Creation and Management of Social Network Honeypots for Detecting Targeted Cyber Attacks

Abigail Paradise, Asaf Shabtai, Rami Puzis, Aviad Elyashar, Yuval Elovici, Mehran Roshandel, and Christoph Peylo

***Abstract*— Reconnaissance is the initial and essential phase of a successful advanced persistent threat (APT). In many cases, attackers collect information from social media, such as professional social networks. This information is used to select members that can be exploited to penetrate the organization. Detecting such reconnaissance activity is extremely hard because it is performed outside the organization premises. In this paper, we propose a framework for management of social network honeypots to aid in detection of APTs at the reconnaissance phase. We discuss the challenges that such a framework faces, describe its main components, and present a case study based on the results of a field trial conducted with the cooperation of a large European organization. In the case study, we analyze the deployment process of the social network honeypots and their maintenance in real social networks. The honeypot profiles were successfully assimilated into the organizational social network and received suspicious friend requests and mail messages that revealed basic indications of a potential forthcoming attack. In addition, we explore the behavior of employees in professional social networks, and their resilience and vulnerability toward social network infiltration.**

***Index Terms*— Advanced persistent threats (APTs), social network security, socialbots.**

I. Introduction

ADVANCED Persistent Threats (APTs) are sophisticated attacks that incorporate advanced methods for evading current security mechanisms. Traditional security solutions, such as intrusion prevention systems and endpoint protection have failed repeatedly to mitigate such threats [1]. Several recent studies have expressed the need for new tools and methods specifically aimed at detecting APTs [2]. Deploying new and versatile technologies for identifying and investigating suspicious activities is the only way to survive in the cyber arms race [3].

APTs usually follow a methodological multistage process for conducting a cyberattack [4]. The defense approaches focus on later stages of an APT, for example: detecting the command and control communication [5], or detecting anomalies caused by the actual attack [6]. Jasek *et al.* [7] suggested using honeypots to detect activities related to APTs after the attacker has already penetrated the organizational network. So far, little attention is payed to the early stages of an attack, when the adversaries collect information about the organization. In this paper, we propose detecting indications of forthcoming attacks by focusing on the reconnaissance stage.

Detecting reconnaissance activities is very difficult since usually it is performed outside of the organization's premises and without direct interaction with the organizational resources. At some point, the reconnaissance phase enables the attacker to find an entry point into the organization leading to the next phases. But some reconnaissance activities do leave traces. For example, advanced attackers make use of online social networks (OSNs) in order to extract useful information and establish contact with company employees as potential entry points into the organization. A recent survey suggests that 74% of the Internet users are now active on social media [8]. According to the 2013 APT Awareness Study, 92% of respondents believe that the use of OSNs increases the likelihood of a successful APT attack [9]. Social media is already ripe with threats: between 8%–10% of all social media profiles are malicious in nature [10].

Information extracted from OSNs may include organizational structure, positions, and roles within the organization, contact information, etc. Once an attacker gains information about an employee, he or she may trick it down to providing access to important assets. Such an employee may receive a specially crafted malicious email (a.k.a. spear phishing) get exposed to the attack through news, status messages, or job postings that lead the user to a subverted Internet resource [11].

For example, consider a cyber-attack originated in Iran and primarily targeted senior U.S. military during 2011–2014 [12]. This attack used artificial profiles in OSNs to build relationships and trust that were later exploited to gain access to sensitive information and deliver malware. Attacks that make use of social media to infiltrate an organization go largely unaddressed by traditional mechanisms. There is a growing need for tools that can be used to detect reconnaissance and initial penetration performed with the help of the social networks.

The main objective of this paper is to gain insight about the deployment process, creation, and management of the social honeypots, as well as their feasibility and authenticity. The contribution of this paper is therefore, threefold. First, we propose social network honeypots as a means of detecting APTs in the early phases of their life cycle. Social network

Manuscript received March 30, 2016; accepted June 8, 2017. Date of publication July 14, 2017; date of current version August 28, 2017. *(Corresponding author: Abigail Paradise.)*

A. Paradise, A. Shabtai, R. Puzis, A. Elyashar, and Y. Elovici are with the Department of Software and Information Systems Engineering, Ben-Gurion University of the Negev, Beer-Sheva 8410501, Israel (e-mail: abigailp@post.bgu.ac.il; shabtaia@bgu.ac.il; puzis@bgu.ac.il; aviade@post.bgu.ac.il; elovici@bgu.ac.il).

M. Roshandel is with Deutsche Telekom AG (T-Systems and Telekom Innovation Laboratories), 10365 Berlin, Germany (e-mail: Mehran.Roshandel@telekom.de).

C. Peylo is with the Bosch Center for Artificial Intelligence, 71272 Renningen, Germany (e-mail: christoph.peylo@telekom.de).

Digital Object Identifier 10.1109/TCSS.2017.2719705

2329-924X © 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

honeypots are used in our research to direct the cyber security officer of an organization to social network profiles which require further investigation. Second, we present a framework for the efficient creation and maintenance of honeypots. This framework includes components that assist in: social network acquisition, generating and managing artificial profiles, wiring them into the network, and monitoring the profiles and associated email accounts. Third, we present a field trial as a proof of concept to demonstrate the suggested framework in practice with its important technical and organizational caveats.

The honeypot profiles deployed during the field trial were successfully assimilated in the social network and received suspicious friend requests and mail messages. In the case study, we derive useful insights regarding the creation of genuine and attractive profiles. For example, we can see that in order to increase the profile credibility and make the honeypots deployment smoother, the respective email accounts should be added to the organizational internal address book.

The rest of this paper is structured as follows. Section II reviews previous related works. Section III reviews the APT and the reconnaissance phase. We present the suggested framework in Section IV and the case study in Section V. Section VI presents legal and ethical considerations. Our conclusions, research limitations, and future works are summarized in Section VII.

II. Related Works

A. Social Network Honeypots

Recently, Virvilis [2] indicated that creating social network avatars may be a good solution to identifying activities related to an APT.

Jasek *et al.* [7] suggested the general concept of using honeypots (not social networks honeypots) to detect activities associated with APTs. In our research, we suggest a solution that specifically targets social networks and takes into consideration their technical and logistical concerns.

Several previous studies [13]–[16] have focused on the identification of spammers that use social honeypots and the creation of classifiers in order to distinguish social spammers from legitimate users. These studies assume that spammers follow certain behavioral patterns that can be identified using machine learning techniques. In contrast to spammers, advanced attackers use deceptive socialbots that are designed to avoid automatic detection and usually require an investigation by a human operator to be discovered.

Zhu *et al.* [17] used honeypots for the purpose of exposing information about botnets by becoming part of a malicious botnet. While this can be used to detect attacks on organizations, it requires detection and infiltration of an existing botnet that has potentially already caused harm.

To the best of our knowledge, artificial social network profiles have not been used so far, to detect activities related to APTs. But they have been extensively used by attackers.

B. Attacking With Artificial Profiles

Several studies presented attack methods of using socialbots for different purposes: infiltration of the organization by maintaining friendships with profiles [18], [19], targeting specific users [20], obtaining personal information [21], or simply gaining influence [22].

