev-Yao attacker model: messages can be eavesdropped, intercepted and faked.

The Isabelle formalization uses inductive definitions. These definitions are contained in Isabelle theory files adding some modularity to the inductive approach. That is, for our application, we can use aspects of the TLS protocol and mainly its underlying theory of cryptographic keys and messages. To ensure end-to-end security, however, we define our own lightweight security protocol that runs within the TLS. The full TLS specification and proved properties can be found elsewhere [12]. Details on formalizing a protocol in the inductive approach are explained below when presenting the protocol.

5.2 End-to-End Security for Smartphone Apps Over JSON and REST

We can now express a simple protocol that supports mutual authentication between the healthapp represented as P (for Patient) and a database server Server. In the process of this communication a shared key is negotiated that can then be used for future encrypted data upload to the server. We focus on two security goals: (A) authentication of the client to the server and of the server to the client (B) symmetric key exchange for future confidential communication.

The protocol we propose assumes public keys to be in place and trusted. This is a realistic simplifying assumption since the communication is usually to fixed institutions and additional public keys of new servers can be added on the smart phone app. An additional public key certification authority protocol in the style of the DNSsec protocol could be used to set this up [13].

Nonces (Numbers only used once) are used for freshness in the key establishment and authentication phase. The goal of the protocol is the establishment of a shared key KP, Server for a future secure communication of private data from the app to the application on the Server. The healthapp is here referred to as P (for Patient) and the application on the server as Server. The protocol is specified as a set of event lists in the following inductive definition; the rules are explained below.

```
inductive_set sucsec :: event list set
where
Nil:  [] ∈ sucsec |
Fake: [| evsf ∈ sucsec; X ∈ synth (analz (spies evsf)) |]
        ⇒ Says Spy A X # evsf ∈ sucsec |
suc1: [| evs1 ∈ sucsec; ~(Nonce N_P ∈ used evs1);
       ~(Key K_P,Server ∈ used evs1) |]
      ⇒ Says P Server (Crypt(pubK Server)
            {Key K_P,Server, Nonce N_P}) # evs1 ∈ sucsec |
suc2: [| evs2 ∈ sucsec; ~(Nonce N_Server ∈ used evs2);
        Says P Server (Crypt(pubK Server)
        {Key K_P,Server, Nonce N_P}) ∈ set evs2 |]
      ⇒ Says Server P (Crypt(pubK P)
            {Nonce N_P, Nonce N_Server, Server})
        # evs2 ∈ sucsec |
suc3: [| evs3 ∈ sucsec; Says Server P (Crypt(pubK P)
        {Nonce N_P, Nonce N_Server, Server}) ∈ set evs3 |]
      ⇒ Says P Server
            (Crypt(pubK Server){Nonce N_Server})
        # evs2 ∈ sucsec
```

This protocol is inspired by the improved version of the Needham-Schroeder public key protocol adding the symmetric session key $K_{P,Server}$ created by the healthapp using for example AES 256. The rule Nil initiates the set with the empty trace representing the point before any protocol session starts. The rule fake is the rule that introduces events created by the agent Spy who can synthesize and play into any event trace evsf messages based on what he analyses from all eavesdropped traffic: synth(analz(spies evsf)).

Note, that he “says” this message to an unspecified agent A which could be Server or P. The rule suc1 requires a fresh Nonce and a fresh symmetric key both created by the healthapp. Freshness of a Nonce or a key is expressed for example as ~ (Nonce NP ∈ used evs1) meaning that this Nonce has not (~) been used in the trace evs1 before. Agent P then sends these items encrypted with the public key of the server process Server. Consequently, those items can only be seen by Server. According to rule suc2, the server process responds with a message in which it packs its fresh Nonce and the unpacked Nonce of P submitted in the previous message thereby proving that it is in the possession of the private key priK Server. This corresponds to server authentication. Finally, rule suc3 is the client authentication in which the healthapp P proves that it is in possession of priK P by unpacking and repacking Nonce N_{Server} from the previous message.

Despite these arguments being seemingly obvious deductions from the protocol steps, they need to be verified to guarantee security. The inductive approach in Isabelle allows formal verification of these and other security properties of the protocol.

The provided abstract specification of the protocol can be implemented as initially mentioned using JSON or XML to encode the transmitted data (messages including keys) over https (thus automatically creating the TLS tunnel between the smart phone and the webserver of the hospital or research institution). The asymmetric cryptography for public and private key pairs can be implemented for example using RSA and the symmetric keys could be AES 256. For both JavaScript libraries exist.

The authentication protocol can be used for different servers. The data that is then sent by the healthapp (either in JSON of XML) can be preprocessed by sanitization of the data (e.g., delete names and address for scientific purposes when connecting to the database of the research center). Following instead the SOAP standard would require the use of XML and this is not suited to our mobile app. The practical ad hoc standard of using the lighter JSON data interchange format and combining it with a RESTful web service is practically sufficient and compatible with JavaScript as target language for the certified code generation of the healthapp as output of the BIP process.

6 Discussion and Conclusions

In this paper, we have given an overview of applying a range of formal techniques to the security and privacy sensitive scenario of healthcare focused on mobile Alzheimer's diagnosis. We only sketched the overall process as we envisage to use it in the CHIST-ERA project SUCCESS but detailed on the use of interactive theorem proving in Isabelle in two stages: (1) for a formal machine-supported analysis of attacks at early development stages and (2) for the formal definition of a dedicated end-to-end cryptographic protocol between a smart phone app and server database applications. Both stages are supported by Isabelle frameworks: (a) the Isabelle Insider framework for human centric infrastructure analysis and (b) the inductive approach for security protocol verification. The combination of both within the Isabelle framework is straightforward. A closer integration to formalize and prove deeper security properties involving both levels has been explored in a different context of auction protocols [7] and has procured interesting insights into collusion attacks and new notions of rational agents.

It seems promising and a future challenge for SUCCESS to explore this integration on privacy of IoT solutions for vulnerable agents. Initial challenges like dynamic extension of the infrastructure graph and local policies (example of the sniffer app download) have already been identified in this paper.

The suggested use of the Bluetooth protocol [9] for the short distance communication in the patients home offers an additional security vulnerability due to symmetric key agreement protocols. However, there is a stronger implementation that uses asymmetric key establishment and that is feasible for certain devices including smart phones [14]. Starting from Bluetooth version 2.1 it is required to use Secure Simple Pairing (SSP) for pairing which is the public key based pairing method. If the attack analysis will show that a Bluetooth based attack is a risk SUCCESS needs to address, then we have to verify whether this asymmetric solution is feasible between the motion sensors and smartphone. Otherwise, we need to integrate weaker mitigation stategies, e.g. enable Bluetooth only when required, in the patient diagnosis policy. This part is addressed in the central part of the formal development of a component based architecture using the BIP methodology and is not covered in this paper.

Acknowledgments. Part of the research leading to these results has received funding from the European Union (CHIST-ERA 2015) under grant agreement no. 102112 (SUCCESS). This publication reflects only the authors' views and the Union is not liable for any use that may be made of the information contained herein.

References

1. SUCCESS: SecUre aCCESSibility for the internet of things. CHIST-ERA (2016). http://www.chistera.eu/projects/success
2. Basu, A., Bensalem, S., Bozga, M., Combaz, J., Jaber, M., Nguyen, T.-H., Sifakis, J.: Rigorous component-based system design using the BIP framework. IEEE Softw. **28**(3), 41–48 (2011)
3. Arnold, F., Hermanns, H., Pulungan, R., Stoelinga, M.I.A.: Time-dependent analysis of attacks. In: Principles of Security and Trust, POST 2014. LNCS, pp. 285–305 (2014)
4. Kammüller, F.: Formal modeling and analysis with humans in infrastructures for IoT healthcare systems. In: 5th International Conference on Human Aspects of Information Security, Privacy, and Trust, HAS 2017, co-located with HCII 2017. LNAI. Springer, Heidelberg (2017)
5. Ben Said, N., Abdellatif, T., Bensalem, S., Bozga, M.: Model-driven information flow security for component-based systems. In: ETAPS Workshop 'From Programs to Systems', FPS@ETAPS, vol. 2014, pp. 1–20 (2014)
6. Kammüller, F., Probst, C.W.: Modeling and verification of insider threats using logical analysis. IEEE Syst. J. **PP**(99), 1–12 (2016)
7. Kammüller, F., Kerber, M., Probst, C.W.: Insider threats for auctions: formal modeling, proof, and certified code. Spec. Issue J. Wirel. Mob. Netw. Ubiquit. Comput. Dependable Appl. (JoWUA) **8**(1), 44–78 (2017)
8. Schneier, B.: Secrets and Lies: Digital Security in a Networked World. Wiley, New York (2004)
9. Wikipedia: Bluetooth. https://en.wikipedia.org/wiki/Bluetooth. Accessed 4 Mar 2017
10. JSON. ECMA-404: The JSON Data Interchange Standard (2017). http://www.json.org
11. OASIS: Web services security: SOAP message security. Working Draft 13, Document identifier: WSS: SOAP Message Security -13, OASIS Open 2002. http://www.oasis-open.org/committees/documents.php
12. Paulson, L.C.: Inductive analysis of the internet protocol TLS. ACM Trans. Inf. Syst. Secur. **2**(3), 332–351 (1999)
13. Kammüller, F.: Verification of DNSsec delegation signatures. In: 21st International Conference on Telecommunication. IEEE (2014)
14. Wong, F.-L., Stajano, F., Clulow, J.: Repairing the bluetooth pairing protocol. In: Security Protocols 2005. LNCS, vol. 4631, pp. 31–45. Springer, Heidelberg (2007)

Feasibility of Leveraging an Adaptive Presentation Layer for Cyber Security Visualizations

Lauren Massey[1(✉)], Remzi Seker[1], and Denise Nicholson[2]

[1] Electrical, Computer Software, and Systems Engineering Department,
Embry-Riddle Aeronautical University, Daytona Beach, FL, USA
masseyl@my.erau.edu, remzi.seker@erau.edu
[2] Soar Technology, Orlando, FL, USA
denise.nicholson@soartech.com

Abstract. The balance between end user and software engineer is important to the usage and development of software. Finding this balance, in which the end user can access needed information without overly complicated displays, a time-consuming labyrinth of clicks, and the engineer can implement the display concisely is difficult. Typically, end users desire complex displays that allow for fluid movement to the answers they need. However, accomplishing this can be time consuming for the engineer because complex displays require hard-coded GUIs. Depending on the amount of unique end – users, these issues can multiply because every user role could need a unique, complex display that will require hard coding from the engineer. However, through the usage of the Service Oriented Architecture (SOA) a solution may exist. This architectural style has been leveraged in developing an “adaptive presentation layer” pattern that allows for complex GUIs to be derived without the need of hard coding. This solution was developed for a domain that needed role specific information for map clients; however, other user interface clients have not been applied to this pattern. Therefore, to examine the viability of this solution, it must be applied in other domains using various UI clients. The cyber security domain provides suitable platform to research this solution because of the necessity to monitor several entities of data concurrently and ensure that those monitoring the data can quickly attain the need information. A successful implementation could provide a viable solution in the development of future cyber security interfaces.

Keywords: Human factors · Human-computer interaction · Cyber security · User interface design · Software architecture

1 Introduction

In software development, there is an estimated 70% failure rate in user adoption for IT projects [1]. Although not completely, user adaptation of the new system is a foremost factor that drives up the failure rate. The lack of user adoption typically stems from a fundamental issue of forgetting the end user and their needs and workflows. Therefore, during the development of software it is important to consider how the end user will utilize the solution. Like many other domains, this issue has plagued the cyber security

D. Nicholson (ed.), *Advances in Human Factors in Cybersecurity*,
Advances in Intelligent Systems and Computing 593, DOI 10.1007/978-3-319-60585-2_13

domain; in which, solutions are developed to adhere to the exponentially, increasingly complex world of technology and ensuring reliable defense measures. However, the users of the solution, cyber analysts, are left with tools that are not conducive to the environment and are not as useful or reliable as the basic tools already in use.

A potential avenue to combat this issue is to allow the user to design the front end for what is best for them. However, typically user interfaces must be hard coded by an engineer and require major effort to ensure the design is useful and accurate. It is impractical and potentially hazardous to give access to the source code to a typical front end user so they can design their own user interface. Although this would allow them to create exactly what they desire, it would also create many more issues than solutions. However, this solution may be viable through implementing a service oriented architecture.

The cyber security domain provides an ideal test case to research the feasibility of using the service oriented architecture to generate visualizations because of the growing need for dynamic, concise interfaces to make critical decisions rapidly.

2 Background

Cyber security analytics is the study of understanding the behavior of computers and computer networks and delving into the cyber data behind that behavior. This branch of cyber security also works to defend the computing infrastructure. The need for skilled cyber security analysts is exponentially increasing mirroring the growth of the need, usage, and complexity of technology. However, the tools used by these analysts leave much to be desired with some ultimately using command-line tools [2]. Although command-line tools can be successful in analyzing and protecting against cyber security threats, they cannot effectively sustain the “high volume and velocity of the data” that must be processed [2]. Therefore, the evolution of the toolbox of a cyber security analyst is essential. To accurately modernize the toolbox, an understanding of the working environment and current tools of a cyber analyst is crucial.

2.1 Understanding the Working Environment of a Cyber Analyst

Overall, a cyber analyst works to find often well-hidden, complex correlations between several data sets. To accomplish this, the professional must work on multiple open investigations concurrently and should “rapidly switch between analytic inquiries, multi-tasking and refining or broadening queries as they investigate potential leads” [2]. This generates an exploratory environment in which multiple windows and tools must be displayed simultaneously. This environment allows for “information foraging” in which the analyst investigates several potential leads without having to lose their overall place [3]. The overall goal of these forages is to pinpoint individual clues that correlate to a root cause. Moreover, arrival at this root cause is not a straight line, but instead a series of “tip-off” points that potentially could lead to the “big game”, the culminating root of the issue [2].

During an observation of cyber analysts by Pacific Northwest National Lab (PNNL), several key points were uncovered about how their job was completed. These included the concept of the "quest for a query" which was noted as a common and effective approach to "data foraging and sensemaking to identify [a] suspicious phenomenon" [2]. Essentially, the analyst examines for a descriptive "query" through a series of complex analyses to return only data that concerns the suspicious behavior in question. This "quest" represents how an analyst forages the data and the development several of tip-off points. This process is tedious and is completed through a series of tools that are used based on the needs of the search. These tools are discussed in the subsequent section.

2.2 The Lopsided Toolbox of a Cyber Analyst

There is a spectrum of tools utilized by cyber analysts ranging in flexibility from basic command line prompt to specialized visualizations such as packet-headers, network-flows, system log files, and IDS alerts. However, this toolbox, though extensive, seems to be unevenly utilized with preference for command line rather than any visualization.

Command-line prompt provides an environment that returns raw data points. This allows for the professional to sift through raw data and make their own conclusions. Consequently, this is the most common tool and used because of its "unparalleled flexibility and expressive power" [2]. Many of the "quests" discussed above originated from command line usage. However, PNNL remarked how this approach could be problematic because it forces a formalized hypothesis [2]. Essentially, to complete a query search, the analyst must construct a command line argument that will be passed into the system. This argument represents the hypothesis of the analyst. However this process performed at the inception of a phenomenon search, has high potential to not result in useful data. Although this is common and an analyst does not expect a hit on the first or even the first several tries, it does delay the overall process. PNNL argued that "at the beginning of an investigation …there is much uncertainty [and] analysts are frequently unsure of what to query …" [2]. PNNL attempted to resolve this issue via usage of visualizations for exploratory means. This "gives the analysts opportunity to begin with an informal hypothesis and gradually increase the rigor of their query" [2]. This approach potentially could cut down on the amount of time spent attempting to pinpoint correlations. However, the usage of visualizations is not widely-accepted for several reasons, often citing "visualizations hide what is going on with the data", do not produce useful results, are seen as a crutch, and lack flexibility [2].

An analyst is typically interested in irregularities in data and visualizations may not store all data points or aggregate them which smooths out "noisy" data. Although visualizations may do this in the essence of efficiency, aggregation generates a significant hindrance. An adversary can now "hide in the noise" because it provides camouflage [2]. This leads to a major mistrust of visualizations within the domain. Often the analysts prefer access to the source data and look through it manually. This is time consuming, but it ensures that they will be able to see issues rather than having them overlooked by visualization.

Moreover, analysts seem to have developed a prejudice against all visualizations because of past experiences of poorly designed displays [2]. Some of this seems to stem from data visualizations implemented that were not optimized for the cyber security domain. PNNL noted that tools must be implemented with cyber security in mind rather than a generic approach. This would prevent important data that is domain specific from being overlooked.

It has also been cited that cyber analysts consider visualizations to be a gentle training tool for aspiring defenders [4]. In turn, the analysts view the "ability to read and manipulate massive of textual cyber data as a hallmark of their expertise" [2]. This mindset can falsely lead to the viewpoint that analysts using visual representation have underdeveloped skills.

Another issue noted with visualizations is they are optimized for a certain type of data and generate "special-purpose representations". This was noted by PNNL stating that the optimization of a "tool for one type of data separates the tool from the context of an overall investigation" and will limit its utility to the analyst [2]. This concept uncovers two main issues.

Frist, conducting an investigation may be limited by how far the tool can be applied. Although a tool could provide information about a certain data set, drilling down in to the visualization was not possible. For instance, if a particular data point was of interest on a visualization, the opportunity to investigate in that medium would be limited by how far (if at all) or how fast the visualization could handle that request. If it could not do it efficiently, the analyst would have to switch mediums to further investigate. This is because the visualizations are sometimes designed to "support preconceived workflows rather than [an] open-ended investigation" [2]. This essentially will cause the tool to not be used or only be used when it is the only option. Otherwise, it will slow down the already rigorous process.

Secondly, visual displays lack flexibility. Expanding from the example above, if the analyst wanted to simply take the data collected on the visual display and apply the output to another investigative tool there is little interoperability with other applications. This leads to time loss of importing and exporting data to and from tools [2]. This issue has been researched with some implementations resulting to overcome this within the Intelligence Community (IC) and is discussed below.

However, from the research conducted by PNNL it seems that visualizations do, in fact, have a place in the cyber analyst's toolbox. During interviews of cyber analysts by PNNL, an interviewee was criticizing the usage of visualizations in his domain. However, amidst this critique, he "casually noticed a feature on a scatterplot visualization", this observation was then crosschecked against a query search via command line revealing a solution to a problem he had been working on for two hours [2]. This revealed that although an analyst may not need visualizations, they may lead to important discoveries. Furthermore, during PNNL's investigation, it was observed that during the workflow an analyst would "find the information they need in a visualization… [then] would cross correlate it with other data manually" [2]. Thus, the visualization was a valuable tool in locating potential "tip-off" points.

2.3 Interoperability of Tools for the Intelligence Community

As discussed earlier, there are limitations surrounding the usage of visualizations in the cyber security domain. It is understood that there are numerous investigative tools used by cyber professionals to accomplish various exploratory tasks, and no one tool can adequately illustrate a full picture. Therefore, much time can be spent integrating various outputs of investigative tools into other cyber tools. The lack of interoperability between other tools causes analysts to deter from using such tools, especially when the latter choice is an all-inclusive tool, command-line prompt. Ultimately, interoperability between investigative tools is crucial for the cyber analyst community. The cyber analyst community is still a growing domain of research. Therefore, to adequately research the usages of interoperability of tools, an expansion to the overarching IC was completed.

Although there are still differences between IC and the subgroup of cyber security, the need of making quick decisions under exceptional time pressures while maintaining a high quality of standards remains relevant. Furthermore, the intelligence community is driven by the need to "filter, distill, and correlate large quantities of [data]" and this need is mirrored in the cyber security analysis domain [5]. Lastly, the information needing to be refined and consumed originates, much like the cyber domain, from many multimedia sources. The need of interoperability is relevant in the IC, and thus, this community can be used as launch point. The programs below were chosen because of their relation to the cyber security domain or their overarching principles Table 1.

Table 1. Various programs in the intelligence community with interoperable toolboxes [5–7]

Name	Developed for	Purpose
STARLIGHT	Intelligence community	A potential solution for the IC to generate a visualization system that integrates various data types from data tables to images and maps
Viage	Data-intensive domains	Prototype a user interface environment for exploring three common information types for data-intensive domains
Snap-Together Visualization	Cyber security domain	Rapidly and dynamically mix and match visualizations and coordination to construct custom exploration interfaces without programming

STARLIGHT. PNNL designed an information visualization system called STARLIGHT. STARLIGHT is an "attempt to address the most relevant problems within the IC by developing a visualization system capable of supporting the integrated analysis of a wide range of information types and structures" [5]. This system does rapid concurrent analysis of structured/unstructured text, geographic information, and digital imagery and generates visualizations based off explicit and implicit relationships between the ingested data [5]. The end goal of this project is to enable analyst to quickly develop and test hypotheses based off STARLIGHT's analysis of interrelationships of the data collected. Also, within this research the unique necessity was to rapidly accommodate new information and quickly generate the "view" of a particular data set appropriate to the immediate task. More recently, STARLIGHT has been leveraged to create cyber displays and is an ongoing area of research [8].

Visage. This tool was developed as a prototype for a user interface to explore and analyze information [6]. Visage consists of three tools: Table Lens, IVEE, and SAGE. Table Lens is a dynamic spreadsheet used for examining large, multidimensional data sets. IVEE is an analysis tool used to create dynamic query that can filter data. SAGE rapidly generates visualizations based off multiple attributes. Together these tools create an information-centric approach that allow for selection and combination of user interface, interoperation between programs, and ability to "drill down" and "roll-up" information as needed. Although this tool was not created specifically for the cybersecurity domain, the relevance of the approach almost mirror the issues pointed out with current cybersecurity displays.

Snap-Together Visualization. Developed by the University of Maryland, the Snap-Together Visualization allows users to rapidly and dynamically mix visualizations of their choice without the need of programming [7]. The process is completed by identifying relations in the chosen visualizations and then coordinating the visualizations based on relationships between the chosen visualizations. For instance, a user would first select visualizations from a finite list. Once all the desired visualizations are chosen the user "snaps them together" by coordinating any relationship between the two visualizations. This action tightly couples the two, or more, visualizations [7]. On the backend, these visualizations are called through the Snap API and remain as independent software programs [7]. Therefore, the software package and its internal architecture, data structures, and visualization outputs remain untouched.

3 Defining the Needs of a Cyber Analyst

Now with an understanding of the type of user, their workflows, current issues, and ongoing research to aid in eliminating those issues, the introduction of a potential new solution is possible. Based on PNNL's study into cyber analyst tools, changing workflow or forcing an adaption of new visualizations may not be the correct approach. This can lead to less adoption from the analysts. Moreover, there are a multitude of tools already utilized by the cyber analyst community. The current tools specially created for the cybersecurity domain do not need to be optimized. Based off PNNL's research, there was no convincing evidence that the current tools were inappropriate for the job. Therefore, introduction of more tools would not solve the current issue and instead, once again, lead to less adoption.

Instead, leveraging the existing tools while providing ways to interoperate data between tool sets would be ideal. This is potentially possible by utilizing the service oriented architecture (SOA) and extending its resources to an adaptive presentation layer

(APL). This implementation would allow all of the current tools to be leveraged while providing an environment in which the data can be shared concurrently between tools. Moreover, much like the Snap-Together Visualization, the users would be able to use a rule editor to choose the needed tools to complete a task. Furthermore, this adaptive presentation layer would allow for the visualization to be modified dynamically based on the user's needs without the need for programming. This creates an environment that is possible to implement on an as needed basis without the need to hardcode each individual portion. The benefits of using this architecture could be the key in realigning the currently lopsided toolbox and ultimately providing solutions to the issues pinpointed by PNNL including [2]: "[designing] a way to provide rich linkages among multiple visualizations tools that better support the entire process of analysis" and "tools that help frame queries built from general interactions with the data rather than SQL statements."

This solution has been applied successfully for Command and Control (C2) systems for the DoD [9]. Leveraging the SOA provides effective, role-relevant displays that could be configured on the fly. To accomplish this, a common operational picture (COP) had to be developed and then adopted by the user [9]. The COP provides a mutual display foundation among users in which to build role-relevant displays [10]. This shared foundation provides display configuration formats and the enterprise service set that can be implemented into the display. In this application, the enterprise service set were various map clients that were leveraged and through XML and JSON based configuration files were implemented into a display [9]. A rule editor was responsible for transforming the needed information from the map clients into usable XML and JSON files. The usage of the rule editor prevented the need to program and allowed front end users to adjust their displays as needed. This rule editor was particularly useful for on the fly alterations; so, that as information and the needs of the user changed, what was illustrated changed. The implementation of this application was successful and provides the framework for future uses in other domains.

Although the above application was successful in implementation of SOA for map clients, it has not been evaluated on other platforms. The workflow of a cyber security analyst provides a suitable test environment for such an analysis because of the multitude of tools used and the current need for interoperability. In order to run this test case a comprehension of the SOA and the extension of the architecture through an adaptive presentation layer must be gained.

4 Understanding Service Oriented Architecture (SOA)

Functionality and flexibility is crucial to the usability of a software program. This can be difficult to attain in the ever-changing high tech market because of additional level of complexity of new services added to programs [11]. This can cause the program to become cumbersome leading to frustration and lack of adaptability as a system from the front-end user. The SOA provides an approach to treat each necessary application in the program as a service. This allows for the new applications to be added to the program without slowing down the software, thus, cutting down on the ever-growing issues of complexity with each additional application needing to be integrated.

4.1 Reference Model of SOA

The general design of a SOA is divided into three layers as seen in Fig. 1. This design is representative of the common three-tier architectural pattern consisting of a data layer, logic layer, and a presentation layer. The overall appearance and methodology of the SOA is exceptionally close to this classic model.

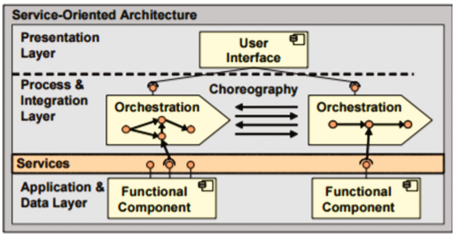


Fig. 1. Typical SOA reference model [12]

The bottom layer is comprised of the existing systems that are conceptually denoted as a functional component. The SOA then exposes the functionality of each of the services and leverages them as reusable services. It is important to note that each legacy system does not need to be adjusted to fit into the model [12]. This cuts down on the amount of effort applied and makes it ideal for systems with several programs integrated into it. After the services of the existing systems are exposed, they can be "orchestrated" by the Process and Integration Layer. During this layer, the services are mapped to accomplish business logic and business processes. During this layer, various services can be integrated allowing for complex metadata to be constructed. The presentation layer invokes this orchestration by either a user interface or a task management system. A task management system would lie between the presentation layer and the process and integration layer. These systems would assign different tasks to the users. This allows for multiple users or organizational units at one time [12].

4.2 Components of the SOA Design

The overall approach of the service oriented architecture is ideal when there is a need for interoperability and flexibility. There are three primary entities of the SOA design: service provider, service consumer, and service registry [12]. The interactions between these three entities can be described as "find-bind-execute" [13]. This is illustrated in the layout of the architecture in Fig. 1 through the services being "found" in the application and data layer, "bound" in the process and integration layer and then "executed" in the presentation layer. Figure 2 depicts how the three entities interact.

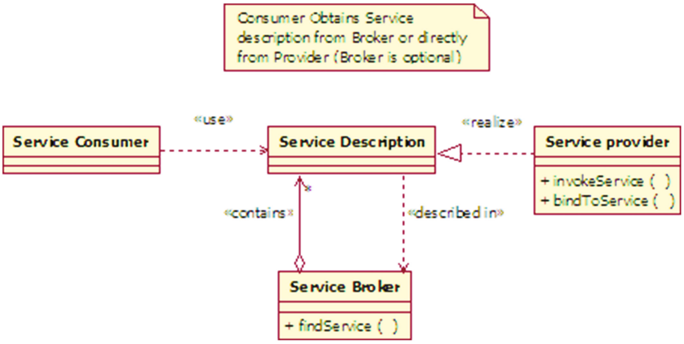


Fig. 2. Components of an SOA design [13]

Service Provider. The “find-bind-execute” principle first needs to have something that can be found. The service provider establishes this entity by providing the implementation of the service and the description of this service. Within an SOA there can be multiple services offered from different service providers. Moreover, the complexity of the architecture of these services is unimportant because the process and integration layer will handle them the same.

Service Consumer. The service consumer makes up the other two portions of the “find-bind-execute” principle in which the consumer must bind the service based on the needs of the business model and then invokes the service. The availability of several services offered allows for this process to be interoperable and reusable. Moreover, if the business logic or process changes this architecture can transform with it. This corresponds to the second layer illustrated in Fig. 1 – process and integration layer.

Service Registry/Service Broker. Lastly, the service registry is the location of the services the service provider offers. Also, a service broker maintains the service registry to ensure its integrity. This is an optional piece but can provide a structure in which to house and maintain the services neatly.

5 Extending SOA with an Adaptive Presentation Layer (APL)

The concept of SOA is “appealing to the presentation layer because quick changing business processes require adjustable, interoperable, and flexible user interfaces” [12]. Therefore, an extension of the service oriented design is desirable and possible to attain through following the same concepts of “find-bind-execute”. However, rather than application towards the business services, the paradigm is applied to presentation services. In Table 2, this shift is illustrated.

Table 2. Extending the SOA "Find-Bind-Execute Paradigm" to the presentation layer [12].

Typical SOA entity	Presentation layer service oriented entity	Description
Service provider	Presentation service provider	Provides presentation components as services. There are typically several service providers offering presentation services
Service consumer	Presentation service consumer	Invokes one or more presentation services in order to integrate them into a specific context. EX: Control Panel. There can be several presentation consumers on the presentation layer, each of which serves a certain business purpose
Service registry/service broker	Presentation service registry	All presentation services made available by service providers are listed and can be found by service consumers

The migration of SOA to the presentation layer is further described in Fig. 3. This shows how the application of the APL is applied visually. As noted in Table 2, the application of the components of the SOA approach to the presentation layer is close to the original SOA. Figure 3, illustrates how the components interact. Essentially, the presentation container acts as the presentation service consumer invoking presentation services provided by the presentation components [12]. To complete this extension, there are two different types of services, as depicted in Fig. 3.

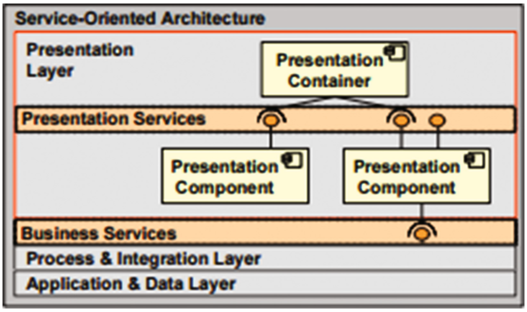


Fig. 3. SOA reference model with an Adaptive Presentation Layer (APL)

The presentation services provide a user interface and allow for direct interaction between the front-end user and the presentation services [12]. The presentation service does not process any business-related data and instead allows the end user to interact with business services. Therefore, this layer acts as a “gateway between the business services and the user” [12]. Since there is a direct user interaction for this service type, it is distinctly different than the business services which focuses on processing data without human interaction. This interaction is also illustrated in Fig. 3 and follows the “request-response model by receiving a request, processing it and generating a response on a programmatic level” [12].

6 Realigning the Toolbox

Overall, the usage of SOA and APL can create a visualization that will allow for fluidity between systems while providing a unique, customizable display. Figure 4 illustrates the full implementation of this approach. Essentially, the individual tools used by a cyber analyst would be held in the application and data layer. These tools would be transformed as needed into particular components that are useful to an analyst. These components in the presentation layer as services, ready to be used as needed by the presentation container.

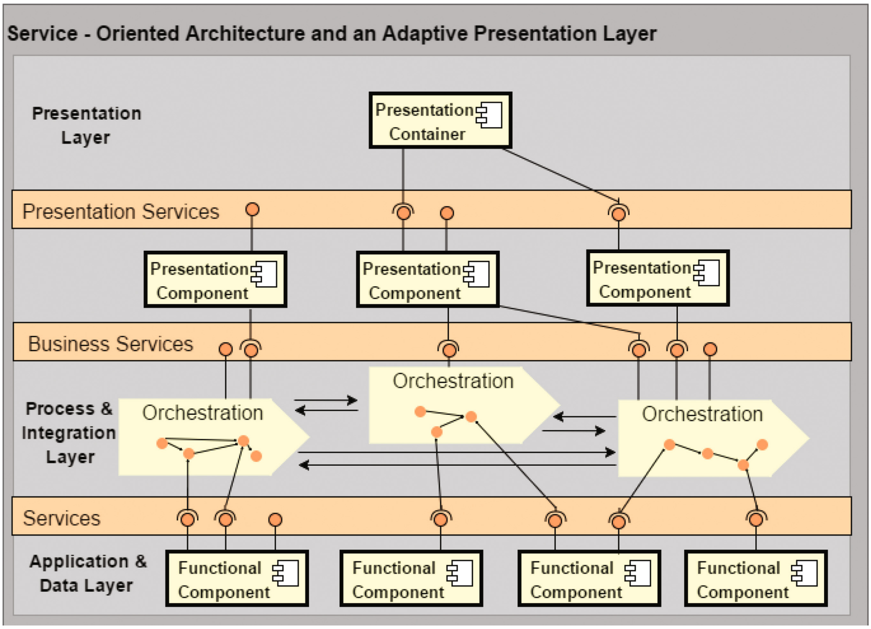


Fig. 4. A full implementation of a SOA with an APL

This approach will allow for the needed manipulation of the tools from the application & data layer to be interoperable through the integration layer. These can then be designed into presentation components that can be accessed by the presentation container. The flexible approach provides a way to give all the tools in the toolbox to the analyst but allow the tools to better work together and provides a solution to the two major issues that PNNL identified during their research [2].

To fully implement this solution, the development of presentation containers that would be useful to an analyst must be researched and designed. This will lead to the development of a common operational picture. As explained in the first iteration of this architecture for Command and Control for the DoD, a common operational picture (COP) must be developed. This becomes the foundation of the presentation container and the final piece to this solution. In order to design a COP, further research into the common layouts and importance of those arrangements are needed.

7 Conclusions

The domain of cyber analytics is still a developing field of study and providing a flexible and user friendly solution for generating visualizations would aid in quicker successful identifications of breakdowns in the defense network. This would alleviate tedious query searching and allow analysts to zero in on interesting irregularities in data sets. However, more research is needed for designing the common operational picture to implement this approach fully.

References

1. Rand Group: Don't Just Build a Solution. Get it Adopted. Rand Group (2017). https://www.randgroup.com/about/methodology/user-adoption-methodology/. Accessed Mar 2017
2. Fink, G., North, C.L., Endert, A., Rose, S.: Visualizing cyber security: usable workspaces. In: 2009 6th International Workshop on Visualization for Cyber Security, Atlantic City (2009)
3. Pirolli, P., Card, S.: Information foraging in information access environments. In: SIGCHI Conference on Human Factors in Computing Systems, Denver (1995)
4. Fink, G., Correa, R., North, C.: System Administrators and their Security Awareness Tools (2005). http://people.cs.vt.edu/~finkga/Research%20Defense/System%20Admin. Accessed Mar 2017
5. Risch, J.S., Rex, D.B., Dowson, S.T., May, R.A., Moon, B.D.: The STARLIGHT information visualization system. In: IEEE Conference on Information Visualization (IV 1997), Phoenix (1997)
6. Roth, S., Lucas, P., Senn, J., Gomberg, C., Burks, M., Stroffolin, P., Kolojejchick, J., Dunmire, C.: Visage: A User Interface Environment for Exploring Information. IEEE, San Fanscisco (2002)
7. North, C., Shneiderman, B.: Snap-Together Visualization: A User Interface for Coordinating Visualizations via Relational Schemata. University of Maryland, Human-Computer eInteraction Lab & Department of Computer Science, College Park (2000)
8. Lavigne, V., Gouin, D.: Applicability of visual analytics to defence and security operations. In: 16th International Command and Control Research and Technology Symposium, Quebec City (2011)
9. Zaientz, J.D., Hultner, M., Ray, D., Hamel, L.: An enterprise service set for adaptive role-relevant operational displays. In: Defense Security and Sensing, Orlando (2011)

10. Satchell, T., Dormish, S., Parker, A.: Creating a Joint Common Operational Picture. Society of American Military Engineer (2017). http://themilitaryengineer.com/index.php/tme-articles/tme-online-exclusive-articles/item/248-creating-a-joint-common-operational-picture. Accessed Mar 2017
11. Channabasavaiah, K., Tuggle, E., Holley, K.: The case for developing a service-oriented architecture. IBM developerWorks, 16 December 2003. https://www.ibm.com/developerworks/library/ws-migratesoa/. Accessed Mar 2017
12. Link, S., Jakobs, F., Neer, L., Abeck, S.: Architecture of and Migration to SOA's Presentation Layer. Cooperation and Management, Universitat, Karlsruhe
13. Arsanjani, A.: Service - oriented modeling and architecture: how to identify, specify, and realize services fot your SOA. IBM (2004)

Human Factors in Cyber-Warfare

Interacting with Synthetic Teammates in Cyberspace

Scott D. Lathrop(✉)

Soar Technology Inc., 3600 Green Court, Ann Arbor, MI 48105, USA
scott.lathrop@soartech.com

Abstract. This paper explores the interaction of humans and autonomous, intelligent agents working together as teammates in cyberspace operations. Though much research has investigated human-machine teams in domains such as robotics, there is a dearth of research into human-agent dynamics in cyberspace operations Some challenges are similar, such as trust between human and agent. Other challenges, such as representation and interface, are unique to cyberspace given that topological, logical, and temporal relationships are first class constructs with different semantic interpretations from their counterpart visual and spatial representations that are prevalent in physical domains. These challenges arise as the software behaves less like a tool and increasingly becomes more like a synthetic teammate.

Keywords: Human factors · Cyberspace operations · Cybersecurity · Human-agent teaming · Knowledge representations

1 Introduction

There has been a plethora of human factors research on human-machines interactions addressing issues such as trust, communication, and user interfaces. For human-machine interaction, the research typically addresses machines that operate in the physical world, such as robotics platforms or training systems, which emulate a physical world. On the other hand, there is a dearth of research regarding how humans interact with a synthetic teammate for cyberspace operations. In fact, there has been very little research in general regarding teaming in cyberspace [1].

Cyberspace is a relatively new domain where concepts such as teaming are being developed as many of the current capabilities used in the domain are built by and for expert cybersecurity professionals for individual purposes rather than for collections of individuals. For cyberspace operations, where military concepts such as fire and maneuver apply, teaming is an inherent requirement.

Consideration must also be given to the velocity and volume of data that must processed and comprehended in cyberspace operations to drive decision-making. Just the shear amount of data one has to understand to make sense of underlying actions demands more automation. Also at play is a well-documented shortfall of a workforce that can scale to can make sense of this data. These shortcomings point towards more autonomy, transferring some of the tactical decision-making to synthetic teammates

D. Nicholson (ed.), *Advances in Human Factors in Cybersecurity*,
Advances in Intelligent Systems and Computing 593, DOI 10.1007/978-3-319-60585-2_14

that can augment humans by supporting them with data analysis, hypothesis generation, and confirming or denying key attributes or indicators of compromise.

When considering such a human-machine teaming construct, representational issues arise as data in the domain describes topological and logical representations that do not always correlate to visual-spatial representations that are first-class constructs in physical domains. A question also surfaces as to the degree to which such a teaming arrangement requires the personification of the synthetic teammate. This is a fundamental question one must answer because it drives the need for whether natural interaction is required or not (e.g. Siri or some form of augmented reality). For cyberspace operations where there are aspects that are similar in physical domains, such as command and control, maneuver, fires, etc., personification may be an important aspect to the design as it helps support explain-ability and ultimately trustworthiness.

We begin by reviewing what is meant by cyberspace operations. We then apply human-machine teaming concepts to cyberspace operations. Following this discussion, we present some of the representational and interface considerations inherent to building trust for human-machine teaming in cyberspace operations. We then conclude with aspects of our future work.

2 Teaming in Cyberspace Operations

Cyberspace operations are actions conducted in cyberspace—the information environment created when we connect computational nodes together through some physical and logical transmission medium such as Ethernet, fiber, or RF [2]. It includes both cyber-pure or cyber-physical systems, which are systems where these compute nodes receive input from sensors in the physical world or compute solutions that cause an effect on an electro-mechanical actuator in the physical world. Examples of cyber-physical systems include automobiles, electrical power plants, robotics platforms, and military weapon systems.

The cyberspace environment also includes a human element—the cognitive and social factors that enable human interaction through this environment. Direct communication is part of this interaction, but the environment supports a much broader array of behaviors between humans. Examples include social meeting places where the exchange of ideas occur, economic activity and transactions, monitoring and controlling of physical systems, and malicious activity such as stealing information or money, or perhaps worse, physical damage to systems [3].

It follows that from a military perspective, cyberspace operations are pro-active actions to defend these cyber-pure and cyber-physical systems from an active adversary in order to retain freedom of maneuver (defensive) while projecting power to achieve military objectives (offensive). The use of the traditional military functions of intelligence, maneuver, fire support, protection, sustainment, and command and control are important in achieving these objectives as well as the integration of cyberspace actions into physical domains (i.e. land, sea, air, space). The integration of these functions and domains demands teaming at tactical, operational, and strategic levels. This research focuses on tactical-level teaming, specifically between agents and humans.

The types of teams in consideration are the Cyber Mission Forces (CMF), which the U.S. Department of Defense established after the standup of U.S. Cyber Command [4]. The CMF is composed of the teams that are the maneuver elements executing cyberspace actions such as reconnaissance, defense, and attack to achieve both defensive and offensive oriented goals.

Large companies increasingly are applying more military-style processes and techniques to drive their cybersecurity operations, so these observations will apply there also. Security operations centers (SOCs), share some similarities with certain CMF teams, where individuals work collectively to maintain persistent observation of the information flowing in and out of that organization while actively searching for potential compromises. This effectively changes these cybsersecurity teams from a reactive security posture to a proactive defensive posture.

An example of this proactive defense are the procedures, techniques, and tools that these teams employ to support cyber threat hunting [5, 6]. The ability to hunt for adversarial threats across networks of enterprise-scale is becoming an increasingly important part of the CMF and a SOC's tactics, techniques, and procedures (TTPs). The goal of hunting is to identify malicious behavior in an organization's network through indicators of compromise (IOC). IOCs include hash values of malicious software; Internet or domain name addresses; host-based (e.g. logs) or network-based evidence (e.g. netflow data); harvested malware binaries or source code (e.g. implants, command and control malware); and, at a more abstract level, adversary TTPs. Identifying adversary tools and TTPs are the most valuable evidence as they are the costliest for an adversary to change.

Threat hunting uses open-source or classified threat intelligence, that when combined with the organization's asset inventory and known vulnerabilities, facilitates generation of hypotheses as to where potential adversaries may, or already have, compromised systems. These hypotheses focus the team's attention on specific aspects of the data to determine if a compromise has occurred and how it might have happened. This information is then feed into an overall representation of the situation generating new hypotheses and repeating the cycle (Fig. 1a).

As this activity is very much conducive to task-decomposition and requires a somewhat persistent presence, cyber-threat hunting is typically carried out by a small team of individuals composed of different skill sets. Figure 1b lists example work roles (operator, analyst, planner, leader) that might make up such a team. As current state of the art for hunting is resource intensive, especially when considering a network on the order of magnitude of 10–100K nodes, there is a need for automated, or autonomous, tools that offload cognitive tasks performed by operators and analysts, enabling them to hunt more efficiently so that measures such as the number of breaches, dwell time (i.e. how long an adversary in the organization's network), and response time can improve. Some of the activity is conducive to automation such as some of the operators and analysts' functions and thus favorable for human-machine teaming.

The ultimate goal of our research is to reduce the workload requirements for cyberspace operators, analysts, and planners so that they can spend more time comprehending and responding to the broader situation. We have demonstrated progress in building autonomous cognitive agent models to support training [8]. These agents work independently of an overall team, avoiding issues such as trust, communication, and human-machine interfaces, although we have incorporated some of the representations described in Sect. 4.

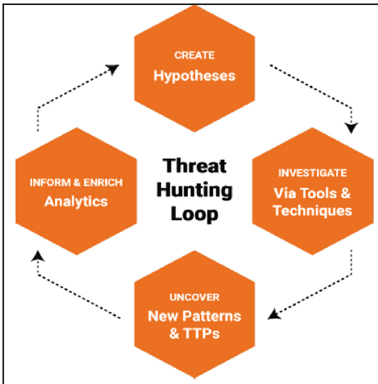
 CREATE Hypotheses INFORM & ENRICH Analytics Threat Hunting Loop INVESTIGATE Via Tools & Techniques UNCOVER New Patterns & TTPs	• **Operator** – Collects and processes information from systems to exploit, locate, and/or track targets of interest. Navigates network and conducts immediate forensics analysis. • **Analyst** – Uses collected information (e.g. threat intelligence, IDS alerts, firewall logs, network traffic logs) to identify, analyze, and report threats that are occurring or might occur. Works closely with operators to help determine their next steps. Works closely with planners to identify and validate requirements for analysis. • **Planner** – Develops detailed plans to satisfy operational requirements. Collaborates with other planners and analysts to identify, validate, and levy requirements for collection and analysis. • **Leader** – Synchronizes activity across the team, executing decision-making authority and establishing vision and direction for an organization's operations. Keeps external organizations and leaders informed of planned and ongoing activities.
(a) Cyber threat hunting loop ([6])	(b) Work roles (adapted from [7])

Fig. 1. Cyber threat hunting loop and workroles

3 Human-Machine Teaming for Cyberspace Operations

There are several aspects of human-machine teaming that have been well studied, many revolving around the issue of trust [9]. The factors associated with trust are also important for human-machine teaming in cyberspace/cybersecurity operations where the state of the practice is transitioning from five-year-old soccer, where teammates bunch around the moving ball, to fourteen-year-old soccer where the teammates play their positions. As organizational structures and processes for cyberspace operations mature, the ability to include autonomous agents as part of the teaming structure to facilitate and reduce human workload becomes more feasible and practical.

For example, Abbass et al. [9] illustrate key components for human-machine teaming for autonomous systems (Fig. 2), connecting desired supporting behaviors with what others [10] argue are the baseline functions for teaming: *information exchange*, *communication*, *shared understanding*, and *communication of human intent* (depicted by the four boxes in the bottom left-hand corner of Fig. 2).

Sycara and Lewis [10] point out that the exchange of information, supported by communication, requires bringing to bear all relevant sources of knowledge given the current situational context (e.g. perceptions, past experiences, current internal state). This communication requires internal semantic representations and interfaces that are both general across many functions but also specific to the domain of interest while

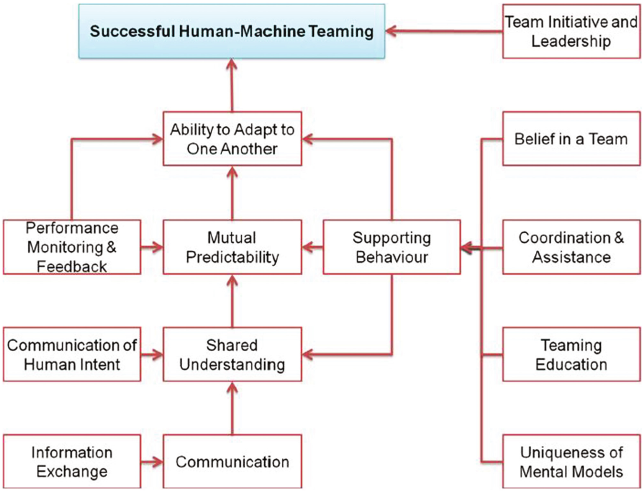

Successful Human-Machine Teaming
Team Initiative and Leadership
Ability to Adapt to One Another
Belief in a Team
Performance Monitoring & Feedback
Mutual Predictability
Supporting Behaviour
Coordination & Assistance
Communication of Human Intent
Shared Understanding
Teaming Education
Information Exchange
Communication
Uniqueness of Mental Models

Fig. 2. Model for successful human-machine teaming [9]

ensuring that noisy data or visual clutter is filtered as much as possible. To support such teaming, they argue architectures should support cognitive processes (e.g. decision-making, planning, reasoning) along with behaviors to support multi-entity relationships (situation assessment and monitoring).

For example, in the cyber threat hunting activity described in the previous section, we have found that an understanding of network topology is important to human operators and analysts [1]. However, the 10–100K nodes and corresponding topological connections that are an organization's network map are not all of equal importance for any particular task. Rather it is typically a few (1–2) compute nodes that are of interest (e.g. a compute node that has been identified with malware) along with a small number of connecting nodes (e.g. 3–10 nodes that are local one-hop connections, organization boundary nodes, and external nodes communicating with the node(s) of interest). There are currently limited ways to communicate a select set of nodes to another application, let alone, to an autonomous agent that may be assisting a hunt activity. Furthermore, the way to internally represent a small subset of a network within an agent's limited capacity for representing knowledge is not well understood.

Also, important in human-machine teaming is the ability to communicate human intent and tasking to an agent [10]. Shared understanding arises between human and machine when the underlying knowledge *representation* supports a two-way dialogue, to include an agent receiving and incorporating a human's intent into its own internal representation and *presenting* the results of its internal processing through a natural interface and in a format that is comprehensible to humans.

An example of where approaches fall short in this regard, are deep learning agents. It is difficult to convey a human's intent to a deep learning architecture in order to direct the system to perform a specific task outside of the task the learned model was trained to recognize. The learned model is trained to classify a particular set of objects. Asking it to classify additional sets of objects or other classes of objects requires retraining the model.

The results of a deep learner's computations and how it inferred its results are not explainable to a human. This is of concern in situations where communication of human intent is important or in adversarial settings where the results may be in question [11]. So, although deep learning performs specific tasks very well (e.g. image recognition), it points to the need for other representations and processes that can support the incorporation of human intentions while explaining recommended actions.

Again, using our cyber threat hunting example, machine learning techniques are now showing up in commercial products to support actions such as anti-malware detection or intrusion detection with limited understanding as to how such systems earn a human's trust and ultimately team with them. There are cases where human analysts have ignored the results of a security product due to a lack of trust in the system's recommendations. For example, during the 2013 exfiltration of credit card data from Target [12], one of the cybersecurity systems warned of the breach but the humans monitoring the system chose to ignore its alerts and turned off its ability to automatically delete malware resulting in 40 million credit card number stolen and a loss of at least 61 million dollars. This is an example where the system was trustworthy, but the humans did not trust its warnings. To help serve as a basis for human-machine teaming in the cyberspace domain, the next section begins to describe some potential avenues to pursue in regards to representation and presentation challenges.

4 Representation and Interfaces for Cyberspace Operations

As stated above, human-machine teaming is centered around trust, with the agent's internal representations and external interface primary factors in supporting information exchange, communication, shared understanding, and communication of human intent. Currently, knowledge representations to support semantics for cyberspace operations is not well understood. Equally important are interfaces that support natural, two-way dialogue and presentation of relevant material, tailored to a human's work role. Within the context of cyber threat hunting, CMF operators and analysts typically prefer command line interfaces with analysts also using web-based tools to support data query, filtering, and prioritization. Visualization of network mapping technology is improving but still rudimentary and displays for situational awareness lacking [1]. Planners and leaders are mostly relegated to presentations and documents to record and convey information—formats that are not conducive to human-machine interaction in a domain such as cyberspace where agility and speed are paramount.

4.1 Knowledge Representations

There has been a plethora of research on knowledge representation to support decision-making, planning, reasoning, and communicating in physical domains. Many symbolic, rule-based systems support knowledge-rich problem spaces where the structure and processing is primarily hand-crafted knowledge based on elicitation from subject matter experts. Such approaches are brittle and do not scale in complex environments where reasoning over concrete representations, such as images or raw malware binaries are necessary. However, these symbolic approaches have been shown to be more explainable and support incorporation of human intent and tasking.

More recently, non-symbolic, deep learning architectures have shown significant progress where the system learns an internal knowledge representation scheme by training it to match its input to a desired output (i.e. supervised learning) or to cluster the input data into groups that are similar (i.e. unsupervised learning). The processing in these systems applies mathematical manipulations by combining affine transformations with continuous functions that are converted to probabilities with a softmax function to classify input. During training of the model, backward processing applies gradient adjustments to weight parameters in order to minimize loss. Despite showing great promise for classification tasks, deep neural networks have limited capacity to reason about their actions and suffer from shortfalls discussed in previously.

Rather than settling on either representation, we have found that support for mixed symbolic and non-symbolic approaches through fixed architectural mechanisms and perceptual interfaces are generalizable across multiple domains [13]. For example, in [14] we demonstrate mixed modality symbolic and non-symbolic representations for visual-spatial domains such as simulations or robotics, where the non-symbolic representations are manifested in the form or mental imagery processing (Table 1).

Amodal, *symbolic* representations are useful for general reasoning and explanations. In physical domains, symbols may denote an object, and visual properties of the object, and qualitative spatial relationships between objects. The first row in Table 1 represents two objects (tree, house) and some qualitative visual and spatial properties (green, left-of).

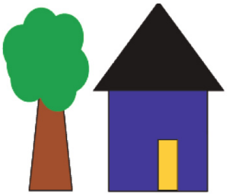
The *non-symbolic, spatial* representation is also amodal, although perceptual-based in that it is an interpretation of senses asserting the location, orientation, and rough shape of objects in space. Spatial processing is accomplished with sentential, mathematical equations. The second row in Table 1 represents the metric location, orientation, and rough shape of a tree and the house. Direction, distances between objects, size, and rough topology can be inferred implicitly from this information.

In contrast to the symbolic and spatial representation, both of which are sentential structures, space, including empty space, is inherent in the visual depictive representation that is based on the raw, perceived or stored data. Computationally, the depiction is a bitmap where the processing uses either mathematical manipulations (e.g., filters or affine transformations) or specialized processing that takes advantage of the topological structure. Both the symbolic and non-symbolic representations have functional and computational trade-offs that specific tasks often highlight. For example, given appropriate inference rules and the symbolic representation in Table 1, one can infer that the green object (tree) is to the left of the blue object (house). However, one cannot infer the distance between the tree and the house or that the top of the house is shaped like a triangle. One can infer these properties from a symbolic representation only when the relevant property is encoded explicitly or when task knowledge supports the inference. Thus symbolic, top-down processing, when augmented with bottom-up, data driven non-symbolic processing provides wider coverage to multiple classes of problems.

Table 1. Symbolic and non-symbolic representations for visual-spatial processing

Representation	Information	Processing	Example

Symbolic	• Object identities
	• Qualitative spatial and visual properties
		Symbolic manipulation/ productions	object(tree)
			color (tree, green)
			left-of(tree, house)

Spatial (non-symbolic)	• Object labels • Quantitative spatial and visual properties ○ Shape ○ Location, Direction, Orientation ○ Size ○ Topology	Mathematical manipulation	tree: location: <-2,4,0> orientation: 0 shape coordinates: <1,3,1>;<2,8,1>;<1,3,0>… house: location: <9,4,0> orientation: 0 shape coordinates: <8,3,1>;<2,3,1>;<4,3,0>…
Visual (non-symbolic)	• Object labels • Visual properties ○ Shape ○ Texture ○ Empty space • Spatial properties ○ Location, Direction ○ Size ○ Topology	Mathematical and depictive manipulations	

When applying this form of symbolic and non-symbolic representation to cyberspace operations some of the semantic interpretation breaks apart. For example, literal metric distance (*spatial*) between two compute nodes in cyberspace has little meaning, but distance in terms of latency (*temporal)* or the number of hops between two nodes (*logical and topological)* has relevant meaning both in a logical and in a geographic sense (i.e. geographic location may be inferred based on latency and other sources of information). In cyberspace operations topological, logical, and temporal relationships are first class constructs. The semantics of the visual and non-visual properties of compute nodes, their logical bindings (e.g. IP addresses), their software artifacts (e.g. files, processes) and spatial relationships with other nodes must be explicitly represented in any knowledge representation scheme.

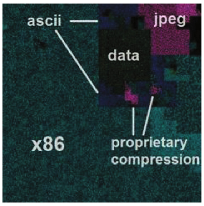
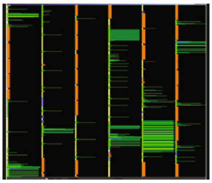
Our hypothesis then is that the symbolic and non-symbolic representations used in physical domains apply in the cyberspace domain, but that the semantic interpretation of these features and relationships differ. Table 2 summarizes some of these differences. For example, symbolic objects in the cyberspace domain might be a hardware compute node, its operating system, applications running on that processing node (to include potential malicious applications, and human users interacting with that node such as normal users, system administrators, and remote adversaries. Topological relationships might include connectivity between nodes or qualitative spatial relationships such as the fact that a certain file is stored on a specific node. Such objects represent the physical, logical, and social-cognitive layers of cyberspace [2]. Note that these representations may not necessarily be stored within the agent's memories but rather may exist on an external system with which an agent interacts.

Quantitative spatial relationships may include distance, as previously discussed, or other relationships such as direction, described by a network interface (*logical*) vice a degrees or orientation (*spatial*). Location might imply a physical, medium access layer numeric (i.e. a MAC address), a logical address (e.g. an IP address), a listening port (e.g. a TCP port), a geographic location, or some combination. As discussed above, such quantitative relationships combine symbolic labels with concrete numeric (non-symbolic) information. Finally, visual representations in cyberspace operations

Table 2. Symbolic and non-symbolic representations for cyberspace processing

Representation	Information	Processing	Example

Symbolic	• Object identities
	• Qualitative spatial, visual, and non-visual properties
		Symbolic manipulation/ productions	node(node-1)
			binary(file-1); on (file-1, node-1)
			connected (node-1, node-2)

Spatial (non-symbolic)	• Object labels • Quantitative spatial, non-visual, and visual properties ○ Shape – not defined ○ Location (physical network) ○ Location (logical network) ○ Location (organization) ○ Location (geolocation) ○ Orientation ○ Size (e.g. file size, packet size) ○ Topology ○ Direction (e.g. network interface)	Mathematical manipulation	node-1: location (net): 192.168.1.1 location (geo): <1,3,1> direction: <eth0> direction: <eth1> node-2: location (net): 192.168.1.2 location (geo): <1,2,1> direction: <eth0> file-1: size: 215KB connection: <node-1, eth1> <node-2, eth0> distance: 5ms
Visual (non-symbolic)	• Object labels • Visual properties ○ Shape ○ Texture ○ Empty space • Spatial properties ○ Location ○ Size ○ Topology ○ Direction	Mathematical and depictive manipulations	 Executable  Network packets

that the agent might use to reason over for functional or efficiency gains or to present to the human user for further analysis could include visualizations of binary data such as executables, network packets, or file types within a directory structure [15].

As we have found in physical domains, our hypothesis is that the use of these hybrid approaches can afford efficient processing and provide additional functionality for a certain class of problems with cyberspace. For example, in cyber threat hunting operations, an agent may need to measure distance between communicating nodes by sending a *ping* request and measure latency. Non-symbolic, deep neural networks may provide some of the sensing infrastructure with the symbolic classifications received as perceptual input to the agent. The mix between symbolic and non-symbolic processing then provides support for decision-making and learning over multiple time scales while providing explanation-based representations in the form of symbolic knowledge. These representations are important not only for an agent's own internal processing but also supports interfacing with human teammates.

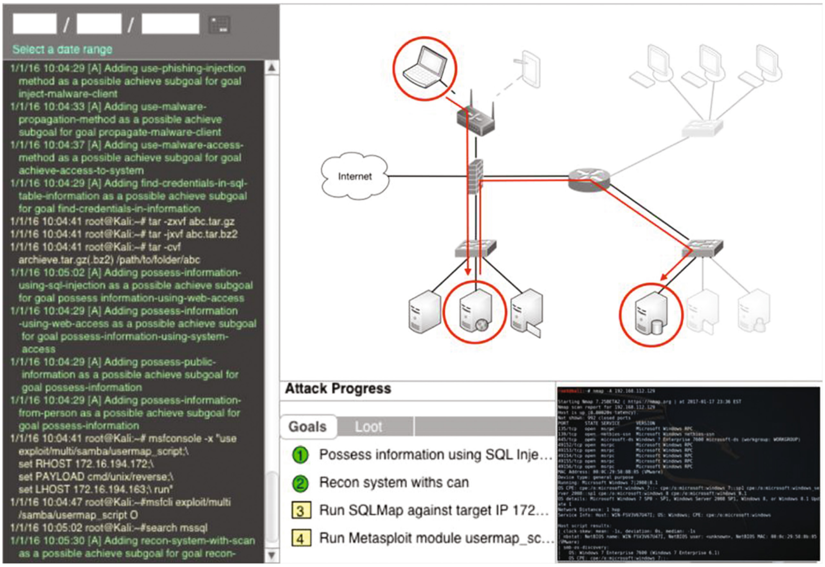
4.2 Interfaces to Support Two-Way Dialogue

We have found that developing usable human-agent interfaces for teaming requires not only an agent's internal knowledge representation as described above, but also maintenance of a model of the user to help understand their current information needs. This requires understanding users in context, making sense of the user's input, translating that input into a representation that the agent can process and store internally, and then taking the results of the agent's decision-making process across multiple time scales and presenting it to the user in natural ways.

Our research has provided much insight into how this interaction occurs in a human-machine teaming scenario involving unmanned systems [16]. However, we have not applied these lessons for cyberspace agents. Our hypothesis is that many of the techniques we have used for robot-human teaming will also apply here. For example, the interactive devices we have prototyped and employed, enable supervisory control providing the user with the ability to issue high-level commands to the robot with the robot providing feedback to maintain the user's situational awareness. These interactions are through natural interfaces, such as speech, gesture, sketch. To support such interaction, interface devices must have their own level of sophistication with modules to support *dialog management*, *human comprehension model,* and *planning and execution*.