Mitter *et al.* [23] classified different attack methods adopted by socialbots within the OSN. The attack strategies using socialbots included: a cross-site profile cloning attack by identifying a victim and creating a new identical profile [21], simulating a data harvesting attack only using the People You May Know [24], or maintaining friendships with employees of an organization [19]. In their initial study, Elyashar *et al.* [19] showed that artificial profiles can be utilized to mine information from a given organization. Next, they presented a method for infiltrating a specific user based on sending friend requests to all the specific users' mutual friends and then to the target [20]. In this paper, we employ artificial profile wiring techniques that were shown to attract organization-targeted socialbots [25], [26].

Several studies have presented strategies for connecting profiles using socialbots and gaining influence on the Twitter social network [27], [28]. Twitter socialbots mainly apply simple strategies such as follows only users that followed the bots and posting tweets about popular and focused topics [29].

C. Socialbot Detection

Several studies have presented techniques to detect socialbots by detecting anomalous behavior [15], [30], classification based on content and network topology [31]–[35] or by the detection of loosely synchronized actions [36], [37].

Deceptive socialbots try to follow the common activity patterns of real users. This allows socialbots to avoid detection and be used for longer period of time. Some OSN analytics can help identify socialbots that penetrate the organization's social cycle [38], however, these methods are mainly effective against non-sophisticated attackers. Paradise *et al.* [25] attempted to detect socialbots that connected to employees in a targeted organization in the reconnaissance phase. Their method was based on intelligently selecting organization member profiles and monitoring and investigating their activity. In contrast to Paradise *et al.* [25], in this paper, we utilize artificial profiles as honeypots in order to avoid the necessity of monitoring profiles of real employees. In Section II-B, we discussed studies that investigate the use of socialbots from the attack perspective. In this paper, we propose turning the tables and using artificial profiles to detect malicious socialbots or/and attackers.

III. Reconnaissance and Penetration Through Social Networks

Reconnaissance is the first phase in the APT attack cycle, involving information collection which is an important preparatory step taken before the more aggressive steps of APT attacks [39], [40]. In this step, attackers identify and study the targeted organization and collect information about the technical environment and key personnel in the organization [41]. This information is often gathered via Open Source INTelligence (OSINT) tools and social engineering techniques. OSINT is a form of intelligence collection from publicly available sources, and nowadays it typically refers to

aggregating information about a subject via free sources on the Internet [42].

One of the most fertile sources of information about an organization is social networks [43], [44]. As attackers have widened the scope of their attacks beyond traditional attack vectors, they have increasingly started to exploit OSNs. They make use of social engineering techniques such as exploiting personal connections and manipulating people through social network interactions [45], [46]. There are also documented incidents of malware distribution through social networks; for example, a Trojan attack infected an estimated 110 000 Facebook users' machines over a two-day period [47], and the W32.Koobface worm that targeting social networks using clever social engineering attacks and the link-opening behavior of social media users [41]. Security researchers at Proofpoint estimated that from 2014 to 2015, there will be a 400 percent increase in malicious social media content [48].

ZeroFOX, a company that provides cyber security solutions, listed top social network threats and attacks in 2015. Attacks included: information leakage, creating fake profiles in order to send phishing links and malware, launching social engineering attacks, and scamming employees [49].

Several cyber security experts have mentioned that attackers actively try to exploit information from social networks in order to gain access to private information that is shared by employees and intended to be seen only by insiders [50]–[52]. As a case in point, recently, friend requests were sent through Viadeo (a professional social media network based in France) to the French offices of Trend Micro. The requests targeted several specific employees, and the profile which sent the requests pretended to be an IT manager from the Trend Micro Australia office who had been with the company for 18 years. Checking the company directory confirmed that there was no employee with that name [51]. Another recent example is the identification of 25 fake LinkedIn profiles as part of a targeted social engineering focused on the mobile telecom sector; the profiles claimed to belong to employees at major organizations. In this case, attackers may have been interested in stealing data or trying to access the telephony networks to intercept communications [53].

These are just a few of the many types of attack activity, this paper aims to address. These examples mentioned demonstrate the increasing use of OSNs as an information gathering tool and means of initial penetration of an organizational network.

IV. Social Network Honeypot

We propose social network honeypot for acquiring indications of forthcoming attacks. The general concept is presented in Fig. 1. The artificial profiles (blue avatars in Fig. 1) are created, integrated into the OSN, and monitored. An attacker operates several OSN profiles (red avatars) to search for relevant employees and connect with them. During this process, the attacker attempts to contact the honeypots (red dashed lines in Fig. 1), for example, by sending a friend request or an email message with a malicious payload. Suspicious friend requests and emails sent to the honeypot's email account are analyzed. The goal of the social network honeypot is to trap the attacker's activity as soon as possible.

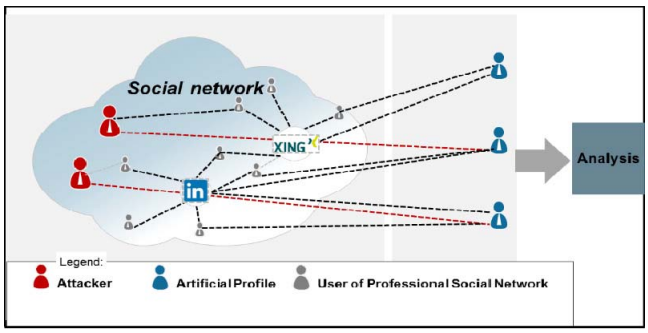

Fig. 1. General concept of the social network honeypot.

Using the proposed framework, we strive to provide the following benefits to the organization.

1) Understanding the extent to which the organization is a target of attacks via OSNs.
2) Understanding which functions in the organization that attackers are interested in (e.g., secretaries versus senior technical personnel).
3) Detecting APT attacks at early phases in the APT life cycle.
4) Providing detection with a minimal false positive rate.
5) Understanding to what extent attackers use the email addresses of employees or the OSNs as an entry point to the organization (e.g., for injecting malicious code).
6) Providing the organization with additional forensic information.

The proposed framework, presented in Fig. 2, supports the creation, maintenance, and monitoring of artificial profiles (i.e., honeypots) in OSNs. We elaborate on its main components in the following sections.

A. Social Network Acquisition

The first component focuses on acquisition of the informal social network of organization employees. It includes a crawler whose objective is extracting user information from profiles of members of the target organization from various OSNs. Such information is utilized by the system for creating reliable artificial profiles. Two main methods can be used for crawling: using (developer) API and Web Scrapping.

Acquiring OSN profiles is a challenging task. Social network services detect and block unsolicited crawling activities and official data acquisition channels are not well established yet. Other technical challenges include varied API and page structures of the different OSNs where employees may have their profiles. Additionally, it is important to normalize the data and mark the missing pieces appropriately. Finally, identification of the employee profiles in the various OSNs is a nontrivial task, especially when the employees are kept uninformed regarding the honeypot deployment in order to minimize the threat of insider data leakage.

In general, the data collection required to create genuine honeypots is very similar to the reconnaissance activity performed by attackers. For example, both parties can employ targeted crawling [54] or homing socialbots [19] to acquire the data. A third approach, suitable for very few organizations,

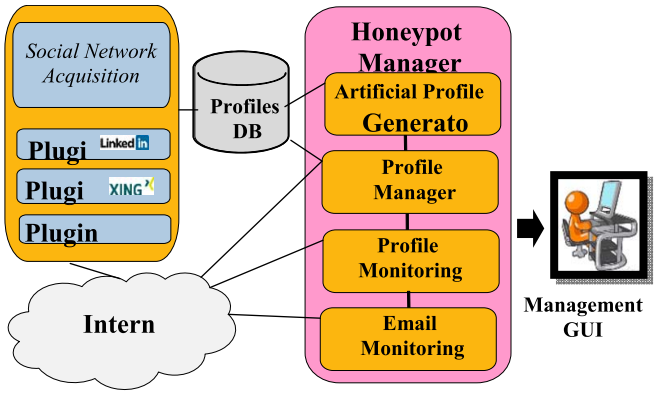

Fig. 2. OSN honeypot framework—main components.

is to oblige the employees to expose their personal OSN profiles to some organizational web application.

In current implementation, we employ targeted crawling.

B. Artificial Profile Generator

This component supports both manual and automatic honeypot generation based on information obtained through the social network acquisition module.

The process of generating the honeypots is supported by a wizard that follows the following workflow: selecting home address, updating basic profile information, inserting work history, inserting education history, and finally reviewing and saving the new artificial profile. Fig. 3 presents the profile review and modification form.

Although the framework mainly targets professional OSNs such as LinkedIn, it is extendible to other OSNs such as Facebook and additional information items (such as groups of interests, and posts) that can be added to the framework.

In order to ease on the operator and create genuine profiles, each phase of the wizard relies on statistics of existing (crawled) profiles and information from previous phases. For example, the wizard will offer first name/last name/city choices most relevant to the country chosen by the operator.

The framework also supports an automatic process of basic information for the honeypot generation. Using HoneyGen [55], we generate high-quality artificial profiles that follow association rules mined from the crawled data. HoneyGen was initially proposed for creating high-quality artificial database records based on real (genuine) records for exporting databases without compromising privacy (for example: for system tests) as depicted in Fig. 4.

The input to the profile generation process includes a database of real profiles (tokens) crawled from the OSN. In the first phase, rules are mined from the database. Then, based on the crawled data and the extracted rules, a large number of artificial records are generated. In the last phase, the generated artificial profiles are sorted by similarity to real tokens in the input database. The similarity score is assigned to each artificial profile using a likelihood rating function that considers the combinations of its values. In the future work, we plan to add the framework, the ability to generate realistic education, and employment history.

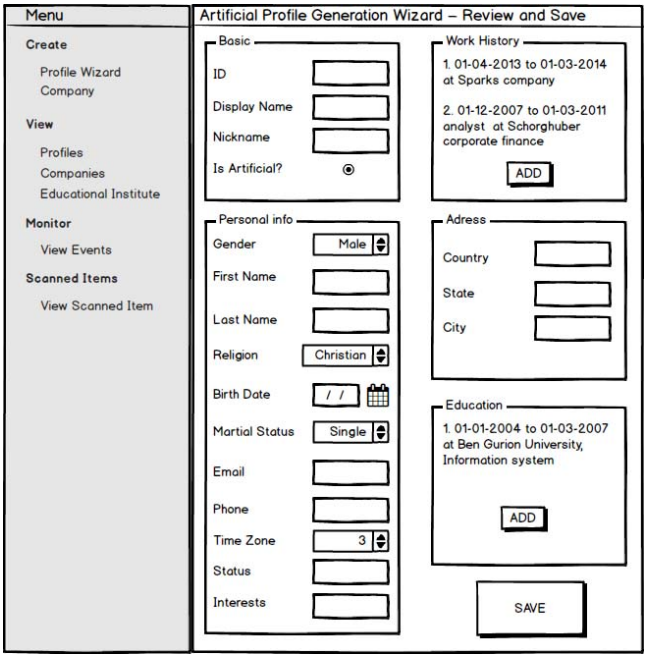

Fig. 3. Artificial profile generation.

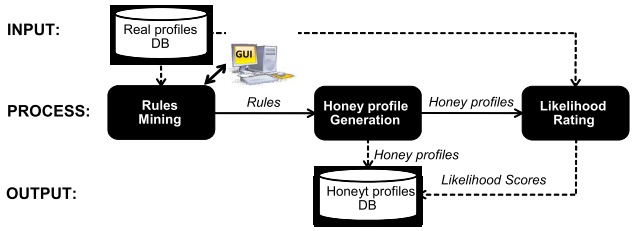

Fig. 4. HoneyGen tool that is used for generating high-quality artificial profiles.

C. Profile Manager

This is the main component of the framework which controls the profile after its creation and supports: accepting/sending friend requests, sending posts and messages, completing the "like" action, and more, depending on the API provided by the specific OSN.

Its primary task is "wiring" the honeypot (i.e., connecting it with other profiles in the OSN in order to increase its reliability). The framework provides a method for identifying profiles that should be approached with a friend request. This is performed based on the "social-bot organization intrusion" strategy [19], [25] which is based on the following assumptions:

1) The more friends a user has, the more likely he or she is to approve the friend request.
2) The more mutual friends a user has with the requester the more likely he or she will approve the friend request.

The wiring algorithm includes these main phases:

Phase 1 *(Connect to Collaborating Employees):* Typically, a group of employees is aware of the honeypot deployment process in order to support it from the IT, Human Resources (HR), and security perspectives. Connecting the

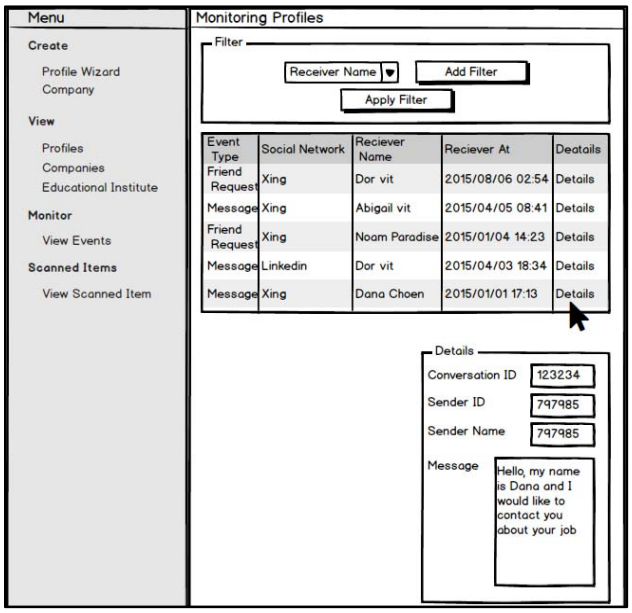

Fig. 5. Screenshot of the artificial profile monitoring module.

honeypots to a subset of these employees, with their consent, will increase the profile credibility and help with further assimilation in the OSNs.

Phase 2 *(Send Requests to External, Highly Connected, Profiles):* This phase helps to further increase the credibility of the honeypot by connecting to profiles with a high probability of approving friend requests.

Phase 3 *(Send requests to insider profiles):* Phase 3 consists of sending friend requests to the employees having the highest probability of accepting the friend requests according to the number of friends a profile has and the number of mutual friends [25].