In many cases the combination of multiple modes can help clarify the situation for the agent and build the human's trust that the agent understands the current task. For example, using a prototype interface in Fig. 3, a human cyber hunt analyst may task an agent by circling a node on a network graph and then stating "search for btw.z in the registry keys and identify any anomalous external nodes that *it* is communicating with". The agent interprets the *it* as the node that was circled by the analyst and, after conducting an DNS name lookup on the node may backbrief the analyst by stating, "searching web-1.acme.com for btw.z and calls to suspicious external nodes." As part of the feedback, that agent may visually show a subset of the network graph and highlight the communication path between the compromised internal node and an external command and control server.

We have also explored the use of augmented reality interfaces for human-machine interaction for robotics and Army battle staffs, finding that these interfaces work best when the overlay of control graphics or non-visible entities is important for the operation. Others have investigated the impact on cognitive workload when using augmented reality for SOCs [17]. Their research found that subjects wearing the devices reported reduced cognitive workload, performing the primary cyber-related tasks more efficiently, and responding to ancillary events more successfully. Such approaches may be useful in continuous monitoring situations where hunt operators or analysts need to move away periodically from the display to check on a physical computer.



```
1/1/16 10:04:29 [A] Adding use-phishing-injection
  method as a possible achieve subgoal for goal
  inject-malware-client
1/1/16 10:04:33 [A] Adding use-malware-
  propagation-method as a possible achieve
  subgoal for goal propagate-malware-client
1/1/16 10:04:37 [A] Adding use-malware-access-
  method as a possible achieve subgoal for goal
  achieve-access-to-system
1/1/16 10:04:29 [A] Adding find-credentials-in-sql-
  table-information as a possible achieve subgoal
  for goal find-credentials-in-information
1/1/16 10:04:41 root@Kali:~# tar -zxvf abc.tar.gz
1/1/16 10:04:41 root@Kali:~# tar -jxvf abc.tar.bz2
1/1/16 10:04:41 root@Kali:~# tar -cvf
  archieve.tar.gz(.bz2) /path/to/folder/abc
1/1/16 10:05:02 [A] Adding possess-information-
  using-sql-injection as a possible achieve subgoal
  for goal possess information-using-web-access
1/1/16 10:04:29 [A] Adding possess-information
  -using-web-access as a possible achieve subgoal
   for goal possess-information-using-system-
   access
1/1/16 10:04:29 [A] Adding possess-public-
   information as a possible achieve subgoal for
   goal possess-information
1/1/16 10:04:29 [A] Adding possess-information-
  from-person as a possible achieve subgoal for
  goal possess-information
1/1/16 10:04:41 root@Kali:~# msfconsole -x "use
  exploit/multi/samba/usermap_script;\
  set RHOST 172.16.194.172;\
  set PAYLOAD cmd/unix/reverse;\
  set LHOST 172.16.194.163;\ run"
1/1/16 10:04:47 root@Kali:~#msfcli exploit/multi
  /samba/usermap_script O
1/1/16 10:05:02 root@Kali:~#search mssql
1/1/16 10:05:30 [A] Adding recon-system-with-scan
  as a possible achieve subgoal for goal recon-
```

1 Possess information using SQL Inje...
2 Recon system withs can
3 Run SQLMap against target IP 172...
4 Run Metasploit module usermap_sc...

```
root@kali:~# nmap -A 192.168.112.129

Starting Nmap 7.25BETA2 ( https://nmap.org ) at 2017-01-17 23:36 EST
Nmap scan report for 192.168.112.129
Host is up (0.00020s latency).
Not shown: 992 closed ports
PORT      STATE SERVICE      VERSION
135/tcp   open  msrpc        Microsoft Windows RPC
139/tcp   open  netbios-ssn  Microsoft Windows netbios-ssn
445/tcp   open  microsoft-ds Windows 7 Enterprise 7600 microsoft-ds (workgroup: WORKGROUP)
49152/tcp open  msrpc        Microsoft Windows RPC
49153/tcp open  msrpc        Microsoft Windows RPC
49154/tcp open  msrpc        Microsoft Windows RPC
49155/tcp open  msrpc        Microsoft Windows RPC
49156/tcp open  msrpc        Microsoft Windows RPC
MAC Address: 00:0C:29:58:08:05 (VMware)
Device type: general purpose
Running: Microsoft Windows 7|2008|8.1
OS CPE: cpe:/o:microsoft:windows_7::- cpe:/o:microsoft:windows_7::sp1 cpe:/o:microsoft:windows_se
rver_2008::sp1 cpe:/o:microsoft:windows_8 cpe:/o:microsoft:windows_8.1
OS details: Microsoft Windows 7 SP0 - SP1, Windows Server 2008 SP1, Windows 8, or Windows 8.1 Upd
ate 1
Network Distance: 1 hop
Service Info: Host: WIN-FSV3V67O47I; OS: Windows; CPE: cpe:/o:microsoft:windows

Host script results:
|_clock-skew: mean: -1s, deviation: 0s, median: -1s
|_nbstat: NetBIOS name: WIN-FSV3V67O47I, NetBIOS user: <unknown>, NetBIOS MAC: 00:0c:29:58:08:05
(VMware)
| smb-os-discovery:
|   OS: Windows 7 Enterprise 7600 (Windows 7 Enterprise 6.1)
|   OS CPE: cpe:/o:microsoft:windows_7::-
```

Fig. 3. Mockup use of sketch, visual, and speech to interface with cyberspace agent

5 Conclusion

This paper explores the interaction of humans with autonomous, intelligent agents working together as teammates in cyberspace operations. The ultimate goal of our research is to reduce workload requirements for cyberspace operators, analysts, and planners so that they can spend more time comprehending and responding to the broader threat.

To support communication, sharing of human intent, and explain ability, symbolic and non-symbolic knowledge representations were explored. Representational challenges unique to cyberspace operations are unlike physical domains where spatial and visual properties and relationships provide concrete interpretations of the world model.

Topological, logical, and temporal relationships are first class constructs in cyberspace, requiring a semantic interpretation of common properties and relationships such as distance, direction, location, and connectedness.

Future work will continue to investigate and prototype these agents with the proposed knowledge representation and natural interaction schemes for cyberspace operations while exploring how malicious adversaries can potential violate these mechanisms in support of their own goals.

References

1. Lathrop, S.D., Trent, S., Hoffman, R.: Applying human factors research towards cyberspace operations: a practitioner's perspective. In: Advances in Human Factors in Cybersecurity, pp. 281–293. Springer (2016)
2. Joint Publication 3–12, Cyberspace Operations (2013)
3. Lee, R.M., Assante, M.J., Conway, T.: Analysis of the cyber attack on the Ukrainian power grid (Traffic Light Protocol (TLP) White). Electrical Information Sharing and Analysis Center (2016)
4. The Department of Defense Cyber Strategy (2015). http://www.defense.gov/Portals/1/features/2015/0415_cyber-strategy/Final_2015_DoD_CYBER_STRATEGY_for_web.pdf
5. Toussain, M.: Home-Field Advantage: Using Indicators of Compromise to Hunt Down the Advanced Persistent Threat. SANS Institute InfoSec Reading Room (2014)
6. Cyber Threat Hunting. https://sqrrl.com/solutions/cyber-threat-hunting/
7. Newhouse, B., Keith, S., Schribner, B., Witte, G.: NIST SP 800-181, NICE Cybersecurity Workforce Framework, National Initiative for Cybersecurity Education (Draft) (2016)
8. Jones, R.M., O'Grady, R., Nicholson, D., Hoffman, R., Bunch, L., Bradshaw, J., Bolton, A.: Modeling and integrating cognitive agents within the emerging cyber domain. In: Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC), vol. 20 (2015)
9. Abbass, H.A., Petraki, E., Merrick, K., Harvey, J., Barlow, M.: Trusted autonomy and cognitive cyber symbiosis: open challenges. Cogn. Comput. **8**(3), 385–408 (2016)
10. Sycara K, Lewis M. Integrating intelligent agents into human teams. In: Salas, E., Fiore, S., (eds.) Team Cognition: Process and Performance at the Inter and Intra-individual Level, Washington. American Psychological Association (2004)
11. Huang, L., Antony, J.D., Nelson, B., Rubinstein, B.I.P., Tygar, J.D.: Adversarial machine learning. In: The 4th ACM Workshop on Artificial Intelligence and Security, Chicago, IL (2011)
12. Riley, M., Elgin, B., Lawrence, D., Matlack, C.: Missed Alarms and 40 Million Stolen Credit Card Numbers: How Target Blew It, Bloomberg.com (2016). http://www.bloomberg.com/news/articles/2014-03-13/target-missed-warnings-in-epic-hack-of-credit-card-data. Accessed 9 Apr 2016
13. Laird, J.E.: The Soar Cognitive Architecture. MIT Press, Cambridge (2012)
14. Lathrop, S.D., Wintermute, S., Laird, J.E.: Exploring the functional advantages of spatial and visual cognition from an architectural perspective. Top. Cogn. Sci. **3**(4), 796–818 (2010)
15. Conti, G.: Security Data Visualization: Graphical Techniques for Network Analysis. No Starch Press, San Francisco (2007)

16. Taylor, G., Purman, B., Schermerhorn, P., Garcia-Sampedro, G., Lanting, M., Quist, M., Kawatsu, C.: Natural interaction for unmanned systems. In: SPIE Defense+Security, pp. 946805–946805. International Society for Optics and Photonics (2015)
17. Beitzel, S., Dykstra, J., Huver, S., Kaplan, M., Loushine, M., Youzwak, J.: Cognitive performance impact of augmented reality for network operations tasks. In: Advances in Human Factors in Cybersecurity, pp. 139–151. Springer (2016)

Valuing Information Security from a Phishing Attack

Kenneth D. Nguyen[1](✉), Heather Rosoff[2], and Richard S. John[1]

[1] Department of Psychology, University of Southern California,
Los Angeles, CA, USA
{hoangdun, richardj}@usc.edu

[2] Center for Risk and Economic Analysis of Terrorism Events,
University of Southern California, Los Angeles, CA, USA
rosoff@usc.edu

Abstract. In most cyber security contexts, users need to make trade-offs for information security. This research examined this issue by quantifying the relative value of information security within a value system that comprises of multiple conflicting objectives. Using this quantification as a platform, this research also examined the effect of different usage contexts on information security concern. Users were asked to indicate how much loss in productivity and time, and how much more money they were willing to incur to acquire an effective phishing filter. The results indicated that users prioritize productivity and time over information security while there was much more heterogeneity in the concern about cost. The value of information security was insignificantly different across different usage contexts. The relative value of information security was found to be predictive of self-reported online security behaviors. These results offer valuable implications for the design of a more usable information security system.

Keywords: Security trade-offs · Phishing · Multi-attribute utility

1 Introduction

Because phishing attacks can result in severe consequences, a number of approaches have been initiated to help users become more aware of and learn to protect themselves from phishing attacks, and increase information security in general. Two of the most commonly used strategies include (1) providing users with information security training and (2) equipping users with technologies designed for information security purposes [1]. However, these approaches have not been very successful in keeping internet users from becoming victims of cyber attacks. For example, about 20% of respondents in a recent survey indicated that they did not know how to protect their information in cyber space [2]. In addition, those lacking necessary skills to implement security technology were also reluctant to pay for services that can improve their security system [3]. Similarly, studies have indicated that conventional security training, in which users simply receive education materials on phishing attacks, do not significantly increase safe online behaviors [4].

D. Nicholson (ed.), *Advances in Human Factors in Cybersecurity*,
Advances in Intelligent Systems and Computing 593, DOI 10.1007/978-3-319-60585-2_15

The limitations of these approaches raise the question of how best to motivate users to engage in safe online behaviors. One approach is to consider values users consider important when adopting new technology for information security purpose. In psychology, "values" have been defined as a cognitive representation of needs [5]. Although individuals differ in how they rank the importance of specific values [6], it is generally agreed that psychological values are important to maintain in the long run [7]. Thus, information security products that offer high value to users will be adopted quickly, whereas products that offer little value to users are unlikely to ever gain acceptance [8]. For example, the effectiveness of the interactive training involving pseudo phishing attacks [9] may be enhanced by learning the values users appreciate and incorporating these values into the training content. Specifically, because users want to access legitimate online material, the training may become more effective by focusing on teaching users to recognize safe online content from phishing attempts as opposed to focusing solely on recognizing cues of phishing attacks.

However, understanding the values and concerns that users have in a complex task usually involves accounting for multiple objectives [10, 11]. The situation becomes even more challenging as these objectives are often conflicting, in the sense that getting more of one means giving up performance on another [12]. A usable product may sacrifice some security features (and vice versa), so both designers and users may be required to make hard choices involving trade-offs between these conflicting objectives. The current research presents an approach to explore the values users may have when adopting a technology designed to increase information security. Our model is based on multi-attribute utility theory (MAUT), a quantitative framework that is often used to evaluate decision problems involving multiple conflicting objectives [11]. The mathematical formulation is expressed as:

$$\mathrm{U}(\mathrm{X}_1, \mathrm{X}_2 \ldots \mathrm{Xn}) = \sum\nolimits_{i=1}^{n} w_i u_i(x_i). \tag{1}$$

"U" is the multi-attribute utility of a decision's outcome and the vector "X" represents an alternative characterized on each of the n attributes, operationalizing each objective. The lower case "x" indicates the scale values for alternative X on each of the respective attributes. The indexed "u" represents the (standardized) utility transformation of the raw scale values, x, to a standardized unit interval accounting for (decreasing or increasing) marginal value and risk attitude; and "w" is a scaling constant representing the exchange rates between conflicting attributes ($\sum \mathrm{w}_\mathrm{i} = 1.0$). A rational decision maker is presumed to select the alternative with the maximum (multi-attribute) utility.

In the current research context, a user value model can be developed to describe different conflicting objectives that users may have when evaluating a technology designed to enhance information security. Such value models are useful in many applied contexts. For example, different users can develop their own models to evaluate and select the best option, accounting for the objective of maximizing information security and other personal priorities, e.g., maximizing convenience. Similarly, information system developers can rely on such models to improve their products or services by examining the discrepancy between their current products and users' desires.

This approach is likely to add more values in their applications, hence increasing the acceptance of their products and services [12].

Certainly, each individual user will have an idiosyncratic set of objectives and each decision problem or application will require a unique set of evaluation criteria. Thus, it would be impossible to describe an exhaustive set of objectives that addresses all of users' concerns. Our focus in this study, therefore, is to describe *in general* how individual decision makers value security protection within a set of conflicting and desirable objectives that are relevant, independent, and relatively complete [13]. Furthermore, the selection of some of the objectives is motivated by the Technology Acceptance Model [14]. For instance, the "usability" factor in the model can be translated in terms of the time it takes users to interact with a security tool, which is the minimizing latency objective in our model whereas the "usefulness" factor is conceptually related to the maximizing productivity objective. The selection of the cost objective is motivated by the fact that commercial phishing filters are available for a cost to users whereas the inclusion of the security objective is an obvious choice.

We created a decision context in which users consider the purchase of a commercial phishing filter tool that guarantees a high detection rate of attacks to quantify the *relative importance of information security* in a value system that consists of multiple conflicting objectives. The decision context is characterized by a choice between two filters described on four attributes: *security, cost, latency, and productivity*, corresponding to four respective objectives: *maximizing security, minimizing cost*, *miming wait time (latency)*, and *maximizing productivity.* Users evaluate how much security protection is worth in terms of reductions in achievement on each of the other three non-security attributes. These trade-offs allow us to quantify the value of information security in multiple metrics, i.e., monetary cost, latency, and productivity. From these quantities, the relative value of information security can be quantified.

2 The Effect of Usage Contexts on Information Privacy Concern

Importantly, we used the valuation as a platform to investigate meaningful research questions. For example, because phishing can occur in different scenarios, an interesting research question is that whether and to what extent internet users value information security under different phishing contexts. Kujala and Väänänen-Vainio-Mattila [15] have argued that the value of a technology does not arise from its properties, but is contingent upon the interactions of users and the product in a particular situation. In other words, this argument implies that the value of information security is contingent upon different usage contexts. On the other hand, there are reasons to believe that value of information security is generalizable. These two perspectives suggest two contradictory hypotheses about the effect of usage context on value of information security. At one extreme, the *generalized security hypothesis* suggests that security premium(s) should be generalizable. This is because the value of a security tool should be judged the same as long as it returns the same benefit(s) regardless of the context in which it is used. For example, if a phishing filter successfully detects 20 phishing attempts out of

100 suspicious online contents, its *objective* value should not be altered whether the online contents are Facebook posts or pop-up alerts.

At the other extreme, the *context-specific hypothesis* posits that security premium(s) should be sensitive to usage context. Boiney [16] underscored the context-based usefulness of technology by suggesting that the same technology can provide different users with distinct benefits in unique settings. For example, users may be more familiar with and perhaps feel more competent in handling spam messages than in dealing with social media phishing attempts since the former have been around for a long time and the latter are a relatively new emerging threat. Thus, these perceptions and feelings may lead users to value their information protection largely in the social media context than in the email context.

The two aforementioned hypotheses suggest two opposite predictions. The generalized security hypothesis predicts that the security premiums are invariant across usage contexts while the context-specific hypothesis suggests otherwise. Methodologically, we explore the effect of various usage contexts by manipulating the context where a phishing attack can occur, including email, web browsing, and social media. Because there are few empirical studies relevant to this research question, instead of making a priori hypothesis, we explore whether the relative value of information security, as found in one usage, can be generalized to other contexts.

3 Method

3.1 Procedure

The experiment began with a four-minute video describing the study and carefully explaining how the attributes were defined. Each respondent was randomly assigned to one of the three phishing attack contexts: (1) email, (2) web browsing, and (3) social media conditions. Respondents completed trade-off assessments for all six possible attribute pairs, with up to three binary choices per assessment (18 in total). The focus of the current study is on the three trade-offs for security: (1) security versus cost, (2) security vs. latency, and (3) security vs. productivity. Respondents also reported the frequency that they engaged in certain types of self-protective security behaviors, responded to a perceived security vulnerability scale, a perceived severity scale, a scale that measures the cost-benefit ratio of implementing online security measures, and a measure of security self-efficacy. The psychometric scales were adapted from a previous study [10]. The inclusion of these scales allows us to explore the effects of individual characteristics. The experiment was hosted on Qualtrics.com and 275 respondents were recruited from Amazon Mechanical Turk (AMT). Previous studies have shown that AMT samples are generally more representative than other convenience samples [17–19]. The sample sizes under each of the three context conditions were 95 (email), 87 (pop up), and 93 (social media). The mean age was 35, and 46.54% of the respondents were female.

3.2 Attribute Definitions

The four user-centric objectives of interest are (1) maximizing information security, (2) minimizing cost, (3) minimizing latency, and (4) maximizing productivity (work and play). Security is defined as the miss rate or the number of phishing attempts that bypass a filter and appear in the users' inbox (web browser/social media platform) per 100 emails (pop-ups/apps) per year. Cost is defined as the monthly payment that respondents have to pay for a (email/web browsing/social media) phishing filter. Latency is operationalized as the time it takes the phishing filter to screen an email/pop-up/app before it is allowed (or disallowed) in the user's email inbox (web browser/social media platform). Finally, productivity is defined as the false alarm rate, or the number of valid contents that are misclassified as phishing attempts and diverted from the users' inbox (web browser/social media platform) per 100 emails (pop-ups/apps) per year.

3.3 Trade-off Elicitation Methodology

The detail of the trade-off elicitation has been described extensively in a previous publication [20], and we briefly describe the general procedure here. In each trade-off assessment that involves any two of the four attributes, respondents (users) were asked to choose between two phishing filter alternatives. The first phishing filter is more attractive than the second phishing filter in the first attribute but the opposite is true in the second attribute. Users were asked to indicate which option is more attractive, or they can indicate "indifference," meaning that they perceive the two alternatives as being equally good. We used this elicitation protocol to estimate users' trade-offs for information security against each of the other three attributes, i.e., cost, latency, and productivity.

Figure 1 graphically illustrates the elicitation procedure. We considered the trade-off between security and cost represented in a series of three binary-choice trials with two phishing filter options, A and B_i (i = 1… 7). Filter A is less effective in detecting phishing emails, but it is inexpensive. On the other hand, B_1 is more expensive but B_1 is more effective in identifying phishing emails. Users are asked to choose either A or B_1. Depending on the decision makers' choice in the first trial, the cost for B_1 is adjusted dynamically while the cost for A is fixed in the next trial; $B_2 > B_1$ if respondents choose B_1; conversely, $B_2 < B_1$ if A is chosen. The procedure is repeated until the respondent is indifferent between the two options or she completes the third trial, at which point the trade-off is bounded.

The dependent variable, a security premium, is determined by taking the difference in cost between two options whenever the respondent indicates indifference. If the respondent does not select the "indifference" option in any of the three trials, the premium for security protection is bounded using an inequality determined from the three trials[1]. For instance, if a respondent selects A in the first trial, B_2 in the second

[1] We could continue the elicitation beyond three choices, but this was deemed unnecessary as the purpose of the study was to bound the premiums. In addition, having up to three trials already allows us to specify fifteen (small) ranges of the premiums that users were willing to exchange for a higher level of information security (see the Appendix for more details).

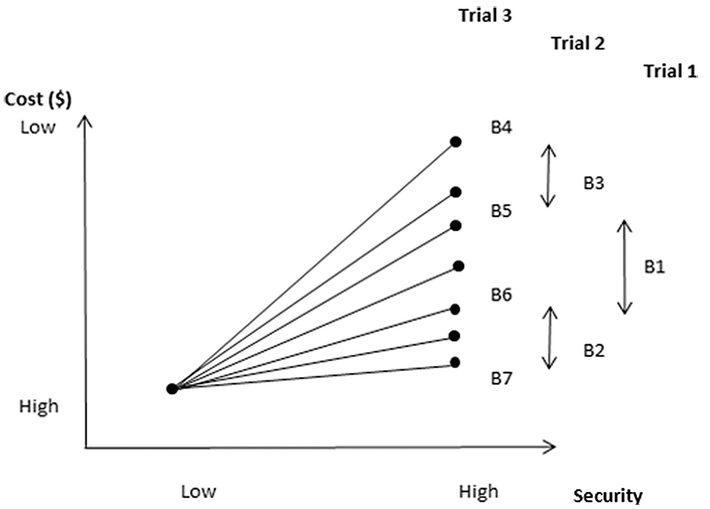


Fig. 1. Graphical illustration of the trade-off method

trial, and B_3 in the third trail, we could infer that the respondent is willing to pay between B_6 and B_1 ($B_6 < B_1$) dollars for a more effective phishing filter, i.e. the security premium in terms of dollars is in the range [B_6, B_1]. Elicitations of security premiums in terms of the other two attributes, latency and productivity, follow the same procedure. The appendix provides details on how the binary choices were constructed.

The following provides a concretize example of the trade-off procedure (see Fig. 2 for the choice presentation). Respondents are presented with two alternatives: filter A costs $5 and identifies 50 out of 100 phishing emails, labeled A ($5, 0.5), while filter B costs $10 and identifies 90 out of 100 phishing emails, labeled B ($10, 0.9). If a respondent prefers A ($5, 0.5) over B ($10, 0.9), then the implication is that she does not consider it worthwhile to pay an additional $5 to reduce the number of phishing emails by 40%. The choice is then repeated, but the cost of option B is reduced to $8 to make option B more attractive, B ($8, 0.9) If the respondent persists in choosing option A (to save money), the cost of option B is further reduced to $6, B ($6, 0.5). If the respondent is indifferent, then it is inferred that he is willing to pay an additional $1 ($6–$5) to reduce the number of phishing emails by 40%, i.e. the security premium in terms of money is $1.

3.4 Experimental Manipulation

The key procedure testing the generalizability versus context-specific hypotheses is the manipulation of the phishing context. As specified, manipulation was achieved by varying the phishing context (email, web browsing, social media). Specifically, phishing attacks can occur when users follow the instruction in a phishing email, which looks and feels as if it were from a valid entity, and submit their sensitive information. Phishing attacks through web browsing are another common context. Attackers carefully craft pop-up window alert messages to ask the recipient to enter sensitive information into a website from which the attacker then collects the data. A third context, social media phishing attacks, has become common recently, in particular through the use of fake applications and feeds on social media sites such as Facebook. Attackers carefully craft fake applications or feeds asking users to "like" or asking users to click on a malicious link. Social media users are then prompted to enter their sensitive information into a website from which the attacker then collects the data.

In this experiment, each respondent was presented with one of the phishing attack contexts in the introductory video. The video included four elements: (1) a general definition of a phishing attack, (2) a description of the phishing attack context, (3) the sequential binary choice task, and (4) definitions of the four attributes representing alternative phishing mitigation alternatives. Respondents received a unique version of the introductory video, corresponding to the context condition to which they were randomly assigned. The three videos differ in their descriptive languages and graphical images. For instance, respondents under the email phishing condition were told that they would be asked to select between *email phishing* filters while respondents under the social media condition were told to select between *social media phishing filters*. The visual images in the three videos are identical except for some modifications to fit each specific context (e.g. replacing an image of an email inbox with a Facebook app). All videos were audio-recorded by the same male research assistant, and the videos are nearly identical in length: (3 min, 54 s).

4 Results

4.1 Calculating Weights for Security, Cost, Latency, and Productivity

Attribute weights were computed for each respondent. The procedure to compute weight is described in [21]. In general, when respondents were indifferent between the two phishing filter options, they implied that the expected utility in option A equals to the expected utility in option B. This implication can be expressed mathematically as:

$$\mathrm{EU(A) = EU(B)}. \tag{2}$$

Where

$$\mathrm{EU(A) = w(attribute\ 1) * Utility(attribute\ 1.A) + w(attribute\ 2) * Utility\ (attribute\ \ 2.A).}$$
$$\mathrm{EU(B) = w(attribute\ 1) * Utility(attribute\ 1.B) + w(attribute\ 2) * Utility\ (attribute\ 2.B).}$$

The utility associated with each option was computed by assuming a linear utility function and from the trade-off values. Because each respondent completed multiple trade-off assessments, each respondent had a unique system of three equations. Using the additional normalization constraint such that the sum of the attribute weights is unity, we could find the scaling constants (or weights) for the four objectives: cost, latency, productivity, and safety.

4.2 Statistical Approach

Figure 2 plots the distributions of weights for the four objectives and across the three experimental groups. Examinations of the distributions revealed non-normality, suggesting a violation of one of the assumptions in using parametric statistics. Thus, all of the analyses were conducted by using non-parametric statistics. First, Kruskal-Wallis, a non-parametric version of the omnibus one-way ANOVA, was used to detect changes in distributions of security weight across the experimental groups. If there is a significant effect, a followed up Kolmogorov–Smirnov (KS) test would be used for pairwise comparison. Correlation analyses were used to explore the predictive validity of security weight whereas multiple regression analyses were conducted to explore the role of individual differences.

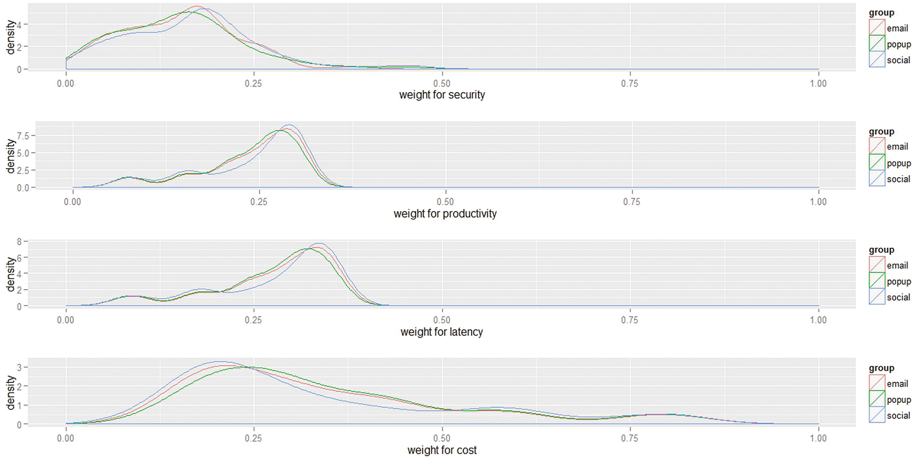


Fig. 2. Distributions of objective weights

4.3 The Effect of the Attacking Contexts on the Information Security Concern

Visual inspection suggests that there was little change in weight for security across contexts. Statistical tests confirmed this visual inspection. Kruskal-Wallis (KW) test returned a non-significant result KW(2) = 1.9805, p = 0.3715. Thus, this finding suggests the nil influence of the attacking contexts on preference for information security.

4.4 The Predictive Validity of Information Security Concern

Respondents were asked about the frequency that they engaged in the following online activities (0 = Never, 5 = Always): scanning computer for viruses (Virus), open unexpected pop-up alerts (Popup), erase cookies (Cookies), open an attachment from an unknown email (Email), report suspicious threats in social media sites (Report), share personal cell number on social media sites (Cell phone), share pictures of home on social media sites (Home), and share employment information on social media sites (Employment). Correlation analyses were conducted to explore the relationships between the relative value of security and the self-reported behavioral responses. The results suggest that security weight is minimally associated with most of the behavioral variables except for Cell Phone, Home, and Employment. In fact, statistical tests revealed that the higher the relative value of information security is, the less likely that users were willing to share their cell phone number in social media sites, 95% CI [−0.28, −0.17], p = .01, and employment information in social networks pages, 95% CI [−0.25, −0.14], p = .02. The correlation between security weight and Home was marginally significant, p = .06.

4.5 Exploring the Role of Individual Differences in Online Behaviors

The aforementioned relationships between the relative value of information security and self-reported protective behaviors could be due to third factors. For example, a respondent value security more also perceives a higher risk of becoming victims of cyber hacks. We applied multiple regression analysis to explore this possibility. First, we created an index of security behaviors by averaging the two items Cell Phone and Home. Second, we regressed this index on the following variables: sex, age, perceived vulnerability, perceive severity, self-efficacy, response cost, and the experimental groups. The model is significant $F(8, 266) = 2.271$, $p = .023$. Table 1 is the summary of the results. Several interesting findings emerged. First, security weight significantly predicted self-reported security behaviors, controlling for the effects of other predictors. The higher the value of security was, the less likely that users were willing to engage in risky behaviors in cyber space. Female users were less likely than male users to engage in risky behaviors, controlling for the effects of other predictors. Interestingly, users who perceived themselves as being vulnerable to cyber security risks were less likely to engage in risky behaviors in cyber space.

Table 1. Multiple regression results

Predictors	Beta	SE	*t*	*p*
Intercept	0.000	0.059	0.000	1.000
Safety weight	−0.131	0.060	−2.167	.031*
Sex (1 = Female)	−0.124	0.062	−1.998	.047*
Group (1 = Social, 0 = Others)	−0.023	0.060	−0.377	.706
Age	−0.084	0.061	−1.381	.168
Severity	0.009	0.063	0.142	.887
Vulnerability	0.165	0.064	2.564	.0109*
Efficacy	−0.089	0.062	−1.435	.152
Response cost	−0.036	0.062	−0.579	.563

Note: * $p < .01$; Standardized estimates are shown

5 Discussion

This study applies Multi-Attribute Utility Theory to conceptualize the multidimensional value of information security. We explored how internet users made trade-offs for an enhanced information security product, a phishing filter, in terms of monetary value, loss in productivity, and wait time. Furthermore, we used these trade-offs as a platform to investigate individual differences regarding information security concern and the effect of usage context on the value of information security.

Figure 2 reveals that the relative value of security is non-zero, suggesting that users concerned about information security. Yet the distribution of security weight is closer to zero relatively to the distributions of productivity and time, implying that users prioritized their desire for faster processing time and productivity over the concern for information security. Interestingly, the distribution of weight for cost is much more spread out, suggesting that some users were much more willing to give up money to increase security (or to achieve a greater level in other objectives) whereas others considered minimizing cost as an ultimate concern.

These results call for a greater research attention on the multidimensional value of security. Previous studies on information security valuation often focus on the economics of privacy [22], i.e. how much is information privacy worth. Yet, users often concern about multiple objectives when making decisions in cyber space [23]. The current study, therefore, broadens the conceptualization of the concept *security value* in a broader context by highlighting the *relative value* of information security within a broad value system that contains multiple conflicting objectives.

This new conceptualization of security value advances our understanding on the *privacy paradox* phenomenon [24]. The paradox highlights the contradictory finding in research such that users often engage in risky behaviors in cyber space despite their sated concern for information security. In other words, the paradox suggests an insignificant relationship between attitude toward information security and security behaviors. On the contrary, results from this experiment suggest that the paradox is not paradoxical at all. This is because users do concern about information security, but this concern is simply weighed less than the concerns for cost, time, and productivity. The implication is that the relative value of security is a better predictor of users' behaviors, compared to the conventional self-reported measure of security concern. Indeed, the regression results suggest that security weight significantly predicted users' online behaviors. The more users valued security, the less likely they were willing to share their personal information in cyber space.

Importantly, using the security valuation as a platform, this study also examined two contradictory predictions regarding the effects of context on information security concern. The context-specific hypothesis predicts that users make different trade-offs for information security across different phishing contexts whereas the generalizability hypothesis suggests otherwise. We found empirical support for the latter hypothesis: the distributions of security weight did not differ across the experimental conditions. On the one hand, any nil experimental effects may be attributed to low power. In fact, the manipulation of the attacking context in this study was very subtle. We simply substituted a few words in the experimental script, and this subtlety may not be salient enough to change respondents' responses. On the other hand, if the concern about information security is generalizable, the finding in this study suggests that future research should explore factors that underlie the general concern for information security.

Most information systems require users to be aware of the potential security and privacy risks and to utilize some form of protective measures against these threats. However, behavioral research consistently demonstrates that people often give up their privacy for other desirable concerns, e.g. to maximize convenience [25]. This ubiquitous finding highlights the importance of integrating users' personal values in the design of cyber security systems because users have distinct multiple and conflicting priorities. Our research findings help to address this issue by considering how users consider security protection in relation to priorities that they highly value. As a result, these empirical findings are pragmatically valuable for the development of a more usable information security system.

References

1. Lwin, M., Wirtz, J., Williams, J.D.: Consumer online privacy concerns and responses: a power–responsibility equilibrium perspective. J. Acad. Mark. Sci. **35**, 572–585 (2007)
2. Paine, C., Reips, U., Stieger, S., et al.: Internet users' perceptions of 'privacy concerns' and 'privacy actions'. Int. J. Hum. Comput. Stud. **65**, 526–536 (2007)
3. Acquisti, A., Grossklags, J.: Privacy and rationality in individual decision making. IEEE Secur. Priv. **3**, 26–33 (2005)
4. Kumaraguru, P., Rhee, L., Acquisti, A., et al.: Protecting people from phishing: the design and evaluation of an embedded training email system. In: 25th Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 905–914. ACM, New York (2007)
5. Verplanken, B., Holland, R.W.: Motivated decision making: effects of activation and self-centrality of values on choices and behavior. J. Pers. Soc. Psychol. **82**, 434–447 (2002)
6. Isomursu, M., Isomursu, P., Ervasti, M., et al.: Understanding human values in adopting new technology—a case study and methodological discussion. Int. J. Hum. Comput. Stud. **69**, 183–200 (2011)
7. Jurison, J.: Perceived value and technology adoption across four end user groups. J. Organ. End User Comp. **12**, 21–28 (2000)
8. Tsai, J., Egelman, S., Cranor, L., Acquisti, A.: The effect of online privacy information on purchasing behavior: an experimental study. Inf. Syst. Res. **22**, 254–268 (2011)
9. Sheng, S., Holbrook, M., Kumaraguru, P., Cranor, L., Downs, J.: Who falls for phish? A demographic analysis of phishing susceptibility and effectiveness of interventions. In: 28th Proceedings of the SIGCHI conference on Human factors in Computing Systems, pp. 373–382. ACM, New York (2010)
10. Workman, M., Bommer, W.H., Straub, D.: Security lapses and the omission of information security measures: a threat control model and empirical test. Comput. Hum. Behav. **24**, 2799–2816 (2008)
11. Keeney, R.L., Raiffa, H.: Decisions with Multiple Objectives: Preferences and Value Tradeoffs. Wiley, New York (1976)

12. Keeney, R.L.: The value of internet commerce to the customer. Manag. Sci. **45**, 533–542 (1999)
13. Eisenführ, F., Weber, M., Langer, T.: Rational Decision Making. Springer, Berlin (2010)
14. Davis, F.D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. Manag. Inf. Syst. Q. **13**, 319–340 (1989)
15. Kujala, S., Väänänen-Vainio-Mattila, K.: Value of information systems and products: understanding the users' perspective and values. J. Inf. Technol. Theor. Appl. **9**, 23–39 (2009)
16. Boiney, L.G.: Reaping the benefits of information technology in organizations: a framework guiding appropriation of group support systems. J. Appl. Behav. Sci. **34**, 327–346 (1998)
17. Buhrmester, M., Kwang, T., Gosling, S.D.: Amazon's mechanical turk: a new source of inexpensive, yet high-quality, data? Perspect. Psychol. Sci. **6**, 3–5 (2011)
18. Mason, W., Suri, S.: Conducting behavioral research on Amazon's mechanical turk. Behav. Res. Methods **44**, 1–23 (2012)
19. Ipeirotis, P.G., Paolacci, G., Chandler, J.: Running experiments on amazon mechanical turk. Judgm. Decis. Mak. **5**, 411–419 (2010)
20. Nguyen, K.D., Rosoff, H., John, R.S.: The effects of attacker identity and individual user characteristics on the value of information privacy. Comput. Hum. Behav. **55**, 372–383 (2016)
21. Kirkwood, C.W.: Strategic Decision Making: Multiobjective Decision Analysis with Spreadsheets. Duxbury Press, Belmont (1997)
22. Acquisti, A., John, L.K., Loewenstein, G.: What is privacy worth? J. Legal Stud. **42**, 249–274 (2013)
23. Dhillon, G., Tiago, O., Susarapu, S., Caldeira, M.: Deciding between information security and usability. Comput. Hum. Behav. **61**, 656–666 (2016)
24. Xu, H., Luo, X., Carroll, J.M., Rosson, M.B.: The personalization privacy paradox: an exploratory study of decision making process for location-aware marketing. Decis. Support Syst. **51**, 42–52 (2011)
25. Glassman, M., Vandenwauver, M., Tam, L.: The psychology of password management: a tradeoff between security and convenience. BT Technol. J. **29**, 233–244 (2010)

Event Detection Based on Nonnegative Matrix Factorization: Ceasefire Violation, Environmental, and Malware Events

Barry Drake(✉), Tiffany Huang, Ashley Beavers, Rundong Du, and Haesun Park

Georgia Institute of Technology, 75 5TH Street NW STE 900, Atlanta, GA 30308-1018, USA
{Barry.Drake,Tiffany.Huang,Ashley.Beavers}@gtri.gatech.edu,
{rdu,hpark}@cc.gatech.edu

Abstract. Event detection is a very important problem across many domains and is a broadly applicable encompassing many disciplines within engineering systems. In this paper, we focus on improving the user's ability to quickly identify threat events such as malware, military policy violations, and natural environmental disasters. The information to perform these detections is extracted from text data sets in the latter two cases. Malware threats are important as they compromise computer system integrity and potentially allow the collection of sensitive information. Military policy violations such as ceasefire policies are important to monitor as they disrupt the daily lives of many people within countries that are torn apart by social violence or civil war. The threat of environmental disasters takes many forms and is an ever-present danger worldwide, and indiscriminate regarding who is harmed or killed. In this paper, we address all three of these threat event types using the same underlying technology for mining the information that leads to detecting such events. We approach malware event detection as a binary classification problem, i.e., one class for the threat mode and another for non-threat mode. We extend our novel classifier utilizing constrained low rank approximation as the core algorithm innovation and apply our Nonnegative Generalized Moody-Darken Architecture (NGMDA) hybrid method using various combinations of input and output layer algorithms. The new algorithm uses a nonconvex optimization problem via the nonnegative matrix factorization (NMF) for the hidden layer of a single layer perceptron and a nonnegative constrained adaptive filter for the output layer estimator. We first show the utility of the core NMF technology for both ceasefire violation and environmental disaster event detection. Next NGMDA is applied to the problem of malware threat events, again based on the NMF as the core computational tool. Also, we demonstrate that an algorithm should be appropriately selected for the data generation process. All this has critical implications for design of solutions for important threat/event detection scenarios. Lastly, we present experimental results on foreign language text for ceasefire violation and environmental disaster events. Experimental results on a KDD competition data set for malware classification are presented using our new NGMDA classifier.

Keywords: Malware detection · Event detection · Perceptron · Clustering · Nonnegative matrix factorization · Adaptive filtering · Hybrid classifier · Topic modeling · Classification

D. Nicholson (ed.), *Advances in Human Factors in Cybersecurity*,
Advances in Intelligent Systems and Computing 593, DOI 10.1007/978-3-319-60585-2_16

1 Introduction

In this paper, we present algorithms for a single layer perceptron (SLP) that combines two algorithms into a new hybrid classification framework and implements locally tuned receptive fields. Our new framework builds on the hybrid architecture first reported by Moody and Darken [1, 2] and generalized by Drake et al. [3]. We use this hybrid classification framework with new algorithms that can be updated on a per-sample or mini-batch (small number of samples) basis and are sensitive to the input domain of the data. Our classification framework has advantages over multilayer perceptron architectures and the support vector machine (SVM) [3]. Our new hybrid framework and algorithms will be presented within the context of three important event detection problems: malware, ceasefire, and environmental disaster event detection. We further extend our Generalized Moody-Darken Architecture (GMDA) framework to the nonnegative domain using the nonnegative matrix factorization (NMF). When combined with a nonnegativity constrained adaptive filter the extended GMDA framework is called the Nonnegative GMDA (NGMDA), which is able to discover discriminative features for classification and demonstrate better performance for nonnegative input data.

Our GMDA classifier uses a clustering method to find the centers of the activation units. The kmeans algorithm has typically been used to determine the activation unit centers. However, for nonnegative input data, we show in this paper experimental results that demonstrate a loss of information in the classifier, which increases the classification error. We extend the GMDA further with an objective function based on a constrained low rank approximation (CLRA) method called the nonnegative matrix factorization (NMF) [3, 4], which we have applied in numerous application domains for text analytics. CLRA methods [4] have played a crucial role as one of the most fundamental tools in machine learning, data mining, image processing, information retrieval, computer vision, signal processing, and other areas of computational science and engineering. Since the NMF objective function is formulated using nonnegative constraints, application of NMF to certain problems produces results that are more interpretable for many types of problems. In [3] we demonstrated the importance of incorporating nonnegative constraints for applications such as image processing, chemometrics, and text analytics. NMF has become a valuable tool for many applications such as clustering, subspace-based topic modeling, general dimension reduction (PCA-like), hyperspectral image processing, and many more. In this paper, NMF is utilized as the first stage of our new NGMDA classifier. In previous work [3] we demonstrated the efficacy of NMF-based GMDA with preliminary results and demonstrated comparable performance to a support vector machine (SVM) classifier. We show that implementing the output layer of the NGMDA with a nonnegative adaptive filter improves classification performance over unconstrained adaptive algorithms. An NMF event detection methodology is demonstrated for foreign language texts in Arabic (ceasefire violation events in Yemen) and Chinese (environmental disaster events). The classification results are demonstrated on event detections for malware (cybersecurity). Experimental results are presented on malware event classification, and event detection from text data. Thus, the experimental results demonstrate NMF for stand-alone applications and as part of a hybrid classification architecture, the NGMDA.

2 Background Material

As a review of the underlying components of the GMDA, we briefly describe the components of the GMDA for the hidden and output layers. First we provide some important mathematical properties of two neural network architectures. The Moody-Darken single layer perceptron (SLP) architecture has advantages over a multi-layer perceptron (MLP) architecture in both a fundamental mathematical sense and performance considerations. Both share the *universal approximation* property, which provides existence proofs of an interpolating set of basis polynomials for arbitrary inputs. One advantage of the SLP over a MLP architecture is the *best approximation* property, which guarantees that there is a set of approximating functions corresponding to all possible choices of the model parameters with one function from the set that minimizes the approximation errors [5–7].

The general equation for the GMDA is

$$\mathrm{y} = b + \sum_{i=1}^{k_{hidden}} w_i f\left(\frac{\exp\|x - c_i\|^2}{2s_i^2}\right) \tag{1}$$

where x is the input data, c_i are the activation unit centers, w_i are the activation unit magnitudes computed by the output layer algorithm, f denotes the radial basis function (RBF) kernel, the summation is over the number of activation units in the hidden layer, b is a bias term, and y is the output. As stated above the c_i are computed using a clustering method (NMF for nonnegative data inputs), which performs a dimension reduction and, thus, reduces the computational complexity by decreasing the amount of data required for the output layer algorithm. The s_i are the estimated standard deviations of the computed clusters, which are based on the normalized sum of the distances between the samples in the cluster to its center and the output layer weights of the GMDA are computed by an adaptive filter [5, 8]. Various adaptive filters are possible for estimating the weights. Examples are LMS, Recursive Least Squares (RLS), and QR Decomposition Recursive Least Squares (QRDRLS) and many more. For this paper, we examine both unconstrained and nonnegative variants of adaptive filters, mainly the nonnegative LMS (NNLMS) [9] and our Adaptive Sequential Coordinate-wise Algorithm (ASCA).

First we review the SLP hidden layer algorithms from [3] followed by the background for the output layer algorithms. Throughout this paper $A \in \mathbb{R}^{mxn}$ where m is the number of features and n are the columns that represent the data. Assuming $rank(A) = r$, the low rank approximation of A, for $k \leq r$, is denoted $rank\left(\widehat{A}\right) = k$, i.e., we approximate the matrix A with factors that, when multiplied together produce a matrix $\widehat{A}$ that is close to A in some norm, usually the Frobenius norm [10], which we use throughout this paper. The Frobenius norm is analogous to the vector 2-norm (L_2) but for matrices. One of the most commonly used low rank approximations is the singular value decomposition (SVD) [10]. A general form of a low rank approximation can be expressed as

$$\min_{W,H} \|A - WH\|_F \tag{2}$$

where F denotes the Frobenius norm, $W \in \mathbb{R}^{mxk}$, and $H \in \mathbb{R}^{kxn}$ and k is the rank of $A \in \mathbb{R}^{mxn}$. Depending on the constraints on W and H, (1) may be a nonconvex optimization problem. However, the SVD gives the optimal global solution when the objective of Eq. (1) is *unconstrained*. Equation (2) will be the basic equation for discussing the clustering methods used in the SLP hidden layer. In the case of NMF, (2) becomes

$$\min_{W,H\geq 0} \|A - WH\|_F \tag{3}$$

where Eq. (3) is the same as (2) but with nonnegativity constraints on W and H, which makes (3) a nonconvex optimization problem, a very difficult problem to solve. Generally speaking, (2) achieves a global minimum while (3) is only guaranteed to achieve a local minimum at best. However, by choosing the best algorithms to solve (3) convergence to a local minimum is guaranteed. The choice of algorithm is extremely important in order to ensure convergence to a stationary point of the minimization surface [4]. In general, Eq. (3) performs clustering of the input data be it text (for documents) or pixels (images). Topic modeling discussed in a latter section is document clustering where each cluster is composed of semantically related terms (words) and k is user chosen as the desired number of clusters (topics) to compute. The W is the cluster *indicator* matrix and H is known as the cluster *membership* matrix. Thus, H can be used to retrieve documents within clusters (topics).

As mentioned above, the output layer of the GMDA is implemented using adaptive algorithms such as adaptive filters [5]. Adaptive filters are commonly found in the signal processing literature and have applications in adaptive beamforming, direction finding, speech processing, and many more. A commonly used adaptive filter is the Least Mean Squares (LMS) algorithm, which is an approximation of the gradient descent method. The LMS algorithm is simple to implement and has some nice statistical properties but may suffer from slow convergence and misadjustment error, i.e., the learning curve (MSE) slowly approaches the asymptotic minimum value. In this paper, we use a NNLMS, which for the first time, is incorporated into a hybrid learning architecture such as the NGMDA for malware event detection.

In general, the learning rule of the output layer can be cast as a model-building problem via regression. Thus, the general form of this learning rule is

$$\min_{X} \|BX - D\|_F \tag{4}$$

where Eq. (4) is the general linear regression problem with multiple right hand sides. We solve (4) using one of the adaptive filter methods mentioned above as each data sample becomes available (adaptive) and for a single right hand side, which computes the output of the hidden layer. Both the hidden layer and the output layer of the GMDA or NGMDA can be updated on a per sample basis with suitably chosen algorithms. For example, for arbitrary data ‘kmeans + LMS’ (GMDA); for nonnegative data ‘NMF + NNLMS’ (NGMDA) is an implementation that can be updated/downdated in both the hidden and output layers.

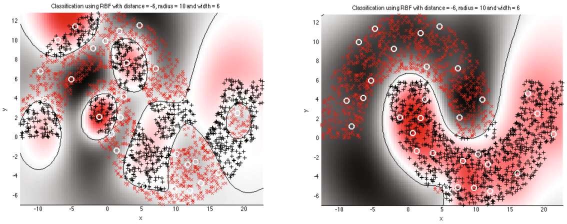


Fig. 1. a. GMDA with LMS adaptive filter as the output layer algorithm; b. GMDA with output layer algorithm RLS. We observe that LMS breaks down in this case while RLS obtains good classification results [1].

Another approach for the output layer, presented in [3], is the ASCA, the adaptive sequential coordinate-wise algorithm, which is an adaptive version of SCA, sequential coordinate-wise algorithm [11], for solving the nonnegative least squares (NLS) problem. We will show that the results using ASCA, rather than NNLMS in the output layer are superior, and that, with NMF as the hidden layer algorithm, NMF-based NGMDA outperforms kmeans for computing the hidden layer activation functions. This serves to emphasize the better numerical properties of our new ASCA algorithm used for the NGMDA output layer as reported in [3] and the superior performance achieved when a properly constrained algorithm for the data is used to compute the hidden layer, i.e., NMF rather than kmeans.

Now we review the importance of choosing the correct algorithm for the data [3]. We briefly review those considerations here with simulated half-moon data in Fig. 1 and simulated chemical detections in Fig. 2 below. These figures illustrate that the choice of algorithms is critical, though some algorithms may have higher computational complexity, the benefits may outweigh the costs. Since the half-moon data was real (+, 0, − values), Fig. 1a, b shows results with kmeans as the hidden layer algorithm with 30 activation units. Figure 1 demonstrates that classification results can depend critically on the selected algorithm.

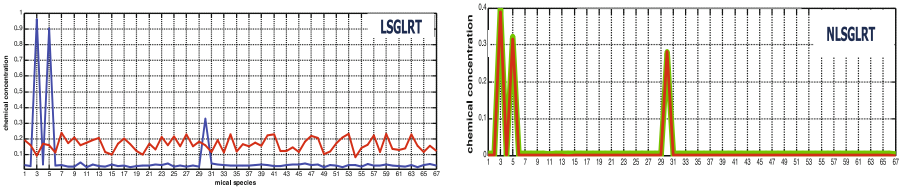


Fig. 2. a. Generalized likelihood ratio test *without* numerical problems (blue) and *with* numerical problems (red). Note that concentrations were not determined correctly and that there are no detections when numerical problems are present (red). b. detection results are clearly corrected with the right computational model: a. uses *standard* least squares; b uses *nonnegative* least squares.

The chemical detection problem, Fig. 2a, b, uses Raman spectroscopy to discover chemical constituents of a chemical sample. For details see [3, 12]. There are two issues: The correct computational model for the data; Numerically robust algorithms that are not degraded by near collinearities (rank deficiency). The figures illustrate both cases for the generalized likelihood ratio test (GLRT) detection algorithm [13]. Note that the chemicals are labeled with species number in the library, which may be traced back to the wavenumber for a specific chemical. The tests were for three chemical species: two closely spaced and one more isolated at 0.3 g/m^2.

With the above results shown in Figs. 1 and 2, and knowing that NMF is the correct computational model for text data, in the next section we demonstrate its use for event detection with a human in the loop process on foreign text. The effectiveness of NMF for this application sets the stage for the development of the NGMDA classifier where we compare experiments with various combinations of computational models.

3 Observer-Based Event Detection from Foreign Language Text Using NMF-Based Topic Modeling

In this section, we demonstrate the use of NMF topic modeling for processing nonnegative text data in order to discover latent information within a text corpus. Topic modeling is a methodology that clusters documents into semantically related words. Within each cluster (topic) the highest frequency keywords reveal the overall concept of the topic. Topic refinement is used to uncover topics that are more relevant to the domain of interest. Our approach uses a human-in-the-loop to interpret the NMF factors in order to accomplish search space reduction; since this approach does not depend on incorporation of complex language considerations, meta data, or a priori identification of target keywords, it can be used in contexts where there has been little natural language processing performed, where reliable meta data is not available, and where keywords are initially unknown. Since NMF is a subspace-based method we can easily access the original documents within certain topics and, if determined not to be relevant, eliminate them to retain only information relevant to our domain questions. After eliminating these documents from the original corpus, a reduced corpus is then analyzed using our topic modeling algorithm. In Fig. 3a, b is shown the process to refine the topic model where Fig. 3a shows the process and Fig. 3b shows the selection and elimination of irrelevant documents.

A relevance score can be computed and a threshold set to determine whether or not a document should be eliminated from the corpus. The relevance score can be computed as

$$g_i > \max\left(g_i^T \times \mathit{threshold}\right) \tag{5}$$

which is computed from the normalized columns of H in

$$\min_{W,H \geq 0} \|A - WH\|_F \tag{6}$$

Recall that H is called the topic (cluster) membership matrix, i.e., each column of the membership matrix tells us which document belongs to which topic and can even indicate how well a document represents the cluster (topic), which is the basis for Eq. (6). These relationships are the key elements for determining the documents to eliminate at each iteration of the refinement process in Fig. 3a.

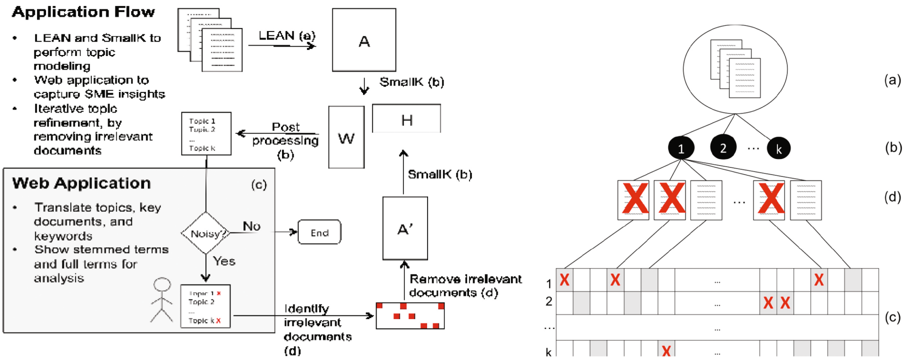
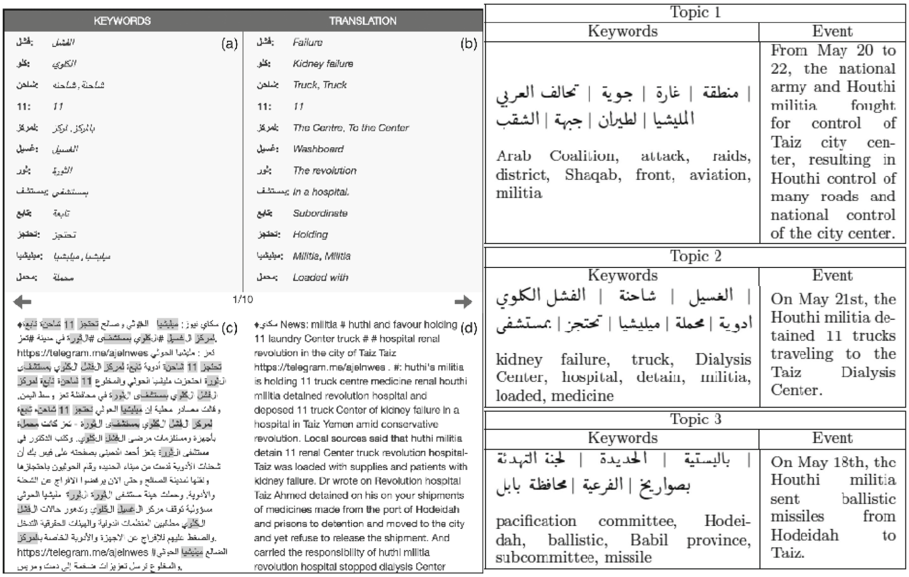

Application Flow
• LEAN and SmallK to perform topic modeling
• Web application to capture SME insights
• Iterative topic refinement, by removing irrelevant documents
LEAN (a)
A
SmallK (b)
W
H
Post processing (b)
Topic 1
Topic 2
...
Topic k
Web Application
(c)
• Translate topics, key documents, and keywords
• Show stemmed terms and full terms for analysis
Noisy?
No
End
Yes
Topic 1 X
Topic 2
...
Topic k X
Identify irrelevant documents (d)
Remove irrelevant documents (d)
SmallK (b)
A'
(a)
1
2
...
k
(b)
(d)
X
(c)
...

Fig. 3. a. Topic refinement process with human-in-the-loop; b. elimination of selected documents based on subject matter expert (SME) input.

In order to demonstrate the effectiveness of this process on a real problem, ceasefire violation events in Yemen will be discovered roughly during the time period around May, 2016. NMF-based topic modeling and the topic refinement processes are used to discover these events from raw Arabic telegrams (social media not available), i.e., the telegrams are *not* translated to English before using our tools to discover topics.



KEYWORDS		TRANSLATION	
فشل:	الفشل	فشل:	Failure
كلو:	الكلوي	كلو:	Kidney failure
شاحن:	شاحنة, شاحنه	شاحن:	Truck, Truck
11:	11	11:	11
لمركز:	بالمركز, لمركز	لمركز:	The Centre, To the Center
غسيل:	الغسيل	غسيل:	Washboard
ثور:	الثورة	ثور:	The revolution
بمستشف:	بمستشفى	بمستشف:	In a hospital.
تابع:	تابعة	تابع:	Subordinate
تحتجز:	تحتجز	تحتجز:	Holding
ميليشيا:	ميليشيا, ميليشيا	ميليشيا:	Militia, Militia
محمل:	محملة	محمل:	Loaded with

♦سكاي نيوز : ميليشيا الحوثي وصالح تحتجز 11 شاحنة تابعة
لمركز الغسيل الكلوي بمستشفى #الثورة في مدينة #تعز
https://telegram.me/ajelnwes تعز : مليشيا الحوثي
تحتجز 11 شاحنة أدوية تابعة لمركز الفشل الكلوي بمستشفى
الثورة احتجزت مليشيا الحوثي والمخلوع 11 شاحنة تابعة لمركز
الفشل الكلوي بمستشفى الثورة في محافظة تعز وسط اليمن.
وقالت مصادر محلية إن ميليشيا الحوثي تحتجز 11 شاحنة تابعة
لمركز الفشل الكلوي بمستشفى الثورة - تعز كانت محملة
بأجهزة ومستلزمات مرضى الفشل الكلوي. وكتب الدكتور في
مستشفى الثورة بتعز أحمد الشميري بصفحته على فيس بك ان
شحنات الأدوية قدمت من ميناء الحديده وقام الحوثيون باحتجازها
ونقلها لمدينة الصالح وحتى الان يرفضوا الافراج عن الشحنة
والأدوية. وحملت جبهة مستشفى الثورة الثورة مليشيا الحوثي
مسؤولية توقف مركز الغسيل الكلوي وتدهور حالات الفشل
الكلوي مطالبين المنظمات الدولية والهيئات الحقوقية التدخل
والضغط عليهم للإفراج عن الاجهزة والأدوية الخاصة بالمركز.
https://telegram.me/ajelnwes الضالع ميليشيا الحوثي
والمخلوع ترسل تعزيزات ضخمة إلى دمت ومريس

♦سكاي News: militia # huthi and favour holding 11 laundry Center truck # # hospital renal revolution in the city of Taiz Taiz https://telegram.me/ajelnwes . #: huthi's militia is holding 11 truck centre medicine renal houthi militia detained revolution hospital and deposed 11 truck Center of kidney failure in a hospital in Taiz Yemen amid conservative revolution. Local sources said that huthi militia detain 11 renal Center truck revolution hospital-Taiz was loaded with supplies and patients with kidney failure. Dr wrote on Revolution hospital Taiz Ahmed detained on his on your shipments of medicines made from the port of Hodeidah and prisons to detention and moved to the city and yet refuse to release the shipment. And carried the responsibility of huthi militia revolution hospital stopped dialysis Center

Topic 1	
Keywords	Event
منطقة | غارة | جوية | تحالف العربي | المليشيا | لطيران | جبهة | الشقب Arab Coalition, attack, raids, district, Shaqab, front, aviation, militia	From May 20 to 22, the national army and Houthi militia fought for control of Taiz city center, resulting in Houthi control of many roads and national control of the city center.
Topic 2	
Keywords	Event
الغسيل | شاحنة | الفشل الكلوي | ادوية | محملة | ميليشيا | تحتجز | بمستشفى kidney failure, truck, Dialysis Center, hospital, detain, militia, loaded, medicine	On May 21st, the Houthi militia detained 11 trucks traveling to the Taiz Dialysis Center.
Topic 3	
Keywords	Event
بالستية | الحديدة | لجنة التهدئة | بصواريخ | الفرعية | محافظة بابل pacification committee, Hodeidah, ballistic, Babil province, subcommittee, missile	On May 18th, the Houthi militia sent ballistic missiles from Hodeidah to Taiz.