Each friendship request proposed by the algorithm should be approved by the system operator before sending the request.

D. Profile Monitor

The goal of this module is monitoring the OSN events related to the honeypot profiles. Fig. 5 presents the monitoring module user interface (UI) and its filtering capabilities.

The monitoring module will collect and aggregate the events in order to present them in a single unified UI.

The module is configured with the set of honeypots that need to be monitored (including target OSN, profile ID, access tokens, etc.) and it collects events related to the honeypot such as friendship requests, incoming messages, and comments. The available events depend on the specific OSN API. This module allows filtering and ordering the events to simplify their exploration. It is developed as a set of plugins, each able to access accounts of a specific OSN, and is therefore extendible to any OSN of interest.

E. Email Monitor

For each deployed honeypot, a shadow organization email account is generated. The goal of this module is monitoring

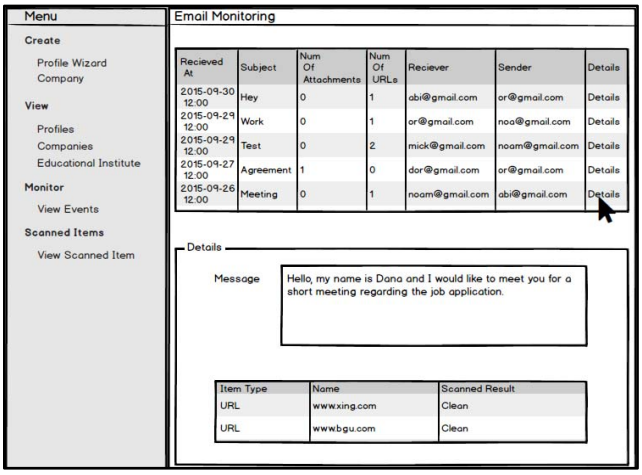

Fig. 6. Screenshot of the email monitoring module.

the honeypot mailboxes. This module is granted permission (as a client application) to collect emails sent to these accounts. It stores the email content, including: timestamp, sender, recipients, subject, content, attachments, and URLs in a database and scans the attachments and URLs for possible exploits. This module supports integration and custom development of new detectors and scanning logic. The results of all scanning engines are stored in the database as well. A management service provides graphical UI for exploring the emails and scanning results (as presented in Fig. 6).

The management service integrates with this module to provide periodical scanning of previous attachments and URLs (e.g., every month). This is an important feature because an attachment containing a zero-day exploit may not be detected by any of the detection engines at the time that an email was sent. However, after a few months, as the exploit becomes known, the updated detection engines will identify the attachment as malicious and the organization will be informed that there was a penetration attempt.

V. Case Study: Applying The Social Network Honeypot Framework

A. Overview

In order to demonstrate the proposed approach and framework, we conducted a field trial with the assistance of a large European company (hereafter referred to as Organization *A*). During eight months, we created, operated, and monitored seven artificial profiles in two OSNs: Xing and LinkedIn. The profiles were generated based on aggregated data provided to us by Organization *A*, and statistical information we obtained through targeted crawling of the OSNs.

It this trial, we limited the number of deployed profiles to seven based on the following reasons: First, each profile was approved by the organization according to a strict selection process that required both time and effort. In addition, incoming communication of every profile during the trial had to be inspected by qualified organization employee. Therefore, in order to reduce the inspection effort the company insisted on a "manageable" number of profiles. The main goal of the

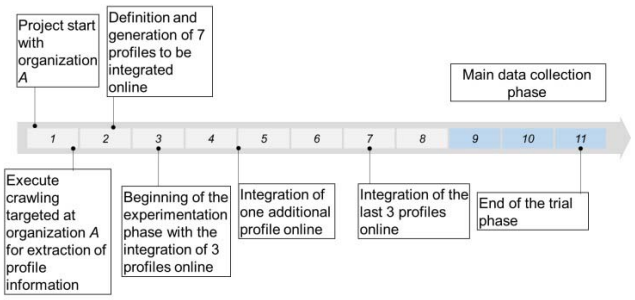

Fig. 7. Field trial timeline.

field trial was to gain insights about the process of creating, integrating, and maintaining social honeypots, a goal that does not necessitate the use of a large number of profiles. Even though, the goal of this paper was not hunting for attacks, we do believe that even a small number of good and convincing profiles reveal valuable information about the attacks. Based on the knowledge gained in this paper, we plan to conduct a long term, trial with the specific goal of identifying targeted reconnaissance activities.

The timeline (in months) and important milestones are presented in Fig. 7.

In the case study, we used the proposed framework that was described in Section IV, and we used the social network acquisition module to extract statistical information about the organization through targeted crawling of OSNs. Due to the small number of profiles in the case study, the creation of the profile's information was done manually, without the need to use the artificial profile generator module; the profile generation will be described in Section V-C. Following the creation of the profiles, each of the profiles underwent a wiring process using the profile manager module as described in Section IV-C. During the trial, the profiles were monitored by scanning the OSN events and incoming emails using the profile monitor module.

B. Case Study Objectives

The objective of this case study was to analyze the suggested method of using artificial profiles as honeypots. In addition, we derive useful insight which relate to the operation and deployment of such social network honeypots. Specifically, we attempt to answer the following questions:

1) How can we create a genuine and attractive honeypot?
2) What should be the honeypot wiring strategy?
3) How often are profiles subject to attacks (suspicious emails or suspicious friend requests)?
4) How can malicious contact attempts be identified?
5) How easily do employees trust and connect with unknown people?
6) Can the proposed framework be executed and operated on real OSNs?

C. Profile Generation

Seven artificial profiles were generated in both LinkedIn and Xing: two female profiles and five male profiles. The profiles were created manually using the Profile Generator wizard and with the help of Organization *A* cyber security department.

TABLE I
PROFILE DETAILS

Profile Name	Gender	Role	Creation Date	Profile Picture
USER_1	Male	Email Systems Service Manager	20/03	yes
USER_2	Male	Internal IT Project Manager	29/03	yes
USER_3	Female	Strategy Manager	29/03	yes
USER_4	Male	Systems Integration	27/04	yes
USER_5	Male	Service Manager Security Systems	19/07	no
USER_6	Male	Email Administrator	20/07	no
USER_7	Female	Business Intelligence Manager	26/07	no

Table I presents the artificial profiles with the following basic details: name, gender, role, and creation date. The creation dates correspond to the timeline in Fig. 7.

Creating the profiles included the following main phases which we took in order to make the profiles more attractive and reliable:

1) Organization *A* provided the names, countries, and cities of the profiles.
2) Additional info was defined based on statistical data as described in Section IV-B.
   a) *Occupation:* Out of the positions suggested by the Profile Generator based on the honeypot residence and crawled data, we selected the most attractive ones with the assistance of Organization *A*.
   b) *Gender:* We used gender distribution to see what the dominant gender in Organization *A* for each position was.
   c) *Age:* We choose the average age of the employees of the specified gender holding the specified position.
   d) *Work Experience and Education:* The collection of work history and education was constructed by combining the data from employees holding similar positions.
3) *Creating Profile Picture*: For the four initial profiles, we used uncopyrighted profile pictures. We did not add a profile picture to the last three profiles.
4) *About Me*: We provided the profiles with an appealing "about me" section with relevant content regarding the job, education history, and occupation.