Fig. 4. a. Topic modeling on Arabic telegram text; b. findings based on topic refinement.

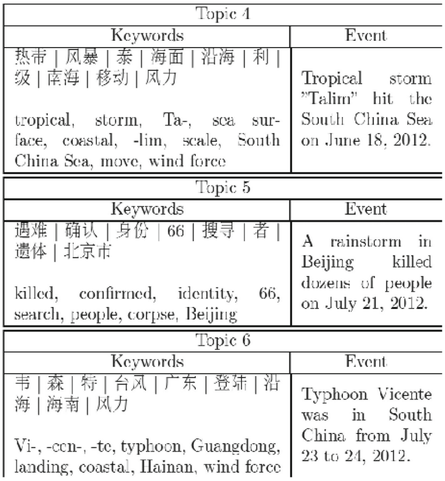
Figure 4a shows some of the intermediate processing steps used to refine the topic modeling. These are illustrated in Fig. 4a, pane (a), (b), (c), and (d). (a) Stemmed keywords followed by the full keywords in order of frequency, e.g., target: targets, targeted, etc.; (b) translations of the full keywords for each stemmed keyword; (c) original text with highlighted keywords; (d) translated text. The relevant topics found are shown in Fig. 4b, which shows 3 topics with events and their dates that indicate ceasefire violations. In Fig. 4, results are shown for the geographical region around Taiz, Yemen. The analysis generally focused on major cities where ceasefire violations were more likely and the data sets were larger and richer in content. The key point is that specific events were discovered using a semi-supervised methodology based on the NMF.

Figures 5a and b below show the results using our refined topic modeling process for environmental events in China during 2012. As above, the raw Chinese text is processed and NMF topic modeling is used to analyze the results. Revealed in Topic 1 is a straw fire that causes severe haze in Jiangsu and Hubei Provinces, which occurred around June 9. Topic 5 reveals a severe rain storm that killed dozens of people in Beijing on July 21.

Topic 1	
Keywords	Event
霾 | 秸秆 | 雾 | 空气 | 焚烧 | 武汉 | 能见度 | pm | 质量 | 污染 haze, straw, fog, air, incineration, Wuhan, visibility, pm, quality, pollution	Straw burning causes severe haze in Jiangsu Province and Hubei Province beginning June 9, 2012.

Topic 2	
Keywords	Event
云南省 | 地震 | 宁 | 丽江 | 彝族 | 盐 | 云南 | 宁乡 | 交界 | 灾区 Yunnan, earthquake, Ning, Lijiang, Yi, Salt, Yunnan, Ningxiang, junction, disaster area	On June 24, 2012, an earthquake occurred at the junction of Sichuan and Yunnan Provinces.

Topic 3	
Keywords	Event
黄河 | 流量 | 陕西省 | 洪峰 | 秒 | 立方米 | 洪水 The Yellow River, flow, Shaanxi Province, flood peak, seconds, cubic meters,flood	Flooding occurred in Shaanxi Province on July 27, 2012.



Topic 4	
Keywords	Event
热带 | 风暴 | 泰 | 海面 | 沿海 | 利 | 级 | 南海 | 移动 | 风力 tropical, storm, Ta-, sea surface, coastal, -lim, scale, South China Sea, move, wind force	Tropical storm "Talim" hit the South China Sea on June 18, 2012.

Topic 5	
Keywords	Event
遇难 | 确认 | 身份 | 66 | 搜寻 | 者 | 遗体 | 北京市 killed, confirmed, identity, 66, search, people, corpse, Beijing	A rainstorm in Beijing killed dozens of people on July 21, 2012.

Topic 6	
Keywords	Event
韦 | 森 | 特 | 台风 | 广东 | 登陆 | 沿海 | 海南 | 风力 Vi-, -cen-, -te, typhoon, Guangdong, landing, coastal, Hainan, wind force	Typhoon Vicente was in South China from July 23 to 24, 2012.

Fig. 5. Topic modeling on Chinese newspaper text; a. topics 1–3; b. topics 4–6.

A point to keep in mind is that these environmental events were discovered in a semi-supervised manner with a human in the loop. The topic modeling algorithm, however, is unsupervised and discovers the topics and related keywords automatically without any human intervention. Once the initial topic modeling is performed the top keywords are determined for each topic, which span the topic concept. A label can be determined for each topic and the topics can become categories for classification. This is consistent with the notion that clustering is often referred to as unsupervised classification.

The above examples of using NMF for topic model event detection and, in the next section, classification, doesn't nearly cover the utility for which this versatile mathematical method is capable: image segmentation, hyperspectral image processing, speckle removal from noisy images, non-stationary speech denoising, motion detection from video sequences, music analysis, bioinformatics applications, chemometrics, and many more.

4 NMF-Based Generalized Moody-Darken Architecture (NGMDA) for Malware Event Detection: Experiments and Results

The GMDA can also be used as the clustering method for a hybrid classification algorithm by configuring the hidden and output layers with algorithms that constrain the solution to the nonnegative domain for nonnegative input data. In this section, the data is not composed of documents but rather network intrusion data, malware. By utilizing nonnegativity constrained low rank approximation in the hidden layer and a nonnegativity constrained output layer adaptation rule, the classifications can be mapped to the domain where interpretation of the results is possible, whereas without these constraints the results may not only be not interpretable, but also fail completely to provide meaningful solutions. This will be further illustrated by using NMF as the hidden layer algorithm for GMDA.

In Figs. 1 and 2 above we illustrated that algorithm selection is an important consideration, which justifies the reengineering of the original Moody-Darken SLP architecture with algorithms that conform to the specifications indicated by the input data, e.g., text, image data, chemical concentrations, etc.

Several experiments with various algorithm combinations implemented in our GMDA framework were applied to the malware dataset[1] that is comprised of 21 attack types on network systems. The data consisted of 125973 training samples and 22543 testing samples with 41 features. For the experiments, the data was further processed so that all types of attacks were gathered into one class; instances of normal network traffic formed another class. Thus, the data was composed of two classes: malware and non-malware.

First, the analysis was focused on the performance of kmeans as the hidden layer with the LMS filter output layer algorithm as in the original Moody-Darken hybrid architecture [1]. The second major testing was performed on the extension of the original architecture to handle nonnegative input data. Thus, extending the original architecture to our GMDA. The third set of experimental results use NMF as the hidden layer algorithm with a nonnegativity constrained output layer adaptive filter, which is the NGMDA. The experiments culminate with NMF as the hidden layer and our ASCA nonnegative least squares implementation as the output layer algorithm NGMDA.

For all of the experiments, the number of hidden nodes or centers for the hidden layer activation units was set to 15, and the number of trials or epochs to train the weight vector for the output layer was fixed at 1000.

The summary of the performance from the experiments is shown in Table 1 below.

Note that by replacing kmeans with NMF in finding the cluster centers, the performance of the GMDA improved significantly in terms of the metrics shown in Table 1 and the receiver operating characteristic (ROC) curves AUC (area under the curve). Another important item to note in Table 1 is that the AIC for kmeans + LMS seems to be acceptable.

[1] The UNB ISCX NSL-KDD malware dataset was obtained from http://www.unb.ca/research/iscx/dataset/iscx-NSL-KDD-dataset.html and a github site for the data used in this work: https://github.com/defcom17/NSL_KDD.

Table 1. Misclassrfication errors (%) of the resting samples, Akaike Information Criterion (AIC) of the training samples, and Area Under Curve (AUC) of the testing samples on the UNB ISCX NSLKDD binary t-Liss dataset for llie proposed GMDA

Algorithms	Miselassification error in %	AIC	AUC
kmeans + LMS	55.98	−222209	0.38
kmeans + RLS	69.15	−5621	0.38
kmeans + NNLMS	43.07	30.0	0.51
NMF + NNLMS	40.59	−100176	0.79
NMF + ASCA	33.07	−293620	0.86

AIC was computed based on the sum of squared errors *(SSE)* from training the particular filler, weighted by the number of training samples (n) and penalized by the number of centers (k). It was computed as: n ln(SSE) – n ln(n) + 2k. Therefore, AIC is interpreted as a measure of model training accuracy, where a smaller value (or most negative value) represented better training accuracy.

However, AIC measures the *training accuracy* only and indicates for this case that, given the input data and the algorithm choices, the algorithms were indeed trained well on the presented nonnegative data. But, the poor classification results underscore the mismatch between algorithm choice (unconstrained) and input data (nonnegative). This is also demonstrated by the kmeans + NNLMS line in Table 1 where the poorest results are shown. Another way to state this is that training the filter on results from an unconstrained dimension reduction algorithm for nonnegative data is not recommended. Both the training accuracy and ROC AUC are very poor. In fact, the AUC indicates that this combination is not much better than tossing a coin for classification.

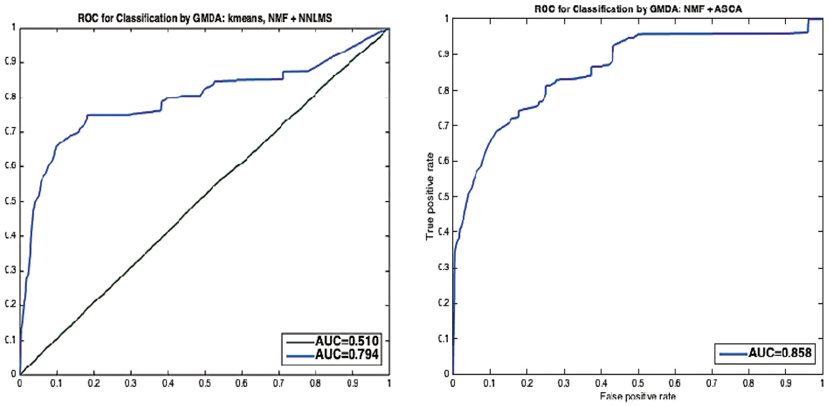


Fig. 6. a. kmeans + NNLMS (green) produced nearly a straight line, AUC = 0.51, while NMF + NNLMS (blue), AUC = 0.79; b. NMF + ASCA, AUC = 0.86 (best result).

The result suggests that NMF, with its increased interpretability, preserved the data relationships for constructing meaningful clusters that could be the basis for classification, but kmeans did not. The performance of NGMDA was improved further by running NMF with ASCA. With a dimension reduction algorithm that preserves cluster structure, such as NMF, and a nonnegative constraint on the adaptive filtering algorithm, such as

NNLMS or ASCA, the performance of NGMDA in terms of classification and model training accuracy is greatly improved. NMF is again shown to be the right computational model for the data.

To illustrate the performance of NGMDA for the analyses, the ROC AUC is shown in Figs. 6a and b. It is evident from these two figures that NGMDA easily outperforms the other methods tested.

5 Conclusions

In this paper, the many examples of using the correct algorithm for the data were reviewed and extended with our experimental results on malware data. We have demonstrated new methods for event detection on a variety of event scenarios. The events were discovered from various text sources for ceasefire violations in Arabic text and environmental events from Chinese text. We also discovered malware events from a KDD data set containing network intrusion data. The text-based events discovered from our topic refinement process were very specific in terms of event type, location, severity of impact, and date. The technology could be used in a number of critical situational awareness applications, especially when embedded in a visual analytics system. Our new NGMDA is promising as a classifier for various nonnegative data sources where online, streaming updates are a requirement. The single layer perceptron architecture does not require the expensive backpropagation algorithm. Data can flow through the NGMDA as each sample or mini-batch of samples is available.

The motivation for these generalizations of the original Moody-Darken architecture has been to apply that architecture to a wider class of problem domains, which may impact social situational awareness, social conflict monitoring from text corpora and other data sources, image understanding and crop monitoring, and deep feature extraction for even better classification results. Like GMDA as in reported in [3], NGMDA inherits from the original Moody-Darken architecture fast training of locally tuned receptive fields, the best approximation property, and the ability to incorporate adaptability to changing data.

The results obtained on the KDD malware data appear to be promising for applying the NGMDA in other domains. The misclassification errors were very good when the computational model consistently utilized nonnegative constraints. Our future work will focus on deep feature extraction from a hierarchical NGMDA. We currently have preliminary results on image classification, again demonstrating that the correct computational model uses nonnegative constraints.

Acknowledgments. The authors thank Dr. Richard Boyd for his significant contributions to SmallK and LEAN, both of which were essential tools used in our topic refinement process. We would also like to thank Dr. Elizabeth Whitaker for the opportunity to present this work and the Information and Communication Laboratory (ICL) at Georgia Tech Research Institute for supporting this research. This work was supported in part by the Defense Advanced Research Projects Agency (DARPA) XDATA program grant FA8750-12-2-0309. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of DARPA.

References

1. Moody, J., Darken, C.: Fast learning in networks of locally-tuned processing units. Neural Comput. **1**(1), 281–294 (1989)
2. Lippmann, R.P.: Pattern classification using neural networks. IEEE Commun. Mag. **27**(11), 47–54 (1989)
3. Drake, B., Huang, T., Cistola, C.: Malware detection based on new implementations of the moody-darken single-layer perceptron architecture: when the data speaks, are we listening? In: Ahram, T., Karwowski, W. (eds.) Proceedings of the 7th International Conference on Applied Human Factors and Ergonomics, Human Factors in Cybersecurity, Orlando, Florida, USA, 27–21 July. Springer (2016, invited paper)
4. Kim, J., He, Y., Park, H.: Algorithms for nonnegative matrix and tensor factorizations: a unified view based on block coordinate descent framework. J. Glob. Optim. **58**, 285–319 (2014)
5. Haykin, S.: Adaptive Filter Theory, 5th edn. Pearson Education, Inc., New Jersey (2014)
6. Haykin, S.: Neural Networks and Learning Machines, 3rd edn. Prentice Hall, Upper Saddle River (2009)
7. Bishop, C.: Neural Networks for Pattern Recognition. Oxford University Press, New York (2005)
8. Drake, B., Luk, F., Speiser, J.M., Symanski, J.: SLAPP: a systolic linear algebra parallel processor: IEEE. Computer **20**(7), 45–49 (1987)
9. Chen, J., Richard, C., Bermudez, J., Honeine, P.: Nonnegative least-mean-square algorithm. IEEE Trans. Sig. Process. **59**(11), 5225–5235 (2011)
10. Golub, G.H., Van Loan, C.F.: Matrix Computations. The Johns Hopkins University Press, Baltimore (2013)
11. Franc, V., Hlaváč, V., Navara, M.: Sequential coordinate-wise algorithm for the non-negative least squares problem. In: Proceedings of the 11th International Conference on Computer Analysis of Images and Patterns (CAIP), pp. 407–414 (2005)
12. Drake, B., Kim, J., Mallick, M., Park, H.: Supervised Raman spectra estimation based on nonnegative rank deficient least squares. In: Proceedings 13th International Conference on Information Fusion, Edinburgh, UK (2010)
13. Kay, S.M.: Fundamentals of Statistical Signal Processing: Detection Theory, vol. 2. Prentice Hall PTR, Englewood Cliffs (1998)

Human Dimension and Visualization for Cybersecurity

Human Behavior Analytics from Microworlds: The Cyber Security Game

Johan de Heer(✉) and Paul Porskamp

Thales Research and Technology T-Xchange, University of Twente,
Westhorst Building 22 – WH226, Drienerlolaan 5,
7522 NB Enschede, The Netherlands
{Johan.deHeer,Paul.Porskamp}@nl.thalesgroup.com

Abstract. Games viewed as socio-technical representations of real world system-of-systems may turn into Microworld research tools to monitor human dynamic decision making. In this paper we illustrate the potential of this methodology focusing on a Cyber Security Dilemma game, and various player models that we can elucidate from them at individual and aggregated levels.

Keywords: Game based learning · Stealth assessment · Human behavior modelling · Cyber security

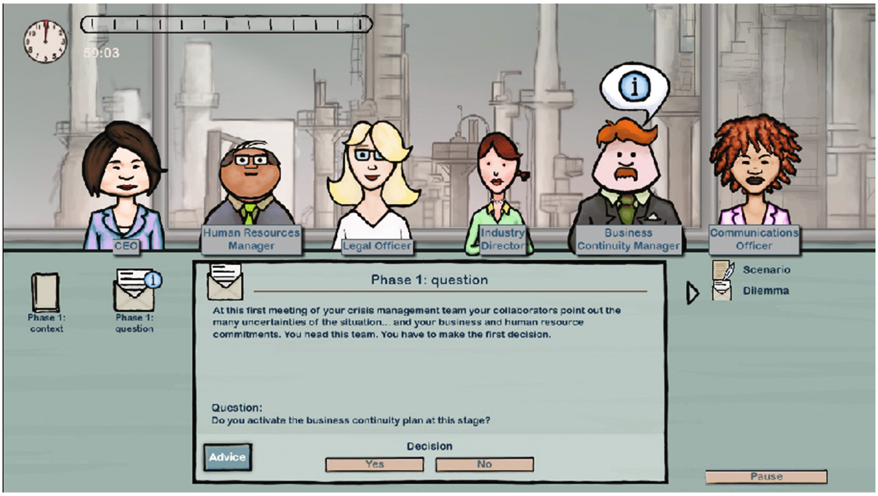
1 Introduction

Making judgments and taking decisions is daily practice for lots of people. Understanding and elucidating the dynamics of human reasoning, however, is an enigma and requires a theory of mind, appropriate theoretical concepts, methods and techniques for studying Dynamic Decision Making (DDM). Let alone, predicting human judgment and decision-making behaviors. This paper sketches a '*game-based-micro-world*' for studying Dynamic Decision Making [1]. Microworlds [2] are used to record, monitor and analyze how people make decisions over time. DDM takes into account [3]: sequences of decisions to reach a goal, interdependence of decisions on previous decisions, dynamics of a changing environment, and that decisions are made in real time (that is, in time pressured situations). We illustrate such a microworld with an example that enables us to study how players in the role of crisis managers make decisions during the unfolding of a cyber security interactive storyline. In addition, we present several type of human behavioral models, including risk taken and avoidance behaviors that can be provided by game statistical and analytical services.

2 Microworld: Cyber Game

We designed and developed a game based microworld (see Fig. 1) that represents the essential real world elements during a cyber crisis from the crisis manager point of view. Note, that it is beyond the scope of this paper to discuss how we designed and developed this model-based and configurable game based microworld. It needs understanding of the

D. Nicholson (ed.), *Advances in Human Factors in Cybersecurity*,
Advances in Intelligent Systems and Computing 593, DOI 10.1007/978-3-319-60585-2_17



Phase 1: question

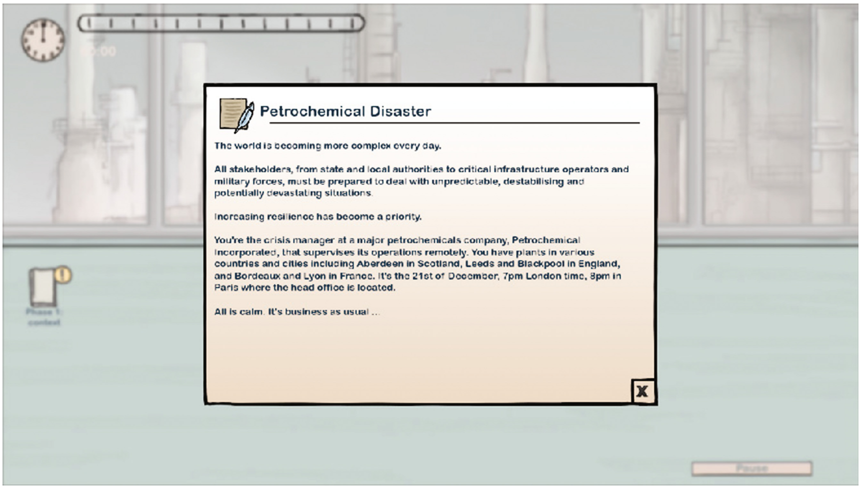
At this first meeting of your crisis management team your collaborators point out the many uncertainties of the situation... and your business and human resource commitments. You head this team. You have to make the first decision.

Question:
Do you activate the business continuity plan at this stage?

Fig. 1. Single player turn taking 2D narrative game.

specific game scenario [4], how to design a game based systems [5], and a thorough understanding of the components that game systems are made of [6].

The game flow of this single player turn-taking narrative game based microworld is as follows. First, the crisis manager – the player - is presented a context scenario, in which the setting is briefly explained (Fig. 2), in this case the occurrence of a petrochemical disaster.



Petrochemical Disaster

The world is becoming more complex every day.

All stakeholders, from state and local authorities to critical infrastructure operators and military forces, must be prepared to deal with unpredictable, destabilising and potentially devastating situations.

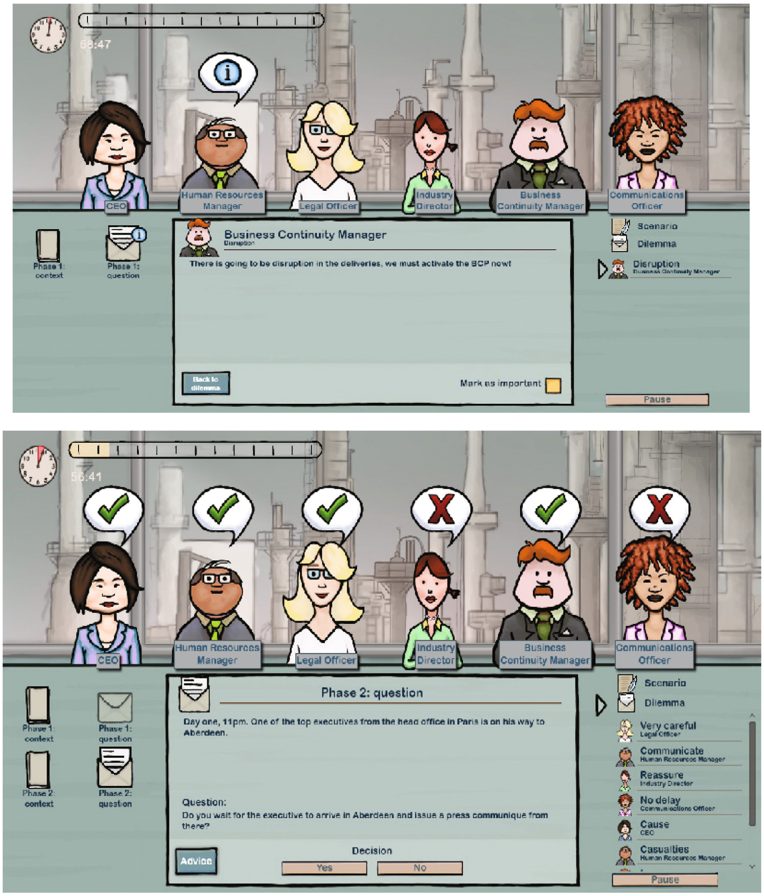
Increasing resilience has become a priority.

You're the crisis manager at a major petrochemicals company, Petrochemical Incorporated, that supervises its operations remotely. You have plants in various countries and cities including Aberdeen in Scotland, Leeds and Blackpool in England, and Bordeaux and Lyon in France. It's the 21st of December, 7pm London time, 8pm in Paris where the head office is located.

All is calm. It's business as usual ...

Fig. 2. Context scenario

Related to this context scenario, a series of six dilemmas are introduced that all end with a question where the player has to make a 'yes' or 'no' decision. The dilemmas – depicted in bottom left corner in the form of envelopes can be opened with a simple mouse click - appear over the course of (playing) time (Figs. 1 and 3). A typical dilemma relates to aspects of uncertainty and ambiguity of a specific crisis phase. The decision to take is for example: 'Do you activate the business continuity plan at this stage?' Note, that this game based microworld embeds dilemmas where there is no right or wrong answer; for each decision a rationale may be found, or a story can be told or argued. In the game (virtual) crisis team members are gathered around a table and may let the player know if they have potential relevant information (depicted by a text balloon above their heads) that may possibly alter the decision - if taken into account by the player. There are the CEO, the operations manager, the communication manager, the legal affairs manager, the Business Continuity Manager, the IT manager of the company, and even a representative of the national security agency, called in because of the unusual nature of the crisis [4]. The player is free to select and read information from his team advisors, and may even ask them for advice what they would decide - indicated by green (voting for a yes decision) and red (voting for a no decision) (Fig. 3).



Business Continuity Manager
Disruption

There is going to be disruption in the deliveries, we must activate the BCP now!

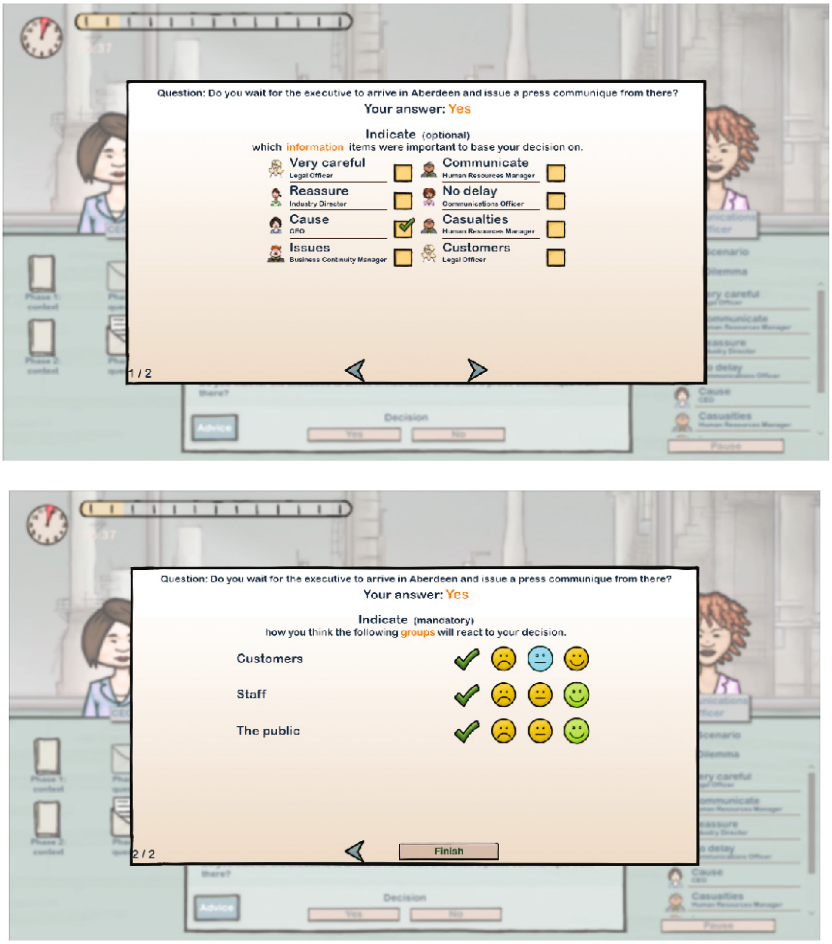
Phase 2: question

Day one, 11pm. One of the top executives from the head office in Paris is on his way to Aberdeen.

Question:
Do you wait for the executive to arrive in Aberdeen and issue a press communique from there?

Fig. 3. Asking information and/or advice

Once, the dilemma has been answered, the game pauses and the player is asked to indicate, which information provided by a virtual team member was taken into account and considered relevant regarding the decision s/he took. Virtual characters start to smile after a while if the player occasionally 'listens' to them, but will look sad if players just 'hear' what they have to say. Secondly, the player needs to indicate his/her perception with respect to the impact of the decision on the customers, internal staff or the general public (see Fig. 4). After the player provides this in-situ input, the player automatically returns to the game.



Question: Do you wait for the executive to arrive in Aberdeen and issue a press communique from there?

Your answer: Yes

Indicate (optional)

which information items were important to base your decision on.

Very careful Legal Officer	□	Communicate Human Resources Manager	□
Reassure Industry Director	□	No delay Communications Officer	□
Cause CEO	☑	Casualties Human Resources Manager	□
Issues Business Continuity Manager	□	Customers Legal Officer	□

1 / 2

Question: Do you wait for the executive to arrive in Aberdeen and issue a press communique from there?

Your answer: Yes

Indicate (mandatory)

how you think the following groups will react to your decision.

Customers	✓	☹	😐	☺
Staff	✓	☹	😐	☺
The public	✓	☹	😐	☺

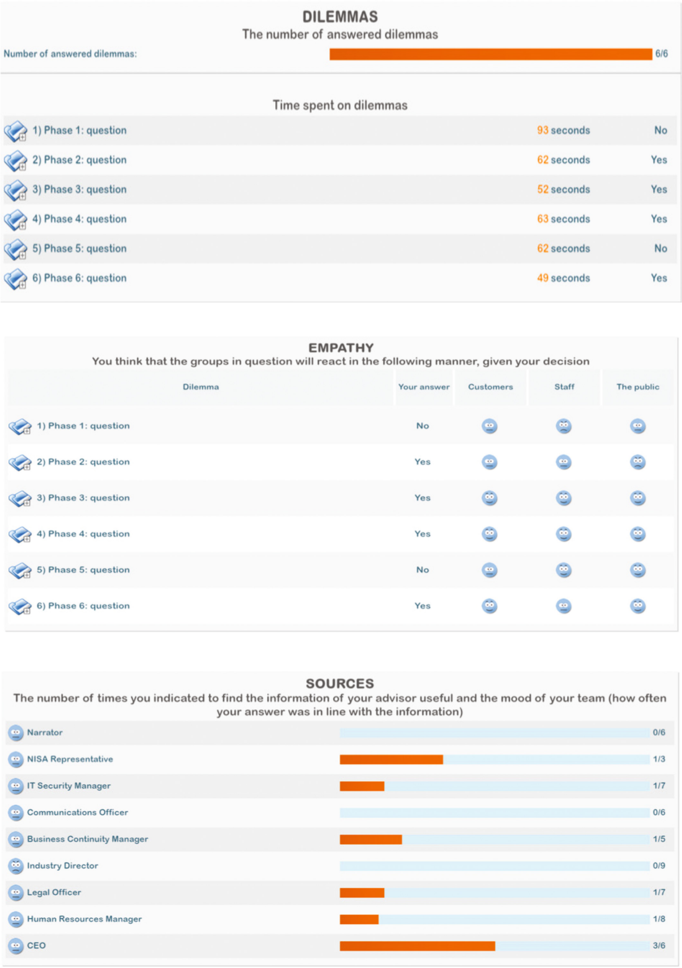
2 / 2

Fig. 4. In-situ input

The game ends when all dilemmas have been answered. The player may read all information items, and even advices what to decide from his/her team members, but it is up to the player to decide if and when s/he uses this information.

3 Game Statistics

First, we generate simple descriptive statistics about the time needed to answer dilemmas, the number of dilemmas answered, the number of times advices of various team members were indicated as important (Fig. 5). This is done on an individual level and provided as feedback to the player. Further analysis is done on aggregating levels based on all game log-files.



DILEMMAS

The number of answered dilemmas

Number of answered dilemmas: 6/6

Time spent on dilemmas

1) Phase 1: question	93 seconds	No
2) Phase 2: question	62 seconds	Yes
3) Phase 3: question	52 seconds	Yes
4) Phase 4: question	63 seconds	Yes
5) Phase 5: question	62 seconds	No
6) Phase 6: question	49 seconds	Yes

EMPATHY

You think that the groups in question will react in the following manner, given your decision

Dilemma	Your answer	Customers	Staff	The public
1) Phase 1: question	No	😐	☹	😐
2) Phase 2: question	Yes	😐	😐	☹
3) Phase 3: question	Yes	🙂	🙂	🙂
4) Phase 4: question	Yes	🙂	🙂	🙂
5) Phase 5: question	No	😐	🙂	🙂
6) Phase 6: question	Yes	🙂	😐	🙂

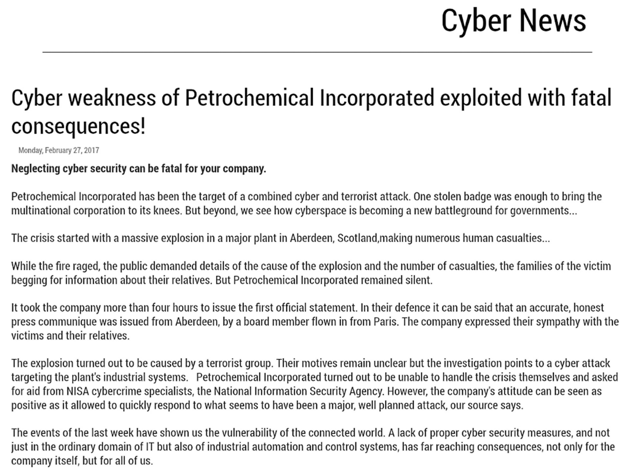
SOURCES

The number of times you indicated to find the information of your advisor useful and the mood of your team (how often your answer was in line with the information)

😐 Narrator	0/6
😐 NISA Representative	1/3
😐 IT Security Manager	1/7
😐 Communications Officer	0/6
😐 Business Continuity Manager	1/5
☹ Industry Director	0/9
😐 Legal Officer	1/7
😐 Human Resources Manager	1/8
😐 CEO	3/6

Fig. 5. Game descriptive statistics

Second, we generate a newspaper article where the narrative is based on the choices the player made during gameplay (Fig. 6).



Cyber News

Cyber weakness of Petrochemical Incorporated exploited with fatal consequences!

Monday, February 27, 2017

Neglecting cyber security can be fatal for your company.

Petrochemical Incorporated has been the target of a combined cyber and terrorist attack. One stolen badge was enough to bring the multinational corporation to its knees. But beyond, we see how cyberspace is becoming a new battleground for governments...

The crisis started with a massive explosion in a major plant in Aberdeen, Scotland,making numerous human casualties...

While the fire raged, the public demanded details of the cause of the explosion and the number of casualties, the families of the victim begging for information about their relatives. But Petrochemical Incorporated remained silent.

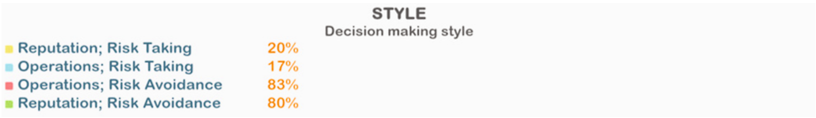
It took the company more than four hours to issue the first official statement. In their defence it can be said that an accurate, honest press communique was issued from Aberdeen, by a board member flown in from Paris. The company expressed their sympathy with the victims and their relatives.

The explosion turned out to be caused by a terrorist group. Their motives remain unclear but the investigation points to a cyber attack targeting the plant's industrial systems. Petrochemical Incorporated turned out to be unable to handle the crisis themselves and asked for aid from NISA cybercrime specialists, the National Information Security Agency. However, the company's attitude can be seen as positive as it allowed to quickly respond to what seems to have been a major, well planned attack, our source says.

The events of the last week have shown us the vulnerability of the connected world. A lack of proper cyber security measures, and not just in the ordinary domain of IT but also of industrial automation and control systems, has far reaching consequences, not only for the company itself, but for all of us.

Fig. 6. Generated newspaper article

Third (see Fig. 7), the players' decisions are related to two different risk taken vs. risk avoidance dimensions, 'risk taken/avoidance behaviors regarding reputational risks' and 'risk taken/avoidance behaviors regarding operational risks'. Reputational risks, often called reputation risks, are risks of loss resulting from damages to a firm's reputation. Operational risks are risks of loss resulting from inadequate or failed internal processes, people and systems or from external events. The scoring is based on an in-game algorithm defined by a subject matter expert with domain knowledge [4, see also acknowledgement].



STYLE

Decision making style

🟨 Reputation; Risk Taking	20%
🟦 Operations; Risk Taking	17%
🟥 Operations; Risk Avoidance	83%
🟩 Reputation; Risk Avoidance	80%

Fig. 7. Risk reputational/operational leadership style

4 Game Analytics

The individual game log data files can be further analyzed into meaningful information to shed light on human reasoning aspects. This makes it possible to examine team and group behaviors across a number of other parameters as well e.g., level of expertise, gender, country, culture, business domain, etc. That activity is still underway and experiments conducted and data gathered will be addressed in future papers. In the following, we are basically pointing to methodological aspects, and the sorts of data and informational patterns we can get out of this type of microworld. We will not provide psychological nor economical interpretations of the statistics and analytics at this point[1].

The analytics provided here are based on the Cyber Security game (cis.txchange.nl) that ran on-line between 2015–2017. The game was accessible on the internet and over this period played 887 times. The web-based game is available in two languages: English and French. Figure 8 show the number of times the game is played across this timespan.

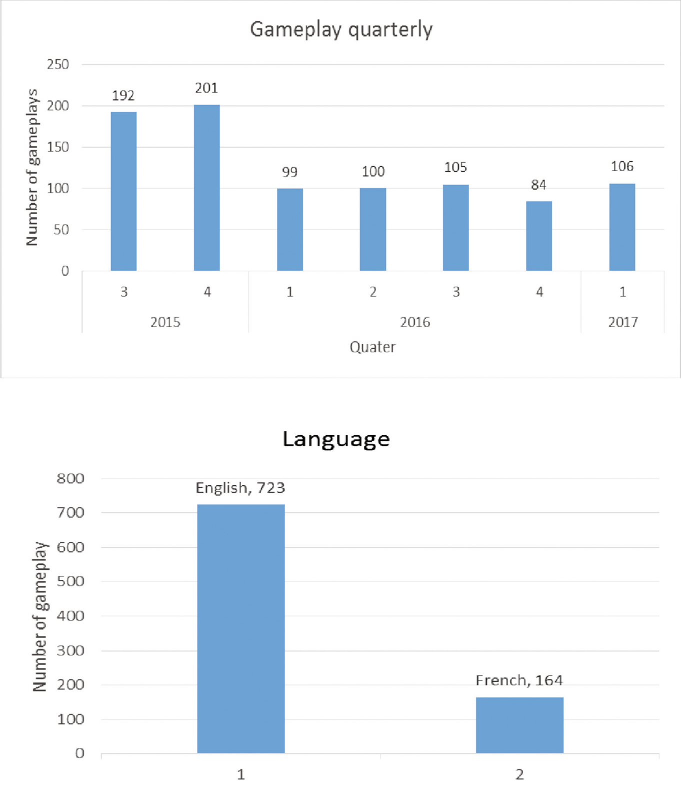


Fig. 8. Total number of game plays over time and between two game variants (English and French version)

For analytical purposes we selected a dataset out of the total of 887 game log files available. We assumed that not all 887 games were played ‘seriously’. We consider a seriously played game when (1) at least 4 out 6 dilemmas were answered, (2) at least for 3 dilemmas minimal 2 information items were opened, (3) that the game play duration at least 7 min took but not longer than 35 min. Thus, a ‘seriously’ played

[1] To falsify your own hypotheses and utilize the data files please contact the authors of this paper.

game utilizes all available game mechanics for several times. Based on these selection criteria we ended up with 377 (out of 887) seriously played games with an average playing time around 10 min. We used these 377 game loggings for further visual inspection and analyses.

Figure 9 illustrates the overall scoring with respect to the playing styles. Reputation risk taking 42% vs. Reputation risk avoidance 58%, Operation Risk taking 29% vs. Operation risk avoidance 71%. These figures are in line with the negativity bias in a plethora of situations related to risk-averse behaviors. Operational risk is the risk arising from execution of a company's business function. And, focuses on the risks arising from people, processes, and systems, including external events that affect a company's operations. Our data indicate that those who played the game are more risk averse regarding operational - than reputational issues. Reputational risk may arise from operational risk but is not, in and of itself, an operational risk.

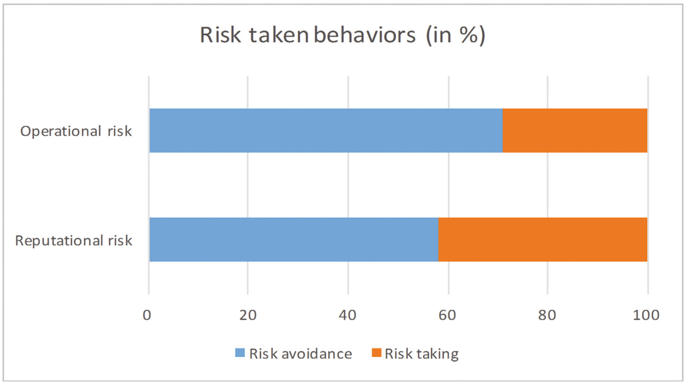


Fig. 9. Risk taken vs. risk avoidance behaviors regarding operational and reputational risks

Figure 10 shows that all scores significantly differ from the 50% change level; using the Nonparametric one-sample Binomial test (significance level is 0.5).

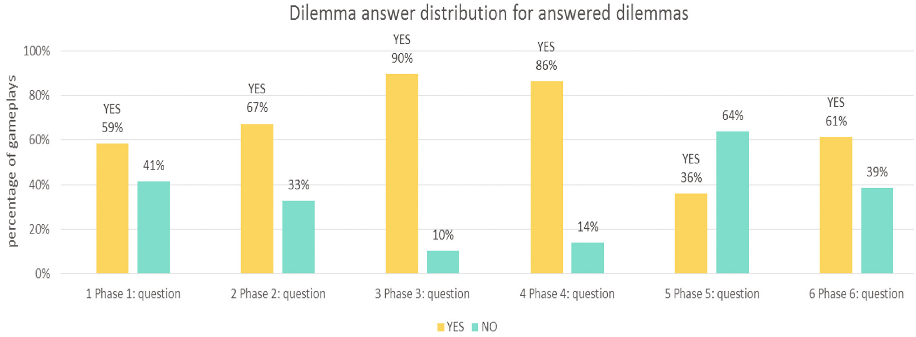


Fig. 10. Yes/no distribution across dilemmas

Figure 11 illustrates the difference in risks behaviors across the different phases during the crisis. The first phase characterized the beginning of the crisis, in the second phase the crisis starts to get going, in the third phase it escalated to reach its climax in the fourth phase, in the fifth phase the company was no longer to target of cyber attacks and the crisis was really over in de sixth phase. In addition, the average decision times for all dilemmas are depicted as well. Note that the no data/scoring was available for reputational risk for the third dilemma.

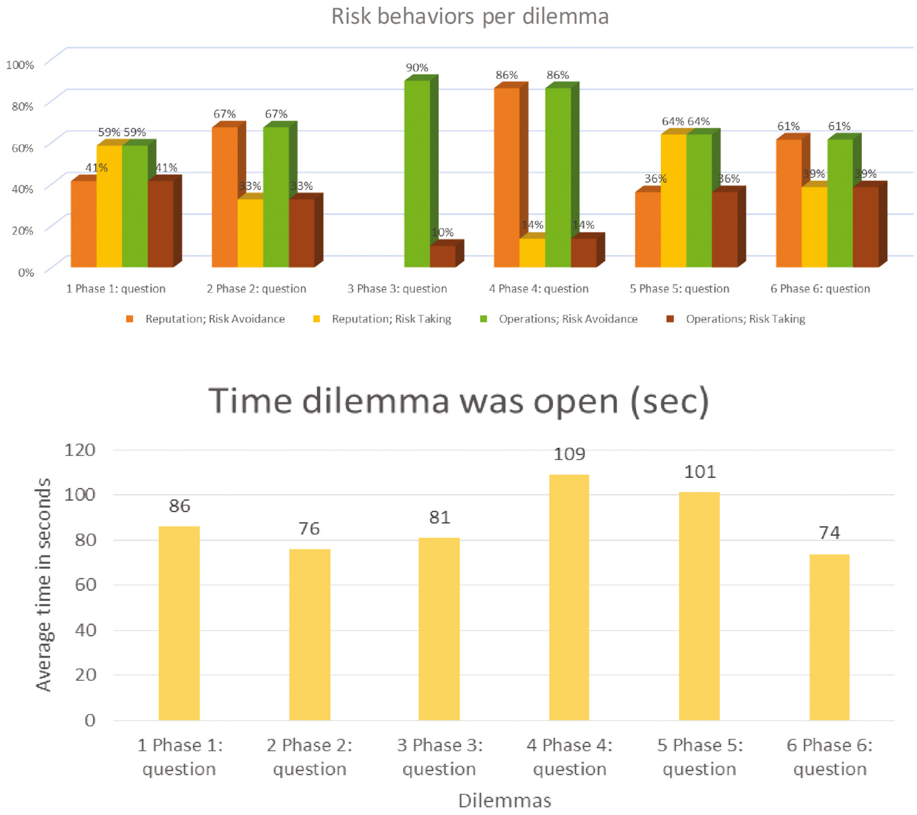


Fig. 11. Risk behaviors and decision-making times per dilemma

Figure 12 shows the percentage that an information item provided by a specific virtual character sitting at the table is opened (in red) and considered important (in blue) by the player. Immediately below the graph regarding the total percentages across the dilemmas.

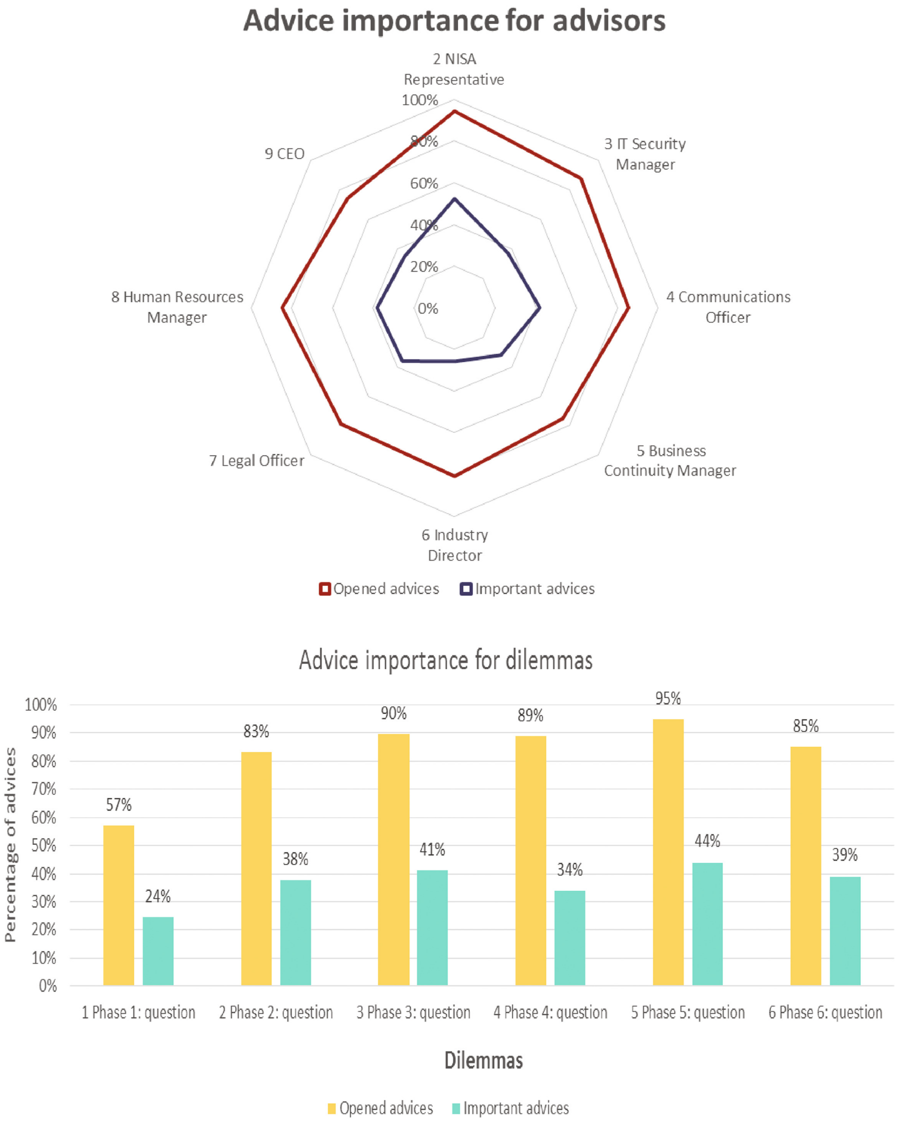


Fig. 12. Advices by virtual characters

Finally, Fig. 13 depicts how many times (in percentages) the player asked for a voting advice (yellow graph) and the times they followed (implicitly) the voting advices by the virtual characters. And in the figure underneath, the average voting advices per dilemma.

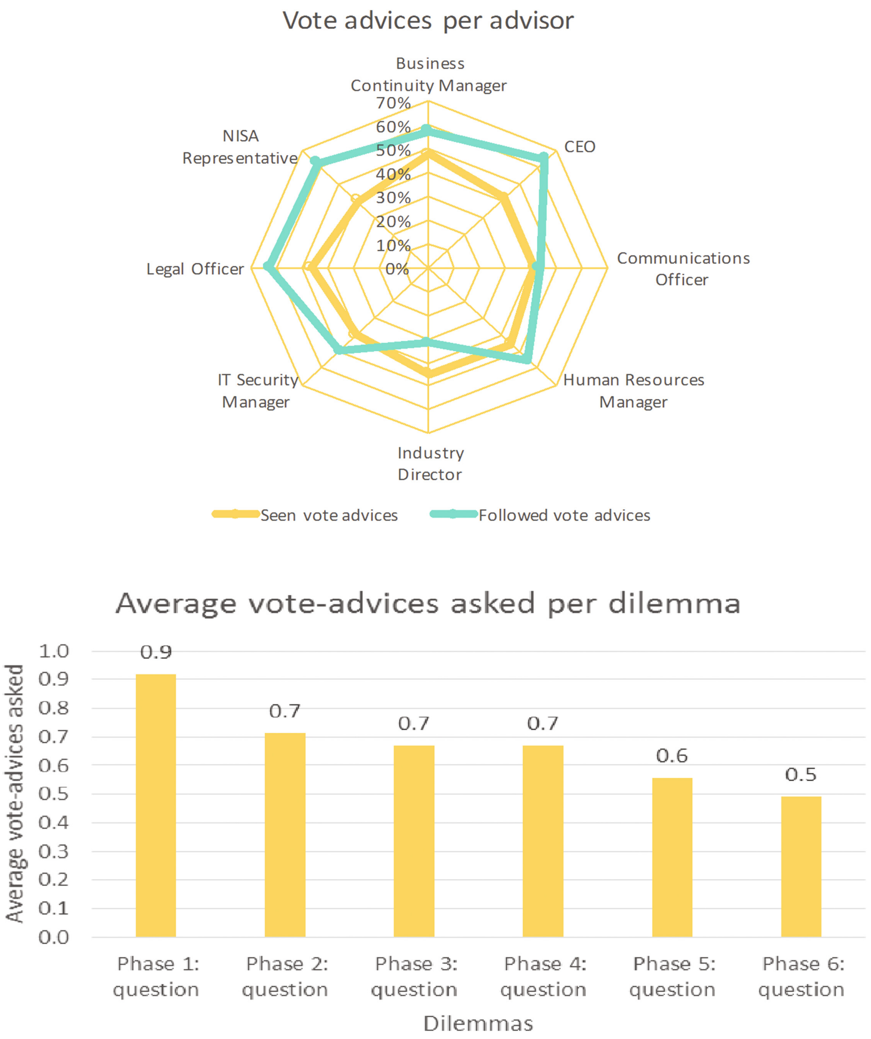


Fig. 13. Voting advices

5 Conclusion

The general goal the present paper was to show that microworlds can provide data and information that can be used for elucidating dynamic decision making models. This was illustrated by risk behaviors during a Cyber attack. We conclude that game based microworlds will bring us statistic and analytics in understanding how we think, reason, and decide. This type of data can be used by researchers to falsify their hypotheses. For example, related to research questions on the type and occurrence of risk behaviors during several crisis situations. Our future work is focusing on the unobtrusive measurement of competency where we explore the combination of several top-down (e.g. Bayesian networks) and bottom-up data mining techniques to analyze and predict human behaviors. We not only focus on competencies but also on preferred playing styles during game flow, in terms of actions, tactics, and strategies for managing the uncertainty and dynamics in the game [7, 8]. The latter is important, since player strategies are suggested as predictors regarding transferability from in game to out of game behaviors [9].

Acknowledgments. The game scenario was developed with Paul Théron [4] a Thales cyber security expert.

References

1. De Heer, J.: How do Architects think? A game based microworld for elucidating dynamic decision-making. In: Auvray, G., et al. (eds.) Complex Systems Design and Management, pp. 133–142. Springer International Publishing, Cham (2016). doi:10.1007/978-3-319-26109-6_10
2. Brehmer, B., Dorner, D.: Experiments with computer simulated microworlds: escaping both the narrow straits of the laboratory and the deep blue sea of the fields study. Comput. Hum. Behav. **9**, 171–184 (2003)
3. Gonzalez, C., Lerch, J.F., Lebiere, C.: Instance-based learning in dynamic decision making. Cogn. Sci. **27**, 591–635 (2003)
4. Théron, P.: Informing business strategists about the cyber threat: why not play serious games? In: Hills, M. (ed.) Why Cyber Security is a Socio-Technical Challenge: New Concepts and Practical Measures to Enhance Detection, pp. 129–158. Northampton Business School, University of Northampton, UK (2016). ISBN 978-1-53610-090-7
5. Klabbers, H.G.: The Magic Circle: Principles of Gaming and Simulation, 3rd edn. Sense Publishers, Rotterdam (2009)
6. Schell, J.: The Art of Game Design: A Book of Lenses, 2nd edn. AK Peters/CRC Press, Natick (2008)
7. Bakkes, S.C.J., Spronck, P.H.M., van Lankveld, G.: Player behavioural modelling for video games. Entertainment Comput. **3**, 71–79 (2012)
8. Ross, A.M., Fitzgerald, M.E., Rhodes, D.H.: Game-based learning for system engineering concepts. In: Conference on Systems Engineering Research, pp. 1–11 (CSER 2014) (2014)
9. Kaser, T., Hallinen, N.R., Schwartz, D.L.: Modeling strategies to predict student performance with a learning environment and beyond. In: Proceedings of the Seventh International Learning Analytics and Knowledge Conference, LAK 2017, pp. 31–40 (2017). ISBN 978-1-503-4870-6

Culture + Cyber: Exploring the Relationship

Char Sample[1(✉)], Jennifer Cowley[1], Steve Hutchinson[1], and Jonathan Bakdash[2]

[1] US Army Research Laboratory, Adelphi, MD, USA
{char.sampl,Jennifer.Cowley,Steve.Hutchinson}@icfi.com
[2] US Army Research Laboratory, Aberdeen, MD, USA
jonathan.z.backdash.civ@mail.mil

Abstract. Distinguished social psychologist Geert Hofstede observed, “This dominance of technology over culture is an illusion. The software of the machines may be globalized, but the software of the minds that use them is not.” The role of culture in the thought process is prevalent, yet unstated, that many cultural beliefs and biases are accepted as truths. Cultural beliefs and biases are incorporated into the thought process where they reveal themselves in patterns of thought. Once the thought patterns are established they may be observed in the digital trail that results from online interactions. Once captured online, the behaviors can be reviewed and examined in multiple ways so that researchers can gain new insights.

Historically, observations have taken place in the physical environment; this talk discusses findings of cultural markers in the cyber realm. The results of evidence-based research exploring the relationship between national culture and cyber behaviors will be discussed. These quantitative, observational studies were the result of researchers mining the raw website defacements found in the Zone-H archives containing over 10 million records. Mining the dataset and evaluating the findings within Hofstede’s cultural framework allowed for research into behaviors, preferences, reasons, imaging, sentiment analysis, and various other aspects of attacker and victim cybersecurity actors. The use of Hofstede’s six dimensional cultural framework to define culture, along with some basic inferential statistics, resulted in specific digital identifiers that were associated with specific cultural dimensions. Over time findings can be trended, allowing for more accurate modeling of cyber actors based on cultural values. The results supported Nisbett’s observation that people “think the way they do because of the nature of the societies they line in”.

This discussion centers on the six dimensions of culture, the values associated with each dimension, and examples of those values in cyber space for victims, attackers and defenders. The six cultural dimensions measure views on self-determination, collectivism, aggression, nurturing, uncertain outcomes, holism, instant gratification, and levels of societal openness. The behavioral traits that associate with the cultural values are behavioural traits that are consistent with cyber behaviors.

Cultural values provide context for individual behaviors by determining the norm for a group. Thus, behavior that may seem perfectly normal in one environment may stand out as odd in a different environment. Cultural difference have been historically used to model adversaries in the kinetic world. Moving this analysis into the cyber realm offers the potential to gain greater insights into all cyber actors.

Keywords: Culture · Cultural dimensions · Attackers · Victims · Cyber · Vectors

D. Nicholson (ed.), *Advances in Human Factors in Cybersecurity*,
Advances in Intelligent Systems and Computing 593, DOI 10.1007/978-3-319-60585-2_18

1 Introduction

The statement by Hofstede et al., that "Culture is everything" [1] still resonates today in the cyber environment, even if this area of research has not be widely studied to date [2]. Hofstede et al., defined culture as the "collective mental programming that distinguishes one group of people from another" (p. 6) [1]. The possibility that a cyber actors could be grouped according to culture has been illustrated through various studies [3, 4, 5, 6, 7]. These studies appear to reinforce Nisbett's (2010) observation that people in different geographic regions see the world differently [8]. This different view informs perception and reactions [9, 10] further reinforcing Hofstede's comment that while computer systems are standardized the manner in which they are used differs according to the mind of the operator [1].

Interactions in the digital environment leave evidence; thus the persistent nature of the public digital environment assures the researcher that observable data is available [11]. Observable cyber data or digital artifacts resulting from benign and malicious online transactions, can be characterized according to three general types of actors; attackers, defenders, and victims. These digital artifacts are available for analysis by various disciplines that may produce new insights to existing problems. This paper consolidates the findings of several published and pre-published cross-discipline studies that characterized different types of humans as cyber actors. Our researchers herein seek to group certain patterns of observed network behaviors by profiles of cultural dimensions [12] from Hofstede's model [1].

2 The Problem

One way to improve cybersecurity modeling is to improve the predictive capabilities of cyber actors. Predicting strategies, tactics and other behaviors offers the opportunity to customize environments and actor responses. Attribution and analysis is time-consuming, expensive, and post-hoc what is needed is a grouping variable across classes of actors that may be a good predictor of the types of strategies and techniques they will use, based on values and priorities in the cyber environment.

Cultural anthropologists use archeology to study the artifacts a particular group of people leaves in an environment. The anthropologist infers how this artifact explains the culture of the creator and user. Similarly, networks produce digital artifacts that require analysts to infer values and priorities of cyber actors. We believe that culture could be a viable predictor in cybersecurity network actor modeling. The purpose of this manuscript is to review prior work on characterizing attackers, defenders and victims to identify literature gaps that if addressed, may enable better culture-based predictions of cyber actor maneuvers.

3 Culture

Hofstede's definitions of culture share constructs with other definitions promulgated over time that share characterized definitions [13, 14]. For example, definitions of culture were characterized with these parameters [15]: structure/pattern, function, process, product(artifacts), refinement (intellect/morality), power or ideology, group-membership

One assumption underlying these definitional characteristics is that definitions assume cultural coherence and uniformity within a collective [13] meaning that social conformance to a set of ideas, beliefs, behaviors, etc. exists to obtain cultural membership. But sociology and other cultural researchers have contradicted that assumption with empirical evidence [1, 16, 17, 18]. Another criticism of cultural research is that cultural definitions are often broad, politicized and folklorist [13] such that the true semantic meaning of the term is obfuscated.

3.1 Cultural Predictions in Past Research

We briefly review what culture has predicted in past research using any theoretical model of culture and then review what Hofstede's cultural model predicts at the national level. First, we understand that culture can predict emotive human states like well-being [18] and trust [19, 20]. We also know that culture can predict the occurrence of behavior and respective outcomes at the individual level [21] like ethical decision-making [22], educational performance [23] interpersonal information exchange [24], and the use and acceptance of information technology [25].

However, Hofstede's cultural dimensions at the national-level have been related to certain preferences and outcomes. For example, Shane [26] low UAI cultures were more preferential to innovation within organizations and high IvC cultures have the lowest level of information-seeking behaviors within their social networks [27]. The purpose of this section is to demonstrate that predictive capability exists. For a more comprehensive meta-analytic qualitative review of this phenomenon that Hofstede's cultural dimensions could predict or explain at the individual, organizational and nation-state level, see Kirkman et al. [28] (p. 299).

3.2 Cultural Values in Cybersecurity

How does Hofstede's definition of culture relate to prior research in cybersecurity culture research? We consider that culture herein defines shared knowledge, values, attitudes, linguistics and respective behaviors that are shared by a group. The possibility that these national collectives could serve as a grouping variable of a particular cyber behavior or preference has been illustrated through the various studies [3, 4, 5, 6, 7]. These studies appear to reinforce Nisbett's (2010) observation that people in different geographic regions see the world differently and respond to it in habitual patterns [8]. Nisbett's view informs the study of human perception and behavioral responses [9, 10] further reinforcing Hofstede's idea that while computer systems standardized the manner in which humans use them, this mind of the operator can also shape the way the he/she interacts with the computer systems [1]. These computer systems are designed with some of the same constructs as the human mind [29], yet their usage differs.

Several decades of research have characterized human thought patterns by national culture. Culture's role in thought has been long documented [1, 8, 12, 30, 31, 32, 33]. Cultural values are transmitted and reinforced throughout society via social institutions.

This transmission extends into digital environment, where virtual social based institutions exist. Social reinforcement of cultural values and norms results in unconscious patterns in behavior and thought [8, 21, 30, 31, 32].

Hofstede used national groupings for his framework. Hofstede et al. [1] has identified cultural values for over 100 countries across six dimensions scored from 0 to 100. Cultural values guide and direct responses to environmental stimuli, along with the tools used to do work and language patterns of communication [33].

The persistent nature of the public digital environment assures the researcher that observable data is readily available for future studies [11, 34, 35, 36, 37]. Thus, observable data or digital artifacts resulting from benign and malicious online transactions, can be characterized according to three general types of actors; attackers, defenders, and victims. This paper consolidates the findings of several published and pre-published studies that characterized different types of cyber actors [3, 4, 5, 6, 7]. Our researchers herein seek to group certain patterns of observed network behaviors by profiles of cultural dimensions [12] from Hofstede's model [1].

3.3 Hofstede's Cultural Theory and the Exploration of Cybersecurity Actors

The six dimensions that Hofstede et al. [1] use to define cultural values are; power distance index (PDI), individualism versus collectivism (IvC), masculinity versus femininity (MvF), uncertainty avoidance index (UAI), long-term orientation versus short-term orientation (LvS) and indulgence versus restraint (IvR). Table 1 provides a brief explanation of values for each cultural dimension, a more complete explanation may be found in Hofstede's publications [1, 6, 7].

Table 1. Hofstede's six dimensions of cultural values

Dimension	Description
PDI	PDI describes a societal belief on the origination of power in that society (top vs. bottom)
IvC	IvC describes the societal preference in the role of the person in the larger society
MvF	MvF describes societal views on gender roles and how that outlook shapes conflict management
UAI	UAI describes the level of anxiety associated with new or unknown objects or events
LvS	LvS describes the level of willingness in a society to wait for rewards on gratification
IvR	IvR describes acceptable levels of expression

Beyond Hofstede's cultural values Nisbett [8] identified cultural differences between Eastern cultures developed in the tradition of Confucius and Westerner cultures developed in the tradition of Aristotle many of which align with Hofstede's cultural value dimensions listed in Table 1. These findings reveal profound differences in language constructs, and perception, community and events [8]. Furthermore, these differences in values match many of the values that Hofstede identified in his cultural dimension framework. Table 2 contains a summary of cultural differences between East and West mapped to the corresponding values found in Hofstede's framework.

Table 2. East/West Cultural Descriptions Mapped to Hofstede's Dimensions

Hofstede's Cultural Values	Nisbett's Cultural Descriptions	
Dimension(s)	East	West
IvC, LvS	Relationship oriented	Object oriented
MvF	Control environment	Manage environment
UAI	Circular Perspective	Linear perspective
IvC, UAI, LvS	Circumstances important	Prone to fundamental attribution error.
PDI, IvC	In group	Individualism
PDI, IvC	Commonality	Exceptionalism
IvR	General knowledge of many subjects	Single subject mastery
PDI, IvC	Respect for authority	Speak truth to power
PDI, IvC	Prefer working in groups, roles well-defined	Prefer working alone
PDI, IvC, MvF	Learn through sharing and assistance	Learn through debate
IvC	Group objects by functions	Categories and taxonomies
PDI, IvC, LvS	Smaller piece of larger whole (ropes in a net)	Egalitarian, self-determination
LvS	Comfortable with complexity	Prefer to deconstruct
UAI, LvS, IvR	Plausible over logic	Logic over plausible
UAI, LvS	Senses inform logic	Logic over senses
IvC, MvF, IvR	Harmonious relationships	Personal goals

Sample (2013) [38] discovered a statistical link between PDI and IvC cultural dimensions and patriotic, political website defacements. Subsequent observational studies continued to support the initial suggestion that national culture can characterize patterns of cyber behaviors. However, we believe that additional research would help to gain additional insights into cyber actor values for future predictive purposes.

A cultural profile of cyber actors values can help explain and characterize cyber norms for groups of people, and assuming that cultural values are susceptible to slow evolution, they may provide good prediction factors. Thus, knowing the cultural values of cyber actors might predict the efficacy of attacking and defending outcomes. Prior research has identified relationships between a profile of cultural values and attackers [3, 4], defenders [7] and victims [5, 6], although, the profile does not explain acts by individuals. Individuals' behaviors remain a research area for psychologists.

4 Studies

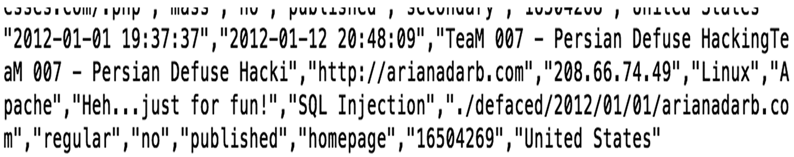
Culture can explain similarities as well as differences [12] within and between groups (attackers, defenders and victims) so we aim to statistically describe new relationships.

The studies examined contain insights into attacker preferences [3, 4], defending strategies [7], and victim analysis [5, 6]. All of the studies were observational in nature, and relied on publicly available data. All of the studies were performed with the goal of determining if statistical similarities existed between the actor groups being studied.

Only one study, relied on data found at the ICANN website (www.icann.org) [35] all other studies relied on subsets of data contained in the Zone-H (www.zone-h.org) [34] archives from the years 2005 – 2014. The use of large sets of data in conjunction with Hofstede's operationalized data (www.geert-hofstede.com) [36], allowed for quantitative analysis for each of the studies. The evidence-based quantitative studies allowed the researchers to reduce cultural biases during the analysis phase, a common problem during analysis [39] and base findings on objective data [40]. The following paragraphs summarize the findings from each of the studies.

4.1 Data

All data sources used for the studies were primary sites. The Zone-H data, used in both attacker and victim studies is a rich dataset comprised of over 13 million records that are a mix of structured and unstructured data. The studies that used Zone-H data, to date have relied on the metadata fields. Figure 1 provides an example of a raw Zone-H record.



```
csses.com/.php","mass","no","published","secondary","16504266","United States"
"2012-01-01 19:37:37","2012-01-12 20:48:09","TeaM 007 - Persian Defuse HackingTe
aM 007 - Persian Defuse Hacki","http://arianadarb.com","208.66.74.49","Linux","A
pache","Heh...just for fun!","SQL Injection","./defaced/2012/01/01/arianadarb.co
m","regular","no","published","homepage","16504269","United States"
```

Fig. 1. Zone-H record example

The studies to date have relied on the third (attacker), fourth (victim), eighth (reason), ninth (method, or attack vector) and last (location) fields. While the Zone-H archive contains over 13 million records, filters for each of the studies reduces the number of records based on the criteria for the study. For example, the *Cultural Exploration of Attack Vector Preferences for Self-Identified Attackers* study used a pool of 466220, but when filtering by attack vectors, the number of records was reduced to 267556 records. Each of the studies used for this paper contain the information on the number of records used and the study parameters.

The Internet Corporation for Assigned Names and Numbers (ICANN) maintains a list of DNSSEC signed zones [35]. This list may be found at http://stats.research.icann.org/dns/tld_report [Ibid]. The 2015 *Culture and Cyber Behaviours: DNS Defending* study relied on values collected in 2014, many top-level domains (TLDs) and secondary domains were not included since this study used only country code TLDs (ccTLDs). Since that time, other TLDs also may have DNSSEC signed their zones.

One other source of data used in all of the studies is Hofstede's website (www.geert-hofstede.com) [36]. This site contains the most recent updates of the cultural values. Hofstede's initial survey included values for 78 countries across four dimensions.

Presently, the site supports findings on over 100 countries with values defined across 6 dimensions. The studies rely on the 2013 datasets.

Attacker Studies. The 2016 *Re-thinking Threat Intelligence* study [3] challenged the assertion of a single hacker culture that follows a single playbook dictated by SANS [41] in two specific areas. The first challenge showed that the preferences did not follow the order included in the SANS [Ibid] and MITRE [42] lists for the year examined. The second, finding, not included in the publication extracted from the data made available, showed that some attackers appeared to randomize attack vectors while others did not[1]. This small finding suggests the need for additional research in exploring the relationship between culture and decision-making in cyber security.

The 2017 study *Cultural Exploration of Attack Vector Preferences for Self-Identified Attackers* [4] presented a larger and longer examination of a small but more diverse group of attackers using 267556 records. This study relied on 10 years of attacks, using attack vectors identified in MITRE Cyber Observables EXpression (CybOX^{TM}) (http://cybox.mitre.org) framework [43] by self-identified attackers. The included attack vectors were, undiscovered or zero-day, brute force (exhaustive combinations), configuration/administrator error (resulting from a misconfiguration of permissions), mail (including phishing using embedded links and other attacks on the mail server), password sniffing, social engineering, and SQL injection (results from entering an unexpected input string into a database creating in an unhandled event resulting in a software failure) [44].

Each of these attack vectors was represented by a sufficient number of countries (14% – 35%) so that comparisons and analysis could be performed with confidence. The findings showed that cultural preferences exist when attackers deploy attack vectors. High PDI strongly associated with zero-day, social engineering and SQL-injection attackers. Masculinity associated with brute force, configuration/administrator error, and password sniffing attackers. High UAI values and restraint also associated with social engineering attackers and SQL-injection also associated with high UAI values. These findings support the researchers' suggestion that some attack vectors appear to seem more attractive to attackers based on their mental software, supporting the assertion of globalized software being used by culturally influenced minds [1].

Victim Studies. The 2016 study *Hofstede's Cultural Markers in Successful Victim Cyber Exploitations*, on victims of SQL-injection attacks found that statically speaking, high PDI values were common among the victim countries [5]. This study relied on 1971 SQL-injection attack records for the years 2011–2014. The resultant finding was consistent on both of the commonly used platforms, IIS and Apache.

The 2017 victims of social engineering attacks study *Cultural Observations on Social Engineering Victims* [6], used a collection of 87218 records for sorting and comparison over the period of 2011–2014. The results of this study showed that the

[1] The small number of records and the scope of the study made this an unreported finding. Of the 20 countries examined 10 showed strong preferences where the frequency of first choice vector was used in over 50% of the cases and 5 showed no preference where the frequency of the first choice and other choices were evenly distributed. The group that showed the median value of the no preference group's long-term orientation to be 20.

victim countries of social engineering attacks tended to be masculine, individualist and long-term oriented when compared against the non-victim countries. The attackers and the victims appear to show differing weaknesses where certain cultural values are present.

Defender Studies. The 2015 Domain Name System Security Extensions (DNSSEC) study, *Culture and Cyber Behaviours: DNS Defending* [7], was performed on global cyber defenders. This observational study examined which top-level domains (TLDs) were digitally signed using DNSSEC. The act of signing a zone, equates with being a good net citizen [45] but does not assure additional security until keys are exchanged; however signing a zone is indicative of plans to share DNSSEC information with other signed zones since this is required for zone sharing information [46], and a signed TLD may hold and manage keys for children domains (Ibid). This study was focused on TLDs. The findings from this study revealed that the countries with signed TLD zones were low PDI, individualistic and long-term oriented.

Consolidation of the Studies. The studies performed to date on cyber actors are summarized in Tables 3, 4 and 5 are the consolidated findings. Table entries with a value of "X" indicate fields where statistically significant findings were not present. The remaining entries indicate the cultural values that relate to the activity.

Table 3. Observed Attacker Preferences

Attacker Preference	PDI	IvC	MvF	UAI	LvS	IvR
[a] Self identification	X	X	Masc.	High	X	X
[b] Vector preference	High	X	X	X	X	X
[b] No vector preference	X	X	X	X	Short	X
[c] Zero day	High	X	X	X	X	X
[c] Brute force	X	X	Masc.	X	X	X
[c] Config./admin error	High	X	Masc.	X	X	X
[c] Mail	High	X	X	X	X	X
[c] Password sniffing	High	X	Masc.	X	X	X
[c] Social engineering	High	X	X	High	X	Restr.
[c] SQL-injection	High	X	X	High	X	X

[a] n=466220 entries, 45 self-identified attackers, MvF and UAI $p <= 0.05$; [b] n=215892, vector preference based on vector frequency PDI median score 79, no vector preference based on vector frequency LvS median score 20; [c] n=267556, 41 countries, $p <= 0.\ 05$

Table 4. Observed victim vulnerabilities

Victim vulnerabilities	PDI	IvC	MvF	UAI	LvS	IvR
[d] SQL-injection	High	X	Masc.	X	X	X
[e] Social engineering	X	Indiv.	Masc.	X	Long	X

[d] n = 1971 records, 51 countries, $p <= 0.05$; [e] n = 2758, 77 countries, $p <= 0.05$

Table 5. Observed defender preferences

Defender preferences	PDI	IvC	MvF	UAI	LvS	IvR
[f] DNSSEC signed zones	Low	Indiv.	X	X	Long	X
[f] DNSSEC unsigned zones	High	Coll.	X	X	Short	X

[f] n = 100 records, p <= 0. 05

4.2 Beyond Hofstede

While prior research elucidated the possibility that cultural dimensions could characterize cyber actors, this provides us with a superficial knowledge of each entity. If Hofstede's profiles of culture are not mutually exclusive to a particular nation state, then attacker attribution could be an array of possible attackers rather than a single entity. To improve understanding of cyber actors, the research community needs more nation specific data on typical digital tools used at various stages of the attack, typical linguistic patterns and other artifacts. Archeology has various methods and processes for cultural inference given an array of artifacts that cybersecurity could use [47, 48, 49]. How could researchers combine Hofstede's cultural values with inferences produced from digital artifacts, to improve cyber actor understanding?

5 Discussion

An understanding of cultural factors in the physical world is limited, even less so for cyber. However, cyber behaviors are observable using digital trails of cognitive and social processes. Examining observable cyber behaviors offers multiple ways to gain new insights into culture in cyber and the physical world. We synthesize previous research and suggest future research directions for culture and cyber behaviors using a large archive of website defacements from Zone-H. The studies examined different aspects of cyber security to include attacking and defending (both successful and unsuccessful).

Attackers generally appeared to exhibit high PDI values, and depending on the attack vector, the PDI values, while high, were variable in the high range suggesting that some of the most hierarchical nations prefer specific vectors. The differences observed in other dimensions between the self-identified attackers suggest a potential for profiling attacks by cultural values. Meanwhile, successful defending also appears to relate to long-term orientation and in some cases feminine and collectivist values. Masculine values appear to associate with a higher instance of victimization but this does not suggest causation, rather this is an observation that aligns with an earlier study [50].

6 Conclusion

The studies examined in this analysis represent the entry point into a new area of research that has not been widely studied [2]. This line of research offers opportunities to learn about values and norms for cyber actors, particularly group affiliated actors in the virtual environment of the cyber battlefield. These post-hoc observational studies, allowed the researchers to observe artifacts from the subject behaving naturally.

The studies to date used the only the structured metadata fields of the Zone-H archives [34]. The actual content of the defaced pages can be viewed in the same manner as street art or graffiti [51] while the medium is different the messaging, sentiment and artistry are consistent with the physical world. Analysis on content, sentiment, imaging, group linkages remain to be discovered. Many of the defacements, especially those by long active groups, may provide insights into changes over time that may suggest unique trends and strategies along the lines of how groups grow, changing relationships between hacker groups, and changes in response to economic, geo-political or other events.

There are many new insights to be gained by examining the Zone-H data archive [34] where attackers express their values, attitudes, believes, likes and dislikes, in short their thoughts. The victims in this large dataset can be examined over time for evidence of cybersecurity weaknesses as suggested in the exemplar victim studies [5, 6] and the attackers can be examined over time may show preferences [3, 4] and cyber defenders can learn from studies on both groups of cyber actors.

Acknowledgments. This research was sponsored by the Army Research Laboratory. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on. Contributing studies for this research were performed with the University of Warwick through the IAS fellowship. Special thanks to Professor Tim Watson, Professor Carsten Maple and Bil Hallaq at the University of Warwick, Dr. Andre Karamanian at Cisco Systems and Marc Kolenko at Price-Waterhouse Coopers for their assistance on the referenced studies and their assistance in attacker identification.

References

1. Hofstede, G., Hofstede, G.J., Minkov, M.: Cultures and Organizations. McGraw-Hill Publishing, New York (2010)
2. Henshel, D., Sample, C., Cains, M.G., Hoffman, B.: Integrating cultural factors into human factors framework for cyber attackers. In: 7th Annual Conference on Applied Human Factors and Ergonomics Conference, Orlando, FL (2016)
3. Sample, C., Cowley, J., Watson, T., Maple, C.: Re-thinking threat intelligence. In: International Conference on Cyber Conflict (CyCon US), pp. 1–9. IEEE (2016)
4. Sample, C., Cowley, J., Hutchinson, S.: Cultural exploration of attack vector preferences for self-identified attackers. In: IEEE 11th International Conference on Research Challenges in Information Science, Brighton, UK (2017, submitted)
5. Sample, C., Hutchinson, S., Karamanian, A., Maple, C.: Cultural observations on social engineering victims. In: 16th European Conference on Cyber Warfare and Security (ECCWS) Dublin, Ireland (2017, accepted)
6. Sample, C., Bakdash, J., Abdelnour-Nocera, J., Maple, C.: What's in a name? Cultural observations on nationally named hacking groups. In: 12th International Conference on Cyber Warfare and Security (ICCWS) Dayton, Ohio, pp. 332–340 (2017)
7. Sample, C., Karamanian, A.: Culture and cyber behaviours: DNS defending. In: 14th ECCWS, Hatfield, UK, pp. 233–240 (2015)
8. Nisbett, R.: The Geography of Thought: How Asians and Westerners Think Differently… and Why. Simon & Schuster, New York (2010)

9. Guess, C.D.: Decision-making in Individualistic and Collectivist Cultures, Readings in Psychology and Culture, 4. http://scholarworks.gvsu.edu/cgi/viewcontent.csg?article=1032&context=orpc
10. Guss, C.D., Dorner, D.: Cultural differences in dynamic decision-making strategies in a non-linear, time-delayed task. Cog. Sys. Res. **12**(3), 365–376 (2011)
11. Bishop, M., Butler, E., Butler, K., Gates, C., Greenspan, S.: Forgive and forget. In: 21st EICAR Annual Conference Proceedings, pp. 151–159 (2012)
12. Wang, Q.: Why we should all be cultural psychologists? Lessons learned from the study of social cognition. Perspect. Psychol. Sci. **11**(5), 583–596 (2016)
13. Rathje, S.: The definition of culture: an application-oriented overhaul. Interculture J. **8**, 35–59 (2009)
14. Birukou, A., Blanzieri, E., Giorgini, P., Giunchiglia, F.: A formal definition of culture, University of Trento, Italy (2009)
15. Faulkner, S.L., Baldwin, J.R., Lindsley, S.L., Hecht, M.L.: Layers of Meaning: An Analysis of Definitions of Culture. Redefining Culture: Perspectives Across the Disciplines, pp. 27–51. Lawrence Erlbaum Associates, Publishers, Mahwah, (2006)
16. Bhabha, H.K.: The world and the home. Cult. Polit. **11**, 445–455 (1997)
17. Roberts Jr., J.M., Moore, C.C., Romney, A.K., Barbujani, G., Bellwood, P., Dunnell, R.C., Green, R.C., Kirch, P.V., Marcus, J., Flannery, K.V., Terrell, J.: Predicting similarity in material culture among new guinea villages from propinquity and language: a log-linear approach [and comments and reply]. Curr. Anthropol. **36**(5), 769–788 (1995)
18. Arrindell, W.A., Hatzichristou, C., Wensink, J., Rosenberg, E., van Twillert, B., Sedema, J., Meijer, D.: Dimensions of national culture as predictors of cross-national differences in subjective well-being. Personality Individ. Differ. **23**(1), 37–53 (1997)
19. Doney, P.M., Cannon, J.P., Mullen, M.R.: Understanding the influence of national culture on the development of trust. Acad. Manag. Rev. **23**(3), 601–620 (1998)
20. Huff, L., Kelley, L.: Levels of organizational trust in individualist versus collectivist societies: a seven-nation study. Organ. Sci. **14**(1), 81–90 (2003)
21. Singelis, T.M., Brown, W.J.: Culture, self, and collectivist communication linking culture to individual behavior. Hum. Commun. Res. **21**(3), 354–389 (1995)
22. Vitell, S.J., Nwachukwu, S.L., Barnes, J.H.: The effects of culture on ethical decision-making: an application of hofstede's typology. J. Bus. Ethics **12**(10), 753–760 (1993)
23. Aguayo, D., Herman, K., Ojeda, L., Flores, L.Y.: Culture predicts mexican americans' college self-efficacy and college performance. J. Divers. High. Educ. **4**(2), 79 (2011)
24. Dawar, N., Parker, P.M., Price, L.J.: A cross-cultural study of interpersonal information exchange. J. Int. Bus. Stud. **27**(3), 497–516 (1996)
25. Al-Gahtani, S.S., Hubona, G.S., Wang, J.: Information technology (IT) in Saudi Arabia: culture and the acceptance and use of IT. Inf. Manage. **44**(8), 681–691 (2007)
26. Shane, S.: Uncertainty avoidance and the preference for innovation championing roles'. J. Int. Bus. Stud. **26**(1), 47–68 (1995)
27. Zaheer, S., Zaheer, A.: Country effects on information seeking in global electronic networks. J. Int. Bus. Stud. **28**(1), 77–100 (1997)
28. Kirkman, B.L., Lowe, K.B., Gibson, C.B.: A quarter century of culture's consequences: a review of empirical research incorporating hofstede's cultural values framework. J. Int. Bus. Stud. **37**(3), 285–320 (2006)
29. Gazzaniga, M., Ivry, R.B., Mangun, G.R.: Cognitive Neuroscience: The Biology of the Mind. W.W. Norton & Company Inc, London (2014)
30. Bargh, J., Morsella, E.: The unconscious mind. Perspect. Psychol. Sci. **3**(1), 73–79 (2008)

31. Buchtel, E.E., Norenzayan, A.: Thinking across cultures: implications for dual processes. J. St. BT Evans & K. Frankish, pp. 217–38 (2009)
32. Evans, J.S.B.T.: Dual-processing accounts of reasoning, judgment, and social cognition. Annu. Rev. Psychol. **59**, 255–278 (2008)
33. Minkov, M.: Cultural Differences in a Globalizing World. Emerald Group Publishing Limited, WA (2011)
34. Zone-H archives: http://www.Zone-h.org
35. ICANN: http://www.icann.org
36. Geert Hofstede: https://geert-hofstede.com/countries.html
37. Hofstede, G.: Cultures Consequences: International Differences in Work-Related Values (5). Sage, Thousand Oaks
38. Sample, C.: Applicability of cultural markers in computer network attack attribution. In: Proceedings of the 12th ECCWS, Finland, pp. 361–369 (2013)
39. Fiske, S.T., Taylor, S.E.: Social cognition: from brains to culture. Sage, Thousand Oaks (2013)
40. Van de Vijver, F., Leung, K.: Methods and Data Analysis for Cross-Cultural Research, vol. 1. Sage, Thousand Oaks (1997)
41. Martin, B., Brown, M., Paller, A., Kirby, D.D., Christey, S.: 2011 CWE/SANS Top 25 Most Dangerous Software Errors. Common Weakness Enumeration (2011)
42. MITRE website: Common Weaknesses Enumerated. http://cwe.mitre.org/top25/#listing
43. MITRE website: Common Cyber Observables. http://cybox.mitre.org
44. Su, Z., Wassermann, G.: The essence of command injection attacks in web applications. ACM SIGPLAN Not. **41**(1), 372–382 (2006)
45. Arends, R., Austein, R., Larson, M., Massey, D., Rose, S.:. DNS Security Introduction and Requirements, Internet Engineering Task Force (IETF) RFC 4033 (2005)
46. St. Johns, M.: Automated updates of DNS security (DNSSEC) trust anchors. In: IETF (2007)
47. Binford, L.R.: Archaeology as anthropology, pp. 217–225. American Antiquity (1962)
48. Watson, R.A.: Inference in Archaeology, pp. 58–66. American Antiquity (1976)
49. Sullivan, A.P.: Inference and evidence in archaeology: a discussion of the conceptual problems. Adv. Archaeol. Method Theory **1**, 183–222 (1978)
50. Sample, C.: Cyber + Culture Early Warning Study, CMU/SEI-2015–SR-025 (2015)
51. Landry, D.: 12 Defensible aesthetics. Graffiti and Street Art: Reading, Writing and Representing the City (2016)

Exploring 3D Cybersecurity Visualization with the Microsoft HoloLens

Steve Beitzel[1], Josiah Dykstra[2], Paul Toliver[1], and Jason Youzwak[1](✉)

[1] Vencore Labs, Basking Ridge, NJ, USA
{sbeitzel,ptoliver,jyouzwak}@vencorelabs.com
[2] Laboratory for Telecommunication Sciences, College Park, MD, USA
jdykstra@LTSnet.net

Abstract. We describe the novel use of the Microsoft HoloLens to assist human operators with computer network operations tasks. We created three applications to explore how the HoloLens may aid cybersecurity practitioners. First, we developed a 3D network visualizer that displays network topologies in varying levels of detail, ranging from a global perspective down to specific properties of individual nodes. The user navigates through the topology views using hand gestures while responding to simulated alarm conditions on specific nodes. Second, we developed an application that simulates a "capture the flag" exercise. Third, we developed an application to test network connectivity. We discuss the benefits, challenges, and lessons learned from developing mixed-reality applications for computer network operations. We also discuss ideas for further development in this area.

Keywords: Cyber security · Network security · Mixed Reality

1 Introduction

The goal of this work is to investigate the feasibility of using Mixed Reality (MR) devices to assist the day-to-day work of network operators. Network operators, who monitor and defend computer networks, are often required to perform several simultaneous tasks that require focused concentration, while also handling interruptions due to emergent high-priority tasks. This places high cognitive load on the operator. One of our primary research goals is to explore ways of incorporating MR devices into the network operations workflow to improve the user experience. In future research, we also plan to perform additional evaluation on how it impacts stress and cognitive load.

In previous research [1], we explored the use of Android-based Augmented Reality (AR) devices with basic capabilities and performed experiments designed to demonstrate the effect of AR devices on user cognitive load. These experiments showed that users expressed a decrease in their cognitive load when using an AR device with limited capabilities to monitor for emergent alerts. In this subsequent effort, we familiarized ourselves and experimented with the HoloLens [2], Microsoft's hardware and accompanying software for Mixed Reality. Compared to the previous AR devices, the HoloLens has more advanced features such as a stereoscopic 3D optical head-mounted

D. Nicholson (ed.), *Advances in Human Factors in Cybersecurity*,
Advances in Intelligent Systems and Computing 593, DOI 10.1007/978-3-319-60585-2_19

display, gaze tracking, spatial mapping, hand gesture navigation, advanced voice commands, and spatial sound. Since the release of the HoloLens, there has been increasing research into its use for a range of visualization applications, including molecular structures and architectural forms [3], augmented reality assisted surgery [4], and using biometric feedback to encourage focused concentration [5]. The work described in this paper is targeted at exploring the possibilities afforded by HoloLens capabilities within the context of computer network operations (CNO).

In this paper we describe the capabilities of the HoloLens and lessons learned in the development process we used to create applications for it. We built several prototype applications that explore CNO activities mapped into 3D environments, including tools for network visualization and network monitoring. We discuss the advantages and limitations of using the HoloLens, and offer ideas for future work.

2 Approach

2.1 Mixed-Reality Headset

The HoloLens is a mixed-reality head-mounted device developed by Microsoft, and marketed for a wide range of applications including gaming, design and engineering, education and training, and data visualization. In Fig. 1 we show the HoloLens device and internal components. The HoloLens contains an Intel 32-bit processor, a custom-built Microsoft Holographic Processing Unit (HPU 1.0), 2 GB RAM, 64 GB flash memory, and network connectivity via Wi-Fi 802.11ac [6]. Using projection-based smart-glasses that utilize optical waveguide technology, 2D and 3D images can be displayed on the HoloLens, overlaid on top of the user's field of view. Depth-sensing and 2D cameras enable spatial mapping and image sensing of the user's environment, allowing the HoloLens to track the user's gaze and place virtual 3D objects at known positions relative to real-world surfaces. A pair of speakers integrated into the headset enables binaural audio to simulate effects such as spatial sounds within the user's environment. Finally, an integrated noise-cancelling microphone enables control of applications via voice commands.

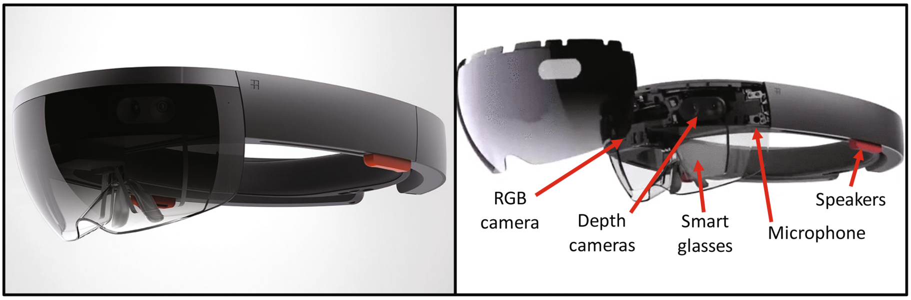


Fig. 1. HoloLens device and major components [2, 7]

2.2 Mixed-Reality Application Software

The HoloLens runs a 32-bit version of Windows Holographic 10 (currently version 14393.693 as of March, 2017) that supports Universal Windows Platform (UWP) apps. The HoloLens supports 2D apps, which are experienced as 2D projections within the user's field of view (e.g. web browser pane on wall), as well as full stereoscopic 3D apps, which fully immerse the user in a rich 3D experience. 3D apps are the focus of this work.

Holographic apps utilize Windows Holographic APIs, which provide a range of building blocks for interfacing with the HoloLens device itself including: (i) world coordinate system, (ii) tracking of user's gaze, (iii) gesture input, (iv) voice input, (v) spatial sound, and (vi) spatial mapping of the user's environment. These building blocks are completely integrated into Unity [8], which greatly simplifies 3D app development. In addition, Microsoft developed the Unity HoloToolkit [9], which provides additional components for developers including: (i) 3D cursors, (ii) display of spatial mapping, (iii) gesture-based object placement, (iv) object scaling and rotation, (v) linking of objects to particular spatial sounds, (vi) and spatial anchoring for coordinate system registration.

3 App Development Process

3.1 Development Workflow

In Fig. 2 below, we illustrate the workflow we used in developing several custom Holographic apps for the HoloLens. We utilized the Unity game engine as our core development platform. Unity manages integration and linking of the component assets used within an app project, including C# scripts, 3D objects, text data, images, audio, and others. In addition, Unity provides its own built-in components for 3D primitives

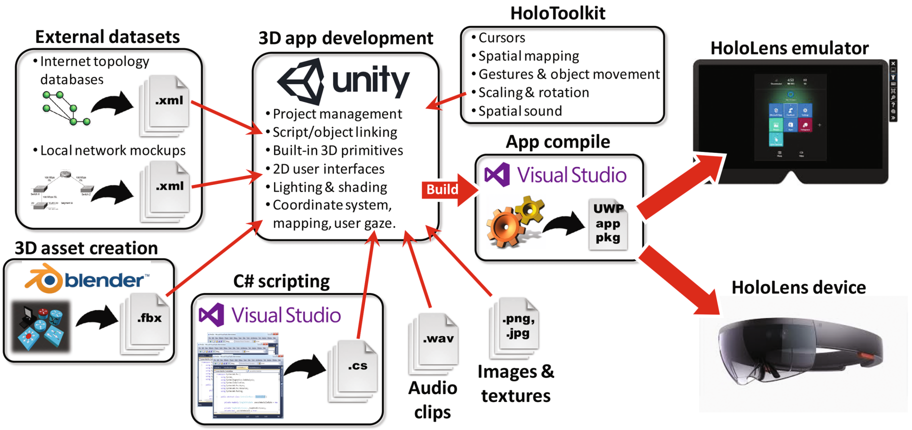


Fig. 2. Workflow we followed for developing HoloLens Apps.

(e.g. spheres, cubes, cylinders, 3D text) and libraries for building 2D user interfaces for use within the 3D world.

We also used other tools and data sources to create several external assets that we imported into our Unity projects. For example, databases describing the global and local network topologies that we display in our network visualizer 3D app are stored in XML files defined in a GraphML [10] format. We used Blender [11], an open-source 3D software tool, to create more complex 3D objects beyond simple spheres and cubes, such as iconography for network elements (routers, machines, etc.). We also developed several custom C# scripts using the Visual Studio editor [12] to interface with the HoloLens device, dynamically create 3D scenes, and enable remote network connectivity. Finally, we imported image files in several formats (e.g. PNG, JPG) for texturing raw 3D objects, and we imported audio clips to allow for specific spatial sound effects.

4 Results

We developed several applications to demonstrate alternate approaches to network security operations leveraging the HoloLens. In particular, we explored using the 3D stereoscopic display to provide a novel method of visualizing network status and the use of hand gestures for navigation. Over the next several sections, we describe applications demonstrating support for 3D network visualization, notional network security training exercises, monitoring of network captures, and testing network connectivity to the HoloLens.

4.1 3D Network Visualizer App

The 3D network visualizer prototype app displays network topologies in two levels of detail, first from a global perspective, showing all networked assets at a high-level; and second, the local area network topology surrounding a user-selected node from the global view. The app simulates intermittent random network "alerts", representing emergent problems that require attention, to draw the user's attention to specific nodes. The user can then gesture-select the node(s) in question to zoom into a more detailed view of the local topology.

The example global network topologies used by the app are based on GraphML databases available from the Internet Topology Zoo [13]. The app geographically displays these topologies on an Earth sphere in the global view. The app parses each GraphML database and dynamically constructs the corresponding 3D network graph within Unity, with spheres representing nodes and links interconnecting the nodes. In Fig. 3 below, we show an example that includes topologies based on the AT&T North America, British Telecommunications (BT) Latin America, BT Europe, and BT Asia-Pacific databases. Two different Earth images are used to illustrate daytime and nighttime texturing based on NASA images [14]. Using the HoloLens' gesture-based input capabilities, the user can set the globe in the network visualizer app to rotate automatically or to respond to real-time control by the user. The globe can also be positioned within the spatially-mapped environment and have its size scaled dynamically using similar gesture-based controls.

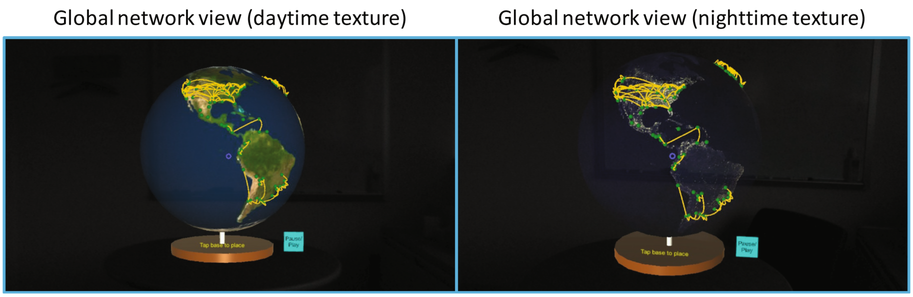

Global network view (daytime texture)
Global network view (nighttime texture)
Tap base to place
Pause/
Play
Tap base to place
Pause/
Play

Fig. 3. Network Visualizer App in global view

To simulate alerts, the app selects random nodes from the global network topology, with random intervals between each alert. The app signifies an alert by changing the affected node's color from green to red and displaying a flag with the node's name as seen in Fig. 4. When the user gazes to the flag and gesture-selects it, the view of the visualizer dynamically changes from a global view to a local network view associated with that particular node.

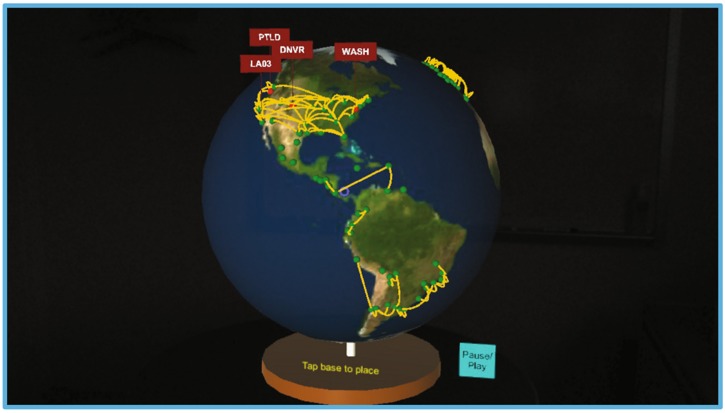


Fig. 4. 3D Network Visualizer App with simulated network alerts in global view

The local topologies displayed in the network visualizer are mockups of networks and associated network elements similar to what might be representative of a point-of-presence [15]. The app defines local networks using a format similar to the Internet Topology Zoo GraphML databases. We constructed 3D iconography for a variety of network component types, including routers, switches, and computers. Each local network topology is imported and dynamically constructed upon global node selection. In Fig. 5 below, we show three different examples of local topologies, each including various 3D network elements as well as their interconnecting links. In like fashion to the global topology, the user can rotate and scale through local network topologies through gesture-based user control.

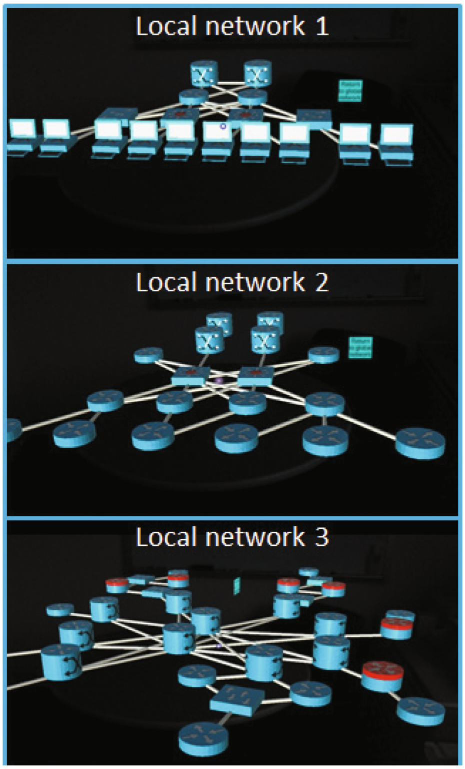


Fig. 5. 3D Network Visualizer App for different local network views.

The HoloLens, tracks the user's gaze and highlights individual links or network elements as the user views the local network topology. When the user manipulates an element with gesture-selection, the HoloLens displays mockups of different network diagnostic windows on top of individual 3D network objects. In Fig. 6, we show some of these, including a mockup command terminal window to illustrate potential HoloLens-based interaction when diagnosing problems on network switches and routers. We also created a

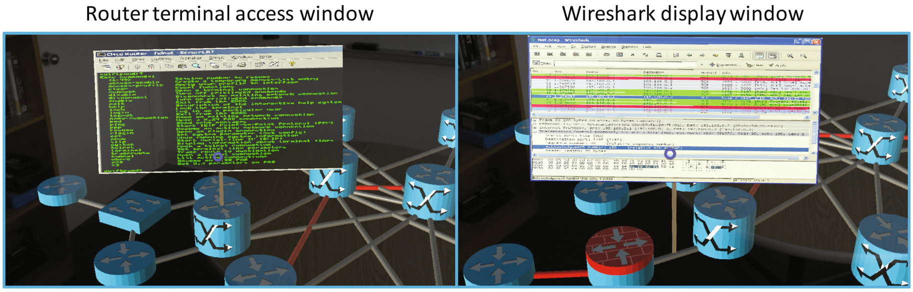


Fig. 6. 3D Network Visualizer App with mockups of gesture-selected display windows for network elements (terminal access window) and network links (Wireshark display)

mockup Wireshark display to illustrate envisioned user interactions when analyzing traffic on network links.

4.2 Capture the Flag via 3D Objects

We built a HoloLens "Capture the Flag" application to simulate a network security exercise. The application requires the user to locate a set of encrypted text files scattered randomly across a simulated file system. The contents of one file decrypt the contents of the next file in a sequence, which continues until all files have been decrypted. While performing the main task of locating and decrypting files, the user also needs to respond to simulated network alerts that occur at random intervals. This app embodies the test scenario we developed for our previous work in this area [1], adapted for use on the HoloLens.

The user navigates through items in a file browser window using HoloLens hand gestures. When the user locates a "flag" (denoted by the filename beginning with the string "flag"), the app creates a colored 3D lock representing that flag displays it in the user's world as shown in Fig. 7. The first flag, "flag 0", is initially unlocked, with the remaining flags locked. Using hand gestures, the user moves "flag 0" onto "flag 1", which unlocks "flag 1". When "flag 1" is unlocked, it yields a 3D key of matching color that unlocks "flag 2" and so on.

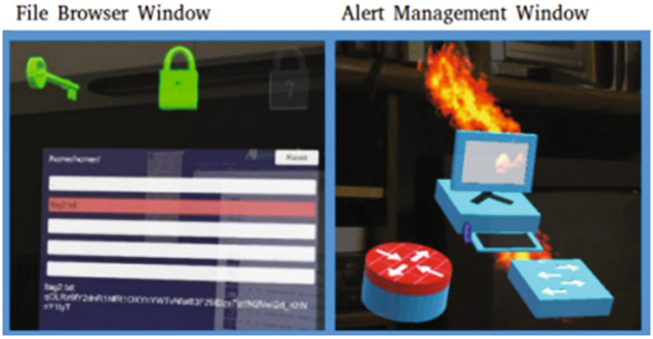


Fig. 7. Capture the Flag App displaying a file browser navigation window and Alert Management area

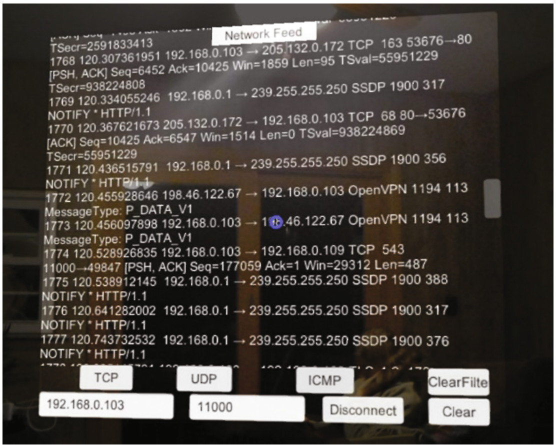
While performing the main task of searching for and unlocking colored flags, the user is interrupted periodically by visual and audible alerts. The user responds to alerts by shifting their gaze to the area of their world where the alerts are displayed. Once selected, the user dismisses the alert by using a hand gesture to apply a simulated countermeasure that clears the alert as shown in Fig. 7.

In previous research [1], user input came from a computer keyboard, a computer monitor was used as the main display, and alerts were displayed on Android-based AR glasses with limited capabilities. To respond to an alert, the user manually typed the IP address of the affected node into a control terminal. In the HoloLens app, the display is seamlessly overlaid on to the user's field of view and all input is supplied through hand gestures. The user interface was also designed to be more intuitive: when an alert occurs, a fire is visually and audibly indicated, and the user applies a countermeasure by tapping on the network object to ‘put out the fire’.

4.3 Network Feed App (from Remote System)

The Network Feed application allows the user to view a real-time network capture obtained from a remote system, conceptually similar to a typical Wireshark session. The HoloLens user can enter the network location of the remote system, connect, and view the text output in a scrolling window.

On the HoloLens, the user enters the network address and port of a host running the network capture service. Once connected, the network capture service sends a live stream to the HoloLens containing the network traffic captured on the remote system. The Network Feed App displays the capture results in real-time in a scrolling window on the HoloLens displays, as shown in Fig. 8. To narrow the focus of the capture, the user can apply some basic filters (UDP, TCP, and ICMP).



```
TSecr=2591833413
1768 120.307361951 192.168.0.103 → 205.132.0.172 TCP  163 53676→80
[PSH, ACK] Seq=6452 Ack=10425 Win=1859 Len=95 TSval=55951229
TSecr=938224808
1769 120.334055246  192.168.0.1 → 239.255.255.250 SSDP 1900 317
NOTIFY * HTTP/1.1
1770 120.367621673 205.132.0.172 → 192.168.0.103 TCP  68 80→53676
[ACK] Seq=10425 Ack=6547 Win=1514 Len=0 TSval=938224869
TSecr=55951229
1771 120.436515791  192.168.0.1 → 239.255.255.250 SSDP 1900 356
NOTIFY * HTTP/1.1
1772 120.455928646 198.46.122.67 → 192.168.0.103 OpenVPN 1194 113
MessageType: P_DATA_V1
1773 120.456097898 192.168.0.103 → 198.46.122.67 OpenVPN 1194 113
MessageType: P_DATA_V1
1774 120.528926835 192.168.0.103 → 192.168.0.109 TCP  543
11000→49847 [PSH, ACK] Seq=177059 Ack=1 Win=29312 Len=487
1775 120.538912145  192.168.0.1 → 239.255.255.250 SSDP 1900 388
NOTIFY * HTTP/1.1
1776 120.641282002  192.168.0.1 → 239.255.255.250 SSDP 1900 317
NOTIFY * HTTP/1.1
1777 120.743732532  192.168.0.1 → 239.255.255.250 SSDP 1900 376
NOTIFY * HTTP/1.1
```

192.168.0.103 11000

Fig. 8. Network monitoring window running on HoloLens

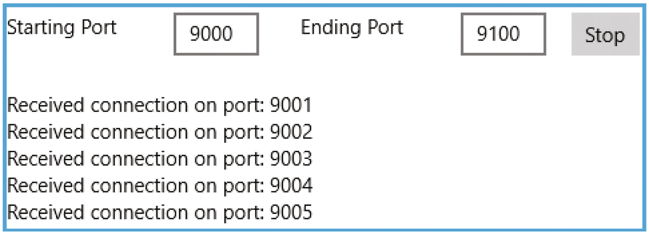
The network capture service on the remote host runs the tshark [16] command in the background and publishes the output to HoloLens clients connected via a network connection. HoloLens client apps also use this communication channel to command and control the program to apply various network filters.

4.4 2D Network Connectivity Tester App

Universal Windows Platform (UWP) applications can run on PC-based versions of Windows 10 and the HoloLens (as well as several other platforms). As mentioned previously, developers build UWP applications in Visual Studio without the need for Unity or DirectX libraries. UWP applications execute in the “main world” of the

HoloLens, and a user can have several UWP apps active at a time. We chose to develop a simple UWP app to explore the user experience of 2D UWP apps running alongside other 2D UWP applications.

The Network Connectivity Tester is a simple application that listens on a range of TCP ports and displays a message when an incoming connection is received, as shown in Fig. 9. This can be used for debugging a network connection, and to determine if there are any port restrictions on the HoloLens.



Starting Port 9000 Ending Port 9100

```
Received connection on port: 9001
Received connection on port: 9002
Received connection on port: 9003
Received connection on port: 9004
Received connection on port: 9005
```

Fig. 9. Network Connectivity Tester

5 Discussion

From a human factors and ergonomics standpoint, the HoloLens headset represents a significant advance over previous augmented reality products. However, the current version does have some limitations that may impact a user's experience, depending upon the application and personal user preferences. In this section, we summarize our opinions of the key capabilities and limitations we observed based on our experiences with the HoloLens to date.

Key capabilities:

- Full-featured MR device with no physical tethering constraints.
- Smart-glasses provide good image displays for 2D and 3D stereoscopic apps.
- Multiple capabilities integrated in headset including spatial mapping, spatial sound, gesture & voice control, and wireless communications (both Wi-Fi & Bluetooth).
- High quality spatial mapping of the environment and stabilization of 3D objects enabled by depth cameras and Microsoft's custom Holographic Processing Unit (HPU).
- Extensive development support, including development tools, library toolkits, examples, tutorials, and active developer forums.
- Powerful testing and debugging tools, including the Emulator and Device Portal.
- Unity provides a framework that allows 3D stereographic apps to be created relatively quickly (compared to developing graphics routines directly in DirectX).

Potential device limitations:

- Relatively large, heavy device, particularly on the front side.
- Relatively narrow field of view (~33°−45°).
- Requires calibration of inter-pupil distance for best display.

- Constrained range for 3D objects and spatial mapping (closest: ~1 m, furthest: ~5 m).
- Gesture-based interface can be slow and limiting in some cases where a physical keyboard would be more efficient.
- No official support for creating custom gestures, although a third-party toolkit does exist [17].

6 Future Work

As part of our future work with the HoloLens, we plan to extend the 3D Network Visualizer app described in Sect. 4.1 beyond its present mockup form. Specific areas for improvements include incorporating more detailed local topology diagrams based on realistic networks. We also plan to extend the diagnostic pop-up windows (currently displayed as static image mockups) to allow for interactive display of realistic network data, such as traffic link analysis. In addition to network traffic, we plan to consider other types of network data, such as reported events and alarm conditions.

We will also be exploring how to extend network data visualization beyond the current 2D diagnostic displays to an immersive 3D world. Options here include dynamically updating the 3D network diagrams based on changing network conditions, such as link traffic volume. This could be represented, for example, by changing physical link size or including additional 3D features. Another option is to utilize other visualization techniques beyond topology diagrams, such as connectivity ring graphs or connection flow diagrams that have previously been explored for 2D network dashboards [18], and extending them to the immersive 3D environment. Finally, we plan to explore how the HoloLens' capability for sharing a common spatial reference frame between multiple users could be applied in the context of collaborative network operations tasks.

In addition to exploring technical improvements, we also plan to perform an evaluation on how the use of the HoloLens impacts stress and cognitive load in performing network operations tasks compared to working in an environment without MR devices.

7 Conclusions

We have presented new concepts for assisting computer network operations tasks through the use of the HoloLens mixed-reality device. We described a range of 3D applications built for testing different user scenarios and summarized the software development process involved in prototyping such apps. In contrast to the 2D limits of traditional computer displays, apps running on the HoloLens device enable both physical and virtual objects to co-exist within the real world environment while allowing the user to move around and interact with the mixed-reality 3D scene. Our prototype apps demonstrated how the HoloLens can augment the network operator's experience by visualizing network topology and status conditions in a 3D space, for example. Plans for future work are focused on extending these apps towards increasingly realistic network scenarios and sources of data, such as network traffic flows.

Acknowledgements. This material is based upon work supported by the U.S. Government under contract HR98230-13-D-0055. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Government.

References

1. Beitzel, S., Dykstra, J., Huver, S., Kaplan, M., Loushine, M.: Youzwak, J: Cognitive performance impact of augmented reality for network operations tasks. In: Nicholson, D. (ed.) Advances in Intelligent Systems, pp. 139–152. Springer, Cham (2016)
2. Microsoft HoloLens. https://www.microsoft.com/microsoft-hololens/en-us
3. Hockett, P., Ingleby, T.: Augmented reality with HoloLens: experiential architectures embedded in the real world. arXiv preprint, arXiv:1610.04281 (2016)
4. Cui, N., Kharel, P., Gruev, V.: Augmented reality with Microsoft HoloLens holograms for near infrared fluorescence based image guided surgery. In: SPIE, vol. 10049. SPIE Publications (2017)
5. Ideals – Focus. https://devpost.com/software/ideals-67c1kl
6. Microsoft HoloLens Hardware Details. http://www.microsoft.com/microsoft-hololens/en-us/buy
7. HoloLens Development - Tony Labs. http://www.tonylabs.com/hololens-development/
8. Unity - Microsoft Windows – HoloLens. https://unity3d.com/partners/microsoft/hololens
9. HoloToolkit. https://github.com/Microsoft/HoloToolkit-Unity
10. The GraphML File Format. http://graphml.graphdrawing.org/
11. Blender. https://www.blender.org/
12. Visual Studio. https://www.visualstudio.com/
13. The Internet Topology Zoo. http://topology-zoo.org/
14. NASA Visible Earth. http://visibleearth.nasa.gov/
15. Service Provider IP System Test. http://www.cisco.com/systemtest/spip2b/index.htm
16. Tshark - The Wireshark Network Analyzer. https://www.wireshark.org/docs/man-pages/tshark.html
17. Gesture Recognition Toolkit. https://github.com/nickgillian/grt
18. Chen, S., Merkle, F., Schaefer, H., Guo, C., Ai, H., Yuan, X., Ertl, T.: VAST 2013 mini challenge 3: AnNetTe collaboration oriented visualization of network data. In: IEEE VIS 2013 (2013)

Cybersecurity Training and Education

Humans as the Strong Link in Securing the Total Learning Architecture

Fernando Maymí(✉), Angela Woods, and Jeremiah Folsom-Kovarik

Soar Technology, Ann Arbor, MI, USA
{fernando.maymi,angela.woods,jeremiah}@soartech.com

Abstract. This paper describes a proposed approach, centered on human factors, for securing the Total Learning Architecture (TLA). The TLA, which is being developed for the United States Department of Defense, will rely on large stores of personal data that could be targeted by sophisticated adversaries. We describe the TLA and its envisioned users at a fairly high level before describing expected classes of attacks against it. We then examine existing and proposed controls that, if properly managed, should allow users and service providers to significantly reduce the risks to the system.

Keywords: Total learning architecture · Cybersecurity · Threat modeling · Human-Systems integration

1 Introduction

The last twenty years have brought significant advances in educational technology, which is now a core element of life-long learning for many children and adults in the western hemisphere. Though learning management systems (LMS) and education management information systems (EMIS) can be credited for many breakthroughs, they are by no means the only sources of learning activities, particularly those related to career advancement. Increasingly, people are turning to a variety of online resources in order to learn new skills, improve or maintain existing ones, or otherwise further their education.

This growing demand for educational technology solutions is driving many organizations to supply a variety of learning activities, most of which exist independently or within proprietary ecosystems. Learners in this environment are forced to maintain multiple online personas and track their progress on site-by-site basis. As the number of learning activity providers that longitudinally track learner skills continues to grow, there is an opportunity to enhance competency mastery by aggregating these information sources and providing tailored recommendations to individual learners. This is the promise of the Total Learning Architecture (TLA).

D. Nicholson (ed.), *Advances in Human Factors in Cybersecurity*,
Advances in Intelligent Systems and Computing 593, DOI 10.1007/978-3-319-60585-2_20

2 Architecture Overview

The TLA is not a software system; rather, it is a set of Application Programming Interface (API) specifications that create a learning framework wherein learning activity providers and others can responsibly share learning data. The learners, providers and other relevant organizations create an ecosystem within which learners are able to avail themselves of new learning opportunities while leveraging all their historical data. For example, if a learner is subscribed to activities at sites X and Y and then chooses to also participate in site Z, the TLA would allow competencies from all three sites to be considered when making recommendations for new activities in all sites, including the newly added site Z.

To illustrate the use of these TLA APIs, we have developed reference implementations of certain components (e.g. the Learning Record Store or LRS). However, the reference implementations themselves are not part of the formal definition of the TLA. Any developer is free to create their own replacement implementation of any component for which a reference implementation is provided, all that is necessary is components that conform to the API specifications that govern that area of the TLA.

2.1 Interfaces

The interfaces are what define the TLA; without them, the ecosystem could not exist. These APIs perform two key functions: they define *what* is shared, and they specify *how* it is shared. The first of these functions is accomplished by enforcing consistent data structures. This creates a shared language that allows information to be unambiguously interpreted by different entities in the TLA. The second function, which deals with how this information is shared, is made possible by standardizing the transfer methods used when exchanging information about learning experiences between entities that comply with the architecture.

The TLA comprises multiple optional APIs that regulate everything from learning activities to the use of different assessment frameworks to learner profiles, just to name a few. Among these, the most developed is the Experience API (xAPI), which is the principal means by which any component can understand exactly what the learner has done in the past or is currently doing. The key object within xAPI is the Statement, which describes fine-grained communication about learner experiences from minute to minute in order to help interpret learner performance in context.

2.2 Data

The learner record store (LRS) is where all the learner experiences are stored. This, together with the activity providers, are the most important components of the TLA; without them, the architecture would not accomplish much. An LRS can be centralized, or it can be distributed. In fact, an activity provider can provide its own LRS, which could then interface with other stores to provide holistic tracking of competencies as well as a richer set of recommended activities for a given learner. The exchange of this data can be regulated by the user or sponsoring organization.

The data in a single xAPI Statement object, which captures a learning experience, can be thought of as a sentence with a subject, verb, object and, optionally, other components such as the outcome. For example, if an activity provider would like to store an experience in the LRS, the statement would add to a historical stream describing what an individual learner has encountered, accomplished, and done in context across all activity providers.

For concreteness, an example xAPI statement might add to Angela's record in the LRS to reflect that she scored a 97% on a graded task to configure a pfSense firewall. The activity sending the statement is recorded along with various metadata. Similarly, there can be a variety of assessments that could be mapped to the result. So, while the format of the Statement is fixed, it can be arbitrarily enriched by activity providers and others for the purpose of interpreting and understanding the learner's experience.

3 Use Cases

In order to better understand the uses and potential vulnerabilities of the TLA, it is helpful to describe some relevant use cases. The sections below provide a limited, but representative sample of cases in which this architecture would be used by learners both in their personal or work spaces.

3.1 Support a Job Task

Since learning On-The-Job (OTJ) is an important aspect of much adult training, the TLA could be used to support Just-In-Time (JIT) delivery of task support when a person needs help with a job task. The delivery device could be an embedded system or a personal mobile device, and the trigger for the TLA to offer help might be performance monitoring of job tasks through the same channels TLA uses to monitor instructional performance. Job task support is valuable for a range of populations including apprentices and novices, people responding to an infrequently occurring emergency, or people who need cognitive support for specific deficits.

3.2 Learn

We expect users of the TLA to spend most of their time learning in a structured, guided environment. The interaction of learners with these environments, which are developed and delivered by a variety of Learning Activity Providers, represents the most likely and frequent use case for the TLA. While these are perhaps best visualized by evolutions of the computer based training with which many of us are familiar, they will also involve novel modalities such as presenting flashcards on a smart watch right before the learner delivers a presentation.

3.3 Monitor Human Performance

A growing number of individuals are interested in monitoring and measuring information about their own daily lives for reasons of health, self-improvement, or simply personal interest [1]. In parallel, Defense researchers are carefully studying human performance and precursors or mediators that contribute to performance [2]. The TLA could help individuals learn about themselves by facilitating the empirical measurement and manipulation of individual experience. For example, a person who wants to compare their personal caffeine intake against their sleep patterns can record both in their own TLA Learner Record Store (LRS). By using the TLA, the person can gradually expand the data they collect as needed to learn more about themselves.

3.4 Integrate with Personal Assistant

At the time of publication, commercial assistants are available in most of the consumer computer and mobile platforms, including Siri, Google Assistant, and Cortana. Information in the TLA about learners could help tailor each of these assistants to individual needs. For example, when the learner engages in informal learning by searching through a commercial assistant, the TLA can help identify the appropriate reading level and the background knowledge the learner has. Of course, the TLA would also benefit from any information the assistants share about the learner's current context and life experience.

3.5 Track Progress

Some learning activities will evaluate the progress of learners automatically, while other activities will rely to some extent on inputs from other people. Instructors, supervisors, and perhaps others could interact with TLA components to directly assess or provide input into the assessment of learners. An instructor could manually grade exercises, whereas a supervisor could validate that a learner was (or was not) able to show evidence of a proficiency. In these cases, select individuals will have access to relevant components in order to track the progress of learners in their purview.

4 Threat Model

A threat model identifies threat sources and methods that can undermine the functionality of a system and result in losses to an organization. Our approach is to identify the classes of threat actors that would have the intent and capability to attack the TLA, and then infer the means by which they would accomplish their goals. What follows is the first threat model developed for the TLA.

4.1 Threat Actors

We focus our discussion of threat sources on four classes of actors: terrorists, nation states, insiders and criminals. These classes emerged from the misuse case analysis that is described in the following section, but we describe them here in order to facilitate our later description of their desired actions. We note that there are numerous other potential classes of threat actors who could attempt to compromise a TLA system; we simply focus on the ones that appear likeliest to threaten the target systems.

Terrorist. Terrorist threat actors could attempt to compromise a TLA system if they think that doing so would allow them to cause death or destruction. A potentially exploitable area are industrial processes that are increasingly automated and digitally connected, and present opportunities to remotely cause physical effects. An example would be a food processing plant in which a computer controls the amount of iron that is added to a popular breakfast cereal. If threat actors were to target industrial systems operators responsible for regulating the iron levels, they could cause iron poisoning on a national or perhaps international scale. The concomitant loss of public trust in the food supply would further magnify the effects of such an attack.

State Actor. We assess that state actors represent the greatest threat to TLA systems. They could be interested in using TLA systems to cause physical destruction (with or without loss of life). We have seen at least one example of this in the 2013 breach of computer systems at the Bowman Avenue Dam in New York. The U.S. government indicted seven individuals who allegedly targeted the dam on behalf of the Iranian government [6]. While they were not able to cause damage, the event is an indicator of increasing proficiency and desire to damage cyber physical systems (CPS).

State actors could also want to alter the data within the TLA in support of information operations (IO), which involve deliberate attempts to influence what a population believes on specific issues. The 2016 compromise of George Soros' organizations [7], attributed to Russia, is a good example of a state actor altering stolen information in support of IO. In that operation, the actors modified some of the files to give the impression that his foundation was funding Russian dissidents [8]. As part of IO, one could imagine state actors implanting false information to influence how users perceive an issue of interest to the actors or for other purposes.

Criminal. Unlike the state actor, criminals are motivated by financial profits. Typically, monetization is accomplished by stealing large volumes of personal data and then selling them on online markets [9]. The most valuable targets for these actors are repositories of personal [10] or financial information [11], credentials (e.g., passwords) [12] and valid email addresses [13]. Depending on the specific system involved, TLA components will almost certainly contain at least one and perhaps all these types of valuable information. It would be reasonable to expect that these systems would almost certainly be targeted by criminal threat actors.

Insider. Insider actors also want to access data, albeit for a different purpose than the other threat actors. The insiders would most likely be interested in reading other users learning records, either out of misguided curiosity [14] or as a form of cyber stalking [15]. In fact, the news media has reported on many cases of employees with access to federated information systems similar to the TLA who have been disciplined or fired for improperly accessing the records of others.