D. Wiring Profiles

1) Connection Types: Each honeypot was connected to three types of profiles:

Collaborators: Profiles of a few employees within the organization who willingly participated in the trial.

Highly connected: Arbitrary popular profiles, which are not affiliated with the organization and have over 500 connections.

Insiders: Social network profiles that contain the organization name but are not collaborators.

TABLE II
PROFILE STATISTICS

Profile Name	Gender	Profile Picture	Num. of friends (Xing)	Num. of friends (LinkedIn)	Acceptance rate (Xing)	Acceptance rate (LinkedIn)
USER_1	Male	yes	214	186	91.27%	74.42%
USER_2	Male	yes	256	243	92.22%	75.00%
USER_3	Female	yes	254	255	94.92%	82.04%
USER_4	Male	yes	221	179	93.53%	77.62%
Average	**-**	**-**	**236.25**	**215.75**	**92.99%**	**77.27%**
USER_5	Male	no	55	63	76.06%	80.28%
USER_6	Male	no	121	76	86.13%	62.50%
USER_7	Female	no	91	90	91.00%	83.16%
Average	**-**	**-**	**89.00**	**76.33**	**84.40%**	**75.31%**

2) Wiring Process: We used the profile manager module as presented in Section IV-C to get recommendations about the next friend requests for each artificial profile. The wiring profiles process included three main phases as was presented in Section IV-C.

3) Termination of Insider Connection Requests: An internal investigation was undertaken following a friend request from one of the honeypots to an insider (non-collaborator). The employee searched for the artificial profile and contacted the HR department for information. Due to this investigation, we were requested to stop sending friend requests to insiders after five month since the beginning of the trial. Therefore, the last four profiles did not send friend request to employees of Organization *A*.

E. Profile Statistics

1) Acceptance of Friend Requests: Table II presents the number of friends and the acceptance rate in each network for each profile. We divided the table into two groups, the four profiles that were created first and the last three profiles.

Observation 1 *(Acceptance Rate):* Table II shows that the acceptance rate is higher for profiles with a picture than for profiles without a picture (ANOVA test with confidence level = 0.05). In addition, the female's profile (USER_3) had higher acceptance rates and more friends in both OSNs.

Discussion 1: Profiles with a picture seem to be more attractive than those without. Although, profiles without pictures were introduced late, we argue that the time difference between the profiles' deployment has no effect on the acceptance rate. Despite the extremely small sample size the acceptance rate difference is significant, confirming observations made in previous research [56].

Observation 2 *(Acceptance Rate):* The acceptance rate is higher in Xing when compared with LinkedIn.

Discussion 2: We speculate that the acceptance rate in Xing was higher than in LinkedIn because Xing report on the profiles' activity levels. During the honeypot wiring process, we avoided sending friend requests to profiles with low activity level contributing to the higher acceptance rate in Xing.

Table III presents statistics of friend request sent to insiders and highly connected users with respect to the total number of unique profiles that we contacted and total friend requests.

TABLE III
FRIEND REQUEST STATISTICS

	Xing			LinkedIn		
	Unique friend requests	Total friend requests	Acceptance rate	Unique friend requests	Total friend requests	Acceptance rate
Insiders	80	112	74.48%	53	58	66.52%
Highly-connected	477	1180	90.78%	712	1182	76.96%

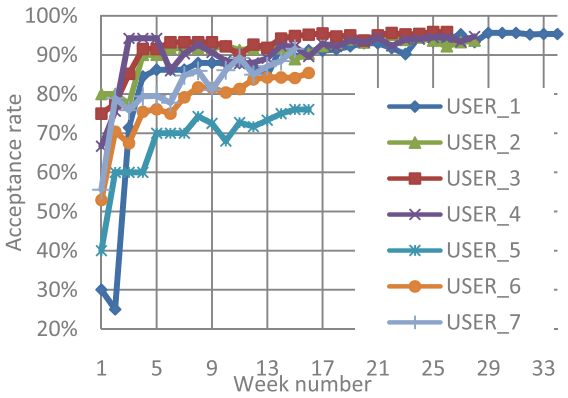

Fig. 8. Aggregated acceptance rate per week for highly connected profiles.

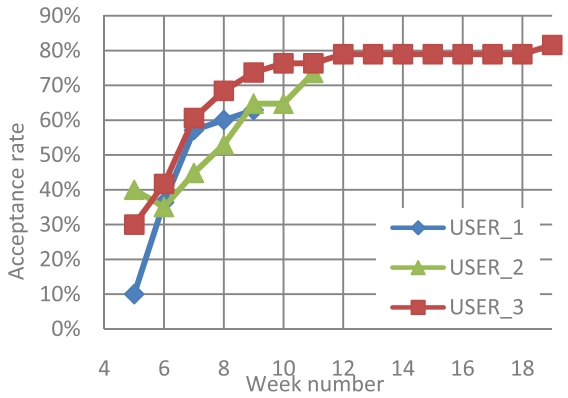

Fig. 9. Aggregated acceptance rate per week for insiders.

Observation 3 *(Acceptance Rate for Highly Connected and Insiders):* The acceptance rate for insiders is lower than for highly connected profiles, but it is still relatively high.

Discussion 3: There is a need to increase awareness of employees to the dangers of accepting requests from OSN members they did not met in real life.

Xing reports the dates of all accepted friend requests. Figs. 8 and 9 show the progression of the acceptance rate in Xing by highly connected and insiders respectively, as a function of time. The week number (X-axis) is relative to the opening day for each profile; since three of the profiles (USER_5, USER_6, USER_7) were created a couple of weeks after the other profiles, their timeline is shorter as presented in Fig. 7. Note that in Fig. 9 only the first three profiles reached phase 3 as explained in Section IV-C.

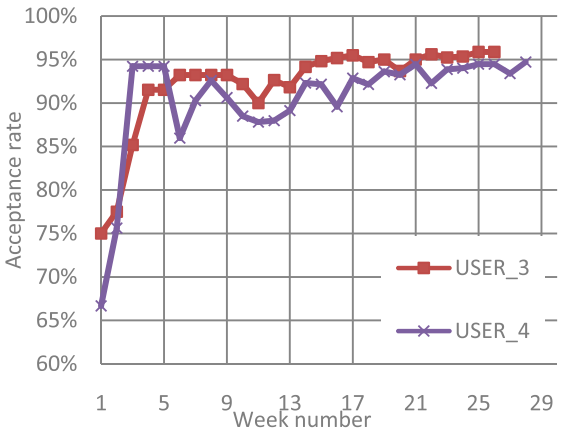

Fig. 10. USER_3 and USER_4 aggregated acceptance rate per week for highly connected users.

Observation 4 *(Acceptance Rate for Highly Connected and Insiders):* Acceptance rate started low but increased over time.

Discussion 4: A profile that existed for a longer period of time is more likely to receive positive responses to his/her friend requests.

Fig. 10 focuses on the acceptance rate in Xing of USER_4 and USER_3 for highly connected users. In the case of USER_4, we sent friend requests to profiles which previously accepted requests from our other artificial profiles.

Observation 5 *(Using an Artificial Profile for Exploration):* USER_4 has a relatively high acceptance rate for a male profile, which can be attributed to the utilization of previous information.

Discussion 5: Artificial profiles can be used to locate profiles with a higher likelihood of accepting friend requests in order to boost acceptance rates of other profiles.

2) Incoming Friend Request Analysis: We accepted all of the incoming friend requests and investigated each one thoroughly as presented in Fig. 11.

Incoming friend requests were examined in the following three phases.

Phase 1 *(Checking the Profile of the Sender):*

In this phase, we checked each request by processing the profile of the sender. Requests from known profiles (collaborators, profiles associated with the organization, and highly connected profiles) were marked as known. Other profiles required further investigation. Fig. 12 shows the distribution of friend requests after phase 1 in LinkedIn and Xing.

Observation 6 *(Incoming Friend Requests):* 28 friend requests were identified as requests that needed further investigation. LinkedIn has a higher number of incoming friend requests. In addition, the female profile has a higher number of incoming friend requests.

Discussion 6: LinkedIn is a larger network with more users compared to Xing which results in a higher number of friend requests. A female profile attracts more incoming friend requests, similar to the higher acceptance rate shown previously.

Observation 7 *(Insider's Incoming Friend Requests):* Artificial profiles received several friend requests from employees

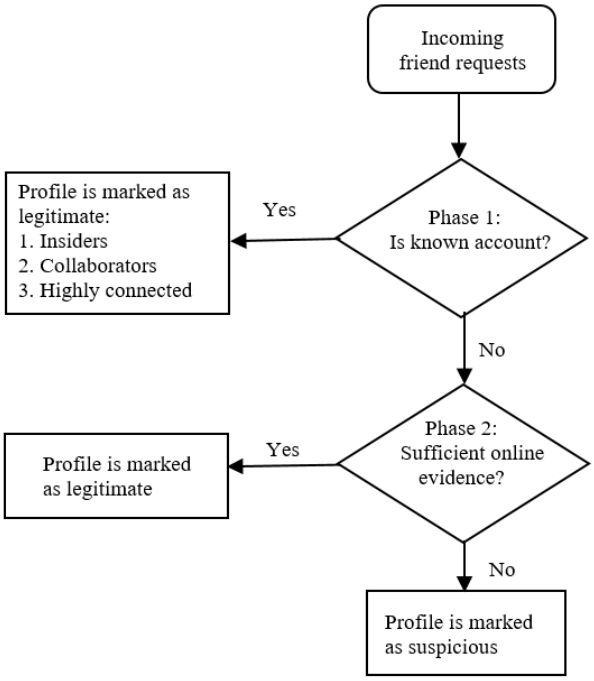

Fig. 11. Incoming friend request processing.

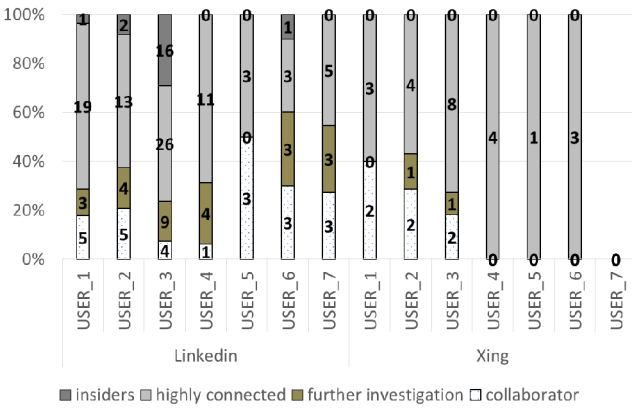

Fig. 12. Incoming friend requests in LinkedIn and Xing.

that were not collaborators (insiders). These requests were investigated by contacting the organization and ensuring they came from valid employees.

Discussion 7: The incoming friend requests from insiders show that our profiles were attractive and genuine enough to attract attention from insiders that were not collaborators.

Phase 2 *(Checking the Profile Online):*

During this phase, we searched for information regarding the name, organization, country, and job mentioned in the profile in order to collect evidence of the person's existence beyond the OSN. We checked the sites on the Internet that contain the profile information; we tried to verify the profile information as follows: we checked to see whether the profile exists in the address book of the organization mentioned in

TABLE IV
SUSPICIOUS PROFILES

Profile Name	Sender Name	Main reasons for suspicion
USER_3	USER_MOS	Common name on web, sent CV file to the artificial profile, no mutual friends
USER_3	USER_SG	No mutual friends, no picture, low information in the profile, no strong evidence on the web
USER_1	USER_AA	No friends, no profile picture, no strong evidence on the web
USER_3	USER_ RA	Common name on web, sent CV file to the artificial profile, no mutual friends
USER_4	USER_ILY	Name is linked to spam and scam messages
USER_2	USER_RS	Name is linked to spam and scam messages
USER_2	USER_AM	Common name on web, sent CV file to the artificial profile, no mutual friends
All 7 profiles	USER_AP	Sent friend requests to all artificial profiles on the same day, listed as a "head of organization" with no extra information, profile picture lead to tutorial on photo editing, no strong evidence on the web

the profile, and we searched for the profile in other OSNs. If this raised any concerns regarding the profile (for example, we did not find the profile in the address book of the organization), we counted the number of sites on the Internet that contain the profile's information; if there were more than 20 appearances, the profile was considered too prevalent, and therefore considered suspicious (for example, it is possible that attackers purposely choose a popular name for the profile in order to increase ambiguity regarding the true identity of the profile). If there were less than 20 appearances, we checked each site and verified that the profile information could be fully identified as a legitimate identity. In addition, if there was no evidence at all on the Internet, the profile was defined as suspicious.

We were able to find strong online evidence for 20 of the friend requests. Eight profiles (14 friend requests) were considered suspicious due to the above reasons. We contacted the profiles that did not show strong global online evidence in order to test how they would respond. Out of the eight profiles we contacted, only four profiles responded to our message. Table IV summarizes the eight suspicious profiles.

We now present observations and discussions on each profile from the table.

Observation 8 *(USER_MOS, USER_RA, USER_AM):* The three profiles claimed to be interested in a job. They sent their CV to the artificial profile mail address. The three PDF files were scanned by a lab that specializes in malware analysis. The scan showed that two of the files were clean and one of the files contained various calls to the operating system, which should not exist in a PDF file.

Discussion 8: APTs usually do not try to send the malware and penetrate during the initial interaction (first contact) with its target but instead slowly build relation and trust to ultimately pick the right time for attack. Therefore, we assume that a longer study would have led to more observations and different outcomes.

Observation 9 *(USER_SG, USER_AA):* The two profiles started out and remained without any details in their profile and zero friends.

Discussion 9: These two profiles were classified as fake profiles.

Observation 10 *(Spammer Profiles):* An online search was performed for USER_ILY, the name used appeared to be linked with spam—therefore, increasing our suspicion it was fake. There was no response to the message we sent. USER_RS also appeared to be linked to spam and scams; we contacted the profile and received a reply that is a common scam message.

Discussion 10: Our online search and interactions with these two profiles were enough cause to classify them as malicious profiles related to scam and spam.

Observation 11 *(USER_AP):* This profile sent friend requests to all seven artificial profiles on the same day. The profile listed the user as a head of an organization without detailing which organization. Also, when searching online for the user's profile picture, we discovered the picture was taken from a website for photo editing.