A less likely but more damaging goal for an insider actor would be the unauthorized modification of learning records. There have been cases in which public officials have been accused or convicted of falsifying such information. The alleged motives range from financial gain [16] to avoiding public relations disasters [17]. As TLA systems become increasingly common, they could present opportunities for insider actors to modify the information contained in them.

4.2 Misuse Cases

A use case is a short story that describes the interaction of one user with the system in order to accomplish a specific task. Collectively, the collection of all use cases describes the entire functionality of the new system. Security professionals have adopted this modeling technique to include not only what authorized system users will do, but what threat actors will want to do also. We employ a common approach to threat modeling called misuse cases [18]. Figure 1 illustrates a partial use case model for the TLA that has been augmented to also show the misuse cases. By convention, the authorized actors and use cases are depicted in white, while the threat actors and misuse cases are shaded.

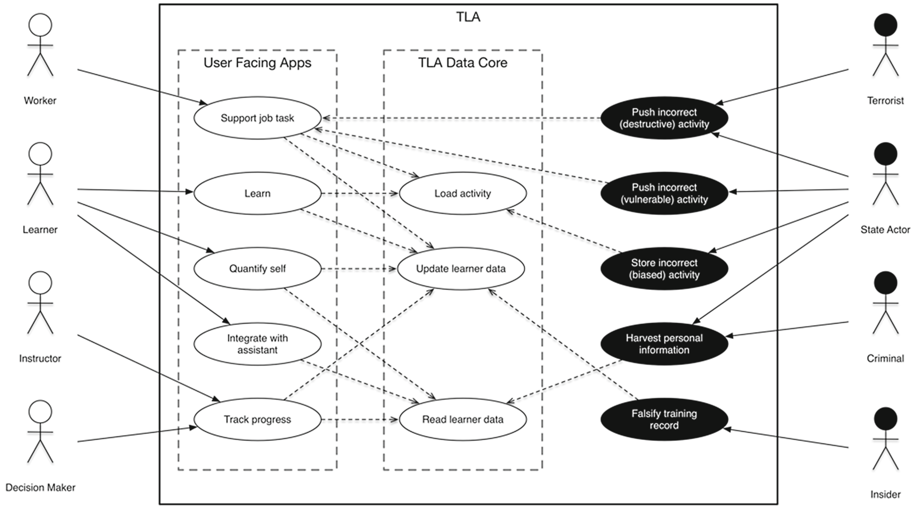


Fig. 1. TLA misuse case diagram showing the benign uses and users of the system filled in white (on the left) and the malicious uses and users of the system shaded black (on the right).

In the diagram we see four threat actors we discussed in the previous section. Their misuse cases encompass their main goals. In the following subsections, we look at how, specifically, these actors could leverage the TLA to accomplish their objectives.

Push Incorrect (Destructive) Activity. One of the features supported by the TLA is providing just-in-time (JIT) training to learners. This is useful when users need to perform a task at which they are not proficient. TLA activity providers would then provide tutorials, step-by-step guides or similar activities to guide the user in performing the task. If the activity had been altered with malicious content, it could be used to direct the user to perform actions that would result in physical destruction or loss of life. An example of this would be an operator of a dam performing an infrequent remote test of the floodgate actuators. If a threat actor modifies the procedure in the JIT activity so that instead of testing it actually causes the floodgates to open, the users actions could result in flooding. Since the learners are unskilled at the task, they would be particularly vulnerable.

In order to modify the activities to contain destructive instructions, the terrorist or state actor would have to gain access to the repository of activities and modify the data without being detected. Stealing credentials would be a feasible way to accomplish this, provided the stolen identity has authority to edit the materials. This would not be so much a TLA attack as one against a specific TLA compliant system. It would also be a broad attack, since everyone who accesses that activity (including knowledgeable authors, administrators or auditors) could notice the modification.

Push Incorrect (Vulnerable) Activity. The illicit modification of JIT activities described above is not limited to destructive purposes. A nation state actor could leverage this feature to induce learners to misconfigure their information systems in order to make it easier for the threat actor to compromise them. An example of this would be a system administrator with limited proficiency in configuring rules for an intrusion detection system (IDS). That administrator could turn to the TLA for JIT training when updating malware signatures. Since an IDS can have cryptic rules, it would be unlikely that the administrator would notice that the rule being typed, instead of generating alerts, would cause the IDS to suppress them. This very simple modification of probably a line or two would make it inordinately easier for the threat actor to successfully attack the target system with very little risk of detection.

Push Incorrect (Biased) Activity. Both of the previous misuses of the TLA provide the learner with intentionally incorrect information. For reasons discussed above, it is preferable for the threat actor to not store the incorrect activity in a legitimate Activity Provider's data stores. There is, however, at least one scenario in which it would make sense for a nation-state actor to store incorrect information so that large numbers of learners have access to it. This scenario involves information operations, which are deliberate activities carried out in order to influence the thinking of target group. A benign example of this is modern marketing practices designed to persuade you to purchase a particular good or service. Less benign examples are misinformation campaigns carried out by oppressive regimes in order to pacify their citizens. Increasingly, however, we are seeing information operations carried out in large scale by nation state actors against citizens of other nations.

The effectiveness of incorrect and biased information is proportional to its reach and volume. This is unlike the previous examples of destructive and vulnerable activities. For this reason, nation state actors would want to store, as opposed to surgically push, this information with multiple Activity Providers and, specifically, for popular activities. These actors can accomplish this objective through a variety of means including stealing credentials for content editors, recruiting legitimate content editors to alter the information, and contaminating sources used by content editors so they contain the incorrect biased information. Note that not all these means are technical, but they all exploit vulnerabilities in the human component of the TLA.

Harvest Personal Information. Exploiting people is often facilitated by gaining access to volumes of personal information. Whether the goal is to recruit foreign agents [19] or to sell personal information online [9], nation state and criminal actors can put a lot of effort into harvesting as much information as possible about their human targets. By providing a means to aggregate a very large set of data on learners (many of them associated with the government) the TLA could provide a lucrative target to these actors.

The likely target within the TLA in this misuse case is the Learner Profile. Whether this data set is contained in one database (as in the prototype implementation) or in a federation of data stores, it is the heart of the TLA and, as such, must be accessible to most other components. This degree of connectivity could present a significant vulnerability if left unattended, so we will provide some recommended controls later in this paper. For now, it is important to keep in mind that the Learner Profile will require a more comprehensive set of controls than other entry vectors discussed in this section.

Falsify Training Record. We conclude our discussion of misuse cases with what is perhaps the least damaging of all: the falsification of learning records. Apart from causing perception and trust challenges, this case is fairly contained both in terms of actions and effects. The likeliest actors to engage in falsification are insiders seeking to modify their own or someone else's records to show proficiencies that are not real. The self-serving version of this act is easier to mitigate by controlling the ability of learners to modify their own records. The technical controls to accomplish this are already part of the TLA.

The challenge is in detecting when an insider inappropriately modifies someone else's records. At issue is the ability of supervisors, trainers and others to certify proficiencies of those under their watch. Technical controls alone are unlikely to prevent this type of misuse because it would require the TLA to differentiate between a legitimate and inappropriate certification by an otherwise authorized user. We will focus on this challenge in our later discussion of procedural controls.

5 Technical Controls

In this section, we address technical controls that can mitigate the risk of compromise in the five misuse cases we have developed. We stress, however, that each technical control's effectiveness can be either undermined or enhanced by appropriate user behaviors. In the discussion that follows, we consider protections to data while it is at rest in some component of the TLA as well as when it is in transit between components.

5.1 Data in Transit

In the first two misuse cases we presented (pushing destructive and vulnerable activities), terrorists or nation states replace or modify a legitimate stream of activity data intended for a specific user. Their goal is to destroy or otherwise exploit a system that the learner is attempting to configure using the JIT training functionality of the TLA. This attack is unlikely to be attempted by modifying the activity data in the providers' data stores (i.e., data at rest). Instead, the attacker would either intercept the learner request and directly provide the malicious activity, or selectively modify portions of the activity as they flow between a legitimate activity provider and the learner.

In the first case the threat actor prevents the learner from connecting to the activity provider and impersonates the latter. Once the learner is connected to the impostor provider, the threat actor is provides tailored content for that learner that results in the destruction of assets or in rendering them vulnerable to follow-on attacks. This case requires a significant amount of preparation since the threat actor must recreate the entire learning environment that the learner expects.

An alternative approach, which is illustrated in Fig. 2, is one in which the threat actors simply inject themselves between the learner and the activity provider. From this position, they can selectively intercept, edit and then forward any part of the activity. The advantage is that the threat actors would not need to replicate the learning environment, but simply ensure all traffic flows through them. When the right content goes across the connection, say the instruction to “turn the knob slightly to the left,” the actors could replace it with “turn the knob fully to the right. Ignore any alarms.”

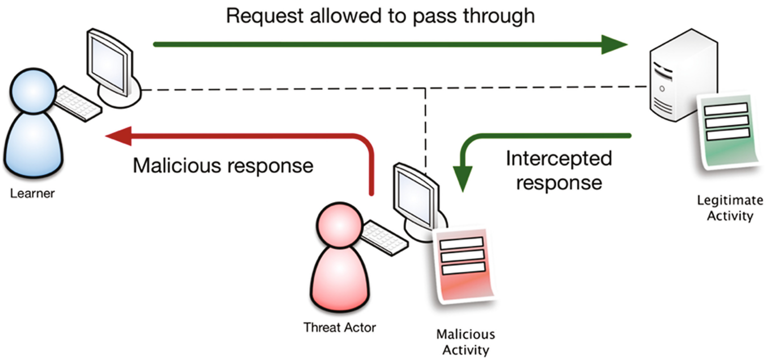


Fig. 2. Threat actor allowing a request from the user to reach the server, but intercepting and modifying the server's response intended for the user.

This type of attack is commonly called a Man-in-the-Middle (MitM), and requires a fair amount of sophistication to succeed. The key to mitigating a MitM attack is to ensure there is a secure link between the learner and provider. An example of this is the establishment of a secure hypertext transfer protocol (HTTPS) session in a client's web browser connecting to a financial institution's server. In order to set it up, the server must present evidence of its identity to the client. This evidence is almost always in the form of a public key infrastructure (PKI) certificate, which is tied to a specific internet domain (e.g., soartech.com). PKI certificates are signed or validated by a trusted third-party certificate authority (CA). The process by which this is done ensures that it is difficult for a threat actor to impersonate a legitimate site over HTTPS. Still, there are ways in which a threat actor might counter this technical control, which makes the user our next line of defense.

Using a Fake PKI Certificate. The PKI certificate is tied to a specific domain, but an actor could use a mismatched certificate (i.e., one whose domain doesn't match the uniform resource locator or URL). The learner would request a connection with the activity provider to which the threat actor would respond with its own certificate. This would cause the learner's browser to show an alert. Unfortunately, many naïve users will simply click "OK" on the warning and proceed with the connection to the bogus site. This risk is best mitigated through security awareness training in which the users are taught to recognize these certificate warnings as a serious threat. Furthermore, we should provide a simple technical means of notifying the appropriate security personnel if any links exhibit this behavior. Lastly, users who accurately report insecure conditions like this one should be recognized or otherwise rewarded for doing so.

Stripping the Secure Connection. The learner may request to connect to an HTTPS server, but it is possible for a threat actor to respond to that request in such a way that the browser accepts an unsecure connection instead. This is known as stripping the connection and is remarkably easy to do. An alert user would notice that the connection is not secure, since the browser would not display the secure icon on or near the address bar. Again, many users would not notice one way or another, but this can be remedied with security awareness, appropriate notification mechanisms, and a system of incentives for reporting anomalies such as this one.

5.2 Data at Rest

The data stored within the TLA could be a target to nation states that exploit information operations (IO). The main purpose of IO is to influence, disrupt, corrupt or usurp adversarial human decision making [20]. The means of carrying out an information operations attack would differ from the preceding discussion on pushing destructive or vulnerable activities. In the prior case, the attack was targeted, while in the case of IO the desire would be to maximize the number of affected individuals. The misinformation would have to be in the activity providers' stores.

A way to accomplish this would be to modify the information in the activity data stores to suit the threat actors' needs. While keeping sophisticated nation state actors from gaining access to a computer network is beyond the means of most organizations, detecting them or their actions is a much more reasonable expectation. Altering large amounts of information would doubtless require a prolonged interactive operation. Implementing best practices for data protection, including extensive logs of information access and modification, would significantly reduce the risk of these activities remaining undetected by content authors and system administrators. In this case, the users who would serve as the strong link would be authors and administrators.

Another way of inserting misinformation would be to target the authors directly or indirectly. A direct means would be to have persons friendly to the threat actor secure employment as content developers. They would then insert the desired content in a way that would be almost impossible to detect through technical means. Alternatively, the threat actor could persuade or coerce legitimate content developers. This would likely not be detected using technical controls either, but alert colleagues could provide early warning. Many government organizations have counter-intelligence programs that aim to identify insider threats. The adoption of the TLA would reinforce the need for these programs in both government and private sector organizations.

Finally, as our misuse cases show, threat actors would be interested in reading TLA information about learners and activities. Whether the actor is a nation state trying to surveil an individual or organization or a criminal trying to sell user data, the personal information within the component systems of the TLA represents a lucrative target to multiple threat actors. This is one class of threats in which we cannot rely on users or content authors for enhanced protection. Already the TLA community is rallying around robust protocols for protecting the confidentiality of learner information within its systems, which will certainly help, but we will rely almost exclusively on systems administrators and security personnel to protect this data at rest.

6 Other Considerations

Apart from the technical and procedural controls we have described for protecting the TLA, there are other considerations that can help protect this ecosystem if properly addressed. Reinforcing the right behaviors among both users and providers can further enable a safe and secure environment that is critical to realizing the promise of the TLA. As the amount of personal data that is stored in networked nodes increases, so must the awareness of the individuals described by that data. Security awareness training programs are intended to help users be aware of threats and how to mitigate them. The TLA has the potential to store very intimate data about its users, which furthers the case for effective security awareness for everyone.

Reducing the amount of personal user data stored in the system naturally creates a tension between functionality and privacy. At this stage in the development of the TLA, the set of requirements that would support a deliberate tradeoff analysis are not specific enough. If the community were to allow these functional requirements to emerge and morph in a naturalistic way, their impacts on privacy would be much more difficult to ascertain. Instead, we propose a deliberate dialog about the tradeoffs that should be considered. This conversation, which has already started informally within the TLA community, should continue as the architecture matures.

7 Conclusion

In this paper, we have presented a detailed threat model for the TLA, together with reasonable controls that could mitigate the risks posed by these threats. While technical controls are always needed in an information system, we have presented procedural counter-measures that can further improve the security of the ecosystem. However, the final and critical layer of protection consists of engaged, aware, and alert users who understand their stake in the process and take appropriate steps to enhance the effectiveness of the security controls that have already been or soon will be built into the TLA.

Acknowledgements. This material is based upon work supported by the Advanced Distributed Learning (ADL) Initiative under Contract No. W911QY-16-C-0019. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Advanced Distributed Learning (ADL) Initiative.

References

1. Swan, M.: Emerging patient-driven health care models. Int. J. Environ. Res. Public health **6**(2), 492–525 (2009)
2. Blackhurst, J.L., Gresham, J.S., Stone, M.O.: The quantified warrior: How DoD should lead human performance augmentation. In: Armed Forces Journal (2012)
3. Kenney, M.: Cyber-terrorism in a post-stuxnet world. Orbis **59**(1), 111–128 (2015)
4. U.S. House of Representatives: Committee on Homeland Security. Hearing on Emerging Cyber Threats to the United States (2016)
5. Cyber terrorism seen as the BIGGEST single future threat. http://securesense.ca/cyber-terrorism-seen-biggest-single-future-threat
6. U.S. charges Iranians for cyberattacks on banks, dam. http://www.cnn.com/2016/03/23/politics/iran-hackers-cyber-new-york-dam
7. Soros hacked, thousands of open society foundation files released online. https://www.rt.com/usa/355919-soros-hacked-files-released
8. Turns out you can't trust Russian hackers anymore. https://foreignpolicy.com/2016/08/22/turns-out-you-cant-trust-russian-hackers-anymore
9. Holt, T.J., Smirnova, O., Chua, Y.-T.: Data Thieves in Action: Examining the International Market for Stolen Personal Information. Springer, New York (2016)
10. Michigan State University hacked, personal information stolen. http://nbc4i.com/2016/11/18/michigan-state-university-hacked-personal-information-stolen
11. How was your credit card stolen? https://krebsonsecurity.com/2015/01/how-was-your-credit-card-stolen
12. Hackers selling 117 million LinkedIn passwords. http://money.cnn.com/2016/05/19/technology/linkedin-hack
13. Hohlfeld, O., Graf, T., Ciucu, F.: Longtime behavior of harvesting spam bots. In: Proceedings of the 2012 ACM Conference on Internet Measurement (2012)
14. Celebrities' medical records tempt hospital workers to snoop. http://www.npr.org/sections/health-shots/2015/12/10/458939656/celebrities-medical-records-tempt-hospital-workers-to-snoop
15. NSA officers spy on love interests. http://blogs.wsj.com/washwire/2013/08/23/nsa-officers-sometimes-spy-on-love-interests/
16. Ky. fire commission investigating southeast bullitt fire department. http://www.wdrb.com/story/24643774/ky-fire-commission-investigating-southeast-bullitt-fire-department

17. Report: Tulsa Sheriff's office falsified training records for reserve deputy who fatally fired gun instead of Taser. https://www.washingtonpost.com/news/morning-mix/wp/2015/04/16/report-tulsa-sheriffs-office-falsified-training-records-for-reserve-deputy-who-fatally-fired-gun-instead-of-taser/?utm_term=.32c56ea0498e
18. Harris, S., Maymi, F.: CISSP All-in-one Exam Guide, 7th edn. McGraw-Hill Inc., San Francisco (2016)
19. Burkett, R.: An Alternative Framework for Agent Recruitment (2013)
20. US DoD: Pub 3-13: Joint Doctrine for Information Operations, vol. 9, pp. 1–9 (1998)

A Team-Level Perspective of Human Factors in Cyber Security: Security Operations Centers

Balázs Péter Hámornik[1](✉) and Csaba Krasznay[2]

[1] Department of Ergonomics and Psychology,
Budapest University of Technology and Economics, Magyar Tudósok körútja,
Budapest 1117, Hungary
hamornik@erg.bme.hu
[2] Institute of E-government, National University of Public Service,
Ménesi út, Budapest 1118, Hungary
krasznay.csaba@uni-nke.hu

Abstract. The paper aims to establish a research framework: encompass various fields of interest that have not been linked previously: the information security, the computer supported collaborative work (CSCW), and team cognition in high-risk situations. Where they meet in practice are the Security Operations Centers (SOCs). These security organization units rely on teamwork of experts and they collaborate under high time pressure. They must react as fast as possible to protect the enterprise assets and data. To understand and support their work the research should focus on them as a team. We are highlighting perspectives to understand the teamwork in SOCs.

Keywords: Human factors · Security Operations Center · Teamwork · Computer supported collaborative work

1 Introduction

The paper aims to point out the synergies between the fields of information security and the psychology of team interaction, especially in a computer-supported setting. Information security has dominated discourse in recent years both in academic research and industry practices. The support of effective security operations requires more than only technical solutions. The human factor should be regarded as key in this industry, the way it is acknowledged in other high-risk fields, too. Research and development in surgery, emergency medicine, aviation, military, and nuclear industry [1–6] is focusing not only on tools and technology but also on the skills, communication, and teamwork of employees.

1.1 Security Operations Centers (SOCs)

Security Operations Centers are defined as both a team and an organization unit, often operating in shifts around the clock SOCs are also a facility dedicated to preventing, detecting, assessing and responding to cybersecurity threats and incidents, as well as to

D. Nicholson (ed.), *Advances in Human Factors in Cybersecurity*,
Advances in Intelligent Systems and Computing 593, DOI 10.1007/978-3-319-60585-2_21

fulfilling and assessing regulatory compliance. This implies many aspects that invite closer examination: the team (in the first place), the organizational unit, and the external expectations that compliance requirements bring. It is important to emphasize that SOCs cover multiple security activities that require different skill sets when it comes to effective teamwork. A fully functional SOC running 24/7 requires a team of minimum eight to 10 people just to maintain two people per shift, working three days on, three days off, four days on and four days off in opposing 12-hour shifts [7]. This requires effective teamwork and competent leadership of such teams. To maintain continuous high quality through the changes in shifts and people, a deep, shared understanding should be developed and kept up-to-date. In addition, the recruitment, selection, and retention of employees is a crucial in SOCs: shift work, time pressure, monotony, and high risk are all demanding on people as they induce stress and fatigue, and are a challenge to work-life balance.

The main activities that a SOC covers are threat hunting and threat intelligence (TI), monitoring functions, detection, triage of alerts, resolution of incidents (by taking actions or escalations), handling of issues (aligned with the internal or external processes required, e.g. ticketing system or reporting).

The last 15 years of the SOC landscape reveal four incremental generations of SOCs developed as responses to increasingly sophisticated attacks [7–9]. The generations distinguish different sets of tools used and ways of working, as well as more and more requirements to comply with. This evolution is visible throughout the research literature from the years 2000s [10]. SOC generations are as follows:

1. First-generation SOC [7]: Security operations are not delivered by the establishment of a formal SOC, but in many cases by an IT operations individual or a team who focuses on a blend of tasks. They cover device and network monitoring, as well as antivirus operation. They rarely work proactively, and the security incident response is not appreciated highly in the enterprise. This initial generation of SOCs does not usually use a centralized system such as a Security Information and Event Management (SIEM) tool.
2. Second-generation SOC [7]: At this stage, SOCs focus on security threat management and event management, which creates the need for SIEM tools. SIEMs aggregate log information from various sources to form events. Events are then correlated to discover the possible relationships between them to help identify a security incident. Incidents are reported and visualized as dashboard alerts to SIEM operators. At this level, SOC activities are integrated with company ticketing systems. The main activity while operating such second generation SOC systems is correlation rule setting and refinement to enable the SIEM tool to capture known or recently discovered threats. It is always a reactive way of working.
3. Third-generation SOC [7]: At this level of evolution, incident response tasks are formalized. Other security services, such as vulnerability management, are linked to SOC operations. This shows a shift toward a more proactive strategy.
4. Fourth-generation SOC [7]: The latest generation is described by the manner SOCs treat data. They can analyze large amounts of data recorded over long periods of time to discover threats and visualize them. This volume of data could also mean big data analytics. The data is enriched using multiple external sources (e.g. geo IP,

DNS, IP and Domain reputation service, threat intelligence feeds). Another key differentiator at this level is the automation of remediation measures (as opposed to manual rule setting processes).

Multiple models of SOCs exist on the market. Among others, there are (1) multi-functional teams serving both Network Operation Center (NOC) and SOC purposes, (2) dedicated, fully functional in-house SOCs, and (3) managed services providing the SOC as an outsourced service.

A SOC facility's physical characteristics (Fig. 1) are inspired by the arrangements found in a network operating room [7]. The aim of the physical setting is to facilitate monitoring, the shared understanding of events, and the collaboration of experts.

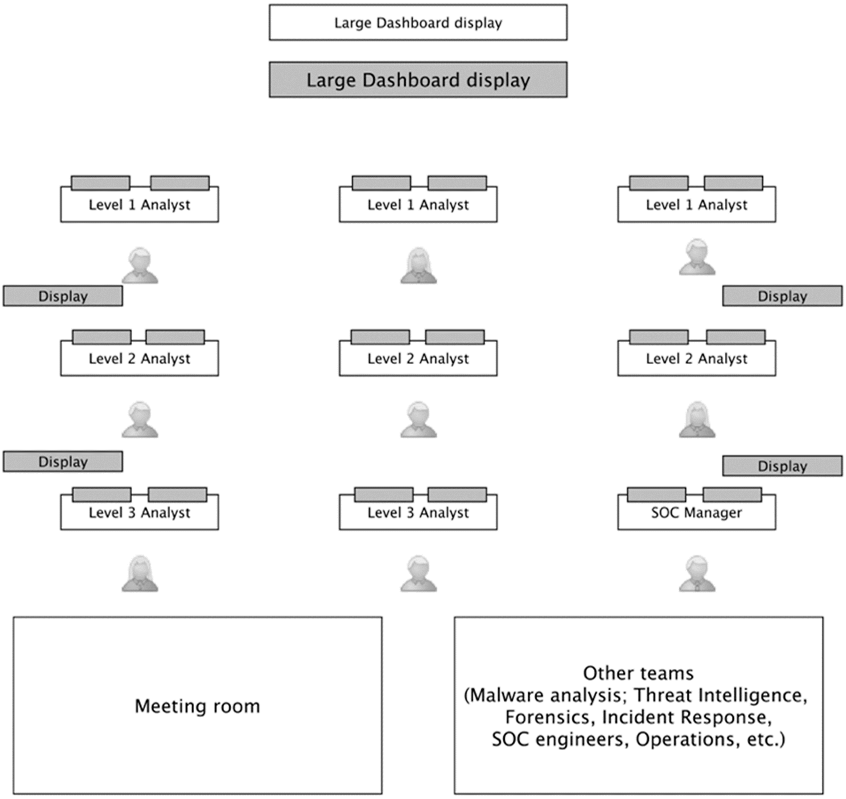


Fig. 1. SOC floor plan

The analysts at the individual workstations with multiple displays are facing toward a large central screen, which shows a dashboard where alerts may appear and where network status is monitored in tables, logs, and charts. The analysts' displays show the same types of information in details. The individual workstations are grouped by the roles of team members. Level 1 analysts, who investigate alerts at the first step, usually sit the closest to the central display. Then level 2 and level 3 come at increasing distances from the center. The adjunct functions of the SOC e.g. TI, forensics, malware analysis may sit separately or even in different facilities. The SOC manager as a team leader is positioned to overlook the whole team in the room to be able to orchestrate their work.

As mentioned previously, the SOC is also defined as a team which has a leader and specialized employees [7, 11]. SOCs are usually led by the SOC manager, who is responsible for the overall leadership. The majority of tasks in the team rely on analysts whose responsibilities can include security event monitoring, incident report investigation, incident handling, threat intelligence, vulnerability intelligence, and reporting. They are organized in escalation levels (1–2–3) or tiers from juniors to seniors. The most advanced experts are doing forensics and malware analysis, which may be somewhat separated from the escalation levels. There are engineering roles (SOC engineers) too, who are responsible for the testing, staging, and deploying of new technology platforms or major releases/updates to those platforms. This also includes the setting and the refinement of correlation and detection rules. Operations roles also exist, focusing on the maintenance and operation of the SOC platforms. Besides these core roles, other support functions can also be represented in the SOC team: project managers, compliance and audit support experts, process/procedure developers, training specialists, communications specialists, etc.

SOCs nowadays face challenges from both internal and external issues. A global shortage of skills and employees constrains the building of SOC teams. From the point of view of external challenges, intensive and ever more complex cyber attacks constantly push SOCs toward applying new technology, and toward a change from a reactive to a proactive way of working, one that is based on threat intelligence and hunting activities [11]. From the perspective of internal challenges, once a SOC has been built, it is not done. Indeed, the operation and the further development of processes, people, and technology remain crucial all along. From the people side, collaboration within the SOC team and with other teams are specifically emphasized in recent market analyses [11].

1.2 Teamwork and Team Cognition

Teams are group of people working together toward reaching a common goal. They work in an interdependent way: every team member's performance contributes to the overall performance and they rely on each other [12]. Their activity is coordinated by a leader who orchestrates the processes and procedures they follow. These are especially valid in the case of high-risk industries – such as aviation, nuclear power-plants (NPPs) or even information security – where an error can lead to fatalities, accidents, or data losses.

Teams are more than the sum of their members: interdependency and collaboration among members produce higher performance than can be reached by the individuals making up the team. This originates from the way knowledge is used and combined in a team. Mental representations that contain information that are applied in the team are named in multiple ways in the literature of psychology. A focused field of applied cognitive and social psychology studies team cognition in multiple industries. Team mental models [13–15] contain the shared knowledge that a team has. This means the up-to-date representation of the internal and external reality, the knowledge that has to be applied during work. It contains the problems and tasks to be solved, the tools to be used, individual knowledge and its distribution (who knows what), the processes to follow on a team or individual level (roles in the team), and the future state that is the aim of the team's activities. The team's mental model functions as a common interpretative frame for the team, which enables them to react effectively to challenges [14]. The team mental model contains decision- and behavioral patterns that can be applied across the team and which enable them to behave coherently.

According to Banks and Millward [16] the procedural knowledge that dictates how to perform a task [7] usually does not have to be owned by every team member (Fig. 2). It is not fully efficient to have the knowledge of procedures represented redundantly: a team should not be a group of one-man-armies. The declarative knowledge containing what to do has to be owned by every team member indeed. It enables the team to keep their focus on their aims, and act in a coordinated way toward the same goal.

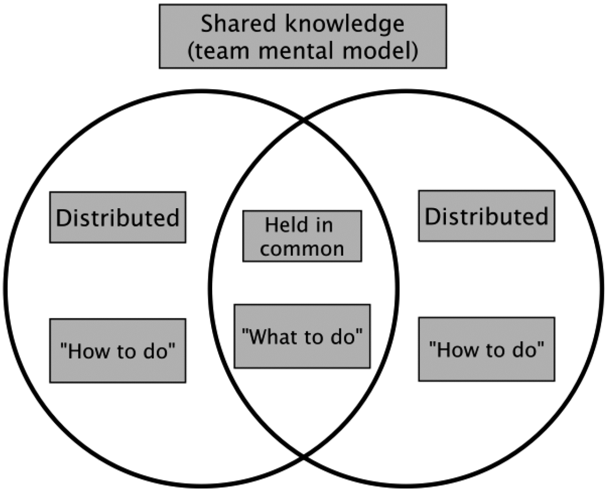


Fig. 2. Team mental model: what is shared and distributed?

The key to using team mental models effectively, as a team-level cognitive process, is communication. Explicit communication enables teams to build and update team mental models [1, 5]. During periods of high pressure, there is often no room to communicate, to explain the background of actions or the context. Thus, communication before actions is crucial to a fully functional team mental model. During an emergency situation, teams perform with a limited communication capacity, coordinating their actions implicitly. This means that they presume that everyone knows what to do and how to perform their roles. The team mental model held in the minds of individuals enables the team to perform effectively. Studies of aviation, surgery and rehabilitation, the nuclear industry, the military, and virtual collaboration discovered similar patterns in team mental models [2, 4–6, 17].

1.3 Computer Supported Cooperative Work (CSCW)

Technology-related research began to focus on teams in the 1980s: teams were defined as people using technology together toward a common work purpose. According to Carstensen and Schmidt [18], the research field of Computer Supported Cooperative Work (CSCW) addresses "*how collaborative activities and their coordination can be supported by means of computer systems.*" Relying on the definition of Carstensen and Schmidt [18] "*computer-based support for cooperative work can be provided by offering better communication facilities, providing improved monitoring and awareness possibilities to the actors, and by aiming at reducing the complexity of the coordination activities to be conducted by the involved actors.*" CSCW focuses on the study of tools and techniques of groupware, *as well as* their psychological, social, and organizational effects [19]. This is aligned with what team cognition research is aiming for in general. The "CSCW matrix" (Fig. 3) considers the work context along two dimensions [20]: space features of collaboration (co-location or geographically distributed) and time features of collaboration (synchronous or asynchronous working). The resulting four cells cover most of the possible ways a team can collaborate or cooperate.

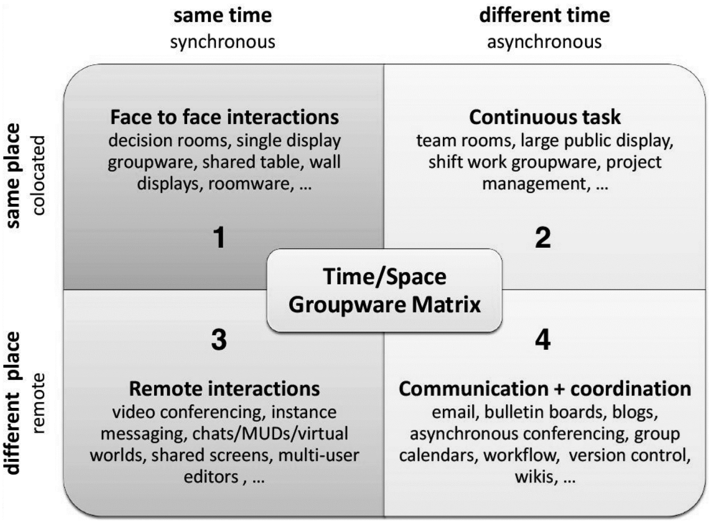

same time
synchronous
different time
asynchronous
same place
colocated
different place
remote
Face to face interactions
decision rooms, single display groupware, shared table, wall displays, roomware, ...
1
Continuous task
team rooms, large public display, shift work groupware, project management, ...
2
Time/Space
Groupware Matrix
3
Remote interactions
video conferencing, instance messaging, chats/MUDs/virtual worlds, shared screens, multi-user editors , ...
4
Communication + coordination
email, bulletin boards, blogs, asynchronous conferencing, group calendars, workflow, version control, wikis, ...

Fig. 3. The CSCW matrix [21]

In order to use our concepts distinctively, we have to define what we mean by collaboration and cooperation. Collaboration is when people work together toward a single shared goal. For example, a band performs a song together as a common shared task and the output is a holistic experience. Cooperation is slightly different: while cooperating, people perform together but also work on their own goals (goals that fit the common aim's direction). Reusing the example of the orchestra, the drum solo is an act of cooperation (the drummer's own performance within the whole). In work situations, often the latter is more common: individuals join forces to reach a common goal, while performing individual actions and reaching individual goals too. Reality is a mixture of collaborative and cooperative situations. The team cognition literature mentioned previously refers to this question by distinguishing knowledge entities that have to be held in common (declarative knowledge) and those to be held distributed (procedural knowledge) [16].

1.4 Aims

In the initial phase, that is Phase 1, of our research, we are aiming to describe the analytic framework of Security Operations Centers from a human factors perspective. For this, the following three assumptions are assessed:

1. Are Security Operations Centers valid fields of research for CSCW? We are aiming to decide whether the concepts and methods of this field could be applied to SOCs.
2. Could team cognition be observed and studied in SOCs? We are aiming to understand whether team mental models and team level cognitive processes appear in SOCs and whether they are key elements in their performance.
3. The most general question is whether SOCs, the people working in SOCs, their tasks, their teamwork could be understood and studied on a team level instead of on the level of the individual. This may affect how we should think about knowledge, skills, abilities and other key attributes that enable people to perform in a SOC. If SOCs can be studied on the team level, then that would assume that the team a SOC team may be more than the sum of experts sitting in the same room in the same shift.

2 Methods

We have conducted 13 semi-structured interviews with industry experts who are operating a SOC or performing tasks related to SOCs. Interviews were focused on the following topics: processes for investigating incidents, roles in the team, tools used, levels in SOCs, time frames and escalations, what information is presented and available for the experts, the physical organization of the workplaces, the largest hindrances among the daily tasks, etc.

The experts interviewed come from the Western-, and Central & Eastern Europe region, and North Africa. They operate in the IT, finance, governmental, and IT security industries. There can be found both in-house SOC operators and managed security service providers (MSSPs) among them. These interviews were combined with two field visits to currently operating SOC departments: one was an in-house SOC of an IT company, the other was a large managed SOC of an MSSP serving clients in Western Europe. This set of interviews and visits were completed with literature review, the analysis of sector-specific market research (e.g. Gartner, Forrester), and two interviews with independent market experts. We handle all the sensitive company information anonymously and present them only as aggregated results here.

3 Results

We have found that SOCs show large differences in maturity levels and SOC models.

On the one hand, there are SOCs owned by enterprises only for compliance reasons, and these are not operated at their full potential. On the other hand, there are highly structured in-house and managed SOCs that focus on proactive security monitoring. The industry trend is to focus on threat intelligence and become more proactive [11] in order to keep up with the rising number of attacks.

We have found that SOC activities are separated from the overall security or operations departments in most cases. Depending on maturity, this means more specified roles, positions, and locations. Summarizing the processes that these SOCs (or SOC-like teams) follow, it is important to emphasize that all of them have dedicated escalation levels with defined time frames to handle an event or incident. This is one main source of time pressure across the teams. The main activity besides event and incident handling is the setting and refining of rules. The core tool used in SOCs is a SIEM that uses rule-based alerting. These rules are reactively made based on previous and recent incidents. The rule set builds up a large knowledge base of security incidents that the SOC can detect and handle. Nevertheless, this rule set requires continuous updating as attacks tend to evolve and change over time. This is the most time-consuming and effortful activity in SOC teams. Following the trends that Gartner [11] revealed, therefore, TI and proactive ways of working (e.g. use of machine learning) are gaining popularity to reduce the hassles caused by manual rule setting.

Regarding human factors, it is important to note that monotony is a strong source of stress for analysts: the monitoring task is repetitive and most of the time no significant incident happens, hence vigilance has to be maintained. This is, among others, a cause that contributes to large fluctuation and a lack of employees in the sector. Shift working in 24/7 is the other factor that contributes to heavy workload and stress. These finding are similar to what research on aviation, NPP, and medical teams found.

The physical work settings, in every case, aim to support the visibility of information: large screens, multiple displays per workstations, and specialized light conditions. The tools analysts use are largely customized based on company requirements, for example, integrated with or even built around the ticketing system.

The roles in SOC teams are highly structured, hence the lack of skilled employees may eventually contribute to more flexibility (e.g. through job rotation) in order to support employee retention. The policies, processes, and even procedures are highly defined to fit company regulation or compliance requirements. This is also similar to the other fields of high-risk teamwork mentioned previously in this paper.

In the SOCs studied, collaboration and cooperation are observable on multiple levels and mediated by several channels. Fist of all, local team members communicate within the team verbally or using email, chat, and the ticketing system (in the case of asynchronous cooperation). The security teams observed in global companies have connections or sub-teams in various locations in the world. There is intensive communication with these remote team members using computer-mediated channels, phone calls, and rarely, face-to-face meetings. Finally, there is cooperation and communication with employees of the company who are not involved in security functions. They are the "ordinary" people who can be both targets of an attack or represent an insider threat. Information exchange with employees mainly happens using email and phone calls. Reaching company employees is especially complicated in the case of a managed security service when the SOC is operated by an external company in a time zone different from the one where the customer to be protected is.

A schematic way of incident handling in a SOC is described by Muniz et al. [7], which is supported by our findings based on our interviews and visits. From a bird's eye view, it consists of the following steps:

1. The security infrastructure collects logs and data all around the network, filtered and correlated by the SIEM. The SIEM produces and displays events that are assessed by the analysts in the SOC.
2. The SIEM rules fire alerts on events that fit any of the known use cases of security threats, indicating that an action needs to be taken to decide whether the alert is a real security incident or a false positive case.
3. The Level 1 SOC analyst reviews the data available in the SIEM in a short time frame (approximately 15 min) and makes a decision: marks the event as a false alarm or escalates it for further investigation to Level 2.
4. The analyst in this next tier (Level 2) of the SOC has more time to investigate the suspicious event and more data available from a tool other than the SIEM. If they can make the decision about the event and take an action required to secure the system, they do so. If more investigation is required either because Level 2 could not decide, or after the action has been taken, more information must be gathered about the nature of the attack (for forensics reasons), Level 2 escalates the case to Level 3.
5. The Level 3 analyst usually has no time limits and often fulfills forensics roles, too. All levels are able to reach out to the employees seemed to be involved in the event to verify the actions observed. Analysts also have to report the incidents to the SOC manager, who reports upwards. Also, the communications representative of the company may be involved if the incident affects customer data.
6. Closing an incident includes collecting forensics data, doing malware analysis (if the incident was caused by malware), refining rules based on the incident in order to make the SIEM capable of capturing further similar attacks.

Besides this chain of actions performed when something out of the ordinary happens, the daily routine of the SOC team consists of prescribed monitoring tasks, threat hunting for suspicious events in data, and rule refinements – these are often monotonous tasks.

4 Conclusion

The findings summarized above prove that the information security activities performed in Security Operations Centers rely heavily on team collaboration, cooperation and on how information is shared and used in teams. These teams use multiple computer-mediated channels for working together and for collecting, displaying, storing information, and taking actions. Examples of these channels are the ticketing system, chat applications, phone calls, and wikis.

Therefore, the quality of computer supported collaborative work may play a key role in SOCs. The SOC team also uses the channels mentioned above for collaboration. Information security teamwork may be studied and supported within the framework of CSCW [22, 23].

The heterogeneous set of cooperative activities in a SOC can be sorted using the cells of the CSCW matrix presented above (Fig. 3).

1. In the case of face-to-face interactions, discussions and meetings take place in the same time and space in the SOC. The large displays of events or alerts happening in real time are visible from all parts of the SOC room and they provide information in a collocated synchronous way.
2. The same large displays, dashboards, and project management tools (e.g. Jira, Slack) supporting continuous work can be used in an asynchronous way too, while analysts are changing shifts or new experts are involved in the incident response.
3. In the groupware matrix of SOCs, remote interactions happen through messaging or chat tools (sometimes integrated in project management applications). The same dashboards and SIEM data are visible in multiple remotely collaborating locations of a SOC. The real-time monitoring of privileged users also fits into this cell of the matrix, that is, when an activity is remotely observed as it happens.
4. Teamwork that is asynchronous in both time and space is related to communication and coordination tools such as the project management or ticketing tools, emails, wikis (e.g. Confluence), and calendars. Threat intelligence and forensics data collection and sharing also fit into this cell.

We suggest the following answers for the three initial questions posed in this paper.

1. SOCs represent a valid field of research within the framework of CSCW: collaborative activities are largely relying on computer-mediated channels covering all cells of the CSCW matrix. This is also aligned with what Goodall, Ozok, Lutters, Rheingans, and Komlodi [23] found.
2. Team cognition can be observed in SOCs because the goal of a SOC team can be reached by the unified effort of all team members. This requires information to be shared and kept up-to-date within the team across time and space. Team mental models may provide an underlying conceptual explanation for these processes similarly to other high-risk fields such as aviation, surgery, and NPP [1, 4, 6, 13, 14].
3. When studying SOCs on a team level, multiple directions can be taken in the future. At first, similarly to team cognition research, studies could be conducted of team mental models, their content, way of development, and their application. The results derived from this direction of research would be applied to support the development of SOC processes and procedures, helping more efficient team cognition through communication and information sharing.

4.1 Analyzing SOCs on Team Level

To understand SOC teams on a higher level, the team cognition literature provides an explanatory framework that fits the phenomena observed in our study. Team cognition, including the building and updating of a team's mental model, may play a key role in SOC teamwork. Knowledge about the network, activities, and the security incidents that had already happened constitute the knowledge to be shared in the team as declarative knowledge [13, 14, 16]. The separate SOC roles (e.g. analysts, operations engineer, forensics) require cooperation but their work procedures should not be shared by everyone: except for job rotation, there is no need to swap positions or tasks. As Banks and Millward [16] states, the sharing of procedural knowledge is not supporting effective collaboration. In the case of SOCs, the sharing of procedural knowledge required for job rotation and career development may have a positive effect on employee retention. Event monitoring activities in a SOC require a constant awareness of the situation, and this keeps the team mental models up-to-date about the security status and the team. This team mental model is developed and updated by both internal and external communication. If the mental models are well functioning, explicit communication and coordination activities may not be required during high-risk incident responses and under high time pressure, similarly to other high-risk fields (NPP, aviation, surgery) [1, 2, 4, 6]. The information provided by face-to-face or computer-mediated means must be available for all team members – this is a key prerequisite for effective team cognition. The communication channels detailed above in the CSCW matrix are all used in building and maintaining the team mental models.

Based on these similarities to team cognition research, we propose future studies on SOCs, measuring team mental models, and the effect of communication, roles, and tasks on them. Measurements should focus on team-level outcomes (which, overall, means security) and on how to break down the subject of our study into observable units. The scenarios of teamwork to be studied include both synchronous and asynchronous cooperation and collaboration using information technology. This characteristic of teamwork in SOCs underlines the importance of CSCW research, and the validity of its assumptions to be involved.

4.2 Summary

In this initial phase, we established the research framework of our further studies of SOCs. We will be treating SOCs as teams that work in a way that can be described as computer supported cooperative work (CSCW) (as defined by the relevant research). As we have seen, SOC experts can be studied as teams that are more than the sum of a set of people, with team cognition being a key to effective security teamwork. The knowledge of SOC teams is supposed to be studied within the applied psychology framework of team cognition in order to capture the team mental models in action. The aim is to contribute to better SOC teamwork and therefore to better information security by applying research results that take into account human factors besides technology.

Acknowledgments. The research is sponsored by the Government of Hungary under Grant No. KÖFOP-2.1.2.-VEKOP-15-2016-00001.

References

1. Juhász, M., Soós, J.K.: Impact of non-technical skills on NPP teams' performance: task load effects on communication. In: 2007 IEEE 8th Human Factors and Power Plants and HPRCT 13th Annual Meeting (2007)
2. Sexton, J.B.B., Helmreich, R.L.L.: Analyzing cockpit communications: the links between language, performance, error, and workload. In: Proceedings of the Tenth International Symposium on Aviation Psychology, pp. 689–695 (1999)
3. Antalovits, M., Izsó, L.: A methodology for assessing and developing teamwork in cognitively demanding jobs. Period. Polytech. Soc. Manag. Sci. **7**, 105–118 (1999)
4. Burtscher, M.J., Wacker, J., Grote, G., Manser, T.: Managing nonroutine events in anesthesia: the role of adaptive coordination. Hum. Fact. J. Hum. Fact. Ergon. Soc. **52**, 282–294 (2010)
5. Hutchins, E.: Cognition in the Wild. MIT Press, Cambridge (1995)
6. Schmutz, J., Hoffmann, F., Heimberg, E., Manser, T.: Effective coordination in medical emergency teams: the moderating role of task type. Eur. J. Work Organ. Psychol. **24**, 761–776 (2015)
7. Muniz, J., McIntyre, G., AlFardan, N.: Security Operations Center: Building, Operating, and Maintaining Your SOC. Cisco Press, Indianapolis (2015)
8. Forte, D.: An inside look at security operation centres. Netw. Secur. **2003**, 11–12 (2003)
9. Ahmad, A., Maynard, S.B., Shanks, G.: A case analysis of information systems and security incident responses. Int. J. Inf. Manag. **35**, 717–723 (2015)
10. Forte, D.: State of the art security management. Comput. Fraud Secur. **2009**, 17–18 (2009)
11. Chuvakin, A.: Design a Modern Security Operation Center. http://blogs.gartner.com/anton-chuvakin/2016/10/11/upcoming-webinar-design-a-modern-security-operation-center-soc/
12. Levi, D.: Group Dynamics for Teams. Sage, Thousand Oaks (2011)
13. Mohammed, S., Klimoski, R., Rentsch, J.R.: The measurement of team mental models: we have no shared schema. Organ. Res. Methods **3**, 123–165 (2000)
14. Klimoski, R., Mohammed, S.: Team mental model: construct or metaphor? J. Manag. **20**, 403–437 (1994)
15. Cooke, N.J., Salas, E., Cannon-Bowers, J.A., Stout, R.J.: Measuring team knowledge. Hum. Fact. **42**, 151–173 (2000)
16. Banks, A.P., Millward, L.J.: Differentiating knowledge in teams: the effect of shared declarative and procedural knowledge on team performance. Gr. Dyn. Theor. Res. Pract. **11**, 95–106 (2007)
17. Hámornik, B.P., Köles, M., Komlódi, A., Hercegfi, K., Izsó, L.: Features of collaboration in the VirCA immersive 3D environment. In: Stanney, K., Hale, K.S. (eds.) Proceedings of Advances in Cognitive Engineering and Neuroergonomics - AHFE 2014, pp. 130–139. The AHFE Conference, Krakow (2014)
18. Carstensen, P.H., Schmidt, K.: Computer supported cooperative work: new challenges to systems design. In: Itoh, K. (ed.) Handbook of Human Factors, pp. 619–636. Asakura Publishing, Tokyo (1999)
19. Wilson, P.A. (Advanced Concepts Branch, Central Computer and Telecommunications Agency, Treasury, Great Britain): Computer Supported Cooperative Work: An Introduction. Intellect, Oxford (1991)

20. Baecker, R.M., Grudin, J., Buxton, W.A.S., Greenberg, S.: Readings in Human-Computer Interaction: Toward the Year 2000, 2nd edn, p. 595. Elsevier, Saint Louis (1995)
21. The CSCW Matrix. https://commons.wikimedia.org/wiki/File:Cscwmatrix.jpg
22. Werlinger, R., Muldner, K., Hawkey, K., Beznosov, K.: Preparation, detection, and analysis: the diagnostic work of IT security incident response. Inf. Manag. Comput. Secur. **18**, 26–42 (2010)
23. Goodall, J.R., Ozok, A.A., Lutters, W.G., Rheingans, P., Komlodi, A.: A user-centered approach to visualizing network traffic for intrusion detection. In: CHI 2005 Extended Abstracts on Human Factors in Computing Systems – CHI 2005. p. 1403. ACM Press, New York (2005)

Utilizing Chatbots to Increase the Efficacy of Information Security Practitioners

Saurabh Dutta(✉), Ger Joyce, and Jay Brewer

Rapid7, 100 Summer St., Boston, MA 02110, USA
{sdutta,gjoyce,jbrewer}@rapid7.com

Abstract. Almost every day, the world hears about a new information security breach. In many cases, this is due to the vast quantity of data generated across millions of connected devices with little insight, and the amount of work that information security practitioners must do to make sense of it all. The lack of skilled information security resources doesn't help. Different approaches are being attempted to fix these issues. However, many approaches are neither cost-effective nor scalable. One potential approach, which is both cost-effective and scalable, is the utilization of chatbots. In this paper, the authors focus on ways in which chatbots can assist information security practitioners, such as security analysts and pentesters, beyond the current human-before-support philosophy. Scenarios include investigations of potentially malicious behavior and team pentest projects, each of which explores how a chatbot might allow the relevant type of information security practitioner to be far more effective and efficient.

Keywords: Chatbots · Information security · Cyber threats · Cybersecurity

1 Introduction

On a regular basis, each of us receive an email, read a blog post, or hear a story on the news about another information security breach. Even to those outside of the information security industry, it would be no surprise to hear that malicious information security attacks are on the rise. Yet, globally there is a severe shortage of skilled information security professionals. In fact, a recent report from Cisco states that there are approximately one million information security job openings around the world [1]. To meet this skills gap, many professionals from adjacent areas, for example Information Technology and application development, are migrating to the security domain. However, even after training, many continue to have questions about the security domain and information security product usage. Further, even information security professionals that have been in the field for years, often have questions about the latest cybersecurity threats, given that the threat landscape changes frequently. This, in turn, leads to a secondary, yet vital, issue—how can the information security industry as a whole support practitioners, regardless of the length of their tenure, in an efficient, effective, and scalable way?

The most obvious option is to increase support personnel, training, and documentation. However, the additional cost related to salaries, software, building space, and so on, might not be a viable choice for many organizations. Another option is to increase

D. Nicholson (ed.), *Advances in Human Factors in Cybersecurity*,
Advances in Intelligent Systems and Computing 593, DOI 10.1007/978-3-319-60585-2_22

the amount of documentation and training content produced. Again, due to the costs involved to produce and potentially to translate, documentation and videos, this might not be an option for every organization.

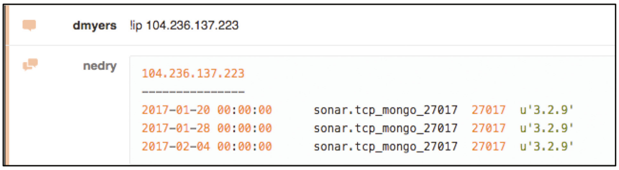
A far more cost-effective and scalable solution, is to utilize chatbots. Chatbots can better communicate on-demand and in users' natural language. Consequently, chatbots have the potential to answer questions from inexperienced, and indeed experienced, information security personnel quickly and accurately. Furthermore, chatbots can individually and contextually communicate on a one-to-many basis. As chatbots can be built inexpensively and require no additional building space, this sounds like a win-win scenario.

2 Future of Chatbots Within Information Security

Moving away from the abstract to the more concrete, let's ask ourselves several questions as to how chatbots might be useful to information security practitioners. The most obvious solution, one that is being implemented widely outside of the industry, is to allow people to interact with a chatbots prior to communicating with support personnel. This is becoming commonplace within industries, such as banking [2], libraries [3], and to assist in recruiting [4].

However, there are other even more creative ways to utilize chatbots. For example, consider how information security practitioners might receive answers to their questions within this inherently complex field, and how might chatbots help information security practitioners to collaborate to ensure organizations are secure? Consider the following scenario, an information security practitioner starts to use a vulnerability management software tool, such as Nexpose from Rapid7. The information security practitioner is reasonably new to the field and still has a lot to learn. The practitioner is aware that new and more complex cyber threats appear frequently, and that they need to be fully aware of such threats if they are to keep their organization secure. As per the information security personas depicted by Bhattarai et al. [5], the practitioner in our scenario is busy and has little time for reading. To that end, the practitioner is not fully aware of the threat landscape, such as the latest cyber threats applicable to the vertical the practitioner works in. Consequently, it is possible that the information security practitioner is focusing on the wrong types of threats. One way to discover the latest threats within the practitioner's vertical, such as healthcare or finance, is to speak with colleagues, read appropriate blog posts and white papers, twitter feeds and so on, yet all of this takes time.

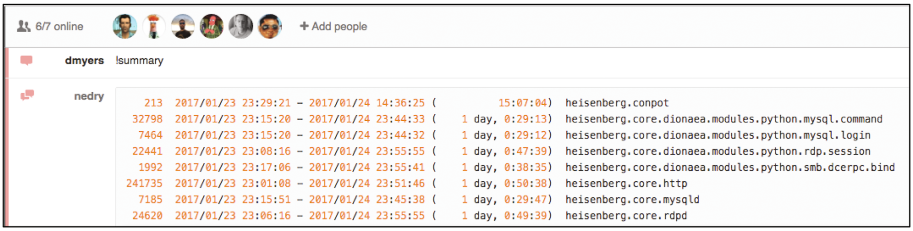
A more efficient approach would be for the information security practitioner to discuss these topics with a chatbot, thus gaining immediate insights. While this scenario might seem futuristic, it is not actually that far off. Two community-run information security projects, led by Rapid7, form the basis of turning big data into big insights, and may revolutionize the way that chatbots interact with information security practitioners. These projects are known as Project Sonar and Heisenberg Cloud, and may form the basis whereby chatbots act as the intermediator between an information security practitioner and big data. This would be made possible as Project Sonar analyzes public networks, and shares the results [6], whereas Heisenberg Cloud focuses on the cloud, specifically attempting to better comprehend what attackers, researchers, and companies are doing within cloud environments [7]. This data is then turned into high quality threat intelligence. Consequently, information security practitioners can then learn about existing and emerging cyber threats in order to improve their organization's security posture. Both data sets can be analyzed in order to surface patterns, trends, and associations, especially in regard to how people behave. Subsequently, should an information security practitioner have a question, they might instead ask a chatbot. For example, an information security practitioner might wonder about a new IP address that was noticed within an investigation of strange behavior within a network. The practitioner can query project sonar about the IP address (Fig. 1), and gather more detailed information from Heisenberg Cloud if needed (Fig. 2). All of this might be possible in the future via chatbot interactions, thus allowing information security practitioners the ability to quickly and easily discover new information that will heighten their organization's security posture.



```
dmyers  !ip 104.236.137.223

nedry   104.236.137.223
        ---------------
        2017-01-20 00:00:00     sonar.tcp_mongo_27017  27017  u'3.2.9'
        2017-01-28 00:00:00     sonar.tcp_mongo_27017  27017  u'3.2.9'
        2017-02-04 00:00:00     sonar.tcp_mongo_27017  27017  u'3.2.9'
```

Fig. 1. Information security practitioner uncovers information about an IP address gathered by Project Sonar.

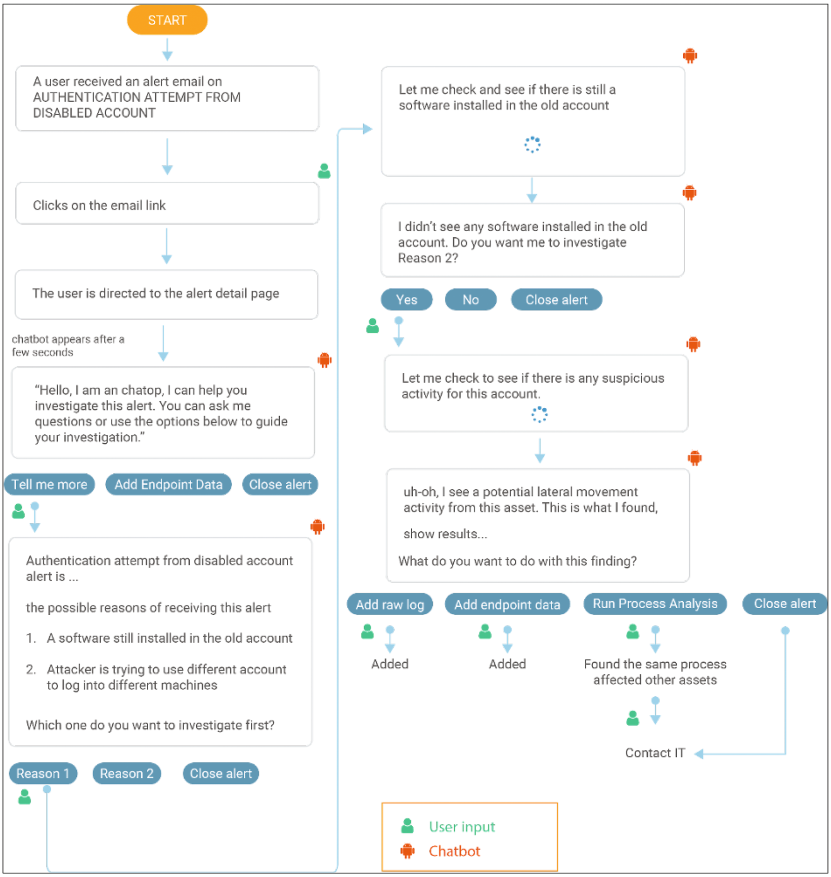


```
dmyers  !summary

nedry      213  2017/01/23 23:29:21 - 2017/01/24 14:36:25 (        15:07:04)  heisenberg.conpot
         32798  2017/01/23 23:15:20 - 2017/01/24 23:44:33 (  1 day, 0:29:13)  heisenberg.core.dionaea.modules.python.mysql.command
          7464  2017/01/23 23:15:20 - 2017/01/24 23:44:32 (  1 day, 0:29:12)  heisenberg.core.dionaea.modules.python.mysql.login
         22441  2017/01/23 23:08:16 - 2017/01/24 23:55:55 (  1 day, 0:47:39)  heisenberg.core.dionaea.modules.python.rdp.session
          1992  2017/01/23 23:17:06 - 2017/01/24 23:55:41 (  1 day, 0:38:35)  heisenberg.core.dionaea.modules.python.smb.dcerpc.bind
        241735  2017/01/23 23:01:08 - 2017/01/24 23:51:46 (  1 day, 0:50:38)  heisenberg.core.http
          7185  2017/01/23 23:15:51 - 2017/01/24 23:45:38 (  1 day, 0:29:47)  heisenberg.core.mysqld
         24620  2017/01/23 23:06:16 - 2017/01/24 23:55:55 (  1 day, 0:49:39)  heisenberg.core.rdpd
```

Fig. 2. Information security practitioner retrieves reconnaissance results from Heisenberg Cloud to add to an investigation

In addition to utilizing chatbots as an intermediator between the information security practitioner, such as a security analyst, and big data during a manual investigation, a chatbot could also be useful in other ways when conducting an investigation. Consider the following scenario whereby an information security practitioner receives an alert from a software application, such as InsightIDR from Rapid7. InsightIDR alerts information security professionals when something seems strange and out of the ordinary occurs within their organization's network. In this case, the information security practitioner receives an alert, clicks on the link to go to the alert details. However, instead of manually attempting to figure out if this is normal or malicious behavior, which takes time out of their already busy day, a chatbot interjects, and assists the information security practitioner with the investigation (Fig. 3). This allows the information security practitioner to close out many more investigations, helping to ensure their organization's information stays safe more efficiently.



A user received an alert email on AUTHENTICATION ATTEMPT FROM DISABLED ACCOUNT

Clicks on the email link

The user is directed to the alert detail page

"Hello, I am an chatop, I can help you investigate this alert. You can ask me questions or use the options below to guide your investigation."

Authentication attempt from disabled account alert is ...

the possible reasons of receiving this alert

1. A software still installed in the old account
2. Attacker is trying to use different account to log into different machines

Which one do you want to investigate first?

Let me check and see if there is still a software installed in the old account

I didn't see any software installed in the old account. Do you want me to investigate Reason 2?

Let me check to see if there is any suspicious activity for this account.

uh-oh, I see a potential lateral movement activity from this asset. This is what I found,

show results...

What do you want to do with this finding?

Fig. 3. An example of how a chatbot can assist an information security practitioner during an investigation

It doesn't stop here. There are many more instances whereby chatbots can assist information security practitioners, no matter how experienced or inexperienced they are. Let's consider a scenario that involves pentesting. Within the world of information security, pentesting is the practice of testing a computer system, network, or web application in order to surface vulnerabilities that an attacker might exploit. Pentesters are generally experienced and have worked within the field of information security for many years. Organizations often hire pentesters on a contractual basis, on the premise that if a sanctioned pentester can find ways to hack into the organization's network, so too can a malicious attacker.

In the first scenario, we looked at how inexperienced information security practitioners might utilize a chatbot. However, would an experienced pentester consider using a chatbot, and if so when? Well, a pentester could use a chatbot as a partner to quickly discover public-facing web assets a client might have. Once those assets were discovered, the pentester could then consider if any open ports were available. Based on this, and on threat intelligence from Project Sonar and Heisenberg Cloud, the chatbot could anticipate which modules and attack technique had the most chance of success within the environment, subsequently recommending these to the pentester (Fig. 4). This approach thereby saves a lot of the pentester's time. This, of course, ensures that the hiring organization increases their security posture based on the recommendations from the pentester much faster and at a lower cost.

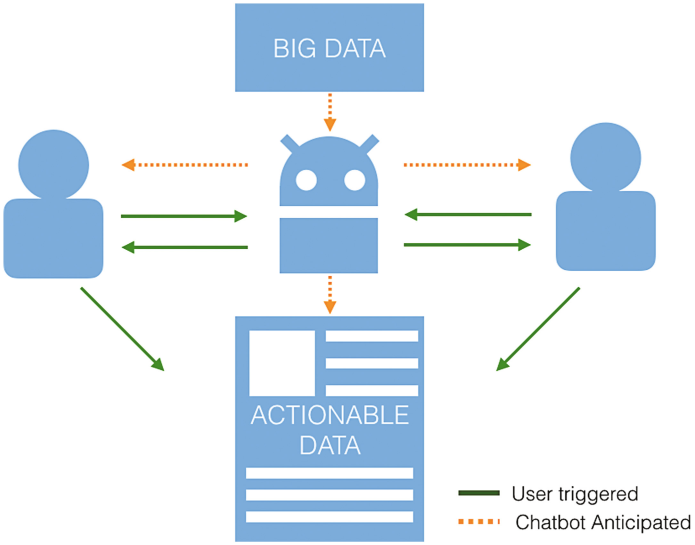


Fig. 4. An example of how a chatbot can help turn big data into actionable data and insights

Chatbots can also assist pentesters during a team engagement. The chatbot can share information among team members, whilst working on traditional menial tasks, such as collecting credentials and taking screenshots of desktops to prove that the pentesters did, in fact, penetrate the organization. The chatbot can use the same shared credentials from all team members to infiltrate other areas of the network, thus allowing multiple pentesters to transparently share meterpreter sessions. This allows the attack to continue without being detected by a machines or networks anti-virus defense [8]. The chatbot can also conduct smart locking/unlocking of shell sessions, sharing of scan data, and shared event logs as a hand-off medium. A timeline of events can be formed based on the pentest tasks performed. To that end, the chatbot acts as a collaborator between multiple pentesters, sharing information for better effectiveness and efficiency in real time. As the individual or team pentest progresses, the chatbot can collate data with a view to producing a final report for the client, including audit trails and interesting findings, once again saving the pentesters a lot of time. Furthermore, the report might not be a report in the traditional sense, such as a PDF. While the report will include an executive summary for a Chief Information Officer (CIO) or Chief Information Security Officer (CISO) at the client organization, the report can also include a pentest DVR or live-action report, which is an audit trail of all replayable pentest actions. This allows in-house security operations personnel the ability to see how a successful pentest occurred, which might be helpful for them to better understand how easy or difficult it might be to exploit vulnerabilities within their organization's environment.

3 Limitations

As with any technology, there will be detractors. Many people distrust the whole idea of chatbots for historical reasons. The term may conjure up images of the intrusive and infamous Microsoft Clippy, seemingly designed to exasperate, not to assist [9]. Nevertheless, Clippy was the (admittedly bumpy) start behind the entire concept. That said, in the decades since Clippy hit our screens, we are not yet where we need to be. While chatbots can communicate in users' natural language, their 'knowledge' is limited. At some stage, people may need to come into play—it is this hand-off between chatbots and humans that can be disruptive. Multiple strategies need to be considered with this type of strategy, namely which type of DATA to be (structured or unstructured), which PROCESS will be followed (rule based or inference based), and finally, what OUTCOME is expected (single correct answer or set of likely answers). Until this is resolved, the ideal, yet elusive, 'seamless switch' cannot happen. This could be detrimental within information security environments where much is at stake. Yet, despite the limitations, there is much fanfare around chatbots, to the point where some feel that chatbots will be the way people will interact with information in the future, no longer needing web or mobile applications [10].

4 Conclusion

The field of information security is fast and exciting. Many different types of practitioners, including Chief Information Security Officers (CISO), security analysts, incident responders, and pentesters strive to stay ahead of malicious attackers each and every day. However, staying on top of the constantly changing cyber threat landscape can be challenging. In addition, many information security practitioners, such as pentesters and security analysts need to undertake manual and tedious work regularly. This level of effort decreases their efficiency.

This paper considers some of the scenarios that information security practitioners may face, and how they can become far more efficient by utilizing chatbots in their day-to-day operations. To that end, this work considers how information security practitioners can be supported in a cost-effective, efficient, and scalable way.

References

1. Cisco: Mitigating the cybersecurity skills shortage (2015). http://www.cisco.com/c/dam/en/us/products/collateral/security/cybersecurity-talent.pdf
2. Business Insider.: The chatbots in banking report: how chatbots can transform digital banking (2017). http://www.businessinsider.com/the-chatbots-in-banking-report-how-chatbots-can-transform-digital-banking-2017-1
3. McNeal, M., Newyear, D.: Chatbots: automating reference in public libraries. In: Edward, I. (ed.) Robots in Academic Libraries: Advancements in Library Automation, pp. 101–114. IGI Global, Hershey (2013)
4. Garimella, U., Paruchuri, P.: ^ 2: An agent for helping HR with recruitment. Int. J. Agent Technol. Syst. (IJATS) **7**(3), 67–85 (2015)
5. Bhattarai, R., Joyce, G., Dutta, S.: Information security application design: understanding your users. In: International Conference on Human Aspects of Information Security, Privacy, and Trust, pp. 103–113, Springer, Cham (2016)
6. Rapid7: Project sonar https://sonar.labs.rapid7.com/ (n.d.)
7. Rapid7: Project heisenberg cloud: cross-cloud adversary analytics (2016). https://information.rapid7.com/rs/495-KNT-277/images/rapid7-project-heisenberg-cloud-research-report.pdf
8. Skape: Metasploit's meterpreter (2004). https://dev.metasploit.com/documents/meterpreter.pdf
9. Meyer, R.: Even early focus groups hated clippy (2015). https://www.theatlantic.com/technology/archive/2015/06/clippy-the-microsoft-office-assistant-is-the-patriarchys-fault/396653/
10. Schlicht, M.: How Bots Will Completely Kill Websites and Mobile Apps (2016)

Understanding Human Factors in Cyber Security as a Dynamic System

Heather Young[1(✉)], Tony van Vliet[1], Josine van de Ven[2], Steven Jol[3], and Carlijn Broekman[1]

[1] The Netherlands Organisation for Applied Scientific Research (TNO),
Kampweg 5, Soesterberg, The Netherlands
{heather.young,tony.vanvliet,carlijn.broekman}@tno.nl

[2] The Netherlands Organisation for Applied Scientific Research (TNO),
Oude Waalsdorperweg 63, The Hague, The Netherlands
josine.vandeven@tno.nl

[3] The Netherlands Organisation for Applied Scientific Research (TNO),
Anna van Buerenplein 1, The Hague, The Netherlands
steven.jol@tno.nl

Abstract. The perspective of human factors is largely missing from the wider cyber security dialogue and its scope is often limited. We propose a framework in which we consider cyber security as a state of a system. System change is brought on by an entity's behavior. Interventions are ways of changing entities' behavior to inhibit undesirable behavior and increase desirable behavior. Choosing an intervention should take into account the dynamic nature of how humans use cyberspace. People are not likely to change old behavior at the drop of a hat. The key is to invent new ways to maintain old behavior in new circumstances. Our framework differentiates three basic pathways of actor behavior that influence the cyber security of a system. The distinction between reflex, habit and thoughtful paths to action does facilitate the endeavor to develop successful interventions.

Keywords: Actor behavior · Human Factors in Cyber Security Framework · Reflex · Habitual · Thoughtful

1 Introduction

Aside from the fact that addressing cyber security from the perspective of human factors is largely missing from the wider cyber security dialogue, when it is considered, its scope is often limited: human factors are considered to be static elements that we can mostly tackle by increasing end-user awareness. Cyber security and the secure management of information, however, are not static, and human behavior rarely changes as a result of awareness alone. To enhance cyber security we need more fit for purpose interventions to change human behavior, which take into account the dynamic and transnational context of the specified cyber system, characterized by the continuously changing nature of attacks, types of perpetrators and victims, and goals of the attacks. This paper

D. Nicholson (ed.), *Advances in Human Factors in Cybersecurity*,
Advances in Intelligent Systems and Computing 593, DOI 10.1007/978-3-319-60585-2_23

describes a framework that integrates human (f)actors in the cyber system. This framework integrates three pillars that can influence the state of cyber security humans, ICT/ digital devices and organization. Its most important yield is a better conceptual and operational understanding of interventions aimed at humans to improve cyber security.

2 Current Insights

Human Factors concerns itself with the design of how humans achieve their work-related goals (Prof Dr Jan Maarten Schraagen, personal communication, January 26, 2017). In the context of cyber security, it addresses factors that influence how individuals interact with information security "systems" [1]. Human beings' counterproductive behaviors that violate information security policies are referred to as "Insider Threats" [2]. For example, a policy can state not to share passwords with other employees, as this increases the risk of a malicious actor obtaining this password. Employees who are non-compliant to this policy, by intentionally or unintentionally sharing passwords, are considered a threat originating from inside the organization. According to a study of ENISA [3], this type of risks has the biggest economic impact of all security threats.

Insiders are "employees or others who have (1) access privileges and (2) intimate knowledge of internal organizational processes that may allow them to exploit weaknesses" [4]. Internal human threats fall into three categories [4]. One, employee violations of security policies may be non-volitional, such as accidentally downloading malicious software. Two, employees might exhibit volitional behaviors that are not motivated by malicious intentions. An example is using cloud computing solutions (e.g. dropbox) while being aware this is not allowed within the organization (the so-called "Shadow IT"). Three are insiders who intentionally violate policies for malicious purposes, such as disclosing classified information to the public. This research focusses on the first two types of insiders, who do not have malicious intentions.

There is not much research to be found on actual descriptions of counterproductive human behaviors. Ilfinedo and Akinnuwesi [5] do provide a list of errors made by Canadian and Nigerian employees, such as responding to spam, downloading unauthorized software, leaving the work-related laptop unattended and sharing passwords with others. However, other commonly made errors, like using unknown USB-sticks, using the work-related laptop to charge a smartphone, creating opportunities for eaves dropping, shoulder surfing etc. are missing in this list. CERT [6] applies a grouping based on the types of data breaches caused by incorrect human behaviors, not on the errors themselves. The categories are as follows:

1. Accidental Disclosure—sensitive information is posted publicly on a website, mishandled, or sent to the wrong party via email.
2. Phishing/Social—an outsider's electronic entry is acquired through social engineering (e.g. phishing e-mail attack, planted or unauthorized USB drive) to acquire an insider's credentials or to plan malware to gain access.
3. Physical Records—lost, discarded, or stolen non-electronic records, such as paper documents.

4. Portable Equipment—lost, discarded, or stolen data storage devices, such as a laptop, smart phone, portable memory device, CD, hard drive, or data tape.

In attacks targeted at specific persons or organizations, the human in the loop plays a crucial role [7]. To obtain access to an organizational network (specified cyber system), hackers apply social engineering activities. Human actors are manipulated to share credentials or download malicious software, often via phishing mails. Other methods include getting employees to use infected USB-sticks, “dumpster diving” or simply looking over someone’s shoulder [8]. It is therefore important to consider the human in the loop in cyber security in order to prevent hacking attacks. Technical employees are a crucial part of this process, as they are the internal IT provider, deliver the software to end-users, monitor the network and configure security tools [9].

Humans differ in the degree they are susceptible to social engineering activities or are prone to making errors. There is a rich body of knowledge on the factors underlying these behaviors. Some researchers focus on the characteristics relevant to a specific threat, such as responding to phishing mails [10] or the usage of shadow IT [11]. Others focus on the characteristics influencing compliance to overall information security policy. According to Safa, Von Solms and Furnell [12] the lack of information security awareness, ignorance, negligence, apathy, mischief, and resistance are the root of users’ mistakes. CERT [13] presents 14 factors, divided into three categories: *demographic* (e.g. age), *organizational* (e.g. job pressure) and *personal* (e.g. lack of attention). Oltramari, Cains, Cains and Hoffman [14] have developed an ontology of factors that influence the behaviors of users, IT personnel and attackers. They distinguish the following types of factors: *social cognitive* (e.g. personality traits), *behavioral* (e.g. motivation), *knowledge* (e.g. security training), *internal stress* (e.g. quality of sleep) and *external stress* (e.g. work-pressure). Furthermore, Kreamer, Carayon and Clem [9] have researched the factors leading to design, configuration and implementation errors of IT systems.

The currently available measures to mitigate the insider threats include organizational, technological and awareness interventions [6]. Organizational measures include top-down initiatives such as the implementation of a solid risk management approach, a visible information security policy, an organizational culture in which security is deemed important [6] and security knowledge sharing [12]. Technical measures include access controls, spam filters and firewalls [6]. If the employees interact directly with the security tool, such as responding to browser warnings, usability should not be neglected. If an employee does not understand the message or when a message interferes too much with the task at hand, it may be ignored [15].

Cyber security awareness training aims to correct employees’ behavior by raising their awareness of cyber security. Awareness raising has been found to be partially successful [12]. Success seems to be linked to repeating the training activities on a continual basis, repeatedly measuring the program’s impact [6, 16] and including “soft skills” in the awareness teams for better internal communication [17]. Other inventions which focus on the individual and the corresponding factors are scarce. The current research puts Human Factors interventions into perspective in a framework that better scopes the system which leads to cyber security, thus allowing for better links with areas for improvement.

3 Human (F)actors as an Integral Part of the Cyber System

3.1 A Dynamic System

Optimal cyber security helps ensure sustainable, lucrative and secure organizations and business processes, in which it is possible to make thoughtful choices regarding risk and mitigation measures. Developing and maintaining optimal state of cyber security cannot be considered in light of individual human or device behaviors that function in a vacuum. Neither can it be seen in terms of a snapshot in time. Instead, cyber security should be considered as a state of a defined cyber system which is in flux, in which humans and devices interact through time. Because the system exists in a context that has to face changes over time – and itself changes over time – the system needs to be resilient.

If we see cyber security as a consequence of dynamic interactions in a system, we can consider this system to have a particular security state at any given time. This state, say at t0, can change as a result of an actor carrying out an action, resulting in a different security state at t1. An actor may be a human actor (e.g. an computer user or a software programmer), an organizational actor (e.g. a company or judicial body that makes legal guidelines), or a technological actor, a device (e.g. a bot or other AI system). Each actor in this system has strengths and vulnerabilities. Finding deeper insights on strengths and vulnerabilities of these actors is fruitful as a means to improve the cyber security state of the system.

3.2 Behavior

The actor's actions shift the system in some way that has an effect on the overall security state of the system, and can vary in terms of its effects and impact. An actor's acts may be undesirable and decrease cyber security: e.g. a deliberate cyber-attack or an accidental click on an attachment in a phishing email. (Note that not all behaviors that decrease cyber security are intentionally malicious, but can be the result of accidents or ignorance.) On the other hand, a behavior can also be desirable and increase cyber security: keeping software updates current, building a robust firewall or developing software that contributes to cyber security and does not degrade performance. In order to influence the actions of humans involved in cyber security their behavior needs to be changed. To successfully change that behavior it is important to understand that there are different types of behavior.

Not all behaviors can be influenced. A behavior that we can influence, for example, is performing software updates. An intervention can be devoted to humans performing software updates more frequently. Retrieving information from memory, e.g. for password management, is a behavior that is difficult to influence. It is possible to provide tips for easier password management. However, improve (or change) the humans' natural information retrieval system, is not feasible (at this moment).

3.3 Intervention Opportunities

A system's cyber security state is dependent on actors' **acts**. Influencing that behavior, inhibiting unwanted and stimulating wanted **acts** can make a system more "cyber secure." Behavior is influenced through the use of interventions. In our approach, interventions can focus on influencing human behavior and/or influencing device behavior. With respect to humans, it is wise to inhibit/stimulate behaviors one can reasonably expect from humans. Interventions should be geared to developing (new) ways to adapt to new and emerging circumstances, given what can be expected of human behavior. Generally speaking, there are three pathways by which behavior comes about (see Fig. 1).

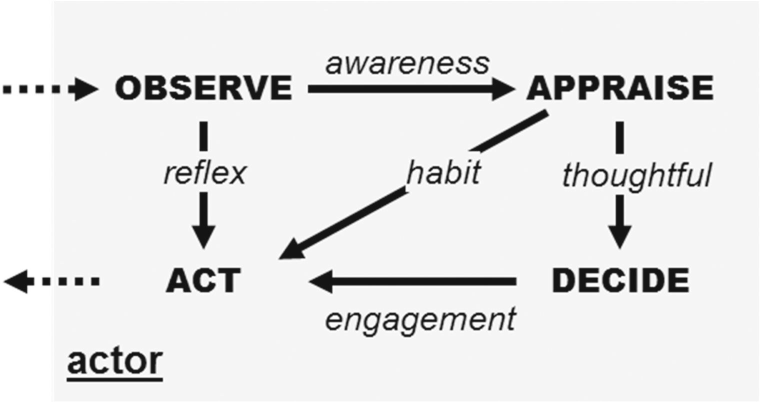


Fig. 1. Three basic pathways by which actor behavior comes about.

Firstly there is ***reflex*** behavior, for example stomping on the brake when the cars in front of you all have highlighted brake lights. ***Reflex*** behavior is the result of cues being perceived by the senses [**OBSERVE**] and being directly carried out [**ACT**] at the command of some of humans' most basic cognitive systems, such as those that detect and protect us from danger.

Secondly there is ***habit*** behavior. These acts need some more cognition; the cues are **APPRAISED** and can be dealt with available heuristics (previously learnt ways of dealing with the cues). For example, agreeing to the conditions of use of newly installed software because the program won't work if you don't agree and you have not experienced negative consequences to agreement in the past.

Finally, there is ***thoughtful*** behavior, which is the result of deliberate consideration and thought processes. For example, reading up on the cyber risks you run, putting in the effort to learn what to do to protect yourself and how and then actually doing it.

Though in principle, increasing cyber secure behavior can occur via all three pathways, we believe that the mechanism most beneficial to change in the context of cyber security is the habitual behavior pathway. By making certain appropriate heuristics salient, the probability of an appropriate response can be increased. An example is the use of the first letters of words in a favorite song line as the characters for passwords.

This heuristic can be learnt through the ***thoughtful*** pathway, but after having applied it a few times can become the salient ***habit*** for generating passwords.

In this light, it is interesting to consider the effectiveness of cyber security awareness campaigns and training aiming to improve cyber security. As discussed in Sect. 2, evidence on the effect of increasing awareness through cyber security awareness trainings is not unequivocal: though there is evidence that it may work [12], there are also considerable findings that it is difficult to use awareness trainings to increase the overall cyber security behavior of employees. Certainly, there is evidence that increasing general awareness is not able to impact this system in the way that is needed to improve cyber security in a sustainable manner.

Thus, to be effective, interventions should be tailored to changing a specific behavior, in a specific context, in a specific way. That is: effective interventions are *fit for purpose*. An intervention must also be *fit for actor*, that is let humans and IT systems do what they are respectively good at, and choose alternative solutions for those things they are not good at. For example, because humans are not good at managing multiple, complex and ever-changing passwords, use two-factor authentication wherever possible. Alternatively, IT systems are not particularly good at complex image recognition, such as distinguishing a door from a window, which is why having users indicate pictures showing one or the other is a good way to tell a computer from a human.

In sum, interventions should be directed at changing the state of a system's level of (a specific aspect of) cyber security, via the influencing of actionable behaviors, thereby contributing to maintaining effective (business) processes. We believe this is most likely to be successful when the focus is on changing habitual behaviors. In the following chapter we will examine how these concepts are integrated into a framework scoping Human Factors in cyber security, which can be used to identify avenues for interventions.

4 Human (F)actors in Cyber Security Framework

Figure 2 shows our Human Factors in Cyber Security Framework. The overall purpose of this framework is to help understand how acts impact the cyber security of the system and to help identify potential intervention leverage points to make the system more cyber secure. This framework is not falsifiable but does help to generate falsifiable hypotheses, which then can be tested.

Inhibiting or stimulating appropriate ACTS is at the core of this improvement and can refer to humans changing their behaviors, technological changes that improve security or changes at the level of the organization. The framework can be used to identify actors in the cyber system and to provide a context for identifying secure and insecure behaviors. This forms the basis for defining behaviors you want to change and the interventions you need to enact that change.

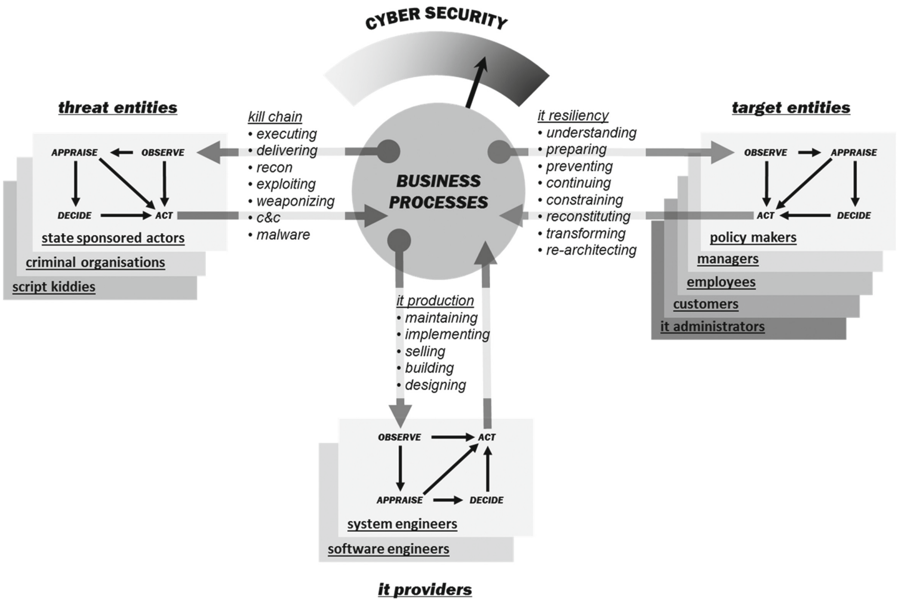


Fig. 2. Human Factors Cyber Security Framework. Note that in our framework threat entities also have a place, however these are not the focus of this paper.

4.1 States

The basic tenet of this framework is that any organization can be considered a system, in which the cyber security, has a particular state vis-à-vis its level of "secureness" (represented by the cyber security meter at the top of the figure). A system can be relatively secure or insecure, or can be in an apprehensive state – an ambiguous state in which the system is unstable and can change in either direction. Any action (behavior) carried out by an actor can influence the secureness of the system's state.

4.2 Actors

In the framework, we define at least three groups of actors: the IT Providers on the bottom, and the target entities on the right and threat entities on the left. These are broadly inclusive groups; other interaction actors can also be included but for the purpose of this paper the depicted entities will suffice.

IT Providers include all those involved in developing digital systems and products, such as architects, software developers and interaction designers.

Target entities are those involved in using digital systems and products, in addition to, for example, system administrators, legal advisors, and managers & directors.

Key is that both IT providers and target entities have distinct and important responsibilities in ensuring an organization's cyber security. Each group performs behaviors that can be positively influenced to increase the overall cyber security of the system.

They can also perform behaviors – either intentionally or unintentionally – that negatively influence the overall cyber security.

Finally, there are **threat entities** who also are actors in the system. For this group the same basic principles hold, but we will not focus our attention on influencing their acts within the scope of this paper.

IT Providers. The actions carried out by IT providers are represented in the model by the stages associated with product life cycles. IT providers design, build, sell, implement and maintain IT systems. At each stage of this process, an IT provider has different possibilities in behavior (related to an action) that affect the overall cyber security of the system. For example, by keeping up to date on new developments in their field and implementing them in their work, or by possessing personality characteristics or by developing skills consistent with those needed to optimally perform their tasks (e.g. pays much attention to detail when configuring a firewall).

A challenge is the product chain, the dependency to other providers. Hardly any company is developing a system including all the individual components by itself. This dependency for components of other providers also creates a dependency to the cyber security of those other providers [18].

Target Entities. Target entities or "Users" refers to both experts and non-experts in the area of cyber security. The actions and corresponding behaviors carried out by this group are represented by the objectives of the Mitre cyber resilience engineering framework (CRE) [19]. The Mitre framework was originally intended to be used "… to meet the challenge of how to evolve architectures, cyber resources, and operational processes to provide cost-effective cyber resiliency." The focus was on system security engineering and business continuity, but did not take explicitly into account the role humans play in cyber resilience. We believe, however, that the objectives of the Mitre framework provide an excellent basis upon which to consider the human's role in cyber security, and a context within which we can identify weak spots and ways to encourage cyber secure behavior.

Consider, for example, the objective "prevent." The CRE framework defines it as "… The Prevent objective is to preclude successful execution of an attack on a set of cyber resources." In the original context of the CRE framework, this takes into consideration cost effectiveness, principles of systems engineering and security controls. An individual human, however, can also carry out specific behaviors to prevent an attack on cyber systems they use (be they personal, public or work-related), such as keeping software up to date, performing regular virus scans and maintaining sound password management. In similar ways, we may extend each objective in the CRE model to apply it to behaviors that a human can carry out in pursuit of that objective on their own personal scale.

As with the IT providers, behaviors can be identified at each level of the Mitre framework, which can affect cyber security: development of poor cyber security policy, poor password or update management, or ignorance of the likelihood and impact of risks. Likewise, the objectives can also help identify strategies to influence behavior such that cyber security improves: training, password requirements, use of biometrics, policy regarding use of open networks, or periodic auditing.

4.3 Behavior Change

It is intuitive to want to make people's behavior result in more security. However, reality shows that increasing awareness often is not sufficient to realize secure behavior. The reality of behavioral change, however, is much more complex. Think of how difficult it is for people to quit smoking, exercise more or eat more healthy despite the abundance of information on the risks associated with unhealthy behavior short-term advantages win. Achieving short-term gratification (e.g., downloading an application), avoiding short-term nuisance (e.g., thinking up a unique password for every new account), coupled with the low perceived likelihood of something going terribly wrong, tend to result in people not taking precautions.

Changing behavior, however, is possible, and is often a question of choosing the right intervention for the job. An intervention to change behavior can directly target the actor (e.g. a training program) or can indirectly affect behavior via technological or organizational solutions. Either way, it is important that the solution fits the problem and any relevant conditions as well as possible: interventions should be both fit for purpose and fit for actor.

Choosing a successful intervention depends on making the desired behavioral change as specific as possible. The list below gives a number of issues to think about:

- Think about behavior in terms of the processes defined in the framework. Consider, for example, if you want IT providers to design a system better versus maintain the system better.
- Target very specific behavior: not better password management, but, more specifically, not sharing passwords with others.
- Target a specific group, rather than simply "people." For example, focus on young people, who may tend to share passwords more than older people.
- Be specific about whether a behavior is to be inhibited (do not open attachments from unknown sources) or encouraged (run regular virus checks on your computer).
- Be explicit on what you expect from the target group, rather than taking for granted the target group is familiar with risks and will take its responsibility.
- Be realistic about what can reasonably be expected from the target group. The (perceived) benefits to the user must outweigh the costs, that is, the user must be willing to improve their behavior. To do this, the user must be able to understand
  - what is expected of them,
  - how to carry out the desired behavior,
  - what is the risk of not performing the desired behavior,
  - how they will benefit, and if this benefit is relevant or desirable.

If these points are not explicit, behavioral change is not likely to occur voluntarily.

In Sum. In all cases, consider which interventions are possible. The ultimate goal is to increase a system's security. The issue is not primarily how to get people to voluntarily change their behavior, rather it is how to choose the best intervention, given the desired goal. As a result, consider how the system benefits from a change, and then consider if that change needs to come from the human in the loop, the technological functionalities of the system, or the organizational context in which the system exists.

5 Some Applications of the Framework

Consider the following situation, targeted individuals can receive phishing mails from some sender. The end-user is perhaps inclined to indiscriminately click on hyperlinks [***reflex***] with unwanted consequences such as instalment of malware. This reflex can be inhibited by having the mail application only allowing plain text reading of the particular mail. The end-user will have a button available to change the format of the mail to html. The act of clicking the read-in-html button not only changes the format but also updates the personal contacts list with a new and trustworthy sender and allows future automatic html reading. This intervention thwarts indiscriminate hyperlink clicking and furthermore raises the readers awareness about the unfamiliarity of the sender. This example illustrates how human behavior can be modified, taking into account the strengths and weaknesses of both types of actors, humans and devices.

A second example is the following. Consider the situation that there is an update of the operating system of your mobile phone patching vulnerabilities. Most devices give some sort of waring that can be clicked away or ignored. To thwart this avoidance behavior [***habit***] we suggest to use an alternative awareness raising intervention. We are suggesting the use of a cracked screen simulation which will allow continuing work but does generate a mild irritation that is harder to ignore. To remove the simulated cracked screen, updating is required. This intervention has the added advantage that coworkers who could glance at the screen are aware of the out of date security state of the device and through the peer pressure sensitivity [***habit***] the user will be more inclined to update the device's security measures.

6 The Way Forward

With this framework in mind we will design interventions that take into account the strengths and weakness of humans and devices and test these in real live systems in order to enhance cyber security. Our take home message is that the system state, cyber security, is dependent on behaviors of humans and devices and that it is worthwhile to identify those human behaviors in context that impede or enhance the state of cyber security. Once these have been identified within the context that these behaviors take place, the significant determinants of those behaviors can be found and allow for hypothesizing and testing of interventions that engage these determinants. The distinction between reflex, habit and thoughtful paths to action does facilitate this endeavor in our experience.

Acknowledgements. We would like to thank the following colleagues: Dianne van Hemert, Helma van den Berg, Roy Mente, Allard Kernkamp, and Tjarda Krabbendam-Hersman.

References

1. Parsons, K., McCornac, A., Butavicus, M., Ferguson, L.: Human factors and information security: individual, culture and security environment. Technical report, DSTO (2010)
2. Marinos, L., Belmonte, A., Rekleitis, E.: ENISA threat landscape. Technical report, ENISA (2016)
3. Tofa, D., Theodoros, N., Darra, E.: The cost of incidents affecting CIIs. Technical report, ENISA (2016)
4. Willisin, R., Warketin, M.: Beyond deterrence: an expanded view of employee computer abuse. MIS Q. **37**(1), 1–20 (2013)
5. Ifinedo, P., Akinnuwesi, B.: Employees' non-malicious, counterproductive computer security behaviors (CCSB) in Nigeria and Canada: an empirical and comparative analysis. In: Proceedings of 2014 IEEE 6th International Conference on Adaptive Science and Technology (ICAST), Lagos, NG (2014)
6. CERT Insider Threat Center: Common sense guide to mitigating insider threats, 5th edn. Technical report, Software Engineering Institute (2016)
7. Krombholz, K., Hobel, H., Huber, M., Weippl, E.: Advanced social engineering attacks. J. Inf. Secur. Appl. **22**, 113–122 (2015)
8. CERT Insider Threat Center: Unintentional insider threats: social engineering. Technical report, Software Engineering Institute (2014)
9. Kreamer, S., Carayon, P., Clem, J.: Human and organizational factors in computer and information security: pathways to vulnerabilities. Comput. Secur. **48**, 509–520 (2009)
10. Sheng, S., Holbrook, M., Kumaraguru, P., Cranor, L., Downs, J: Who falls for phish? A demographic analysis of phishing susceptibility and effectiveness of interventions. In: Proceedings of the 28th International Conference on Human Factors in Computing Systems, pp. 373–382. ACM Press, New York (2010)
11. Kopper, A., Westner, M.: Deriving a framework for causes, consequences, and governance of shadow it from literature. In: Proceedings of MKWI 2016 (2016)
12. Safa, N.S., Von Solms, R., Furnell, S.: Information security policy compliance model in organiz tions. Comput. Secur. **56**, 70–82 (2016)
13. CERT Insider Threat Team: Unintentional insider threats: a foundational study. Technical report, Software Engineering Institute (2013)
14. Oltramari, A., Henshel, D.H., Cains, M., Hoffman., B: Towards a human factors ontology for cyber security. In: Proceedings of the Tenth Conference on Semantic Technology for Intelligence, Defense, and Security, Fairfax, VA, pp 26–33. (2015)
15. Lampson, B.: Privacy and security usable security: how to get it. Commun. ACM **52**, 25–27 (2009)
16. Caldwell, T.: Making security awareness training work. Comput. Fraud Secur. **2016**(6), 8–14 (2016)
17. Rudis, B., Hayden, L., Kretschmer, G., Sasse, A., Becker, A., Homer, J.: Security awareness report. Technical report, SANS (2016)
18. Cyber Security Assessment Netherlands (2016). https://www.ncsc.nl/english/current-topics/Cyber+Security+Assessment+Netherlands. Accessed 8 Mar 2017
19. Bodua, D.J., Graubart, R.: Cyber resiliency engineering framework. Technical report, Mitre Corporation (2011)

Privacy and Cultural Factors in Cybersecurity

Preserving Dignity, Maintaining Security and Acting Ethically

Scott Cadzow(✉)

Cadzow Communications Consulting Ltd., 10 Yewlands, Sawbridgeworth, CM21 9NP, UK
scott@cadzow.com

Abstract. Humans design, operate and are the net beneficiaries of most systems. However humans are fallible and make mistakes. At the same time humans are adaptable and resourceful in both designing systems and correcting them when they go wrong. In contrast machines have in the main been designed to follow rules and are often constrained to produce the same output for the same input over and over again. Ethical decisions require that different outputs arise from apparently identical appearing inputs as the wider context for the decision has changed. Humans make ethical decisions almost automatically but as we move towards an increasingly machine led society those aspects of dignity, ethics and security which are managed by humans will be addressed by machines. The aim of this paper is to give an overview of the state of the art in security standardization in machine to machine and IoT systems, for the use cases of eHealth and autonomous transport systems, in order to outline the new ethics and security challenges of the machine led society. This will consider progress being made in standards towards the ideal of each of a Secure and Privacy Preserving Turing Machine and of an Ethical Turing Machine.

Keywords: Human factors · Security · Ethics · Machine ethics

1 Introduction

The purpose of this paper, and its accompanying presentation material, is to open a debate around the developing paradigm of "ethical by design". The leading paradigms in high level security design are those of "Secure by default" and "Privacy by design". It is not suggested that either of these paradigms is complete and that every product is both secure by default and privacy protecting by design, however even when privacy is protected and security is assured the need for systems to act ethically and to treat their affected users with dignity needs to be assured too. The role of ethics - doing the right thing - in design is not yet clear as it is also not clear in real life. However as more and more decision making is moved into the machine world the need for machines and systems of machines to make the right decision is going to arise more and more. The consideration of dignity is perhaps even harder to quantify but again in machines interacting with humans there is often a need to treat the recipient with a certain degree of dignity, and furthermore to let the human actor to hold their dignity intact.

D. Nicholson (ed.), *Advances in Human Factors in Cybersecurity*,
Advances in Intelligent Systems and Computing 593, DOI 10.1007/978-3-319-60585-2_24

In looking to use cases there are two very obvious areas where machine ethics will be critical. In the domain of Intelligent Transport Systems (ITS) the operation of autonomous vehicles will be increasingly divorced from human control, even if the law demands that a licensed driver has to take control in an emergency, in the short to medium term the likelihood is that an autonomous enabled vehicle will act autonomously for the vast majority of its journeys, thus at the point when a crash is inevitable the vehicle has to be able to react in a way that minimizes injury to both the occupants and to anyone or anything in the local area. There is no rationale for the vehicle to disavow itself of all responsibility and pass control to the local human - almost inevitably this will be too late for the human passenger to become useful as a driver.

The second critical domain is that of health - the classical source of the Hippocratic Oath and its modern interpretation in the World Medical Association International Code of Medical Ethics[1] for which the extract for admission into the profession is given below:

- AT THE TIME OF BEING ADMITTED AS A MEMBER OF THE MEDICAL PROFESSION:
  - I SOLEMNLY PLEDGE to consecrate my life to the service of humanity;
  - I WILL GIVE to my teachers the respect and gratitude that is their due;
  - I WILL PRACTISE my profession with conscience and dignity;
  - THE HEALTH OF MY PATIENT will be my first consideration;
  - I WILL RESPECT the secrets that are confided in me, even after the patient has died;
  - I WILL MAINTAIN by all the means in my power, the honour and the noble traditions of the medical profession;
  - MY COLLEAGUES will be my sisters and brothers;
  - I WILL NOT PERMIT considerations of age, disease or disability, creed, ethnic origin, gender, nationality, political affiliation, race, sexual orientation, social standing or any other factor to intervene between my duty and my patient;
  - I WILL MAINTAIN the utmost respect for human life;
  - I WILL NOT USE my medical knowledge to violate human rights and civil liberties, even under threat;
  - I MAKE THESE PROMISES solemnly, freely and upon my honour.

So if we choose to consider a machine as a member of the medical profession where do we place the ethical liability? The underlying concern is when machines act in the functionality of tele-diagnosis and tele-intervention the ethical issues surrounding do no harm are going to be very real.

Technically the need to manage ethics as well as the base function, of transport in autonomous vehicles and the supporting ITS infrastructure, and of health or wellness in the eHealth domain, becomes more and more important.

How can we program ethics into machines though? This is not just an AI program, or not just a Universal Turing Machine, but an Ethical Universal Turing Machine. The starting point is clarity about interoperability to reduce the risk of being in ethically difficult domains, and then to apply things like game theory to the results.

[1] http://ethics.iit.edu/ecodes/node/4233.

The concept of dignity is closely related to moral, ethical and legal behaviors but in the context of this paper is considered in the way in which machines have to react and interact with humans.

The role of standards in this endeavor should not be understated. As engineers and scientists we need to respect issues such as scientific method, repeatability, ethical behavior and presentation of results, and we need to be as objective as possible - presenting facts and evidence that support any claim. This paper will assert that standards, when used correctly, underpin scientific method and can be used to give greater assurance to users that a product will not be a liability with regards to security, dignity or ethics.

2 Human Fallibility

From the assertion that humans design, operate and are the net beneficiaries of most systems we can also assert that humans are the net losers when systems go wrong. However progress in system design is such that the code we execute is mostly created by a higher layer machine. The complexity of most modern systems is such that it is close to impossible to determine which line of code is at fault if something does go wrong. Taking a car as an example it is estimated that a modern, fairly sophisticated car, will contain around 100 million lines of code[2]. When we finally map the human genome and use it for health processing it is estimated to contain the equivalent of 3.3 trillion lines of code. Debugging a system as large as that is bound to be difficult and if we compare for arguments sake to a suite of libraries each containing 1 million lines of code each of which interacts with at least 100 other libraries then the problem of identifying where something is wrong is going to be difficult. Quite simply we are very lucky that we work at all and even if we can see every line of code (equivalent) in the human genome that doesn't actually tell us anything about the content of memory, experience and behavior. However on top of this our societal systems are made up of the interaction of many people and their environments and decoding that to determine a simple predictive model of how things will work in any given any situation. The complexity means we need to learn how to behave and that will apply to complex machines just as much as to humans. In normal human society our education by family, by peers, by our society and in formal places of learning serve in large part to define our ethical framework - what we deem to be the correct choice of action when dilemmas are put before us.

In practice it is unlikely that our learning will anticipate all possible circumstances and we “wing it” based on past, similar, experiences, and we often learn by failing a little and compensating using simple feedback mechanisms. However what is the consequence of the programmed intelligence of the system making a mistake? Are ethical choices right or wrong? Is affording dignity a programmable trait? If the failures we need to learn through are in the security systems then trust in the system can disappear.

Given that humans are fallible and make mistakes then we can design systems and processes to cope with errors in such a way that they learn without leading to an end

[2] http://www.informationisbeautiful.net/visualizations/million-lines-of-code/.

game of catastrophe. One of the roles of security engineers is to recognize this fallibility and to be up front about what can and cannot be done with respect to countering threats that limits the damage of such fallibility. In doing this it is essential to also recognize that humans are adaptable and resourceful in both designing systems and correcting them when they go wrong. These characteristics mean that humans can be both the strongest and the weakest link in system security. It also means that there is an incentive to manage the human element in systems such that those systems work well (functionality matches the requirement), efficiently (don't overuse resources), safely and securely. Thus human centric design, even for mostly machine based systems, is essential.

In recognizing that it is the human factor that generally identifies risk and maps out the functionality of a system - its goal in other words - it is clear that this strength can be undermined by fallibility.

3 Security Controls? Security Awareness?

The set of Critical Security Controls (CSC) published by the SANS Institute [1] (see list below) are proposed as key to understanding the provision of security to systems, however selling the benefits of such controls, and the threat modelling that underpins many security programmes, including Common Criteria [2] and ETSI's Threat Vulnerability Risk Analysis (TVRA) [3] method to the end user is difficult and more often appears to induce fear rather than contentment that the experts understand their work.

Misapplication of the Critical Security Controls by human error, malicious or accidental, will lead to system vulnerabilities. The importance of such controls has been widely recognized and they can be found, either duplicated or adopted and adapted for sector specific spaces, in ETSI, ISO and in a number of industry best practice guides.

The first of the CSC requires that organizations make an Inventory of Authorized and Unauthorized Devices. On the face of it this is relatively simple - identify the devices you want to authorize and, those you don't. However this introduces the Rumsfeld,[3] conundrum "… there are known knowns… there are known unknowns… there are also unknown unknowns…", it is not possible to identify everything. The second of the CSC to prepare an Inventory of Authorized and Unauthorized Software also has the Rumsfeld conundrum at the root of its problem.

The more flexible a device is the more likely it is to be attacked by exploiting its flexibility. We can also assert that the less flexible a device is it is less able to react to a threat by allowing itself to be modified.

The use of the Johari Window [4] to identify issues is of interest here (using the phrasing of Rumsfeld) in Table 1.

[3] "Reports that say that something hasn't happened are always interesting to me, because as we know, there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns – the ones we don't know we don't know. And if one looks throughout the history of our country and other free countries, it is the latter category that tend to be the difficult ones." Attributed to Donald Rumsfeld on 12-February-2002.

Table 1. Security concerns in Johari window style with Rumsfeld phrasing.

	Known to self	Not known to self
Known to others	Known knowns BOX 1	Unknown knowns BOX 2
Not known to others	Known unknowns BOX 3	Unknown unknowns BOX 4

The human problem is that the final window, the unknown unknowns, is the one that gives rise to most fear but it is the one that is not reasonable (see movie plot threats below). The target of security designers is to maximize the size of box 1 and to minimize the relative size of each of box 2 and box 3. In so doing the scope for box 4 to be of unrestrained size is hopefully minimized (it can never be of zero size).

We can consider the effect of each "box" on the spread of fear:

- **BOX 1:** Knowledge of an attack is public and resources can be brought to bear to counter the fear by determining an effective countermeasure.
- **BOX 2:** The outside world is aware of a vulnerability in your system and will distrust any claim you make if you do not address this blind spot.
- **BOX 3:** The outside world is unaware of your knowledge and cannot make a reasonable assessment of the impact of any attack in this domain and the countermeasures applied to counter it.
- **BOX 4:** The stuff you can do nothing about as as far as you know nothing exists here.

The obvious challenge is thus to bring tools such as the 20 controls from SANS to bear to maximize box 1 at the same time as using education and dissemination to minimize the size of boxes 2 and 3. Box 3 is characteristic of the old, mostly discredited, approach of security by secrecy, whereas Box 1 is characteristic of the open dissemination and collaborative approach of the world of open standards and open source development. Box 1 approaches are not guarantees of never having a security problem. Generally speaking we expect problems migrate from box 4 to boxes 2 and 3 before reaching box 1 and, hopefully, mitigation.

In the security domain we can achieve our goals both technically and procedurally. This also has to be backed up by a series of non-system deterrents that may include the criminalization under law of the attack and a sufficient judiciary penalty (e.g. interment, financial penalty) with adequate law enforcement resources to capture and prosecute the perpetrator. This also requires proper identification of the perpetrator as traditionally security is considered as attacked by *threat agents,* entities that adversely act on the system. However in many cases there is a need distinguish between the threat source and the threat actor even if the end result in terms of technical countermeasures will be much the same, although some aspects of policy and access to non-system deterrents will differ. A *threat source* is a person or organization that desires to breach security and ultimately will benefit from a compromise in some way (e.g. nation state, criminal organization, activist) and who is in a position to recruit, influence or coerce a threat actor to mount an attack on their behalf. A *Threat Actor* is a person, or group of persons, who actually performs the attack (e.g. hackers, script kiddy, insider (e.g. employee), physical intruders). In using botnets of course the coerced actor is a machine and its recruiter may itself be machine. This requires a great deal of work to eliminate the innocent threat actor and to determine the threat source.

The technical domain of security is often described in terms of the CIA paradigm (Confidentiality Integrity Availability) wherein security capabilities are selected from the CIA paradigm to counter risk to the system from a number of forms of cyber attack. The common model is to consider security in broad terms as determination of the triplet {threat, security-dimension, countermeasure} leading to a triple such as {interception, confidentiality, encryption} being formed. The threat in this example being interception which risks the confidentiality of communication, and to which the recommended countermeasure (protection measure) is encryption.

Application of the CIA paradigm works for Box 1 problems and will work reasonably well to mitigate problems from Boxes 2 and 3. One of the big problems in the real world, particularly for ethics is that many of the problems are either in Box 4 or at the limits of Boxes 2 and 3 - ethics problems are almost never in box 1.

The very broad view is thus that security functions are there to protect user content from eavesdropping (using encryption as the known counter to eavesdropping) and networks from fraud (authentication and key management services as the known counters to masquerade and manipulation attacks). What security standards cannot do is give a guarantee of safety, or give assurance of the more ephemeral definitions of security that dwell on human emotional responses to being free from harm. Technical security measures give hard and fast assurance that, for example, the contents of an encrypted file cannot, ever, be seen by somebody without the key to decrypt it. So just as you don't lock your house then hang the key next to the door in open view you have to take precautions to prevent the key getting into the wrong hands. The French mathematician Kerchoff has stated "A cryptosystem should be secure even if everything about the system, except the key, is public knowledge". In very crude terms the mathematics of security, cryptography, provides us with a complicated set of locks and just as in choosing where to lock up a building or a car we need to apply locks to a technical system with the same degree of care. Quite simply we don't need to bother installing a lock on door if we have an open window next to it - the attacker will ignore the locked door and enter the house through the open window. Similarly for a cyber system if crypto locks are put in the wrong place the attacker will bypass them.

It may be argued that common sense has to apply in security planning but the problem is that often common sense is inhibited by unrealistic threats such as the movie plot scenarios discussed below.

4 Movie Plot Threats

Bruce Schneier has defined movie plot threats as "*... a scary-threat story that would make a great movie, but is much too specific to build security policies around*"[4] and rather unfortunately a lot of the real world security has been in response to exactly these

[4] https://www.schneier.com/blog/archives/2014/04/seventh_movie-p.html.

kind of threats. Why? The un-researched and unproven answer is that movie plots are easy to grasp and they tend to be wrapped up for the good at the end.

The practical concerns regarding security and the threats they involve is that they are somewhat insidious, like dripping water they build up over time to radically change the landscape of our environment.

Taking Schneier's premise that our imaginations run wild with detailed and specific threats it is clear that if a story exists that anthrax is being spread from crop dusters over a city, or that terrorists are contaminating the milk supply or any other part of the food chain, that action has to be taken to ground all crop dusters, or to destroy all the milk. As we can make psychological sense of such stories and extend them by a little application of imagination it is possible to see shoes as threats, or liquids as threats. So whilst Richard Reid[5] was not successful and there is no evidence to suggest that a group of terrorists were planning to mix a liquid explosive from "innocent" bottles of liquid, the impact is that due to the advertised concerns the policy response is to address the public fears. Thus we have shoe inspections and restrictions on carrying liquids onto planes. This form of movie theatre scenario and the response ultimately diverts funds and expertise from identifying the root of many of the issues.

Again taking Schneier's premise the problem with movie plot scenarios is that fashions change over time and if security policy is movie plot driven then it becomes a fashion item. The vast bulk of security protection requires a great deal of intelligence gathering, detail analysis of the data and the proposal of targeted counter measures. Very simply by reacting to movie plots the real societal threats are at risk of being ignored through misdirection.

Movie plot derived security policy only works when the movie plot becomes real. If we built out bus network on the assumptions behind Speed we'd need to build bus stops for ingress and egress that are essentially moving pavements that don't allow for the bus to ever slow down, and we'd need to be able to refuel and change drives also without slowing the bus. It'd be a massive waste of money and effort if the attackers did a Speed scenario on the tram or train network or didn't attack at all.

A real problem is that for those making security policy, and for those implementing the countermeasures, they will always be judged in hindsight. If the next attack targets the connected vehicle through the V2I network, we'll demand to know why more wasn't done to protect the connected vehicle. If it targets schoolchildren by attacking the exam results data, we'll demand to know why that threat was ignored. The answer "we didn't know…" or "we hadn't considered this…" is not acceptable.

The attractiveness of movie plot scenarios is probably hard to ignore - they give a focus to both the threat and the countermeasures. In addition we need to consider the role of Chinese Whispers[6] in extending a simple story over time.

We can imagine dangers of believing the end point of a Chinese Whispers game:

- Novocomstat has missile launch capability
- Novocomstat has launched a missile

[5] https://en.wikipedia.org/wiki/Richard_Reid => The "shoe bomber".

[6] https://en.wikipedia.org/wiki/Chinese_whispers => A parlor game that passes a message round introducing subtle changes in meaning with each re-telling.

- Novocomstat has launched a bio weapon
- Novocomstat has launched a bio weapon at Neighbourstat
- Neighbourstat is under attack
- Neighbourstat is an ally and we need to defend them
- We're at war with Novocomstat because they've attacked with the nuclear option

As security engineers the guideline is to never react without proof. Quite simply acting on the first of these Chinese Whispers is unwarranted, and acting on the 6th is unwarranted unless all the prior statements have been rigorously verified, quantified and assessed. The various risk management and analysis approaches that exist (there are many) all come together by quantifying the impact of an attack and its likelihood. In recent work in this field in ETSI the role of motivation as well as capability in assessing risk has been re-assessed and now added to the method [3]. The aim in understanding where to apply countermeasures to perceived risk requires analysis. That analysis requires expertise and knowledge to perform. In the approach defined by ETSI in TS 102 165-1 [3] this means being able to quantify many aspects of carrying out a technical threat including the time required, the knowledge of the system required, the access to the system, the nature of the attack tools and so forth.

What movie plot scenarios do allow though is the provision of a playground to examine ethical scenarios. So whilst such scenarios ought to be dismissed from setting policy they ought to be embraced as learning tools.

5 Intelligent Gaming and Game Theory in Machine Ethics

Ethical decisions are often both time critical and time variant. What is "right" in one context may be "wrong" in another context, where context may include the players, the time, the location or any other variable. An ethical problem often needs solved at the time it arises - there can be no delay without the problem resolving itself or any solution being invalid. Thus the trolley problem is one oft cited example (see Fig. 1).

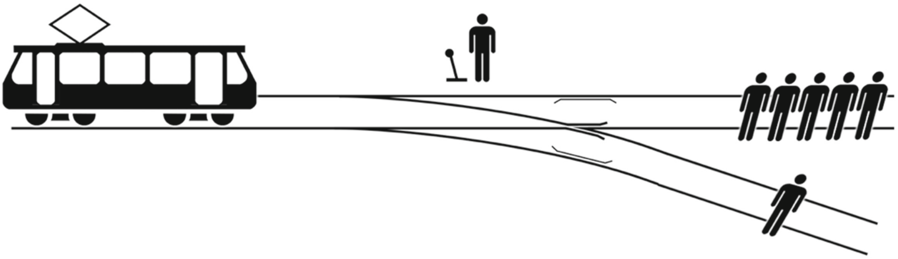


Fig. 1. The Trolley Dilemma (from McGeddon - Own work, CC BY-SA 4.0, https://commons.wikimedia.org/w/index.php?curid=52237245)

The problem is often phrased as a runaway train carriage at speed whilst ahead, on the track, there are five people tied up and unable to move. The train is headed straight for them. You are standing some distance off in the train yard, next to a lever controlled junction. If you pull this lever, the train will switch to a different set of tracks. However, you notice that there is one person on the side track. You have two options:

1. Do nothing, and the trolley kills the five people on the main track.
2. Pull the lever, diverting the trolley onto the side track where it will kill one person.

Which is the most ethical choice? If the choice is to be made by a machine how is the machine programmed? There is no correct choice of course and that is a problem of ethics - the right answer is almost wholly contextual and the deciding actor has limited perspective so can only see the 5 versus 1 conundrum. It is kind of assumed that all alternative avenues have either been tried and failed or are simply not available. How do you win and kill nobody? You can't without changing the problem and modifying the ethical argument.

An alternative view is that presented by the classical prisoner's dilemma but for the general case of co-operation. In moving away from the binary choice in the trolley dilemma the number of actors involved can be expanded such that actors can collude to define the ethically preferable outcome. In the trolley dilemma for example can the trolley itself become involved in the decision? Can it take actions that alter the set of possible outcomes? If we take the prisoner's dilemma where the temptation payoff (T) is greater than the Reward payoff (R) which is greater than the Sucker payoff (S) and which is greater than the Punishment payoff (P) we want to be able to get the actors to work in such a way that with or without collusion they always choose to receive R on the assumption that mutually beneficial strategies are better over the long term.

Game theory is suggested as one way in which ethical issues can be considered. However in order to make such tools work effectively there are a number of pre-conditions that need to be met. The assertion of this paper is that many of the pre-conditions require a commitment to standards to assure interoperability and this is explored more below.

6 The Role of Standards

Standards are peer reviewed and have a primary role in giving assurance of interoperability. Opening up the threat model and the threats you anticipate, moving everything you can into box 1, in a format that is readily exchangeable and understandable is key. The corollary of the above is that if we do not embrace a standards view we cannot share knowledge effectively and that means we grow our box 2, 3, 4 visions of the world and with lack of knowledge of what is going on the ability of fear to grow and unfounded movie plot threats to appear real gets ever larger.

Let us take health as a use case for the role of standards in achieving interoperability. When a patient presents with a problem the diagnostic tools and methods, the means to describe the outcome of the diagnosis, the resulting treatment and so on, have to be sharable with the wider health system. This core requirement arises from acceptance that more than one health professional will be involved. If this is true they need to discuss the patient, they need to do that in confidence, and they need to be accountable for their actions which need to be recorded. Some diseases are "notifiable" and, again, to meet the requirement records have to be kept and shared. When travelling a person may enter a country with an endemic health issue (malaria say) and require immunization or medication before, during and following the visit. Sharing knowledge of the local environment and any endemic health issues requires that the reporting and receiving entities share understanding.

Shared understanding and the sharing of data necessary to achieve it is the essence of interoperability. A unified set of interoperability requirements addresses syntax, semantics, base language, and the fairly obvious areas of mechanical, electrical and radio interoperability.

Syntax derives from the Greek word meaning ordering and arrangement. The sentence structure of subject-verb-object is a simple example of syntax, and generally in formal language syntax is the set of rules that allows a well formed expression to be formed from a fundamental set of symbols. In computing science syntax refers to the normative structure of data. In order to achieve syntactic interoperability there has to be a shared understanding of the symbol set and of the ordering of symbols. In any language the dictionary of symbols is restricted, thus in general a verb should not be misconstrued as a noun for example (although there are particularly glaring examples of misuse that have become normal use, e.g. the use of "medal" as a verb wherein the conventional text "He won a medal" has now been abused as "He medalled"). In the context of eHealth standardization a formally defined message transfer syntax should be considered as the baseline for interoperability.

Syntax cannot convey meaning and this is where semantics is introduced. Semantics derives meaning from syntactically correct statements. Semantic understanding itself is dependent on both pragmatics and context. Thus a statement such as "Patient-X has a heart-rate of 150 bpm" may be syntactically correct but has no practical role without understanding the context. Thus a heart-rate of 150 bpm for a 50-year old male riding a bike at 15 km/h up a 10% hill is probably not a health concern, but the same value when the same 50 year old male is at rest (and has been at rest for 60 min) is very likely a serious health concern. There are a number of ways of exchanging semantic information although the success is dependent on structuring data to optimize the availability of semantic content and the transfer of contextual knowledge (although the transfer of pragmatics is less clear).

Underpinning the requirements for both syntactic and semantic interoperability is the further requirement of a common language. From the eHealth world it has become clear that in spite of a number of European agreements on implementation of a digital plan for Europe in which the early creation of 'e-health' was eagerly expected the uneven development of the digital infrastructure has in practice made for differing levels of initiative and success across the member states. These led to a confusing vocabulary of terms and definitions used by e-health actors and politicians alike. The meaning of the term e-health has been confused with 'tele-health' which in turn is confused with 'm-health;' 'Telemedicine,' a term widely used in the USA has been rejected in Europe in favor of 'telehealth.' There is general agreement that for these terms to be effective we need to redefine them in their practical context. Without an agreed glossary of terms, it will be hard to improve semantic interoperability - a corner stone for the effective building of e-health systems. The vocabulary is not extensive but at present it fails to address the need for clarity in exchange of information in the provision of medical services.

Standards therefore enable and assert interoperability on the understanding that:

$$\text{Interoperability} = \mathit{Semantics} \cup \mathit{Syntax} \cup \mathit{Language} \cup \mathit{Mechanics} \tag{1}$$

Quite simply if any of the elements is missing then interoperability cannot be guaranteed. However we do tend to layer standards on top of one another, and alongside each other, and wind them through each other. The end result unfortunately can confuse almost as much as enlighten and unfortunately the solution of developing another standard to declutter the mess often ends up with just another standard in the mess.

In the security domain understanding that we need interoperability is considered the default but simply achieving interoperability is a necessary but insufficient metric for making any claim for security. As has been noted above the technical domain of security is often described in terms of the CIA paradigm (Confidentiality Integrity Availability) wherein security capabilities are selected from the CIA paradigm to counter risk to the system from a number of forms of cyber attack. The common model is to consider security in broad terms as determination of the triplet {threat, security-dimension, countermeasure} leading to a triple such as {interception, confidentiality, encryption} being formed. The threat in this example being interception which risks the confidentiality of communication, and to which the recommended countermeasure (protection measure) is encryption.

The very broad view is thus that security functions are there to protect user content from eavesdropping (using encryption) and networks from fraud (authentication and key management services to prevent masquerade and manipulation attacks). Technical security, particularly cryptographic security has on occasion climbed the ivory tower away from its core business of making everyday things simply secure.

7 Conclusions

As stated in Sect. 6 of this paper the approach to better understanding of the ethical dimensions in IoT, M2M, ITS and eHealth is wider acceptance of shared knowledge, shared understanding and willingness to educate each other about what we know and what we may not know. The role of standards in giving assurance of interoperability as the key to a solution where more than one stakeholder is involved is difficult to argue against. The nature of the standard is unimportant - it simply has to be accepted by the stakeholders. If the stakeholders are global and largely unknown then an internationally accepted standard is most likely to be the way forward. If, however, the stakeholders are members of a small local team the standard could be as simple as a set of guidance notes maintained on a shared file.

Spreading of fear through a combination of movie plot threats and Chinese Whispers is an inevitable consequence of human curiosity and imagination. But as long as we use the movie plot idea and restrict it to the learning domain we can explore ethical problems. What we should not do is transfer movie plots, and ethical games, into security and societal policy.

Standards are at the root of sharing a common syntactical and semantic understanding of our world. This is as true for security as it is for any other domain and has to be embraced.

Acknowledgments. Contributions made by the author in development of this paper have in part been supported by EU projects i-locate (grant number 621040), SUNSHINE (grant number 325161) and UNCAP (grant number 643555).

References

1. CIS Critical Security Controls - Version 6.1192. http://www.cisecurity.org/critical-controls/
2. The Common Criteria. www.commoncriteriaportal.org
3. ETSI TS 102 165-1: CYBER; methods and protocols; Part 1: method and proforma for Threat, Vulnerability, Risk Analysis (TVRA). https://portal.etsi.org/webapp/WorkProgram/SimpleSearch/QueryForm.asp
4. Luft, J., Ingham, H.: The Johari window, a graphic model of interpersonal awareness. In: Proceedings of the Western Training Laboratory in Group Development, Los Angeles (1955)

Human Factors in Information Security Culture: A Literature Review

Henry W. Glaspie(✉) and Waldemar Karwowski

Department of Industrial Engineering and Management Systems,
University of Central Florida, Orlando, FL 32816-2993, USA
hank@knights.ucf.edu

Abstract. Information security programs are instituted by organizations to provide guidance to their users who handle their data and systems. The main goal of these programs is to foster a positive information security culture within the organization. In this study, we present a literature review on information security culture by outlining the factors that contribute to the security culture of an organization and developing a framework from the synthesized research. The findings in this review can be used to further research in information security culture and can help organizations develop and improve their information security programs.