Discussion 11: The profile was classified as a fake profile but its intent remains unknown.

We quantified the effectiveness of our framework using the discounted cumulative gain (DCG) measure. DCG is a measure used to measure the effectiveness of algorithms. Using the decision of each friend requests as suspicious or not, DCG provides an alternative measure to area under the ROC curve (AUC); a higher DCG is indicative of identifying suspicious cases earlier (in chronical order)

$$DCG = r[1] + \sum_{i=2}^{n} \frac{r[i]}{\log_2 i} \tag{1}$$

where $r[i]$ is 1 if the ith friend request was defined as suspicious (after phase 2 in Fig. 11) or 0 if the ith friend request was defined as legitimate, and n is the number of total incoming requests that required further investigation (moved on to phase 2 in Fig. 11). Tables V and VI present the DCG for each profile with regard to friend requests received in Xing and LinkedIn, respectively. For example, in Table VI for USER_6 the DCG measure was calculated as follows: there were three investigated requests, and the two first requests were defined as legitimate ($r[1], r[2] = 0$), and the last was defined as suspicious ($r[3] = 1$)

$$DCG_{\text{USER6}} = 0 + \frac{r[2]}{\log_2 2} + \frac{r[3]}{\log_2 3} = 0.631. \tag{2}$$

Observation 12 (DCG Measure): The average DCG in both networks (for a profile with a number of investigated requests greater than zero) is 1.336

Discussion 12: A profile that existed for a longer period of time is more likely to receive suspicious requests and get a higher DCG.

TABLE V
SUSPICIOUS FRIEND REQUESTS IN XING

Profile Name	Total number of incoming requests	Number of investigated requests (phase 2)	Number of suspicious requests (after phase 2)	DCG
USER_1	5	0	0	-
USER_2	7	**1**	**1**	1
USER_3	**11**	**1**	0	0
USER_4	4	0	0	-
USER_5	1	0	0	-
USER_6	3	0	0	-
USER_7	5	0	0	-

TABLE VI
SUSPICIOUS FRIEND REQUESTS IN LINKEDIN

Profile Name	Total number of incoming requests	Number of investigated requests (phase 2)	Number of suspicious requests (after phase 2)	DCG
USER_1	28	2	2	2
USER_2	24	4	2	2
USER_3	**55**	**9**	**4**	**2.394**
USER_4	16	4	2	2
USER_5	6	1	1	1
USER_6	10	3	1	0.631
USER_7	11	3	1	1

F. Email Analysis

1) Email Statistics: Table VII shows email distribution for each of the profiles, we used "VirusTotal"[1] to locate spam/malware. The distribution includes the total number of emails, the amount of spam and suspicious emails, emails received from the OSNs, and other messages such as messages from Organization *A*. We did not obtain USER_6's email data, since Organization *A* did not forward his/her emails to us.

Two of our profiles were exposed to spam. For a period of about a month we registered with several websites (websites offering free services, dating websites, etc.). The purpose of this exposure was to test if it would increase the visibility of the profile and make it more accessible to an attacker, specifically an APT.

Observation 12 (Exposure to Spam): Profiles which were exposed to spam received significantly more spam compared to the other profiles. We noticed that USER_2 received mails from spam websites we did not sign up for.

Discussion 12: Exposure to spam provides profiles with increased online presence and can improve the legitimacy and attractiveness to the attacker. In addition, USER_2's mail was shared among spam databases.

[1]http://www.virustotal.com/

TABLE VII
EMAIL ANALYSIS

Profile Name	Total emails received	Spam and suspicious	Xing	LinkedIn	Other	Comment
USER_1	430	7	235	106	82	
USER_2	4718	801	184	156	3577	Exposed to spam
USER_3	250	3	103	119	25	
USER_4	2819	662	52	50	2055	Exposed to spam
USER_5	270	10	111	136	13	
USER_7	414	14	163	218	19	

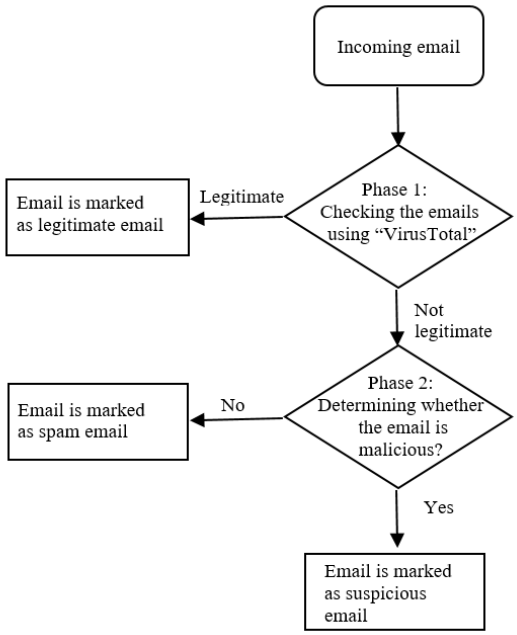

Fig. 13. Incoming email processing.

Fig. 13 presents the classification for incoming emails.

Phase 1 *(Checking the Emails Using "VirusTotal"):* We scanned each email using "VirusTotal" to find malware and spam. Emails without malware or spam were marked as legitimate during this phase.

Phase 2 *(Identification of Suspicious Emails):* We classified the emails that were determined to not be legitimate in phase 1 into two groups: spam emails and suspicious emails. Unsolicited messages without malicious content or attachments were considered spam emails, and unsolicited messages with malicious content or attachments were considered suspicious emails. The spam received was either advertisements for products, software, or dating, or related to adult content. The suspicious emails received were either phishing emails or emails with a URL that led to a malware or virus.

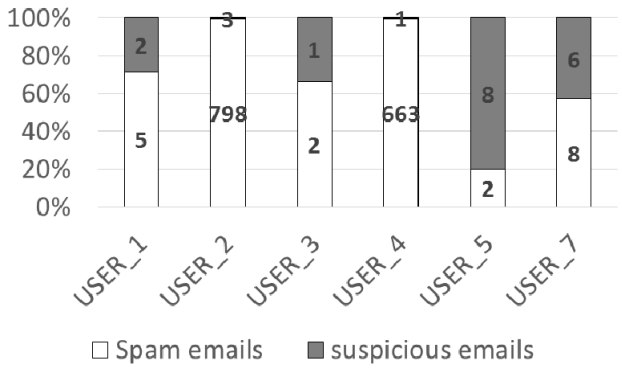

Fig. 14. Spam and suspicious emails.

TABLE VIII
SUSPICIOUS EMAILS

Profile Name	Total number of incoming emails	Number of incoming emails classified as not legitimate by "VirusTotal" (after phase 1)	Number of suspicious emails (after phase 2)	DCG
USER_1	430	7	2	0.743
USER_2	**4718**	**801**	3	0.231
USER_3	250	3	1	0.630
USER_4	**2819**	**664**	1	0.270
USER_5	270	10	**8**	4.534
USER_7	414	14	6	3.948

Fig. 14 presents the number of emails that "VirusTotal" classified as spam or suspicious emails (during phase 1) for each profile. We received a total of 21 suspicious emails.