Keywords: Information security culture · Security programs · Incentives and deterrence · Training and awareness

1 Introduction

In today's environment, organizations collect, transmit, and use data to perform a variety of business-related functions. These functions affect communications, finance, commerce, higher education, and government. Their proliferation of data makes them fertile targets for cyber criminals. The cyber criminals (or hackers) could be working independently, for other organizations, or nation-state actors [1]. The threat of cyber-attack has resulted in large investments in secure data storage, networks, and cyber-defense systems [2]. In spite of these investments, cyber-crime is still very prevalent with massive breaches being reported almost daily in the news media. Over the past few years, cyber-crime and information security incidents have seen an exponential annual increase. According the 2015 IBM Cyber Security Intelligence Index, there were nearly twice as many cyber security incidents than in 2014 [3].

Despite of the significant budgetary expenditures in tools and systems to fight cyber-attacks, there is very little comparative investment in human factors and security culture. Information security is not solely a technical issue. An organization's investment in just technology does not eliminate the many security challenges. Among cyber security practitioners, it is well known that humans are the weak link in information security [4] and many human factors affect information security management [5]. Information system user's undesirable behavior is a direct reflection of the culture of information security in the organization [6]. The aforementioned IBM report states that 9 out of 10 information security incidents were caused by some sort of human error. This is a 10%

D. Nicholson (ed.), *Advances in Human Factors in Cybersecurity*,
Advances in Intelligent Systems and Computing 593, DOI 10.1007/978-3-319-60585-2_25

increase in human involvement reported over a two year span [7]. In spite of this, organizations have still continued to focus their cybersecurity investments in the area of technology infrastructure [8]. There is an obvious gap in information security among organizations that only consider technology aspects of security and forego the human aspects.

Human errors can be the result of negligence, accident, or deliberate action. Because of this, organizations need to invest in building an information security culture that is inclusive of all personnel and leadership [9]. Organizations that have a security culture minimize the risk posed to information privacy [10]. Prevalent research highlights that a positive information security culture can increase security policy compliance, strengthen the overall information security posture, and reduce the financial loss due to security breaches.

This paper provides a review of the factors which affect the human side of information security and promote a desirable security culture. Information security culture has been found to have a positive effect on employee adherence to policy and security behavior [11, 12]. Alhogail and Mirza (2014) define information security culture as the "collection of perceptions, attitudes, values, assumptions, and knowledge that guide the human interaction with information assets in an organization with the aim of influencing employees' security behavior to preserve information security'' [13].

This study utilized 50 pertinent publications in the field of human behavior information security. The query for applicable articles was achieved by conducting a structured search from the years 2010 onward in academic databases such as Compendex, ProQuest, Web of Science, and EBSCO. Additional searches were made in relevant information systems and information security journals and conferences. Prior works have focused on solely on behavior [14] or human behavioral theories [15]. None of these studies have specifically analyzed the factors that influence the overall information security culture and the challenges that leadership faces in cultivating the culture. Additionally, a conceptual model is created from the relevant literature to highlight each factor's role and its contributions.

2 Information Security Culture Model

In synthesizing the research, a conceptual model of the components that impact information security culture was developed. The model [16] includes the following factors of: (1) information security policy; (2) deterrence and incentives; (3) attitudes and involvement; (4) training and awareness, and (5) management support (see Fig. 1).

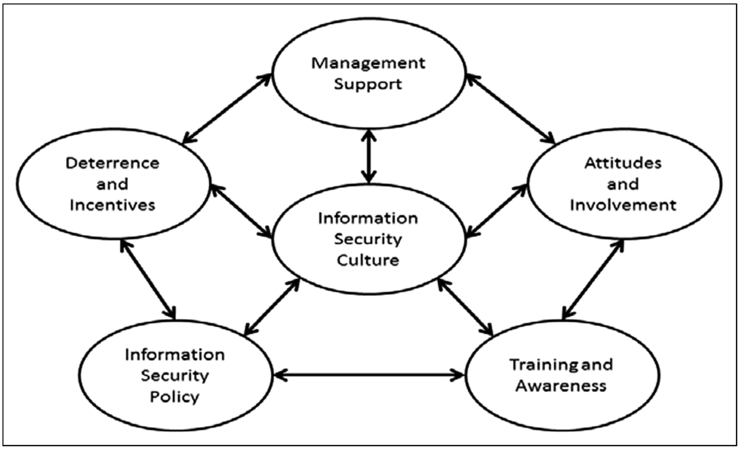

Management Support
Deterrence and Incentives
Attitudes and Involvement
Information Security Culture
Information Security Policy
Training and Awareness

Fig. 1. Model of the factors that influence and cultivate an information security culture

3 Information Security Policy

As the focus of information security measures shift from technology to human factors, many authors have investigated the influence and effect that information security policies have on the overall information security culture. Most organizations are required to have some sort of information security policy in place in the organization. This is usually mandated by a regulatory authority (federal, state, local, accreditation, or auditor) as a condition of certification. The policies set mandatory guidelines to influence favorable organizational behavior when using systems or working with data [17]. All information security policies should comply with and emphasize the organization's objectives [18]. With this in mind, security policies are created to communicate security protocols, assign clear roles and responsibilities, and provide employees with guidance for acceptable usage to ensure security behaviors during the performance of their jobs [19]. The roles, responsibilities, and guidelines also give clarity to who should be contacted and how information security incidents are handled [18]. When policies are complex, ambiguous, complicated, vague, or difficult for users to understand, attitudes towards compliance are negatively affected. Organizations should make their policies as understandable, relevant, and accessible as possible to all employees [20].

The research by Haeussinger and Kranz (2013) shows that the creation and promotion information security policies is the foundational element of any information security management program and has a positive influence on employee awareness [21]. Research by Safa et al. (2015) also notes that an organization's information security policy has an enormous influence on security conscious care behavior [2]. Choi, Levy and Hovav (2013) examined how user awareness of security policies contributed to their initiative skill and action skill, and computer skill related to security [22]. The results showed that awareness of security policies showed significant effects on "action skill" or compliance with policies and procedures.

Having a security policy alone does not ensure employee compliance. An organization must have a comprehensive information security policy to achieve a meaningful impact on the information security culture. A comprehensive policy assimilates technology systems security and security culture [4]. This is supported in research by Chen, Ramamurthy and Wen (2015) showing that awareness alone contributes little to the organization security culture [23]. Hu et al. (2012) noted that a stronger positive attitude towards information security compliance leads an increased intention towards policy compliance [24]. Organizations that actively encourage their employees to comply with their policies see an increase in overall information security [12]. Therefore, management must make sure that employees fully understand the policies and favorably perceive them.

4 Deterrence and Incentives

Most information security policies contain language that informs the applicable parties about the penalties of noncompliance. This is the formal deterrence against negative employee behavior. In many organizations this punishment could range from remediation to termination. Several studies have discussed the link between an employee's willingness towards security policy compliance and their perceived benefit or cost of compliance versus noncompliance. Results from a study by Parsons et al. (2015), reveal that organizations with higher severity in punishments for noncompliance were more likely to have a healthy organizational information security culture [25]. Without clear and consistent consequences for noncompliance, users are likely to demonstrate risky or noncompliant behavior.

Moreover, it has been shown that these perceptions are based on expected outcomes or assessment of consequences. The employees' beliefs about benefit of compliance versus the cost of noncompliance impact their intentions to comply [19]. But differing opinions have been presented on how to motivate employees to adhere to the organization's security policies. Chen, Ramamurthy, and Wen (2012), showed that both the severity level of punishment and the level of reward significantly affect compliance intention [26]. This is reinforced when there is a high level of certainty that the reward or punishment will be enforced. Also, the impact of punishment on intention to comply is greater when there is low reward.

Aside from severity of punishment and formal sanctions for noncompliance, there have been studies on the effect that informal sanctions have on compliance intention. D'Arcy and Devaraj (2012) reported on the informal sanctions, or social and self-imposed costs, as the need for social approval and acceptance through culturally appropriate and acceptable behavior [27]. The authors also presented evidence that these self-imposed costs are significant determinants of compliance intention and are shown to have more significance than formal sanctions. This shows that moral beliefs and social pressures are considered when employees make compliance decisions. Results by Hu et al. (2011) contradict the notion of deterrence as the biggest factor in policy compliance [11]. Their findings suggest that deterrence has no influence on an individual's intent to comply with policy. The authors further state that perceived benefits and intrinsic satisfactions more influential in compliance decision making. In their study population, reward or benefit is a larger motivating force.

In studying how rewards or incentives contribute to the information security culture, Farahmand, Atallah and Spafford (2013) pointed out that not all incentives positively influence performance and caution against using incentives that are not efficient [28]. Efficient incentives persuade a large number of heterogeneous users to act for the common cause. Acting with a common purpose or in an organizationally prosocial context is the basis of research by Thomson and van Niekerk (2012). In their work, the authors state that when prosocial behavior is cultivated in an organization, the need for punishments or rewards to influence compliance is eliminated [29]. In a prosocial environment, employees are not apathetic to the organizations policies. The organizational goals of information security are accepted without the thought of consequences or expectations of rewards. Vance, Siponen and Pahnila (2012) stated that rewards can negatively affect compliance intention if the perceived benefit of noncompliance is greater than the perceived incentive from the organization [30]. Employees that see intrinsic incentives, such as saving time or ease of use, as a benefit of policy noncompliance, are more likely to exhibit inappropriate security behavior. In conclusion, regardless of deterrence or incentives, adherence to information security policy is a major factor in cultivating an information security culture.

5 Attitudes and Involvement

Positive employee attitudes about information security compliance and their involvement in the process is another factor that impacts the information security culture in an organization. Ifinedo (2014) describes attitude as the employee's positive or negative feelings towards a behavior [31]. Research has shown that user's experience and involvement influences their perceptions or attitudes about information security [2]. Lebek et al. (2014) highlights the fact that there is a direct relationship between attitude and behavioral intent [15]. Furthermore, an employee's attitude towards organizational compliance and the perceptions of their colleagues in the workplace greatly affects secure behavior [32].

When employees participate in activities that are focused on a commitment to the organizations security goals and engage with like-minded colleagues in such matters, there is a positive effect on information security compliance [31]. This emphasizes the importance of active employee involvement. Parsons et al. (2014b) noted that employee knowledge, attitudes, and behaviors are influenced by organizational factors [33]. They draw a conclusion that increased knowledge of policy and procedure is highly correlated with a positive attitude towards the organization's policy and procedure. Safa et al. (2016) extend this theme with findings that show that the sharing information security knowledge and security collaboration and mediation, between the organization and its employees, greatly effects compliance [34].

Employee attitudes and involvement are also influenced by experience. Chen and Zahedi (2016) showed that once users' perceive or have experienced a cyber threat, they are more likely to take protective actions [35]. Results from a study by

Öğütçü et al. (2016) confirm these findings by highlighting that the more users perceive threats and increase their awareness of the technology, the more productive their security-focused behavior becomes [6]. Awareness and perception is a positive result of comprehensive information security training. The user's own personal experience or knowledge of incidents that happen to familiar environments eliminates the thought that it "won't happen here" or "won't happen to me" [36].

Guo et al. (2011) demonstrated that attitudes towards security behavior are also influenced by the effect on job performance, workgroups norms, and perceived identity match [37]. Most users want to achieve advantages to increase job performance will engage in any action that improves productivity and efficiency. In the same way they avoid actions that are seen as hindrances. With respect to workgroup norms, employees will adopt the attitudes, opinions, and practices of their work teams in the absence of expertise. In this way group attitudes drive the behavior of individuals. Perceived identity match influences security behaviors based on self-identity. If users believe that following policies is an important part of their self-image in their profession, they will more likely adhere to the policies. Employee attitudes can also be influenced due to changes information systems, workspace, regulatory and compliance rules, and job roles or responsibilities. All of these changes can effect employee satisfaction. Failure of management to recognize these changes and how they affect employees could lead to a negative security culture [38].

When users evaluate an information system, satisfaction is the most commonly used measurement. Shropshire, Warkentin and Sharma (2015) note that information system satisfaction is equated to perceptions of ease of use and system usefulness [39]. The research shows this to be a significant predictor of system security intention. Montesdioca and Maçada (2015) concluded that user displeasure with security measures can be a risk for information system security [40]. The authors assert that a way to change the negative perception is through involving users in the development of security practices. Developing consistent policies, systems that meet users' needs, and training in how to use the systems efficiently, can increase employee productivity and satisfaction.

6 Training and Awareness

Training and awareness is a foundational piece of all thriving information security cultures. It provides employees with the requisite knowledge needed for proper use of systems, compliance with policies, and handling of data. Information security managers must implement training and awareness programs focused on policies, roles, and responsibilities. Employees that lack proper awareness and training can expose the organization to security risks. Organizations need to devote resources towards building information security skills across all levels of the personnel and management [1]. Those that receive training have been shown to adhere to and exhibit a more positive information security culture [41]. No matter the hardware or software system investment, the untrained or unaware employee becomes the vector for cyber-attack [42]. Inadequate skills and awareness can lead to intentional or unintentional errors that can be a liability to security. Computer users who possess the adequate knowledge of information security concepts, exhibit more positive attitude towards information security, which then results in more positive behavior [43]. Organizations need to provide employee information system and security training that is sufficient enough to eliminate errors.

Lack of awareness of cyber-attacks against the human factors of information security contributes significantly to breaches caused by human behavior. Management has the responsibility to make sure their awareness programs benefit employees by promoting consistent review and understanding of the importance of handling data and systems and prevalence of threats against them. Also, the content of the training needs to be constantly reviewed [5]. The awareness programs should be customized using the language and jargon specific to the business objectives and environment [44].

Information security training should not be delivered to users in a “technocratic” or fact-based broadcast. This type of training fails to bridge the gap between the organizations security policies and business objectives and the needs of the audience. Information security training should be focused on the formation of habits in relation to the user’s perceptions and the procedural options available to them. Training that provide relevant and immersive activities to show the steps involved in information security, or the impacts of an incident, are shown to be effective in increasing awareness. They give the employee an avenue to retain the experience, rather than the procedural information.

McBride, Carter and Warkentin (2012) noted that the same training scenarios and illicit different reactions among employees with different personality traits [45]. Because of this, the authors imply that information security training must be varied to accommodate individual employee personality types. The data from this study show that different personality types also react differently to threats and sanctions. As a result, organizations must maintain a nuanced and tailored approach so that their training and awareness programs reflect those differences. Furthermore, organizations under specific regulatory authority should take special care to increase employee awareness of the authority’s policies. These trainings should be on a regular recurring basis to keep up with changes in business processes, standards and regulations [46]. Information security awareness and training should be included in risk assessment strategies to enhance mitigation. Research has shown that despite the threats of cybercrime and insider breach that organizations face, the employee awareness levels are still lacking. Adoption of comprehensive information security awareness programs fosters a culture of security compliance in an organization [47].

7 Management Support

Management support is an important factor in cultivating an information security culture despite the fact that there has not been a lot of research in this area. Consistent top management support is essential to creating a supportive environment in the organization and providing the necessary support. This support includes budget, technology, and human capital. Support and leadership from management are key contributors to successful implementation of information security efforts. Top management must advocate and deliver a clear message of its information policies and goals to the rest of the organization [5]. In order for an information security management program to be effective, management must define the organization’s information security goals and objectives. It is imperative that managers develop a strategic for protecting assets and formulate budgets that incorporate information security to negate the risk of damage caused by possible attacks [40]. Senior management must be actively involved in the planning and decision-making processes. It is also at this level that the policies and guidelines are developed [48]. When management is engaged in the process, employees have more positive attitudes towards compliance. The emphasis that leadership places on information security drives the culture [38]. According to research by Said et al. (2014), top management support makes the strongest contribution to information security knowledge sharing [49].

The leadership should think strategically about developing the policies, objectives, plans that make up the information security strategy. It is then their responsibility to convey clarity and consistency in messages to employees about acceptable behavior and the sanctions for negative actions [16]. There are correlations between management support and security awareness which are strengthened by the security culture [50]. It is further stated that in environments where the employee tasks are highly dependent upon other employees, management needs to emphasize security awareness and training programs as these employees tend to police each other. In organizations where this type of task dependence doesn’t exist, lack of co-worker monitoring and poor attitudes towards compliance may be prevalent. Therefore, management needs emphasize and closely monitor security policy and attitudes for compliance.

The group dynamic is also emphasized in research by Safa et al. 2016. The authors found that management can affect compliance attitudes by facilitating cross-training, knowledge sharing and security collaboration [34]. Employees who are share security knowledge raise awareness on a whole and those who work together on common security goals show a positive attitude towards compliance. Finally, management plays an important role in building proper organizational structures to support the culture. This structure makes sure that the business strategy and the security function remains aligned. Employee attitudes can be negative towards security if the policies and programs are seen as a hindrance to the employees task function. This can lead to noncompliance out of the necessity to efficiently complete tasks [51].

8 Future Work

The presented literature review identified the empirically validated factors that must exist to promote an information security culture. As previously mentioned, information security is an organization-wide issue rather than a technical issue. Without a deep change in its information security culture, security effectiveness will not be achieved. Quite a few studies have highlighted the effects of human behavior factors on the security culture. Very few studies have discussed how the technology facets of information security affect the human factors of security culture.

With many empirical studies showing the relationships between the factors of information security policy, deterrence and incentives, attitudes and involvement, training and awareness, and management support, future research can focus on why there is a failure in organizational commitment to enhance these factors. This is especially pertinent with the consistent rise in security breaches and cyber-crime. Additions to the body of knowledge can also be made by focusing on how an organization's information security technology posture relates to their security culture. For example, what are the culture perceptions of users in an environment with mature or weak enterprise security capabilities? These capabilities include: Privacy, Threat mitigation, Transaction and Data Integrity, Identity and Access Management, Application Security, Physical Security, and Personnel Security.

9 Conclusion

Information security is important in every organization that looks to protect is data, information systems, and assets from cybercrime. Technology-based safeguards alone won't achieve this goal. Humans, in the form of management, employees, and users, play a vital role information security. An organization's information security program success depends on appropriate user behavior. All human contributions to the effort are dependent on the factors that contribute to the information security culture. Organizations across the world invest large amounts of money to ensure the security of their data and information systems, but often lack the human security culture needed at the foundation of their information security efforts. In order to have a positive information security culture, organizations must ensure a mix of technical systems and human behavioral aspects of information security management. This is the exclusive path to cogently address the security challenges. The results from this literature review can support organizations in their information security processes by offering insights into how human factors can enhance the information security culture. The role of management, employees, and users in the organization and how they influence the culture demonstrates the need for further study.

References

1. Adams, M., Makramalla, M.: Cybersecurity skills training: an attacker-centric gamified approach. Technol. Innov. Manag. Rev. **5**(1), 5–14 (2015)
2. Safa, N.S., Sookhak, M., Von Solms, R., Furnell, S., Ghani, N.A., Herawan, T.: Information security conscious care behaviour formation in organizations. Comput. Secur. **53**, 65–78 (2015)
3. IBM: The 2015 IBM cyber security intelligence index. IBM Security Service (2015)
4. Acuña, D.C.: Effects of a comprehensive computer security policy on computer security culture. In: MWAIS 2016 Proceedings, Paper 10 (2016)
5. Alavi, R., Islam, S., Jahankhani, H., Al-Nemrat, A.: Analyzing human factors for an effective information security management system. Int. J. Secure Softw. Eng. (IJSSE) **4**(1), 50–74 (2013)
6. Öğütçü, G., Testik, Ö.M., Chouseinoglou, O.: Analysis of personal information security behavior and awareness. Comput. Secur. **56**, 83–93 (2016)
7. IBM: The 2013 IBM cyber security intelligence index. IBM Security Services (2013)
8. Hershberger, P.: Security Skills Assessment and Training: The "Make or Break" Critical Security Control. SANS Institute InfoSec Reading Room (2014)

9. Guo, K.H.: Security-related behavior in using information systems in the workplace: a review and synthesis. Comput. Secur. **32**, 242–251 (2013)
10. Da Veiga, A., Martins, N.: Information security culture and information protection culture: a validated assessment instrument. Comput. Law Secur. Rev. **31**(2), 243–256 (2015)
11. Hu, Q., Xu, Z., Dinev, T., Ling, H.: Does deterrence work in reducing information security policy abuse by employees? Commun. ACM **54**(6), 54–60 (2011)
12. Tang, M., Zhang, T.: The impacts of organizational culture on information security culture: a case study. Inf. Technol. Manag. **17**, 1–8 (2016)
13. Alhogail, A.R.E.E.J., Mirza, A.: A framework of information security culture change. J. Theoret. Appl. Inf. Technol. **64**(2), 540–549 (2014)
14. Abraham, S.: Information security behavior: factors and research directions. In: AMCIS 2011 Proceedings – All Submissions, Paper 462 (2011)
15. Lebek, B., Uffen, J., Neumann, M., Hohler, B., Breitner, M.H.: Information security awareness and behavior: a theory-based literature review. Manag. Res. Rev. **37**(12), 1049–1092 (2014)
16. AlHogail, A.: Design and validation of information security culture frame-work. Comput. Hum. Behav. **49**, 567–575 (2015)
17. D'Arcy, J., Hovav, A., Galletta, D.: User awareness of security countermeasures and its impact on information systems misuse: a deterrence approach. Inf. Syst. Res. **20**(1), 79–98 (2009)
18. Sari, P.K.: A concept of information security management for higher education. In: International Conference on Technology and Operation Management, 3rd Bandung, pp. 469–477 (2012)
19. Bulgurcu, B., Cavusoglu, H., Benbasat, I.: Information security policy compliance: an empirical study of rationality-based beliefs and information security awareness. MIS Q. **34**(3), 523–548 (2010)
20. Renaud, K.: Blaming noncompliance is too convenient: what really causes information breaches? IEEE Secur. Priv. **10**(3), 57–63 (2012)
21. Haeussinger, F., Kranz, J.: Information security awareness: its antecedents and mediating effects on security compliant behavior. In: 34th International Conference on Information Systems (2013)
22. Choi, M., Levy, Y., Hovav, A.: The role of user computer self-efficacy, cybersecurity countermeasures awareness, and cybersecurity skills influence on computer misuse. In: Proceedings of the Pre-International Conference of Information Systems (ICIS) SIGSEC–Workshop on Information Security and Privacy (WISP), December 2013
23. Chen, Y., Ramamurthy, K., Wen, K.W.: Impacts of comprehensive information security programs on information security culture. J. Comput. Inf. Syst. **55**(3), 11–19 (2015)
24. Hu, Q., Dinev, T., Hart, P., Cooke, D.: Managing employee compliance with information security policies: the critical role of top management and organizational culture. Decis. Sci. **43**(4), 615–660 (2012)
25. Parsons, K.M., Young, E., Butavicius, M.A., McCormac, A., Pattinson, M.R., Jerram, C.: The influence of organizational information security culture on information security decision making. J. Cogn. Eng. Decis. Mak. **9**(2), 117–129 (2015)
26. Chen, Y., Ramamurthy, K., Wen, K.W.: Organizations' information security policy compliance: stick or carrot approach? J. Manag. Inf. Syst. **29**(3), 157–188 (2012)
27. D'Arcy, J., Devaraj, S.: Employee misuse of information technology resources: testing a contemporary deterrence model. Decis. Sci. **43**(6), 1091–1124 (2012)
28. Farahmand, F., Atallah, M.J., Spafford, E.H.: Incentive alignment and risk perception: an information security application. IEEE Trans. Eng. Manag. **60**(2), 238–246 (2013)
29. Thomson, K., van Niekerk, J.: Combating information security apathy by encouraging prosocial organisational behaviour. Inf. Manag. Comput. Secur. **20**(1), 39–46 (2012)

30. Vance, A., Siponen, M., Pahnila, S.: Motivating IS security compliance: insights from habit and protection motivation theory. Inf. Manag. **49**(3), 190–198 (2012)
31. Ifinedo, P.: Information systems security policy compliance: an empirical study of the effects of socialisation, influence, and cognition. Inf. Manag. **51**(1), 69–79 (2014)
32. Ifinedo, P.: Understanding information systems security policy compliance: an integration of the theory of planned behavior and the protection motivation theory. Comput. Secur. **31**(1), 83–95 (2012)
33. Parsons, K., McCormac, A., Butavicius, M., Pattinson, M., Jerram, C.: Determining employee awareness using the human aspects of information security questionnaire (HAIS-Q). Comput. Secur. **42**, 165–176 (2014)
34. Safa, N.S., Von Solms, R., Furnell, S.: Information security policy compliance model in organizations. Comput. Secur. **56**, 70–82 (2016)
35. Chen, Y., Zahedi, F.M.: Individuals'internet security perceptions and behaviors: polycontextual contrasts between The United States and China. MIS Q. **40**(1), 205–222 (2016)
36. Davinson, N., Sillence, E.: It won't happen to me: promoting secure behaviour among internet users. Comput. Hum. Behav. **26**(6), 1739–1747 (2010)
37. Guo, K.H., Yuan, Y., Archer, N.P., Connelly, C.E.: Understanding nonmalicious security violations in the workplace: a composite behavior model. J. Manag. Inf. Syst. **28**(2), 203–236 (2011)
38. Dhillon, G., Syed, R., Pedron, C.: Interpreting information security culture: an organizational transformation case study. Comput. Secur. **56**, 63–69 (2016)
39. Shropshire, J., Warkentin, M., Sharma, S.: Personality, attitudes, and intentions: predicting initial adoption of information security behavior. Comput. Secur. **49**, 177–191 (2015)
40. Montesdioca, G.P.Z., Maçada, A.C.G.: Measuring user satisfaction with information security practices. Comput. Secur. **48**, 267–280 (2015)
41. Da Veiga, A., Martins, N.: Improving the information security culture through monitoring and implementation actions illustrated through a case study. Comput. Secur. **49**, 162–176 (2015)
42. Badie, N., Lashkari, A.H.: A new evaluation criteria for effective security awareness in computer risk management based on AHP. J. Basic Appl. Sci. Res. **2**(9), 9331–9347 (2012)
43. Parsons, K., McCormac, A., Pattinson, M., Butavicius, M., Jerram, C.: A study of information security awareness in Australian government organizations. Inf. Manag. Comput. Secur. **22**(4), 334–345 (2014)
44. Metalidou, E., Marinagi, C., Trivellas, P., Eberhagen, N., Giannakopoulos, G., Skourlas, C.: Human factor and information security in higher education. J. Syst. Inf. Technol. **16**(3), 210–221 (2014)
45. McBride, M., Carter, L., Warkentin, M.: Exploring the role of individual employee characteristics and personality on employee compliance with cybersecurity policies. Technical report, RTI International (2012)
46. Hipsky, S., Younes, W.: Beyond concern: K-12 faculty and staff's perspectives on privacy topics and cybersafety. Int. J. Inf. Commun. Technol. Educ. (IJICTE) **11**(4), 51–66 (2015)
47. Chan, H., Mubarak, S.: Significance of information security awareness in the higher education sector. Int. J. Comput. Appl. **60**(10), 23–31 (2012)
48. Narain Singh, A., Gupta, M.P., Ojha, A.: Identifying factors of "organizational information security management". J. Enterp. Inf. Manag. **27**(5), 644–667 (2014)
49. Said, A.R., Abdullah, H., Uli, J., Mohamed, Z.A.: Relationship between organizational characteristics and information security knowledge management implementation. Procedia-Soc. Behav. Sci. **123**, 433–443 (2014)

50. Knapp, K.J., Ferrante, C.J.: Information security program effectiveness in organizations: the moderating role of task interdependence. J. Organ. End User Comput. (JOEUC) **26**(1), 27–46 (2014)
51. Flores, W.R., Antonsen, E., Ekstedt, M.: Information security knowledge sharing in organizations: investigating the effect of behavioral information security governance and national culture. Comput. Secur. **43**, 90–110 (2014)

The Gender Turing Test

Wayne Patterson(✉), Jacari Boboye, Sidney Hall, and Maalik Hornbuckle

Howard University, Washington, DC, 20059, USA
wpatterson@scs.howard.edu, Jacari.boboye@bison.howard.edu,
sidneybhall@gmail.com, maalikhorn@gmail.com

Abstract. In our Behavioral Cybersecurity course at Howard University in last spring (2016), students for their final exam were asked to write an opinion on the following question: "We know, in general in the US as well as at Howard, that only about 20% of Computer Science majors are female. Furthermore, of those CS students choosing to concentrate in Cybersecurity, fewer than 10% are female. Can you suggest any reason or reasons that so many fewer female computer scientists choose Cybersecurity?" In the course of reviewing the answers, it became clear that the challenge of determining the gender of the writer was a difficult problem. To that end, a sample of approximately 50 readers have analyzed the students' texts and tried to determine the gender of the writers. The distribution of answers, to be presented in the full paper, has provided interesting options for further development of this research. In some aspects, the challenge of determining gender from a source absent of physical signals is similar to the challenge of the original Turing Test, which Turing formulated in order to present the challenge of determining whether or not machines could be said to possess intelligence.

Keywords: Cybersecurity · Gender differences · Alan turing · Turing Test · Gender Turing Test · Behavioral Cybersecurity · Human factors in cybersecurity

1 Introduction

The primary objective of this research has been to determine the possibility of detecting the gender of a writer, such as a hacker in a computing environment.

It has been noted in the current research that posing the Turing Test challenge in the context of gender determination is in fact the manner by which Turing himself chose to explain the concept of his test to an audience that might have been challenged by the idea that machines could conduct a dialogue with human interrogatories. As Turing wrote in Mind [1]: "the problem can be described in terms of a game we call 'The Imitation Game'… it is played with a man, a woman and an interrogator."

In the current instance, the development of a test comparable to what Turing proposed has arisen in the conduct of a new course in the Cybersecurity curriculum at Howard University, offered as both CSCI 456 and CSCI 656, and called Behavioral Cybersecurity [2, 3].

The full paper will present not only the results of the initial responses, disaggregated by the various categories defined above; but also the results of the training program we

D. Nicholson (ed.), *Advances in Human Factors in Cybersecurity*,
Advances in Intelligent Systems and Computing 593, DOI 10.1007/978-3-319-60585-2_26

describe to improve the ability of a respondent to identify tendencies in the use of language that may be more often attributable to one gender than the other.

2 Rationale for the Curricular Development

Given the confluence of external events: the power of the Internet, increasing geopolitical fears of "cyber terrorism" dating from 9/11, a greater understanding of security needs and industry, and economic projections of the enormous employment needs in Cybersecurity have caused many universities to develop more substantial curricula in this area, and the United States National Security Agency has created a process for determining Centers of Excellence in this field [4].

Howard University offers courses in Cybersecurity at the bachelor's, master's, and doctoral levels. Its program has been designated as a Center of Academic Excellence through the process with the National Security Agency as described above. The undergraduate enrollments have been increasing to full capacity. However, as with many universities, there is a gap in the Cybersecurity curriculum that we decided to address.

At the 1980 summer meeting of the American Mathematics Society in Ann Arbor, Michigan, a featured speaker was the distinguished mathematician, the late Peter J. Hilton[1]. Dr. Hilton was known widely for his research in algebraic topology, but on that occasion he spoke publicly for the first time about his work in cryptanalysis during World War II at Hut 8 in Bletchley Park, the home of the now-famous efforts to break German encryption methods such as the Enigma.

The first author was present at that session and has often cited Professor Hilton's influence in sparking interest in what we now call cybersecurity. Hilton at the time revealed many of the techniques used at Bletchley Park in breaking the Enigma code. However, one that was most revealing was the discovery by the British team that, contrary to the protocol, German cipher operators would send the same message twice, something akin to, "how's the weather today?" at the opening of an encryption session. (This discovery was represented in the recent Academy-Award nominated film, "The Imitation Game." [5]) Of course, it is well-known in cryptanalysis that having two different encryptions of the same message with different keys is an enormous clue in breaking a code. Thus it is not an exaggeration to conclude that a behavioral weakness had enormous practical consequences, as the Bletchley Park teams have been credited with saving thousands of lives and helping end the War.

3 A Final Exam-Question

In the offering of this course in the spring semester 2016, there was considerable discussion about the identification of the gender of a potential hacker or other computer user. This led to the question on the final examination which asked students as follows:

[1] Peter Hilton was portrayed in "The Imitation Game," only called "Peter" in the dialogue, but listed in the credits for the actor Matthew Beard in the role of "Peter Hilton."

"We know, in general in the US as well as at Howard, that only about 20% of Computer Science majors are female. Furthermore, of those CS students choosing to concentrate in Cybersecurity, fewer than 10% are female. Can you suggest any reason or reasons that so many fewer female computer scientists choose Cybersecurity?"

4 While Grading

In the course of grading this examination, the first author, as he read each answer, questioned himself as to whether or not he could determine whether the author of the answer was one of his male or female students, based on the tone, language choice, use of certain keywords, and the expected perception of the point of view of the author. Thus as he found that this was very often difficult to determine, it seemed that it might be interesting for other persons of widely varied backgrounds, for example: gender, age, profession, geographic location, first language–to be posed the same questions.

Consequently a test was constructed from the student responses. Three variations were added: the first author himself wrote one of the responses in an attempt to deceive readers into thinking the writer was female; and two responses were repeated in order to validate whether or not the responders could detect the repetition, thus showing that they were concentrating on the questions themselves.

5 Turing's Paper in *"Mind"*

It was noted that there was a certain similarity between the administration of this test and the classic Turing Test originally posed by Turing to respond to the proposition that machines (computers) could possess intelligence.

The Turing Test supposes that a questioner is linked electronically to an entity in some other location, and the only link is electronic. In other words, the questioner does not know whether or not he or she is corresponding with a human being or a computer. The test is conducted in the following way: the questioner may ask as many questions as he or she desires, and if at the end of the session, the questioner can absolutely determine that the invisible entity is human, and in fact it is a computer, then the other entity can reasonably be said to possess intelligence.

There is a large body of research on this topic since Turing first posed the question around 1950, beginning with the development of the artificial "psychologist" named Eliza originally developed by Joseph Weizenbaum [6], and leading all the way to IBM supercomputer Watson, that was able to beat two human experts on the game show Jeopardy in 2013 [7]. It is generally accepted, however, that no computer however powerful has been able to pass this Turing Test.

It has been discussed throughout the history of Computer Science as to whether this test has been satisfied or indeed if it could ever be satisfied.

6 "The Imitation Game"

It seems very interesting, in the context of the Gender Test as described in our course, that in many ways it draws historically from Turing's thinking.

Many readers may note that the recent film, as indicated above, addressing both Turing's life and his efforts in breaking the Enigma Code in the Second World War was called "The Imitation Game." Turing published an extremely important article in the May 1950 issue of *Mind* [1] entitled "Computing Machinery and Intelligence." More to the point, he called the section of this paper in which he first introduced his Turing Test "The Imitation Game," which evolved into the title of his biographical film.

It is significant, in our view, that in order to explain to a 1950 s audience how to establish whether or not an entity possessed intelligence that to describe the test in terms of a human and machine would be incomprehensible to most of his audience, since in 1950 there were only a handful of computers in existence.

Consequently, in introducing the nature of his test, he described it as a way of determining gender as follows:

I propose to consider the question, 'Can machines think?' This should begin with definitions of the meaning of the terms 'machine' and 'think'. The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous. If the meaning of the words 'machine' and 'think' are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, 'Can machines think?' is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words.

The new form of the problem can be described in terms of a game which we call the 'imitation game'. It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the other two the man is and which is the woman. He knows them by labels X and Y, and at the end of the game he says either 'X is A and Y is B' or 'X is B and Y is A'. The interrogator is allowed to put questions to A and B thus:

C: Will X please tell me the length of his or her hair?

We thus view the initiative that we have developed from the Behavioral Cybersecurity course as a descendant in some small way of Turing's proposition.

7 Respondents

In order to understand the ways in which persons interpret written text and try to assign gender to the author–in effect a version of the Gender Turing Test (henceforth, GTT) described by Turing in the paper cited above, a number of individuals from varying backgrounds, genders, ages, first language, country, and profession were given the test in question.

There have been 55 subjects completing this test, and a description of their demographics follows (Table 1).

Table 1. Demographics of participants in the Gender Turing Test

Gender	Female	Male	Total
	24	31	55
Age	Older	Younger	Total
	14	41	55

Nationality	Cameroon	Caribbean	Canada	Mexico	
	10	3	1	2	
	Puerto Rico	Eastern Europe	Saudi Arabia	USA	Total
	1	2	2	34	55

Profession	Anthropology	Computer Sci	Engineering	
	1	32	4	
	Linguistics	Psychology	Student	Total
	1	5	12	55

The participants were selected as volunteers, primarily at occasions where the first author was giving a presentation. No restrictions were placed on the selection of volunteer respondents, nor was there any effort taken to balance the participation according to any demographic objective.

The voluntary subjects were (except on one occasion) given no information about the purpose of the test, and are also guaranteed anonymity in the processing of the test results. There was no limit on the time to take the test, but most observed respondent seem to complete the test in about 15 min.

8 Summary of Results

The responses were scored in two ways. First, the number of correct answers identifying the student author was divided by the total number of questions (24) in the complete test. Alternatively, the score was determined by the number of attempts. Since in only two of 19 instances the difference between the two exceeded 2%, it was decided to utilize the second set of response scores, which are presented below (Table 2).

Observations of the results of these responses from the 55 participants in this study and their very diverse experiences that they have brought to the response to this test yields some very interesting questions to ponder:

First, female respondents were more accurate in the identification of the gender of the students by a margin of 56.89% to 51.02%.

Table 2. Responses to the Gender Turing Test questions

Respondents by Gender	Female	Male
Correct Percentage of Responses	56.89%	51.02%

Respondents by Age	Older	Younger
Correct Percentage of Responses	57.60%	51.77%

Respondents by Nation	Cameroon	Caribbean	Canada	Mexico
Correct %	49.17%	56.67%	45.83%	54.35%
	Puerto Rico	Eastern Europe	Saudi Arabia	United States
Correct %	50.00%	66.67%	33.33%	54.79%

Respondents by Profession	Anthropology	Computer Sci	Engineer
Correct %	64.71%	52.91%	58.33%
	Linguist	Psychologist	Student
Correct %	66.67%	54.17%	50.03%

Next, older respondents were more accurate in their identification than younger responses by a similar margin of 57.6% to 51.77%. This might be a more surprising result since for the most part the older respondents were not as technically experienced in computer science or cyber security matters than the younger responders, who for the most part were students themselves.

One very clear difference is that the Eastern European respondents scored far higher in their correct identification of the students' gender, averaging 66.67% with the nearest other regional responses being fully 10% less. The number of respondents from Eastern Europe was very small, and so generalizations might be risky in this regard. However, the Eastern Europeans (from Romania and Russia) were not first language speakers of English, although they were also quite fluent in the English language. Each of them also tied for the highest percentage of correct answers of anyone amongst all 55 respondents.

There were a fairly large number of Spanish-speaking respondents, and a number of them were not very fluent in English. Nevertheless, the overall score of the Spanish-speaking respondents was above the average for all respondents from English-speaking countries–including Cameroon, the Caribbean, Canada and United States. Both Cameroon and Canada are bilingual French- and English-speaking countries, but all of the respondents in this case were from the English-speaking parts of these two countries. In addition, the Caribbean respondents were also from the English-speaking Caribbean.

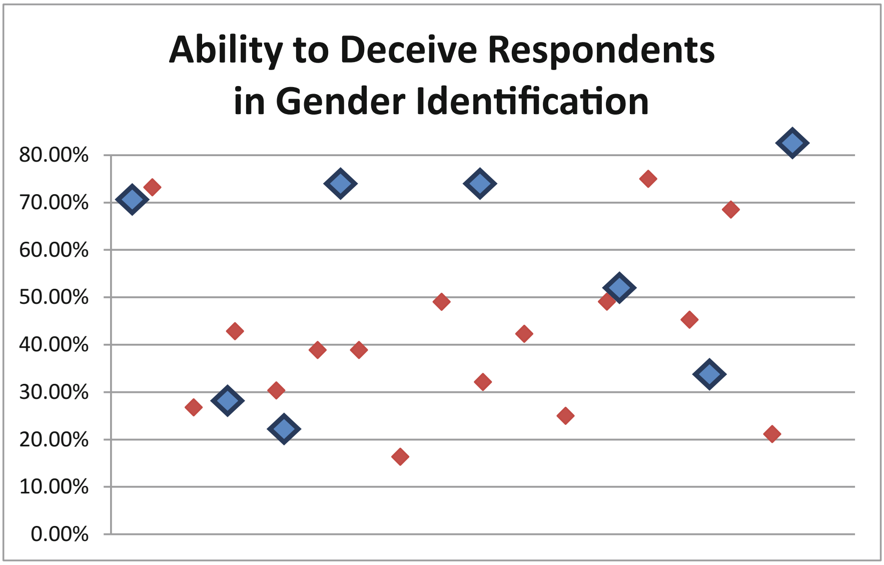
The Saudi Arabian respondents, for whom Arabic is their first language, had greater difficulty in identifying the correct gender. It is possible that these differences could have arisen from the lack of fluency of these respondents in English.

Of the respondents from the various disciplines, the linguist, anthropologist, the engineers, and the psychologists all fared better than the computer scientists–and lowest of all were the students who took the test (as opposed to the students who wrote the original answers).

It is possible, of course, to view the entire data set of responses to this test as a matrix of dimension 24 × 55, wherein the students who wrote the original exam–and thus in effect, created the GTT–represent the rows of the matrix, and the gender classification by the 55 responders as the columns. If we instead examine the matrix in a row wise fashion, we learn of the writing styles of the original test-takers, and their ability (although inadvertent, because no one, other than the first author, planned that the writings would be used to identify the gender of the writer.

Thus it is perhaps more informative than the assessment of the ability of the respondent to determine the gender of the test takers, to note that several of the original test takers were able, unconsciously, to deceive over two thirds of the respondents. Fully one-quarter (six of 24) of the students reach the level of greater than two thirds deception. Of these six "High Deceivers" three were female and three were male students.

At the other end of the spectrum, one third of the students were not very capable of deception–fooling less than one third of the respondents. Of these eight students, six or male in only two were female. On the whole, averaging the level of deception by the male and female students, on average the female students were able to deceive 52.5% of the respondents, while the male students were only able to accomplish this with 42.2% of the respondents. The following chart shows a scatter plot of the student takers ability to fool the respondents (Fig. 1).



Individual Female Student	🔷
Individual Male Student	♦

Fig. 1. Scatterplot of female and male students' success in deceiving respondents.

9 Review by Students

All of the respondents described above had simply been given a test with only the simple instruction described in the attachment, without any prior preparation or understanding on the part of the respondent as to possible techniques for identifying the gender of a writer or author.

Consequently, we determined that it would be useful to see if persons could be given some training in order to try to improve their ability to improve their results on the GTT. We attempted to identify a number of keys that would assist a reader in trying to improve their scores on the GTT or related tests.

Our next objective was to see if a subject could improve on such text analysis in the case of distinguishing the gender of a writer, by looking for certain clues that could be described. Several of the authors of this article identified a number of techniques to identify the gender of an author that they had used themselves (successfully) in performing an analysis of the questions in the original GTT:

1. Examine how many pronouns are being used. Female writers tend to use more pronouns (I, you, she, their, myself).
2. What types of noun modifiers are being used by the author? Types of noun modifiers: A noun can modify another noun by coming immediately before the noun that follows it. Males prefer words that identify or determine nouns (a, the, that) and words that quantify them (one, two, more).
3. Subject matter/Style: the topic dealt with or the subject represented in a debate, exposition, or work of art. "Women have a more interactive style," according to Shlomo Argamon, a computer scientist at the Illinois Institute of Technology in Chicago.
4. Be cognizant of word usage and how it may reveal gender. Some possible feminine keywords include: with, if, not, where, be, should. Some of the other masculine keywords include: around, what, are, as, it, said. This suggests that language tends to encode gender in very subtle ways.
5. "Women tend to have a more interactive style," said Shlomo Argamon, a computer scientist at the Illinois Institute of Technology in Chicago [8]. "They want to create a relationship between the writer and the reader."
   Men, on the other hand, use more numbers, adjectives and determiners–words such as "the," "this" and "that"–because they apparently care more than women do about conveying specific information.
6. Pay attention to the way they reference the gender of which they speak. For example, a female may refer to her own gender by saying "woman" rather than girl.
7. Look at the examples that they give. Would you see a male or female saying this phrase?
8. A male is more likely to use an example that describes how a male feels.
9. Women tend to use better grammar, and better sentence structure than males.
10. When a person of one gender is describing the feelings/thoughts of the opposite gender, they tend to draw conclusions that make sense to them, but will not provide actual data.

It should be noted that some prior work includes the development of an application available on the Internet (Gender Guesser), developed by Neil Krawetz based on [8] and described at the location http://hackerfactor.com/GenderGuesser.php [9].

This application seems to depend on the length of the text being analyzed, and in comparison to the responses of our human responses, does not perform as well is normally the application indicates the text is too short to give a successful determination of gender.

However, because the overall objective of this research is to determine if a GTT can be used in a cyber security context, it is likely that an attacker or hacker might only be providing very short messages–as for example a troller on the Internet trying to mask his or her identity in order to build a relationship, say with an underage potential victim.

10 Future Research

The questions that have been raised by this research have also open to the potential of devising other such tests to determine other characteristics of an author, such as age, profession, geographic origin, or first language. In addition, given that the initial respondents to the test as described above themselves are from a wide variety of areas of expertise, nationality, and first language, a number of the prior participants have indicated interest in participating in future research in any of these aforementioned areas.

References

1. Turing, A.M.: Computing machinery and intelligence. Mind Q. Rev. Psychol. Philos. **LIX**(236), 433–460 (1950)
2. Patterson, W., Winston, C. E., Fleming, L.: Behavioral Cybersecurity: a needed aspect of the security curriculum. In: Proceedings of the IEEE SoutheastCon 2016, Norfolk, VA, March 2016
3. Patterson, W., Winston, C. E., Fleming, L.: Behavioral Cybersecurity: human factors in the cybersecurity curriculum. In: Proceedings of the 2nd International Conference on Human Factors in Cybersecurity, Orlando, FL, July 2016
4. National Centers of Academic Excellence in Information Assurance Education, National Security Agency (2017). https://www.nsa.gov/ia/academic_outreach/nat_cae/index.shtml
5. SONY Pictures Releasing, The Imitation Game (Film) (2014)
6. Weizenbaum, J.: ELIZA—a computer program for the study of natural language communication between man and machine. Commun. Assoc. Comput. Mach. **9**, 36–45 (1966)
7. Baker, S., Jeopardy, F.: The Story of Watson, the Computer That Will Transform Our World. Paperback – Houghton Mifflin Harcourt, Boston, 27 March 2012
8. Argamon, S., Koppel, M., Fine, J., Shimoni, A.R.: Gender, genre, and writing style in formal written texts. Text **23**(3), 321–346 (2003)
9. Krawetz, N.: Gender Guesser. http://hackerfactor.com/GenderGuesser.php

Do You Really Trust "Privacy Policy" or "Terms of Use" Agreements Without Reading Them?

Abbas Moallem(✉)

UX Experts, LLC, Cupertino, CA 95014, USA
Abbas@uxexperts.com

Abstract. An online survey was administered to college students asking them whether they read the terms of use and privacy policy when using services or applications, and if not, why. Also, when apps ask to have access to their location, contacts, or camera, do the students allow access or not, due to security concerns. One hundred and seventy students have completed the survey. Results suggest that 62% of participants "Agree" to not reading the terms of use or privacy policies, with the most common explanation being that the text is "too long." For the question "Have you ever rejected a mobile app request for accessing your contacts, camera or location?" the answers are more encouraging. Ninety-two percent of those surveyed express that they "Yes," have rejected access if they believe the app does not need to access the camera or contacts.

Keywords: Privacy policy · Trust · Application design · User behavior

1 Introduction

With the Internet and smart mobile devices now an essential part of our daily activities, people uses multiple web applications and mobile apps to complete essential tasks. One of the first steps in using all apps or web applications is agreeing to their "terms of use or service" and "privacy policy."

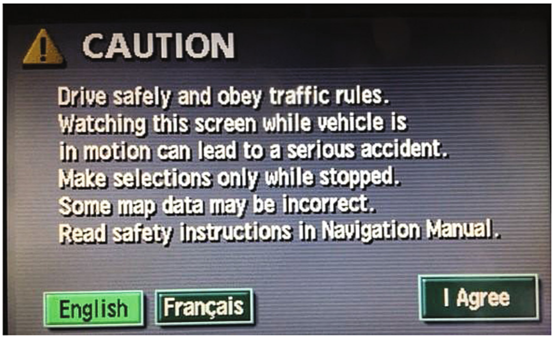
The Terms-of-Service Agreement [1] is used for legal purposes by applications and internet service providers that save users' personal data. A Terms-of-Service Agreement is legally binding and may be subject to change. Terms-of-Service Agreements serve as a contract between the providers an app or web application and the users. The agreement defines the rules the user must agree to before using the application.

A Privacy Policy is a legal document that discloses the ways a party gathers, uses, discloses, and manages users' data according to the existing privacy laws [2, 3] if any. The policy explains how a company collects, stores, and uses data.

Both of these documents, although vital for users, are designed to serve and protect only providers.

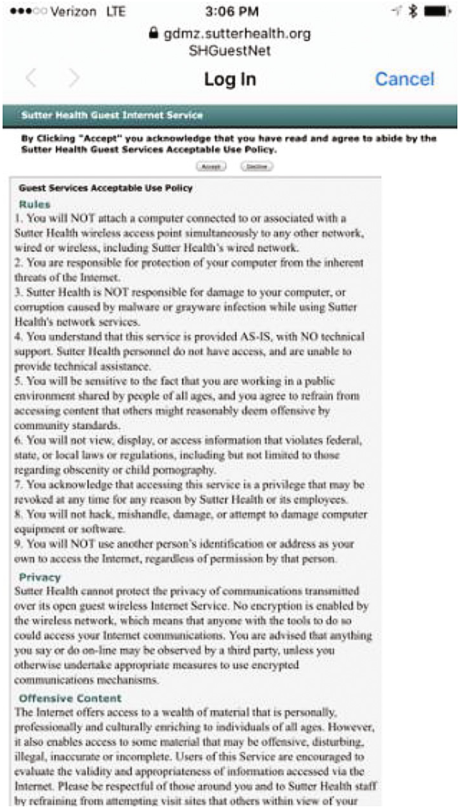
Currently, general protocol dictates that when using apps, Wi-Fi services, or web applications, the user must first agree to a long list of legal agreements. Sometimes even

D. Nicholson (ed.), *Advances in Human Factors in Cybersecurity*,
Advances in Intelligent Systems and Computing 593, DOI 10.1007/978-3-319-60585-2_27



Drive safely and obey traffic rules.
Watching this screen while vehicle is
in motion can lead to a serious accident.
Make selections only while stopped.
Some map data may be incorrect.
Read safety instructions in Navigation Manual.

Fig. 1. Driver must agree with the caution each time the car starts (Lexus SC430 Car)



By Clicking "Accept" you acknowledge that you have read and agree to abide by the Sutter Health Guest Services Acceptable Use Policy.

Guest Services Acceptable Use Policy

Rules

1. You will NOT attach a computer connected to or associated with a Sutter Health wireless access point simultaneously to any other network, wired or wireless, including Sutter Health's wired network.
2. You are responsible for protection of your computer from the inherent threats of the Internet.
3. Sutter Health is NOT responsible for damage to your computer, or corruption caused by malware or grayware infection while using Sutter Health's network services.
4. You understand that this service is provided AS-IS, with NO technical support. Sutter Health personnel do not have access, and are unable to provide technical assistance.
5. You will be sensitive to the fact that you are working in a public environment shared by people of all ages, and you agree to refrain from accessing content that others might reasonably deem offensive by community standards.
6. You will not view, display, or access information that violates federal, state, or local laws or regulations, including but not limited to those regarding obscenity or child pornography.
7. You acknowledge that accessing this service is a privilege that may be revoked at any time for any reason by Sutter Health or its employees.
8. You will not hack, mishandle, damage, or attempt to damage computer equipment or software.
9. You will NOT use another person's identification or address as your own to access the Internet, regardless of permission by that person.

Privacy

Sutter Health cannot protect the privacy of communications transmitted over its open guest wireless Internet Service. No encryption is enabled by the wireless network, which means that anyone with the tools to do so could access your Internet communications. You are advised that anything you say or do on-line may be observed by a third party, unless you otherwise undertake appropriate measures to use encrypted communications mechanisms.

Offensive Content

The Internet offers access to a wealth of material that is personally, professionally and culturally enriching to individuals of all ages. However, it also enables access to some material that may be offensive, disturbing, illegal, inaccurate or incomplete. Users of this Service are encouraged to evaluate the validity and appropriateness of information accessed via the Internet. Please be respectful of those around you and to Sutter Health staff by refraining from attempting visit sites that others within view of your

Fig. 2. User must agree to the terms of use each time using Sutter health guest services

car navigation systems require to agree to terms of use when starting the car, or even the Wi-Fi accessed in a coffee shop or hotel which can require an agreement to such terms each time it's accessed (Figs. 1 and 2).

Also, when people install an application, it has become common practice to require users to grant the permissions requested by the application or else not install (Hobson's choice [4]). Some applications require users to allow the application access to such information like their contact list or access their camera and camera roll. For example, job-searching applications look to access contact lists, as required by LinkedIn applications when a user profile is created.

The research shows that a user's willingness to accept an agreement is related to the degree that the user trusts an application or the company that provides the application.

General observations show that users agree to terms of use and privacy policies without reading the content and just click "Agree" [5]. At the same time, a study [6] reveals that 97% of the people surveyed expressed concern that businesses and the government might misuse their data. Privacy issues also ranked high; 80% of Germans and 72% of Americans are reluctant to share information with businesses because they "just want to maintain [their] privacy." So consumers worry about their personal data—even if they do not know what they are revealing.

In this study, we have administered an online survey among college students asking them whether they read the terms of use and privacy policy when using services or applications, and if not, why. Also, when apps ask users to have access to their location, contacts, or camera do they allow the apps to do so, or do they block access because of security concerns.

2 Method

One hundred and seventy students (51% female and 49% male) participated in this study. 51% of students aged 18 to 24 and 48% were between 24 to 44 years old. They completed an online survey using a Qualtrics survey application. All participants were undergraduate, and graduate level college students were taking HCI or Human Factors courses. The survey was administered during the year of 2016. Participants were asked the following questions:

- Have you ever read a privacy policy when installing or using an application or online service?
  - If your answer to the previous question (privacy policy) was "No," please explain why.
- Have you ever rejected a mobile app request for accessing your contacts, camera or location?
  - If your answer to the previous question (access request) was "Yes" please explain why.

3 Results

Results suggested that 62% (106 participants) "Agree" that they accept without reading the terms of use or privacy policy with the general reason expressed being that the text is "too long" (81% of 'agree' answers). For the question "Have you ever rejected a mobile app request for accessing your contacts, camera or location?” the answers are more encouraging. 92% (153 participants) of those surveyed express that they “yes” have rejected access if they believe the app does not need to access the camera or contacts. This result is in line with a previous study by Haggerty (2015), who found that 74.1% of iOS users would reject the app permissions list [7]. However, in many instances, users do accept granting permissions requested by the majority of applications.

The results of this survey raise the question on how if people do not read these documents and do not read the several notifications they receive about the changes made by companies, then what purpose do these agreements achieve from the user’s perspective? The study attempts to analyze the usefulness of these procedures besides simply being a legal formality. Is there another more effective way, using user interface design, to better inform users about terms of use and privacy policies?

Some studies suggest an improvement in privacy rules and language used might help. For example, an empirical study [8, 9] conducted with 36 users who were novices in privacy policy authoring tools worked to evaluate the quality of rules created and user satisfaction with two experimental privacy-authoring tools and a control condition. The results show that users were able to author significantly higher quality rules using either natural language with a privacy a simple way to guide tool or a structured list tool as compared to an unguided natural language control condition (Figs. 3 and 4).

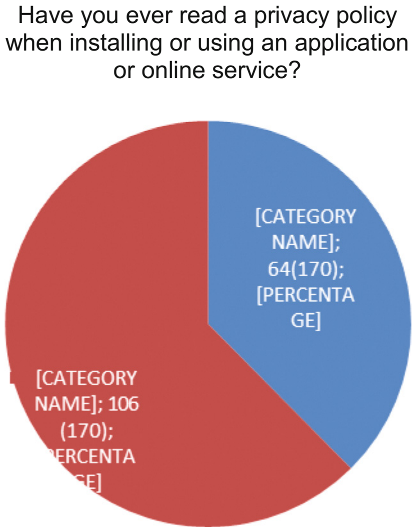


Fig. 3. Percentage of participants “Agree” to never reading the terms of use or privacy policy.

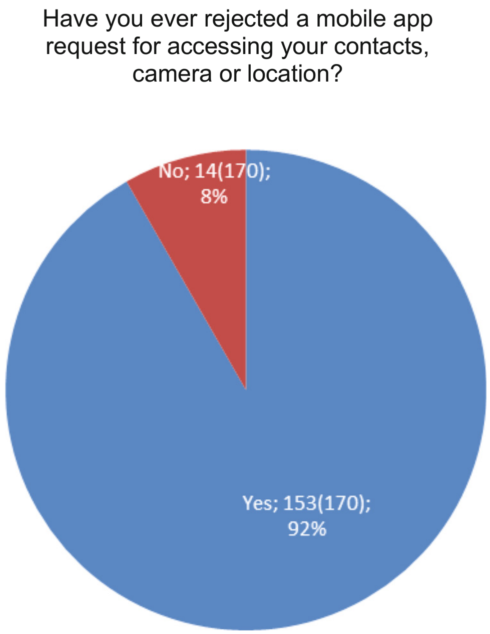


Fig. 4. Percentage of participants "Reject" a mobile app request for accessing contacts, camera or location

4 Conclusion

The results of this study illustrated that most people do not read the privacy policy and terms of use and agreed without knowledge of what they had agreed to. Not reading does not mean that the users do not care about these policies or their privacy, but instead shows that these agreements structured in a language and format that makes it difficult to read and understand. In fact, a highly significant number (over 81%) reported that they do not read because of the lengthy time it would take, and because the agreements are not easy to read. Then we asked participants "Have you ever rejected a mobile app request for accessing your contacts, camera or location?" People most often (92%) have made a judgment in denying the request. Consequently, we can assume that in the case of the terms of use and privacy policy, if they are presented in an easy way underlining what users must give up in their privacy information, then they can make a conscious decision as to whether or not to use a service.

One might question why the "term of usage" or "privacy policy" are too long to read. It is written in a language that people cannot easily understand. Also, it is delivered in a sort of hidden UI. Since it is possible to present them in a simple language, easy to understand and very simple UI, can we assume that the reason the language, length, and access are all highly difficult because software/service providers prefer users not to read them?

This study illustrates in numbers a tendency everybody already may know. However, this study tends to provide evidence and further explore the causes of the trend through a self-reporting survey.

References

1. PC Magazine, Encyclopedia. http://www.pcmag.com/encyclopedia/term/62682/terms-of-service
2. Attorney General Kamala D. Harris Announces Privacy Enforcement and Protection Unit, 9 July 2012. https://oag.ca.gov/news/press-releases/attorney-general-kamala-d-harris-announces-privacy-enforcement-and-protection
3. L0046, Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML
4. tHobson's Choice. Wikipedia. https://en.wikipedia.org/wiki/Hobson%27s_choice
5. Terms of Services, Didn't Read. https://tosdr.org/
6. Morey, T., Forbath, T.H., Schoop, A.: Customer data: designing for transparency and trust. Harv. Bus. Rev. (2015). https://hbr.org/2015/05/customer-data-designing-for-transparency-and-trust
7. Haggerty, J., Hughes-Roberts, T., Hegarty, R.: Hobson's choice: security and privacy permissions in Android and iOS devices. In: Tryfonas, T., Askoxylakis, I. (eds.) Human Aspects of Information Security, Privacy, and Trust, HAS 2015. LNCS, vol. 9190. Springer, Cham (2015)
8. Karat, C.M., Karat, J., Brodie, C., Feng, J.: Evaluating interfaces for privacy policy rule authoring. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2006, pp. 83–92 (2006)
9. Kumaraguru, P., Cranor, L.F., Lobo, J., Calo, S.B.: A survey of privacy policy languages. In: Workshop on Usable IT (2007). http://precog.iiitd.edu.in/Publications_files/Privacy_Policy_Languages.pdf

Users' Attitudes Towards the "Going Dark" Debate

Aseel Addawood[1(✉)], Yixin Zou[2], and Masooda Bashir[3]

[1] Illinois Informatics Institute, University of Illinois at Urbana-Champaign,
Urbana, USA
Aaddaw2@illinois.com

[2] Department of Advertising, University of Illinois at Urbana-Champaign,
Urbana, USA
yzoul5@illinois.com

[3] School of Information Sciences, University of Illinois at Urbana-Champaign,
Urbana, USA
Mnb@illinois.com

Abstract. This study sought to investigate the attitude and behavior of people toward the issue of privacy and national security. The online survey was carried administered to 243 online users. Participants were randomly assigned to evaluate three statements, namely, "*Citizen Privacy takes precedence over national security*," "*Governments should have access to all encrypted data*," and "*Individual privacy is a human right*." For each statement, we measured participants' level of agreement using a 5-point Likert scale. Using a one-way analysis of variance (ANOVA), we examined if privacy attitudes were different among user characteristics such as gender, religions belief, field of study and educational level. The results showed that most people have negative attitude toward government access to private data, but this view is divided along the religious, gender and field of study.

Keywords: Human factors surveillance · National security · Individual privacy · Human rights · Legislation · Terrorism · Security threat

1 Introduction

The global community is divided by a debate about the trade-off between personal privacy and security needs. The Recent battle between Apple company and the US government over a court order requiring Apple company to help FBI to hack an iPhone belonging to killed terrorist to get the terrorism-related information has reignited the more than a century-long debate about the gray area between the individual rights to privacy and national security [1]. In 2013, Edward Snowden (former CIA employee) leaked previously classified information which indicated the extent to which USA government employs widespread surveillance of the citizens. In its defense on the claim that it was infringing on the individual privacy rights, the U.S. government claimed that the information collected plays an important role in fighting against crime and terrorism [2].

D. Nicholson (ed.), *Advances in Human Factors in Cybersecurity*,
Advances in Intelligent Systems and Computing 593, DOI 10.1007/978-3-319-60585-2_28

Civil rights advocates argue that government surveillance of individual privacy is characteristic of dystopia where free thinking and individualism are persecuted. This view is anchored on the notion that we live in the world where information is used for sinister ends. What is agreeable in the debate is that there is a need to give weight to both the national security and the privacy. The government is under the responsibility to take all the necessarily steps to make sure that it has lived up to one of its primary mandates, by guaranteeing security [3]. Some of the measures taken to live up to this role can to some extent get information that contains private information. Human rights limit the extent to which the government can infringe on the individual affairs. Various surveys indicate that members of the public from all over the world are not in favor of government collecting their individual data, that is too sensitive or that can be used for other purposes other than the intended one [4].

There is a growing threat of terrorism in the world given the emergence of more terrorism groups and affiliates organizations in various parts of the world [5]. This creates a need to have a formidable security strategy that will neutralize this threat while at the same time preventing infringement of the rights of the individuals who should be protected. In light of this, the findings of this study will play an imperative role in shaping the opinion of policy makers in matters of security. Specifically, the study will provide vital information relating to the public behavior and attitude toward getting private information for purposes of security [6]. The information will play an important role in helping the security policy makers to know the extent to which the general public will perceive the collection of various kinds of information and thus help them to come up with security strategies that will be acceptable on the social setting and effective in its objective.

Various events have shaped the intensity of security measures taken by various national and international security organs [7]. In order to address the daily security needs of the public, the security agencies have sought to fill the information gap by collecting as much information as they deem necessary for reasons of safety. This has added fuel to the heated debate about the trade-off between national security and individual rights to privacy. Therefore, this study endeavors to answer the question: what are individuals' attitudes towards information privacy and security? The research questions to be answered in this paper are:

RQ1: *What are people's attitudes towards the issue of individual privacy versus national security?*
RQ2: *How do individual characteristics (i.e. gender, field of study, religious belief) influence these attitudes?*

2 Literature Review

In 2013, Pew Research carried out a survey, asking its respondents whether they have greater concern for antiterrorism policies as far as restricting the civil liberties is concerned. The research followed the September 2001 terrorist attack on the USA that heightened the sense of insecurity among Americans. The study revealed a balance of opinion in favor of protection policies. 47% of the respondents said they were more concerned that the security policies by the government have not gone too far in assuring adequate security and protection for the country. On the other hand, 32% of the respondents considered the government to have gone too far when it comes to restricting the personal civil liberties. However, this research gathered the opinion of the people for years close to 9/11 attack, and in recent years it was carried out during the commemoration of 9/11 attack. This skewed the findings in favor of protection policy [8]. The study revealed that as the time goes on since the attack, the opinion for strict government policies are reducing with people increasing unwilling to sacrifice their civil liberties to combat the threat of terrorism [9]. To ensure that the opinions are not skewed, or shaped by the people's emotion, this study will examine the debate over national security versus civil liberty at a time not close to September 11 commemoration.

The shifting balance between security and privacy variables forced Wilson [10] to reexamine the way in which Americans citizens want their government to protect them and what they are willing to sacrifice to ensure their wellbeing. The security agencies in the USA admittedly sacrificed some of the individual privacy, following the September 2001 attack, and the study sought to investigate whether Americans must sacrifice their civil liberties to ensure security. The study found out that most people were of the opinion that even though the government has the responsibility of protecting its citizens, its reaction after September 2001, have been an unnecessary violation of privacy with the inflated threat of terrorism. The study relied much on empirical data and legal cases that have tried to challenge the extended protection measures by the government. This failed to integrate the view of various groups (religion and other social groups and affiliations) in the society. To get information that will fill this gap, my study will analyze the opinions of general public factoring in views of various opinion drivers in the society such as religion.

According to Knowles [11], the national governments have the responsibility of protecting its resources and citizens against any form of security threat. Members of the public value their security and they at times sacrifice some liberties to protect it. The growth in technology has necessitated close monitoring of all potential threats. For this reason, the government is under obligations to collect all forms of information that would assist to identify the threats before they materialize into life threatening situation. Members of the public hold their government responsible for protecting their wellbeing and life and as such, they should be able to trust that the government does not have any ill motivate when it collects data that will help to live up to this mandate [12]. Even though majority of the citizens in the USA are against various measures taken by the government to try to legalize its surveillance on individuals, it is agreeable that terrorism can take any form and deterrence measures that will not result in death or injury of any person should be pursued as far as possible [3]. As such, by collecting individual private data and surveillance on the activities of its citizens, the government reinforces its authority and helps to ensure social stability is maintained and that any form of internal or external aggression is prevented.

Governments collect phone logs and Internet data from the citizens as part of its mission to maintain national security. However, in a study carried out by Betts and Sezer [13] shows that 57% of Americans believe that surveillance programs carried out by the government would not influence the ability of the government to prevent terms attacks. In as much as the government tries to stop all the avenues that bleed terrorism including financial transactions and communication, it is unethical to tap on everything that people do on the internet or say through their phones [14]. Given the mammoth size of the government, it is likely that the information may be misused by those collecting it which may cause more harm and damage to the victim than lack of the information would have caused to the efforts of fighting terrorism [15].

There was a heightened public outcry following the release of the information in 2013 that the America's intelligence collecting organization NSA was gathering metadata on all customers of Verizon Through its top secret program (PRISM), NSA was discovered to have been gathering data from the largest information technology companies including Apple, Google, Microsoft and Facebook, directly from their servers [16]. Even though NSA defended its action by claiming that specific requests for customer data from the IT companies were subject to the legal controls the opposition to such measures created a sense of fear among most Americans regarding the extent to which their conversations can be considered private. According to Abdulhamid et al., [17] collecting general public's private data nullifies the concept of privacy which is entitled to any person both within and without U.S.A legally and ethically [18]. As such, the government should take all the necessary steps to ensure that privacy rights are respected while at the same time they should collect only those data that are relevant to their security policy.

3 Method

This section presents the study population and sampling, data collection instruments, methods and technique use to analyze the data.

3.1 Study Population and Sampling

The study used quantitative and qualitative data that was collected from people from different backgrounds and geographical locations. The sample of the study is made up of 243 online users. This is a sufficient sample that will help to depict the characters and behavior of the entire population [19]. The sample has been selected randomly.

3.2 Research Design

The study will tell us the cross-sectional design where all the data will be collected at one point in time [16]. This design is useful as it helps to save on time and resources required to carry out the research. The data collected will be both primary and secondary. The primary data will be the main type of data collected from the surveys. The secondary data will be used to complement the primary data and will be collected from the reliable sources such as journals and previous studies.

3.3 Data Collection and Analysis

The data will be collected through the online survey. The participants were randomly assigned to evaluate three statements, which are: "Citizen Privacy takes precedence over national security," "Governments should have access to all encrypted data," and "Individual privacy is a human right." For each statement, we measured participants' level of agreement using a 5-point Likert scale. Using a one-way analysis of variance (ANOVA), we examined if privacy attitudes were different among four user characteristics: gender, religious belief, field of study and educational level. In our study, we divided participants into religious (Catholic, Christian, Muslim, Jewish, Unitarian) and non-religious (Agnostic, Atheist and none).

4 Analysis and Findings

4.1 Participants

This survey was distributed using two methods. The first method was Amazon's Mechanical Turk (MTurk), an online crowdsourcing system. We compensated MTurk participants $0.10 USD per survey. The survey was available only to U.S. residents with at least a 95% approval rating (a screening option that MTurk provides). We received a total of 197 surveys from MTurkers, 172 of which we considered valid. The second method was a survey that was distributed through the Facebook and Reddit social media platforms. Participants were not paid for their contribution due to the need to preserve their anonymity. We eliminated empty responses and responses that did not contain complete answers. We received a total of 120 responses, 71 of which were complete. The total number of responses from both was 243.

The data showed that 125 (52.1%) of the respondents were males, 110 (45.8%) were females while 5 respondents preferred not to say. Also, 76.1% of the respondents are whites. Furthermore, the average age of the respondents is 32.28 ranging from 18 to 67. 139 (58.2%) of the respondents had never married, 85 (35.6%) of the respondents were married, 10 respondents were divorced, only one respondent is widowed while 4 respondents preferred not to say. The respondents consist of mostly white people (76.1%) with only 21, 14 and 9 Asians, Hispanic/Latino and blacks respectively. Also, majority of the respondents are educated with only one respondent having less than high school education, while 67, 28, 92 and 24 respondents had some college, high school education, bachelor's degree and professional degree respectively.

4.2 Results

To address the first research question regarding participants' attitudes towards the three statements related to individual privacy and national security, we performed descriptive statistics analysis on their levels of agreement, using a 5-point Likert scale ranging from 1 (Strongly disagree) to 5 (Strongly agree = 5). These three statements are:

- Statement 1: *"Citizen Privacy takes precedence over national security"*
- Statement 2: *"Governments should have access to all encrypted data"*
- Statement 3: *"Individual privacy is a human right"*

Table 1. Descriptive analysis of agreement for each topic

	Statement 1	Statement 2	Statement 3
N (number of participants)	108	126	121
M (Mean)	3.11	1.89	4.12
SD (Standard deviation)	1.105	1.097	0.993

Table 1 summarized our findings about participants' view towards these three statements. Specifically, the majority of participants considered individual privacy as a basic human right (M = 4.12), and held a doubtful view about government's access to encrypted data (M = 1.89). Their opinions were more neutral when it comes to the debate between citizen privacy and national security (M = 3.11). Meanwhile, participants' opinions about privacy as a human right were relatively uniformed (SD = 0.993), compared with the situation when the government (SD = 1.097)/national security (SD = 1.105) is involved.

Table 2. Correlations of agreement levels for all three statements

	Statement 1	Statement 2	Statement 3
Statement 1	–	−0.321	0.354**
Statement 2	−0.321	–	−0.427**
Statement 3	0.354**	−0.427**	–

**Correlation is significant at the 0.01 level (2-tailed)

Correlations of agreement levels for all three statements is shown in Table 2. The results in the table revealed that "*Individual privacy is a human right*" was positively correlated with "*Citizen Privacy takes precedence over national security*" (r = 0.354, p = 0.007, N = 57). In contrast, "*Individual privacy is a human right*" was negatively correlated with "*Governments should have access to all encrypted data*" (r = −0.427, p = 0.001, N = 58), whereas there was no significant correlation between agreement level of "*Governments should have access to all encrypted data*" and that of "*Citizen Privacy takes precedence over national security*". This suggested that the more participants considered individual privacy as a basic human right, the more they were likely to put it as a priority above national security, and the less they were likely to see government's access to encrypted data as justified movement.

4.3 Factors Influencing Attitudes Towards the "Going Dark" Debate

In order to find out what individual characteristics may affect people's views about the debate between individual privacy and national security, we tested potential differences in factors such as gender, religious belief, field of study, and educational level.

Gender. Aseries of t-tests were performed to find out whether gender differences exist in participants' attitudes. Male participants had statistically significant lower agreement levels of the statement "*Governments should have access to all encrypted data*" (t(124) = −2.007, p = 0.047). No significant gender difference was found in the other two statements (Fig. 1 and Table 3).

Table 3. Agreement level of participants of different gender

Statement	Gender	Mean	S.D.
Statement 1	Male	3.08	1.214
	Female	3.14	1.008
Statement 2*	Male	1.72	1.003
	Female	2.11	1.181
Statement 3	Male	4.11	1.008
	Female	4.12	0.982

*p < 0.05

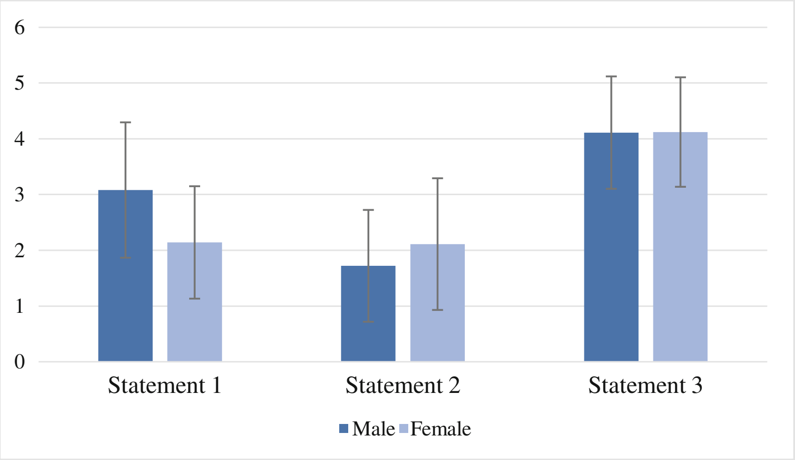


Fig. 1. Agreement level of participants of different gender

Religious Belief. T-test results revealed significant religious belief differences in two out of three statements. Compared to non-religious participants, religious participants showed a stronger support to government's access to encrypted data (t(117) = 2.010, p = 0.047) and remained a relatively conservative attitude towards citizen privacy's precedence to national security (t(95) = −2.236, p = 0.028) (Fig. 2 and Table 4).

Field of Study. To analyze how users' major field of study may affect their privacy-related attitudes, a series of ANOVAs were performed with field of study and agreement levels of each statement as independent and dependent variables respectively. Groups with fewer than two cases were eliminated when performing the analysis. Statistically significant differences across different fields of studies were found in

Table 4. Agreement level of religious vs. non-religious participants

Statement	Religious belief	Mean	S.D.
Statement 1*	Religious	2.74	1.182
	Non-religious	3.29	1.092
Statement 2*	Religious	2.06	1.216
	Non-religious	1.67	0.900
Statement 3	Religious	4.17	0.892
	Non-religious	4.10	0.979

*p < 0.05

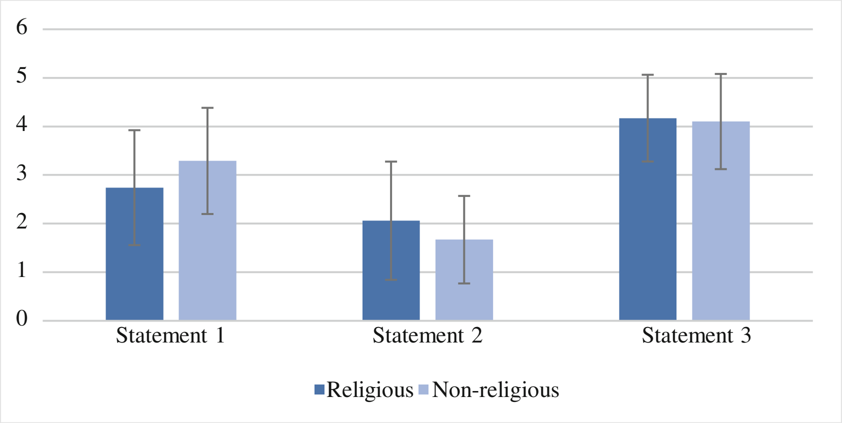


Fig. 2. Agreement level of religious vs. non-religious participants

Table 5. Field of study differences in agreement level of each statement

Field of study		Statement 1	Statement 2*	Statement 3
Social sciences, humanities, liberal arts (N = 46)	Mean (S.D.)	3.64 (1.129)	2.21 (1.228)	4.21 (0.833)
Science, technology, engineering, or Math (N = 80)	Mean (S.D.)	3.00 (1.000)	1.59 (0.858)	4.12 (0.940)
Accounting, business and informatics (N = 31)	Mean (S.D.)	2.73 (1.163)	1.67 (0.856)	4.21 (1.051)
Legal, law enforcement, criminology (N = 5)	Mean (S.D.)	3.00 (0.000)	3.00 (1.826)	3.00 (2.000)
None (N = 40)	Mean (S.D.)	3.00 (1.000)	2.25 (1.152)	4.17 (0.985)
Other (N = 12)	Mean (S.D.)	2.60 (0.894)	2.13 (1.553)	4.25 (1.500)

*p < 0.05

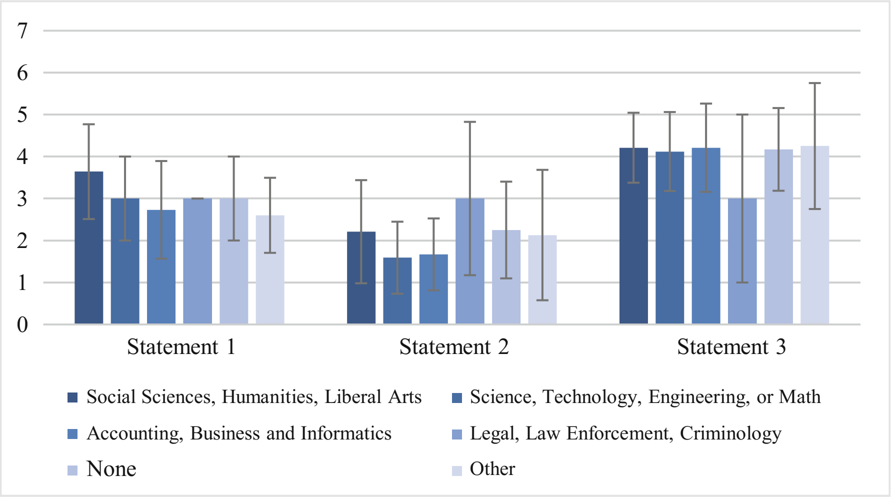


Fig. 3. Field of study differences in agreement level of each statement

the statement “Governments should have access to all encrypted data” (F = 2.744, p = 0.022) (Fig. 3 and Table 5).

Educational Level. One-way ANOVA analysis revealed a statistically significant differences of the agreement level regarding “Governments should have access to all encrypted data” among participants of different educational levels (F = 2.792, p = 0.029). A Tukey post hoc test further revealed that the agreement level of participants with doctoral degrees (3.67 ± 1.528) was significantly than that of participants with bachelor’s degrees (1.008 ± 0.126, p = 0.023) and professional degrees (0.870 ± 0.241, p = 0.027). There was no statistically significant difference between the other groups. Groups with fewer than two cases were eliminated when performing the analysis (Fig. 4 and Table 6).

Table 6. Educational level differences in agreement level of each statement

Educational level		Statement 1	Statement 2*	Statement 3
High school graduate (N = 40)	Mean (S.D.)	3.000 (1.333)	2.00 (1.202)	4.09 (1.136)
Two-year degree – Associate’s degree (N = 117)	Mean (S.D.)	3.18 (1.035)	2.03 (1.124)	4.00 (1.082)
Four-year degree – Bachelor’s degree (N = 147)	Mean (S.D.)	3.13 (1.090)	1.73 (1.008)	4.17 (0.944)
Professional degree (i.e. Law, MBA, etc.) (N = 38)	Mean (S.D.)	2.93 (1.385)	1.62 (0.870)	4.27 (0.647)
Doctoral degree (N = 7)	Mean (S.D.)	3.50 (0.707)	3.67 (1.528)	4.50 (0.707)

*p < 0.05

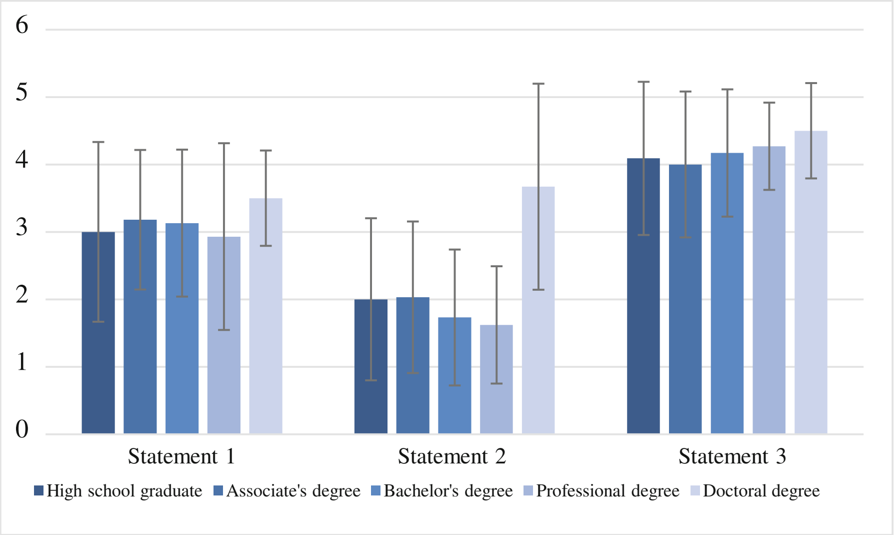


Fig. 4. Educational level differences in agreement level of each statement

5 Discussion

The tension between individuals' right to privacy and national security has long been debated in the US and other parts of the world. One point of view in this debate is that citizens who have "nothing to hide" should not fear government surveillance, and law enforcement should have access to their information whenever necessary [20]. On the other side of the debate, privacy advocates who believe in an individual's right to privacy want to limit government surveillance and access to personal information. People's attitudes and behaviors related to privacy are highly contextualized in the digital age. While many scholars have conceptualized information privacy in various disciplines, investigations of individual users' attitudes and behaviors towards information privacy and security remain limited [21]. We argue that examining user attitudes and opinions on this debate may help scholars, law enforcement officials, and policy makers develop better privacy policies and guidelines. Moreover, it may provide engineers and designers with new ways of improving the design of current privacy-enhancing technologies. In this study, we aimed to fill in this gap by conducting a user survey to assess users' attitudes. The online survey was administrated to 243 online users. Participants were randomly assigned to evaluate three statements, namely, "Citizen Privacy takes precedence over national security," "Governments should have access to all encrypted data," and "Individual privacy is a human right." For each statement, we measured participants' level of agreement using a 5-point Likert scale. Our results indicate that most participants agreed that information privacy is a human right (74%) and most participants disagreed with government's access to encrypted data (78%). However, users had distinct opinions regarding the decision between individual privacy and national security, with only 37% agreeing with the statement, 32% disagreeing, and 31% expressing no opinion. Using a one-way analysis of variance (ANOVA) we examined if privacy attitudes were different among four user characteristics: gender, religious belief, field of study and educational level. In terms of gender, the t-test showed that men had significantly lower agreement with government's access to encrypted data compared to women. For religion analysis, the t-test showed there was a significant variation in attitudes towards citizen privacy and national security between religious (Catholic, Christian, Muslim, Jewish, Unitarian) and non-religious (Agnostic, Atheist and none) participants. In addition, ANOVA analysis revealed that the attitudes towards government's access to encrypted data were varied among participants with different educational level and field of study. Post-hoc analysis further indicated that participants with doctoral degrees showed strong supports to government's access compared to those with bachelor's degrees and professional degrees. This study provides a preliminary understanding of public attitudes towards information privacy and security and how users' individual differences may influence these attitudes. Overall, our results may provide engineers and designers with new ways of improving current privacy-enhancing technologies. They may also help lawmakers develop better privacy-related regulations and policies.

6 Conclusion and Future Work

The study indicates that most of the participants concurred with the statement that privacy constitutes human right and participants did not agree with the government access to the encrypted data. However, there was a distinct opinion between when it comes to the issue of national security and individual privacy. The study also indicates that concerning gender men has a big disagreement with the government's access to the private data in comparison to the women. The issue of privacy and individual rights is highly contentious among different religion divides. In addition, a vast divide was noted in opinion among the various field of study. Individuals studying in the field of Legal, Law Enforcement, Criminology have higher agreement with the government's access to the private data in comparison to other field of study. In conclusion, this study implies people's opinions about the issue of privacy versus national security, government's collection of private data and continuing surveillance on the life of citizens, are influenced by various factors which should be taken into consideration. It is recommended that the policy makers and government strategies should consider these factors that influence the opinion of the public before coming up with security efforts that conflict with human rights.

References

1. Solove, D., Schwartz, P.: Privacy, Law Enforcement and National Security. Wolters Kluwer, New York (2015)
2. Gregory, A.: American Surveillance: Intelligence, Privacy, and the Fourth Amendment. The University of Wisconsin Press, Madison (2016)

3. Namoglu, N., Ulgen, Y.: Network security vulnerabilities and personal privacy issues in healthcare information systems: a case study in a private hospital. In: 2014 18th National BIYOMUT (2014)
4. Marcovici, M.: The Surveillance Society: The Security vs. Privacy Debate. Books on Demand, Norderstedt (2013)
5. Kassahun, D.: Rainwater Harvesting in Ethiopia: Capturing the Realities and Exploring Opportunities. Forum for Social Studies, Addis Ababa (2007)
6. Stefoff, R.: Security vs. privacy: open for debate. Marshall Cavendish Benchmark, New York (2008)
7. Kleinig, J., Mameli, P., Miller, S., Salane, D., Schwartz, A.: Security and Privacy: Global Standards for Ethical Identity Management in Contemporary Liberal Democratic States. ANU E Press, Acton, A.C.T. (2011)
8. Hatfield, E., Cacioppo, J.T., Rapson, R.L.: Emotional contagion. Curr. Dir. Psychol. Sci. **2**(3), 96–100 (1993)
9. Pew Research Center: Balancing act: national security and civil liberties in Post-9/11 era (2013). http://www.pewresearch.org/fact-tank/2013/06/07/balancing-act-national-security-and-civil-liberties-in-post-911-era/. Accessed 12 Feb 2017
10. Wilson, R.B.: A new balance: national security and privacy in a post 9-11 world (2014)
11. Knowles, R.: National security law: up close and personal, an introduction. Valparaiso Univ. Law Rev. **50**(2), 415–417 (2016)
12. Melanson, P.: Secrecy Wars: National Security, Privacy, and the Public's Right to Know. Brassey's, Washington, D.C (2001)
13. Betts, J., Sezer, S.: Ethics and privacy in national security and critical infrastructure protection. In: Proceedings of the IEEE 2014 International Symposium on Ethics in Engineering, Science, and Technology, p. 49. IEEE Press, May 2014
14. Solove, D.: Nothing to Hide: The False Tradeoff Between Privacy and Security. Yale University Press, New Haven Conn (2011)
15. Thomson Reuters: Protecting the Homeland: Balancing National Security and Individual Privacy Interests: Leading Lawyers Weigh National Security Concerns and the Rights of Citizens. West Academic Publishing, Saint Paul (2016)
16. Bertino, E.: Data security and privacy: concepts, approaches, and research directions. In: 2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC), vol. 1, pp. 400–407. IEEE, June 2016
17. Abdulhamid, S.M., Ahmad, S., Waziri, V.O., Jibril, F.N.: Privacy and national security issues in social networks: the challenges (2014). arXiv preprint arXiv:1402.3301
18. Elmaghraby, A.S., Losavio, M.M.: Cyber security challenges in smart cities: safety, security and privacy. J. Adv. Res. **5**(4), 491–497 (2014)
19. Barlett, J.E., Kotrlik, J.W., Higgins, C.C.: Organizational research: determining appropriate sample size in survey research. Inf. Technol. Learn. Perform. J. **19**(1), 43 (2001)
20. Gal, C., Kantor, P., Lesk, M.: Protecting persons while protecting the people. In: Second annual Workshop on Information Privacy and National Security, ISIPS 2008, New Brunswick, NJ, USA, 12 May 2008. Springer, Berlin (2009). Revised selected papers
21. Angwin, J.: Dragnet nation: a quest for privacy, security, and freedom in a world of relentless surveillance. Times Books, Henry Holt and Company, New York (2014)

Identifying Relevance of Security, Privacy, Trust, and Adoption Dimensions Concerning Cloud Computing Applications Employed in Educational Settings

Tihomir Orehovački[1](✉), Snježana Babić[2], and Darko Etinger[1]

[1] Department of Information and Communication Technologies,
Juraj Dobrila University of Pula, Zagrebačka 30, 52100 Pula, Croatia
{tihomir.orehovacki,darko.etinger}@unipu.hr
[2] Polytechnic of Rijeka, Trpimirova 2/V, 51000 Rijeka, Croatia
snjezana.babic@veleri.hr

Abstract. Cloud computing applications are nowadays commonly used in various aspects of human endeavour, and the education is no exception. Although cloud computing applications bring numerous advantages, their adoption could be significantly reduced due to users' concerns related to security, privacy, and trust. This paper introduces a research framework that captures the essence of security, privacy, trust, and adoption in the context of cloud computing applications when used in educational environment. Drawing on an extensive literature review, a finite set of items was determined and consequently employed for the design of the measuring instrument in the form of a post-use questionnaire. With an aim to examine psychometric features of the measuring instrument, an empirical study was carried out. Participants in the study were students from two higher education institutions who employed cloud-based applications for the purpose of creating, sharing, and organizing educational artefacts. Study findings helped us determine the relevance of security, privacy, trust, and adoption dimensions in the context of cloud computing applications as perceived by users who apply them for educational purposes.

Keywords: Cloud computing · Education · Adoption · Trust · Security · Privacy · Post-use questionnaire · Empirical findings

1 Introduction

The need for cutting edge technologies in learning and teaching processes resulted in the introduction of cloud computing applications to the educational settings. Software as a Service (SaaS) is one of cloud service models in which users can only run and use the software on a cloud infrastructure [1]. The cloud computing applications are accessible from various client devices through web browser or a program interface and the consumer does not control the underlying cloud infrastructure (e.g. network, servers, operating systems, storage, or even individual application capabilities) [2]. Such applications enable institutions without their own technical resources required for operation to get access to needed services on demand [3]. Higher education institutions have

D. Nicholson (ed.), *Advances in Human Factors in Cybersecurity*,
Advances in Intelligent Systems and Computing 593, DOI 10.1007/978-3-319-60585-2_29

adopted cloud computing due to next two reasons [4]: (1) cost savings, and (2) scalable and flexible IT services. The most important advantages of using cloud services in educational field are facilitated communication and collaboration among users, enhanced users' productivity, knowledge accessible from any device (e.g. computers, tablet, and mobile phones), reduced time and cost, and encouraged knowledge sharing [5, 6]. Many leading IT companies (e.g. Microsoft, Google, Amazon, and IBM) have adopted the trend of educational cloud computing and have provided tools for learning in cloud to support educational institutions [4, 5]. Majority of cloud computing applications employed in the educational environment like productivity suites (e.g. Google Apps and Office 365) and storage services (e.g. Microsoft OneDrive and Google Drive) are provided as Software as a Service that operate in the public cloud thus serving as extensions to conventional Virtual Learning Environments.

Success of the cloud computing application depends on the set of factors affecting the users' adoption of cloud computing applications. Numerous researchers have successfully modified various existing models (e.g. Technology Acceptance Model (TAM) [7–9] and Unified Theory of Acceptance and Use of Technology (UTAUT) [10]) and have identified many factors that influence the adoption of cloud computing applications. The benefits derived from the advantages and effectiveness of cloud computing services could be mitigated by possible trust, privacy, and security concerns. First, the expanding quantity of personal data in the cloud environment increases the complexity of risk assessment. Second, the issues regarding privacy, security, and trust are the significant barriers to adoption of cloud-based applications in education. Bora and Ahmed [3] found that security concerns relate to risk areas such as external data storage, dependency on the public networks, lack of control, multi-tenancy, and integration with internal security. According to Guilloteau and Mauree [11], when implementing privacy by design concept, objectives such as data minimization, controllability, transparency, user-friendliness, data confidentiality, data quality, and use limitation should be considered. Mutkoski [12] emphasized that data ownership, confidentiality, data privacy, and data protection rights are critical contract terms in many segments where cloud computing is being deployed, and the educational sector is not an exception. The same author argue that data protection and data privacy issues are commonly related to the placement of a very large amount of students', teachers', and institutional data into the hands of a third-party service provider.

The aim of this paper is to identify relevance of security, privacy, trust, and adoption facets with respect to cloud computing applications. Drawing on an extensive literature review, a research framework and corresponding post-use questionnaire were introduced. With an aim to examine their psychometric features, an empirical study was carried out. The analysis of collected data uncovered the relevance of security, privacy, trust, and adoption dimensions in the context of cloud computing applications as perceived by users who apply them for educational purposes. The remainder of the paper is structured as follows. Theoretical foundation of our work is briefly described in the following section. Employed research framework is introduced in the third section. Study findings are outlined in the fourth section. Concluding remarks and future work directions are provided in the last section.

2 Background to the Research

Storage services are one of the most widely used cloud computing applications because they support deposit of all users' important data and facilitate backup of files [13]. Based on the analysis of empirical studies, Meske et al. [14] found that sharing with others, full text search functionality, and simultaneous editing are the most important reasons for using cloud storage services (e.g. Google Drive, Dropbox, SkyDrive, and Amazon Cloud Drive) in higher education. On the other hand, the most common cause for rejecting these services is low confidence in data protection with respect to both privacy and security [14]. Svantesson and Clarke [15] argue that cloud computing is associated with serious risks related to the privacy and rights of consumers. Adrian [16] emphasize that individual's control over distribution of his/her personal information protects the individual's integrity and dignity in a manner that information in not being used in ways which are damaging or embarrassing to the individual. According to Mollah et al. [17], main data security challenges in the context of mobile cloud computing include data loss, data breach, data recovery, data locality, and data privacy where data loss and data breach violate two security requirements such as integrity and confidentiality. In their theoretical framework on the dimensionality of Internet users' information privacy concerns (IUIPC) Malhotra et al. [18] proposed control, awareness, and collection as new facets of privacy concerns. According to Arpaci [19], key attributes of security are confidentiality (prevention of data access by unauthorized users), integrity (protecting personal data from unauthorized modification, deletion, or fabrication), and availability (accessibility and usability of the services and data when needed).

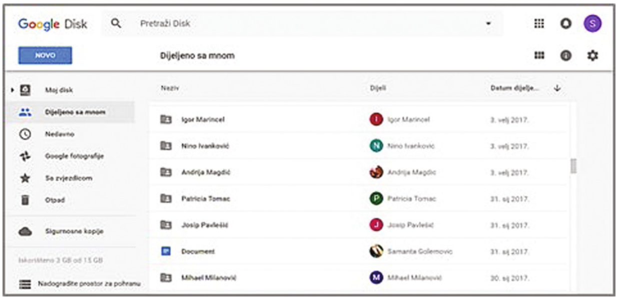
Yang and Lin [13] found that users' perceived usefulness was positively influenced by their continuance intention to use cloud storage services (e.g. Google Drive and Dropbox) while risks related to privacy protection and privacy policy had negative moderating effects on the perceived usefulness and the continuance intention. Drawing on the UTAUT model, Hashim and Hassan [10] uncovered that the most positive effect on the behavioral intention to adopt cloud computing services (e.g. Google Apps) has performance expectancy followed by effort expectancy, social influence, security, and trust. By extending TAM model, Changchit [7] revealed that perceived usefulness, ease of use, security, speed of access, and cost of usage positively affect adoption of cloud computing services. Results of the study conducted by Arpaci [19] imply that 52% of the variance in trust can be explained with combinatory effects of perceived security and perceived privacy and that perceived usefulness, trust, and subjective norm have a significant positive effect on students' attitudes which in turn is a significant predictor of intentions in using mobile cloud storage services (e.g. Dropbox, iCloud, SkyDrive, and Google Drive). By analyzing the students' adoption of cloud computing application Google Docs, Nakayama and Taylor [20] confirmed that perceived risks have negative impact on trust in cloud computing applications while privacy concerns are not significant driver in that respect. In addition, the same authors discovered that users' satisfaction significantly increases their trust in cloud computing applications which in turn positively affects users' intention to use cloud computing applications, but that perceived risks have a negative impact on the increase of users' trust in cloud technology. By exploiting Theory of Planned Behaviour (TPB), Arpaci et al. [21] uncovered that security and privacy have a strong significant influence on the students' attitudes towards using cloud services (e.g. Google Drive and Dropbox) in educational settings. Based on the framework composed of dimensions from TPB, TAM, computer learning theories, and social and economic exchange theories, Li and Chang [8] discovered that perceived security and privacy concerns related to cloud computing applications have a positive impact on perceived risk regarding the interaction with these applications which has a negative impact on the users' attitude toward cloud computing applications such as Office Web Apps and Google Docs. Through the integration of various dimensions originating from service quality, self-efficacy, a motivational model, TAM, the Theory of Reasoned Action (TRA), TPB, and Innovation Diffusion Theory (IDT), Shiau and Chau [9] found that perceived usefulness has the strongest positive effect on user's intention to employ cloud computing applications for educational purposes which was followed by attitude, cloud service quality, perceived behaviour control, result demonstration, visibility, and cloud self-efficacy. On the other hand, the same authors reported that perceived ease use, perceived playfulness, application service quality, compatibility, subjective norm, trialability, and voluntariness do not have significant impact on students' intentions to use cloud computing classroom. According to results of study conducted by Flavián and Guinalíu [22], trust has a positive effect on individual's loyalty to a Web site and also plays significant mediating role between perceived security and loyalty to a Web site.

Current studies in the field are mostly focused on exploring product-oriented security (e.g. [23–29]) and examining law-based privacy policies (e.g. [7, 16, 30]). On the other hand, user-centred studies dealing with an interplay of security, privacy, and trust dimensions in the context of the adoption of cloud based applications are rather rare and mainly treat security, privacy, trust, and adoption as one-dimensional constructs (e.g. [13, 19, 21, 31]). Moreover, the extant body of knowledge lacks studies regarding the security, privacy, and trust concerns when cloud computing applications are applied in the educational settings. All the aforementioned motivated us to initiate a research whose design is described in the following section.

3 Research Design

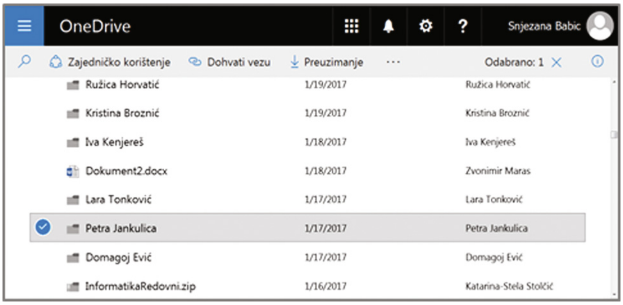
3.1 Procedure

The study was carried out during the winter semester of the academic year 2016/17 in controlled lab conditions and was comprised of two parts: (1) scenario-based interaction with two cloud computing application designed for artefacts management and (2) the employment of a post-use questionnaire for the purpose of evaluating facets of security, privacy, trust, and adoption in the context of the aforementioned cloud computing applications. Upon arriving to the lab, the participants were welcomed and briefly familiarized with the study. At the beginning of the scenario performance session, the form containing a list of 12 representative steps of interaction was given to each participant. Study subjects were asked to complete all scenario steps twice – first by means of Google Drive and then using the Microsoft OneDrive (depicted in Figs. 1 and 2, respectively). After finishing all the scenario steps with both cloud computing applications, the participants were asked to fill in the post-use questionnaire. At the end of the study, respondents were debriefed, and thanked for their participation. The duration of the study was 40 min.



Naziv	Dijeli	Datum dijelje...
Igor Marincel	Igor Marincel	3. velj 2017.
Nino Ivanković	Nino Ivanković	3. velj 2017.
Andrija Magdić	Andrija Magdić	3. velj 2017.
Patricia Tomac	Patricia Tomac	31. sij 2017.
Josip Pavlešić	Josip Pavlešić	31. sij 2017.
Document	Samanta Golemović	31. sij 2017.
Mihael Milanović	Mihael Milanović	30. sij 2017.

Fig. 1. Google Drive



Ružica Horvatić	1/19/2017	Ružica Horvatić
Kristina Broznić	1/19/2017	Kristina Broznić
Iva Kenjereš	1/18/2017	Iva Kenjereš
Dokument2.docx	1/18/2017	Zvonimir Maras
Lara Tonković	1/17/2017	Lara Tonković
Petra Jankulica	1/17/2017	Petra Jankulica
Domagoj Ević	1/17/2017	Domagoj Ević
InformatikaRedovni.zip	1/16/2017	Katarina-Stela Stolčić

Fig. 2. Microsoft OneDrive

3.2 Apparatus

The post-use questionnaire was administrated online by means of the KwikSurveys questionnaire builder. The questionnaire was composed of 16 items related to participants' demography and 82 items designed for measuring dimensions of security, privacy, trust, and adoption. Items on security and privacy were adopted from Cheung and Lee [32], Flavián and Guinalíu [22], Janda et al. [33], O'Cass and Fenech [34], and Ranganathan and Ganapathy [35], items designed for measuring trust were adopted from Kumar et al. [36], Siguaw et al. [37], and Roy et al. [38], items meant for evaluating satisfaction and confirmation of expectations were adopted from Bhattacherjee [39], items created for assessing perceived ease of use, perceived usefulness, social norms, and frequency of use were adopted from Venkatesh and Bala [40] and Venkatesh et al. [41], whereas items for measuring attitude towards use and playfulness were adopted from Moon and Kim [42]. Responses to the post-use questionnaire items were modulated on a five point Likert scale (1 – strongly agree, 5 – strongly disagree). The psychometric features of attributes meant for measuring aspects of perceived security and privacy were explored and reported in [43]. Internal consistency of scales was evaluated with Cronbach's Alpha coefficient. Differences between evaluated cloud computing applications were examined with Wilcoxon Signed-Rank Tests. The reason why we have employed this non-parametric equivalent of the dependent t-test is because results of Shapiro-Wilk Tests revealed that at least one of the variables in a pairwise comparison significantly deviates from a normal distribution ($p < .05$). Consequently, all the reported results are expressed as the median values. The relevance of each identified significant difference was analyzed by means of effect size (r) indicator. It was estimated by dividing Z-value by square root of number of observations. The values of .10, .30, and .50 denote small, medium, and large effect size, respectively [44].

3.3 Framework

The research framework is composed of 17 constructs aimed for measuring various facets of adoption, security, privacy, and trust with respect to cloud computing applications used in educational environments. *Adoption* refers to the extent to which: users commonly employ cloud computing application (Frequency of Use), majority of people that are important to the user think that he/she should employ cloud computing application (Social Norms), users like the idea of employing cloud computing application (Attitude Towards Use), the employment of cloud computing application enhances users' performance in managing artefacts (Perceived Usefulness), is easy for users to become proficient in interaction with cloud computing application (Perceived Ease of Use), cloud computing application is capable to hold the users' attention and stimulate their imagination (Playfulness), interaction with cloud computing application has met users' expectations (Confirmation of Expectations), users are content with employing the cloud computing application (Satisfaction), users are willing to continue to use cloud computing application and recommend it to others (Loyalty). *Trust* denotes the degree to which: cloud computing application takes care about interests of its users and is characterized by clarity of the services it offers to users (Benevolence and Honesty), cloud computing application is receptive to the needs of its users and has all resources required to successfully perform its activities (Receptiveness and Competence). *Security* refers to the level to which cloud computing has implemented high-quality mechanisms that prevent unauthorized access to users' account (Integrity) and unwarranted use and modification of users' data and artefacts (Confidentiality). *Privacy* denotes the extent to which: users are concerned about the privacy of their data and artefacts stored on cloud computing application (Concerns), cloud computing applications take care about privacy protection of its users (Protection), users believe is risky to provide cloud computing application with their personal data (Risks), users think they have control over who has access to and is using their personal data (Control).

4 Results

4.1 Participants

A total of 318 respondents (67.30% male, 32.70% female), aged 21.03 years (SD = 4.197) on average, participated in the study. At the time study was carried out, majority of them (50.31%) were students at Juraj Dobrila University of Pula, Department of Information and Communication Technologies, while remaining 49.69% were enrolled in one of study programs at Polytechnic of Rijeka. Most of the study participants (80.50%) were full-time students. When the computer literacy is taken into account, study subjects are proficient users of both computers and the Internet. Namely, they have between 2 and 29 years (M = 11.82, SD = 3.559) of experience in interaction with computers and between 2 and 20 years (M = 9.76, SD = 3.092) of experience in using the Internet. Furthermore, 74.21% and 82.08% of participants believe that their computer skills and Internet skills, respectively, are at least very good. When the frequency of using the Internet for different purposes is considered, 69.50% of respondents is employing it for communication at least 11 h per week, 60.06% of students is using the Internet for educational purposes between 4 and 20 h per week, 71.07% of participants is using the Internet for fun more than 11 h per week, and 41.82% of students is using the Internet for business purposes at least one hour per week. Study participants had also been loyal users of popular Web 2.0 applications. More specifically, 65.55% respondents have been socializing on Facebook for more than 6 years, 52.86% of them have been podcasting on YouTube for more than 7 years, whereas 67.96% of students have been sharing their moments with a community for less than 2 years. Regarding the length of using Google Drive and Microsoft OneDrive, 49.16% of participants have been using them for more than one year, whereas 12.04% have not used these cloud computing applications prior to this study.

4.2 Findings

Values of the Cronbach's Alpha coefficient were in range from .723 (in the context of measuring the Receptiveness and Competence of Microsoft OneDrive) to .945 (in the case of evaluating Loyalty of Google Drive) thus indicating that internal consistency of scales was deemed adequate [30]. Results of assessing the reliability of constructs that constitute the post-use questionnaire are presented in Table 1.

The analysis of collected data uncovered that respondents are used to employ Google Drive (Mdn = 16) significantly (Z = −9.756, p = .000, r = −.29) more often than to use Microsoft OneDrive (Mdn = 20). It was also found that significantly (Z = −4.549, p = . 000, r = −.18) more persons that are important to study subjects believe they should use Google Drive (Mdn = 9) rather than employ Microsoft OneDrive (Mdn = 10). Study findings indicate that significantly (Z = −6.213, p = .000, r = −.25) more users are feeling positive about employing Google Drive (Mdn = 8) than using Microsoft OneDrive (Mdn = 9). Wilcoxon Signed-Rank Test revealed that Google Drive (Mdn = 14) enhances study participants' performance in managing artefacts to significantly (Z = −5.680, p = . 000, r = −.23) higher extent than Microsoft Drive (Mdn = 15) does. In addition, it appeared that is significantly (Z = −7.531, p = .000, r = −.30) easier for respondents to become proficient in employing Google Drive (Mdn = 11) than applying Microsoft OneDrive (Mdn = 12). Outcomes of data analysis are also implying that Google Drive (Mdn = 19) is capable to hold the users' attention and stimulate their imagination to a significantly (Z = −6.094, p = .000, r = −.24) greater degree than Microsoft OneDrive (Mdn = 20) is. It was also discovered that Google Drive (Mdn = 8) has met users' expectation to a significantly (Z = −5.811, p = .000, r = −.23) higher level than Microsoft

OneDrive (Mdn = 10) has. Furthermore, significantly (Z = −6.054, p = .000, r = −.24) more respondents are pleased with using the Google Drive (Mdn = 10) than with the employment of Microsoft OneDrive (Mdn = 12). Likewise, significantly (Z = −6.943, p = .000, r = −.28) more study participants reported they are willing to continue to use Google Drive (Mdn = 11) and recommend it to others than they would do the same in the case of Microsoft OneDrive (Mdn = 15). All the aforementioned suggests that significantly (Z = −7.838, p = .000, r = −.31) more students would adopt Google Drive (Mdn = 110) than they would accept Microsoft OneDrive (Mdn = 122.50).

Table 1. Reliability of scales

	Number of items	Cronbach's α[a]	
		Google Drive	Microsoft OneDrive
Adoption			
Frequency of Use	5	.899	.919
Social Norms	3	.908	.916
Attitude Towards Use	4	.889	.905
Perceived Usefulness	7	.905	.898
Perceived Ease of Use	6	.926	.936
Playfulness	6	.770	.752
Confirmation of Expectations	4	.856	.850
Satisfaction	5	.920	.926
Loyalty	5	.945	.935
Trust			
Benevolence and Honesty	6	.820	.839
Receptiveness and Competence	4	.750	.723
Security			
Integrity	7	.929	.934
Confidentiality	4	.787	.792
Privacy			
Concerns	4	.858	.864
Protection	5	.863	.862
Risks	5	.903	.901
Control	2	.880	.861

[a] Threshold value in exploratory research [30] > .600

Study findings are implying that perceived trust in Google Drive (Mdn = 22) is significantly (Z = −5.486, p = .000, r = −.22) higher than in Microsoft OneDrive (Mdn = 23). Namely, it appeared that Google Drive (Mdn = 12) is significantly (Z = −5.420, p = .000, r = −.22) more concerned with the interests of its users than Microsoft OneDrive (Mdn = 13) is. It was also discovered that Google Drive (Mdn = 9) was perceived by respondents as significantly (Z = −4.295, p = .000, r = −.17) more competent for conducting its activities and receptive to users' needs than Microsoft OneDrive (Mdn = 9) was.

According to the results of data analysis, Google Drive (Mdn = 24) was perceived by study subjects as significantly (Z = −4.132, p = .000, r = −.16) more secure cloud computing applications than Microsoft OneDrive (Mdn = 25) was. Namely, it was discovered that quality of mechanisms that protect unauthorized access to users' artefacts is significantly (Z = −3.952, p = .000, r = −.16) higher in the case of Google Drive (Mdn = 16) than those integrated in Microsoft OneDrive (Mdn = 17). Wilcoxon Signed-Rank Test also uncovered that quality of mechanisms that protect unauthorized use and modification of users' data is significantly (Z = −3.643, p = .000, r = −.15) better in the context of Google Drive (Mdn = 8) than it is in the case of Microsoft OneDrive (Mdn = 8).

Table 2. Outcomes of data analysis (note that a lower score of median values indicates a better result)

	Z	Effects in size (r)	Median values	
			Google Drive	Microsoft OneDrive
Adoption	**−7.838**	**−.31**	**110.00**	**122.50**
Frequency of Use	−7.278	−.29	16.00	20.00
Social Norms	−4.549	−.18	9.00	10.00
Attitude Towards Use	−6.213	−.25	8.00	9.00
Perceived Usefulness	−5.680	−.23	14.00	15.00
Perceived Ease of Use	−7.531	−.30	11.00	12.00
Playfulness	−6.094	−.24	19.00	20.00
Confirmation of Expectations	−5.811	−.23	8.00	10.00
Satisfaction	−6.054	−.24	10.00	12.00
Loyalty	−6.943	−.28	11.00	15.00
Trust	**−5.486**	**−.22**	**22.00**	**23.00**
Benevolence and Honesty	−5.420	−.22	12.00	13.00
Receptiveness and Competence	−4.295	−.17	9.00	9.00
Security	**−4.132**	**−.16**	**24.00**	**25.00**
Integrity	−3.952	−.16	16.00	17.00
Confidentiality	−3.643	−.15	8.00	8.00
Privacy	**−2.263**	**−.09**	**46.00**	**46.00**
Concerns	−.791	N/A	12.00	12.00
Protection	−2.913	−.12	11.00	12.00
Risks	−.553	N/A	15.00	15.00
Control	−2.015	−.08	6.00	6.00

[a] Google Drive > Microsoft OneDrive

The analysis of collected data suggests that perceived privacy of Google Drive (Mdn = 46) is significantly (Z = −2.263, p = .023, r = −.09) higher than those of Microsoft OneDrive (Mdn = 46). More specifically, study participants reported that Google Drive (Mdn = 11) takes care of the privacy of its users to significantly (Z = −2.913, p = .004, r = −.12) higher extent than Microsoft OneDrive (Mdn = 12) does. It was also found that when employing Google Drive (Mdn = 6), respondents have significantly (Z = −2.015, p = .044, r = −.08) more control over who has an access to their personal data than they have when they are using Microsoft OneDrive (Mdn = 6). However, no significant difference between evaluated cloud computing applications was discovered with respect to the extent to which users are concerned about the privacy of their personal data when employing them (Z = −.791, p = .429) nor to the degree to which users believe is risky to disclose personal information to them (Z = −.553, p = . 580). Reported study findings are summarized in Table 2.

5 Discussion and Concluding Remarks

The aim of the work presented in this paper was to examine relevance of various dimensions of security, privacy, trust, and adoption in the context of cloud computing applications commonly employed in educational ecosystem. For that purpose, an empirical study was carried out during which a within-subjects research design contrasting two cloud-based services was adopted. Findings of the study suggest that composite measures which represent a sum of participants' responses revealed 9.52% medium in size and 71.43% small in size differences between evaluated cloud computing applications at different levels of granularity in a research framework. When medium in size differences are considered, the highest difference between Google Drive and Microsoft OneDrive was found in terms of composite measure that reflects the overall adoption of the aforementioned applications. This is mainly influenced by the extent to which they differ in the context of the effortlessness of their employment. Regarding the identified small in size differences, the highest among them belongs to the composite measure that denotes the level to which users frequently employ cloud computing applications whereas the lowest small in size difference was determined with respect to the composite measure that indicates the extent to which cloud computing applications are concerned about privacy protection of their users. It was also discovered that difference between examined cloud applications was below the .10 threshold for small effects in size in the case of composite measure that reflects overall perceived privacy as well as in terms of composite measure that evaluates the level to which users have control over who can access their personal data and artefacts. Finally, it appeared that composite measures meant for exploring privacy concerns and privacy risks have not revealed significant differences between cloud computing applications that were involved in the study. All the set forth speaks in favor of validity of the introduced research framework and employed post-use questionnaire which makes them both applicable as a foundation for future theoretical advances in the field as well as for measuring and improving facets of security, privacy, trust, and adoption of cloud computing applications. Taking into account that reported findings are a constituent part of an ongoing research, our future work will be focused on the assessment of psychometric characteristics of the conceptual model that will reflect interplay among drivers of the proposed research framework.

References

1. Aldossary, S., Allen, W.: Data security, privacy, availability and integrity in cloud computing: issues and current solutions. Int. J. Adv. Comput. Sci. Appl. **7**(4), 485–498 (2016)
2. Mell, P., Grace, T.: The NIST Definition of Cloud Computing. National Institutes of Standards (NIST) Special Publication, 800-145 (2011)
3. Bora, U.J., Ahmed, M.: E-learning using cloud computing. Int. J. Sci. Mod. Eng. **1**(2), 9–13 (2013)
4. Alharthi, A., Yahya, F., Walters, R.J., Wills, G.: An overview of cloud services adoption challenges in higher education institutions. In: Emerging Software as a Service and Analytics, Lisbon, Portugal, pp. 1–8 (2015)
5. Al-Jebreen, B., Dahanayake, A., Syed, L.: Advances in higher educational resource sharing and cloud services for KSA. Int. J. Comput. Sci. Eng. Surv. (IJCSES) **6**(3), 25–40 (2015)
6. Hashim, H.S., Hassan, Z.B., Hashim, A.S.: The benefits and challenges of cloud computing adoption on Iraqi Universities: results from an empirical study. J. Appl. Sci. Res. **11**(13), 14–21 (2015)
7. Changchit, C.: Students' perceptions of cloud computing. Iss. Inf. Syst. **15**(1), 312–322 (2014)
8. Li, Y., Chang, K.: A study on user acceptance of cloud computing: a multi-theoretical perspective. In: Proceedings of AMCIS 2012 (2012). Paper 19
9. Shiau, W.L., Chau, P.Y.: Understanding behavioral intention to use a cloud computing classroom: a multiple model comparison approach. Inf. Manag. **53**(3), 355–365 (2016)
10. Hashim, H.S., Hassan, Z.B.: Factors that influence the users' adoption of cloud computing services at Iraqi Universities: an empirical study. Aust. J. Basic Appl. Sci. **9**(27), 379–390 (2015). UTUT
11. Guilloteau, S., Mauree, V.: Privacy in cloud computing, ITU-T technology watch report (2012). http://www.itu.int/dms_pub/itu-t/oth/23/01/T23010000160001PDFE.pdf. Accessed 28 Nov 2016
12. Mutkoski, S.: Cloud computing, regulatory compliance, and student privacy: a guide for school administrators and legal counsel. John Marshall J. Inf. Technol. Priv. Law **30**(3), 511–534 (2014)
13. Yang, H.-L., Lin, S.-L.: User continuance intention to use cloud storage service. Comput. Hum. Behav. **52**, 219–232 (2015)
14. Meske, C., Stieglitz, S., Vogl, R., Rudolph, D., Öksüz, A.: Cloud storage services in higher education–results of a preliminary study in the context of the Sync&Share-Project in Germany. In: International Conference on Learning and Collaboration Technologies, pp. 161–171. Springer International Publishing (2014)
15. Svantesson, D., Clarke, R.: Privacy and consumer risks in cloud computing. Comput. Law Secur. Rev. **26**(4), 391–397 (2010)
16. Adrian, A.: How much privacy do clouds provide? An Australian perspective. Comput. Law Secur. Rev. **29**(1), 48–57 (2013)
17. Mollah, M.B., Azad, M.A.K., Vasilakos, A.: Security and privacy challenges in mobile cloud computing: survey and way ahead. J. Netw. Comput. Appl. **84**, 38–54 (2017)
18. Malhotra, N.K., Kim, S.S., Agarwal, J.: Internet Users' Information Privacy Concerns (IUIPC): the construct, the scale, and a causal model. Inf. Syst. Res. **15**(4), 336–355 (2014)

19. Arpaci, I.: Understanding and predicting students' intention to use mobile cloud storage services. Comput. Hum. Behav. **58**, 150–157 (2016)
20. Nakayama, M., Taylor, C.: The effects of perceived functionality and usability on privacy and security concerns about adopting cloud application adoptions. In: Proceedings of the Conference on Information Systems Applied Research (2016). ISSN 2167-1508
21. Arpaci, I., Kilicer, K., Bardakci, S.: Effects of security and privacy concerns on educational use of cloud services. Comput. Hum. Behav. **45**, 93–98 (2015)
22. Flavián, C., Guinalíu, M.: Consumer trust, perceived security and privacy policy: three basic elements of loyalty to a web site. Ind. Manag. Data Syst. **106**(5), 601–620 (2006)
23. Islam, T., Manivannan, D., Zeadally, S.: A classification and characterization of security threats in cloud computing. Int. J. Next-Generation Comput. **7**(1), 1–17 (2016)
24. Khan, M.A.: A survey of security issues for cloud computing. J. Netw. Comput. Appl. **71**, 11–29 (2016)
25. Popović, K., Hocenski, Ž.: Cloud computing security issues and challenges. In: Proceedings of the 33rd International MIPRO Convention, pp. 344–349. IEEE, Opatija (2010)
26. Ryan, M.D.: Cloud computing security: the scientific challenge, and a survey of solutions. J. Syst. Softw. **86**(9), 2263–2268 (2013)
27. Singh, S., Jeong, Y.-S., Park, J.H.: A survey on cloud computing security: issues, threats, and solutions. J. Netw. Comput. Appl. **75**, 200–222 (2016)
28. Subashini, S., Kavitha, V.: A survey on security issues in service delivery models of cloud computing. J. Netw. Comput. Appl. **34**(1), 1–11 (2011)
29. Zissis, D., Lekkas, D.: Addressing cloud computing security issues. Future Gener. Comput. Syst. **28**(3), 583–592 (2012)
30. Hair, J.F., Ringle, C.M., Sarstedt, M.: PLS-SEM: indeed a silver bullet. J. Mark. Theory Pract. **19**(2), 139–151 (2011)
31. Gupta, P., Seetharaman, A., Raj, J.R.: The usage and adoption of cloud computing by small and medium businesses. Int. J. Inf. Manag. **33**(5), 861–874 (2013)
32. Cheung, C.M.K., Lee, M.K.O.: Trust in internet shopping: instrument development and validation through classical and modern approaches. J. Glob. Inf. Manag. **9**(3), 23–35 (2001)
33. Janda, S., Trocchia, P., Gwinner, K.: Consumer perceptions of internet retail service quality. Int. J. Serv. Ind. Manag. **13**(5), 412–431 (2002)
34. O'Cass, A., Fenech, T.: Web retailing adoption: exploring the nature of internet users web retailing behaviour. J. Retail. Consum. Serv. **10**(2), 81–94 (2003)
35. Ranganathan, C., Ganapathy, S.: Key dimensions of business-to-consumer web sites. Inf. Manag. **39**(6), 457–465 (2002)
36. Kumar, N., Scheer, L.K., Steenkamp, J.-B.E.M.: The effects of supplier fairness on vulnerable resellers. J. Mark. Res. **32**(1), 42–53 (1995)
37. Siguaw, J., Simpson, P., Baker, T.: Effects of supplier market orientation on distributor market orientation and the channel relationship: the distributor perspective. J. Mark. **62**, 99–111 (1998)
38. Roy, M., Dewit, O., Aubert, B.: The impact of interface usability on trust in web retailers. Internet Res. Electron. Netw. Appl. Policy **11**(5), 388–398 (2001)
39. Bhattacherjee, A.: Understanding information systems continuance: an expectation confirmation model. MIS Q. **25**(3), 351–370 (2001)
40. Venkatesh, V., Bala, H.: Technology acceptance model 3 and a research agenda on interventions. Decis. Sci. **39**(2), 273–315 (2008)
41. Venkatesh, V., Thong, J.Y.L., Xu, X.: Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q. **36**(1), 157–178 (2012)

42. Moon, J.-W., Kim, Y.-G.: Extending the TAM for a World-Wide-Web context. Inf. Manag. **38**(4), 217–230 (2001)
43. Orehovacki, T., Etinger, D., Babic, S.: Perceived security and privacy of cloud computing applications used in educational ecosystem. In: Proceedings of the 40th Jubilee International Convention on Information and Communication Technology, Electronics and Microelectronics, pp. 823–828. IEEE, Opatija (2017)
44. Cohen, J.: A power primer. Psychol. Bull. **112**(1), 155–159 (1992)

Author Index

D. Nicholson (ed.), *Advances in Human Factors in Cybersecurity*,
Advances in Intelligent Systems and Computing 593, DOI 10.1007/978-3-319-60585-2