2) Suspicious Emails: We used the DCG measure in order to measure the effectiveness of our framework also in detecting suspicious emails received. Table VIII presents the DCG for each profile with regard to emails received. To compute the DCG, we used (1) where $r[i]$ is 1 if the ith email was defined as suspicious (after phase 2 in Fig. 13) or 0 otherwise, and n is the total number of incoming emails that moved to phase 2 in Fig. 13.

Observation 14 (DCG Measure): The average DCG in both networks is 1.726. The number of emails that were classified as not legitimate by "VirusTotal" was extremely high for profiles that were exposed to spam compared to unexposed profiles.

Discussion 14: The DCG score shows that it is inadvisable to expose profiles to spam. The average DCG is 2.464 for emails, excluding the profiles that were exposed to spam (1.726 including the profiles that were exposed to spam).

On USER_7's first day, six suspicious emails were received.

Observation 14 (USER_7): We investigated the profiles that we contacted that day, and they were all legitimate profiles.

Discussion 14: There is a strong connection between the existence of a profile in the OSN and its (active) exposure to suspicious emails.

USER_5 received 5 suspicious emails before the profile was opened.

Observation 15: (USER_5): The email accounts were opened before the creation of the profiles in the OSN.

Discussion 15: This may be proof that the organization was targeted by attackers via email accounts and that a scan for valid email accounts was performed.

G. Challenges

During the phase of sending friend requests to insiders, we received two messages from insiders that replied they could not find us in the organization's address book. An internal investigation began by the insider. The employee contacted the HR for information. As a result of this investigation, we stopped sending friend requests to insiders. From this event, it is possible to conclude that there is a need to add the artificial profiles to the organization address book and involve HR (and other relevant departments) in order to give the profiles more credibility. In addition, there is a need to increase the employee awareness regarding OSN hazards, and checking profiles of strangers in the organization's address book can be a good solution.

H. Lessons Learned

1) Acceptance Rate: The acceptance rate for the profiles of females is higher than that of males, their friend requests are approved at a higher rate.

Xing provides information about how active a profile is on the network. The results show that users with higher activity percentages are more likely to approve incoming friend requests. The acceptance rate of all profiles increased over time, leading to the conclusion that there is a more positive attitude from users toward profiles that have existed for longer periods of time.

2) New Wiring Method: The results showed that profiles that accepted a friend request from one of our profiles were prone to accepting requests from our other profiles as well. Using this information, it is possible to use an artificial explorer profile with the sole purpose of locating profiles that are likely to approve friend requests, thus increasing our chances for higher acceptance rates.

3) Employee Behavior on Social Networks: The case study allowed us to examine the behavior of employees in regard to accepting friend requests from strangers. On average 70% of the employees accepted our friend request. We also received 20 incoming friend requests from insiders. There is a need to increase the employee awareness to OSN hazards.

One possible solution for the awareness problem is checking profiles of strangers in the organization's address book. It should be encouraged that employees make use of such tools before accepting friend requests from profiles claiming to be part of the company.

4) Further Actions to Increase the Artificial Profiles' Credibility: As a result of investigations performed by the employees, we recommend adding the artificial profiles to the organizations address book in order to increase their credibility and involve other relevant departments.

5) Profile Attractiveness: Our artificial profiles received incoming friend requests from strangers, especially when looking at the incoming friend requests from organization employees that were not collaborators. The artificial profiles received suspicious friend requests on average once a month which indicates that our profiles were attractive enough to attract people.

The profiles had very high acceptance rates among both insiders (70%) and highly connected (90.78% in Xing and 76.96% in LinkedIn). Profiles with pictures had higher acceptance rates. The female profile was more attractive with higher acceptance rates and more incoming friend requests.

6) Suspicious Emails: The unsolicited emails contained suspicious and harmful links; however, those received were not directed or targeted toward the user and were more general in nature.

The case of receiving suspicious mails on the first day of opening one of the profiles indicates that the mail received was a result of opening the account. There is a strong connection between the existence of a profile in an OSN and its (active) exposure to suspicious emails.

The DCG measure shows that it is not advisable to expose profiles to spam, because the number of emails that were classified as suspicious by "VirusTotal" was extremely high for those profiles

7) Survival in the Real Social Networks: 100% of the artificial profiles survived within LinkedIn and Xing for the duration of the case study, three profiles were deployed for six months, one profile was deployed for four months, and the last two profiles were deployed for two months. We believe that the OSNs did not flag our profiles as suspicious due to their similarity to other profiles within the organization and our deliberate limited interaction with other social network users (e.g., limiting the number of concurrent friend requests and receiving and accepting friend requests from collaborators inside the organization). This case study showed that it can be implemented and executed within OSN.

I. Summary

We clustered our conclusions into features inspired in [57]. We defined the following three clusters.

Effectiveness: is the measure for deciding whether our framework provides the desired output or not. Being effective means producing the right decision in terms of the emails or friend requests that were identified as suspicious. We used the DCG in order to measure the effectiveness of our framework to detect suspicious activity at an early point in time.

Survivability: Reflects how well did the honeypot profiles survive in the social networks and avoided being blocked.

Attractiveness: Reflects how genuine and attractive the honeypot profiles were.

Table IX summarizes our main conclusions.

TABLE IX
CONCLUSION

	Conclusion
Effectiveness	**Average DCG for friend requests**: 1.336. **Average DCG for emails**: 1.726. The DCG measure shows that it is not advisable to expose profiles to spam, because the number of emails that were classified as suspicious by "VirusTotal" was extremely high for those profiles. **Average DCG for emails excluding the profiles that were exposed to spam**: 2.464.
Survivability	**Deployment** – Seven high quality profiles were deployed and maintained: three profiles were deployed for six months, one profile was deployed for four months, and the last two profiles were deployed for two months. **Survival in real social networks** – All deployed profiles survived within the social networks.
Attractiveness	**New wiring method** – We implemented an artificial explorer profile with the purpose of locating profiles that were likely to approve friend requests. **Employee behavior on social networks** – On average, 70% of the employees accepted our friend request. There is a need to increase employee awareness to OSN hazards. **Further actions to increase the artificial profiles' credibility** – We recommend adding the artificial profiles to the organization's address book in order to increase their credibility. **Profile attractiveness** – The artificial profiles received a number of incoming friend requests and emails from strangers. There were a number of suspicious requests and emails among the incoming friend requests. The profiles had very high acceptance rates among both insiders and highly-connected profiles

VI. Legal and Ethical Considerations

The creation of artificial profiles raises a number of legal and ethical challenges and dilemmas.

A. Legal Considerations

We believe that a minimal violation of terms is the most effective way to reliably estimate the feasibility of an attack and determine how organizations can protect themselves against threats involving social networks. Our framework provides insight into the attacks and greater understanding of the defenses required. Users will benefit from an increase in security by developing a detective and protective mechanism to defend against malicious acts [58]–[60].

The creation of honeypot profiles may lead to the violation of the user terms of a social network, but we believe that our study benefits both users and social network providers. Between 8% and 10% of all social media profiles (approximately 150 million profiles) are malicious in nature [10]. This enormous number should emphasize the acute problem that we are facing and demonstrate the need for further study and solutions, particularly since social network providers have repeatedly failed to mitigate such threats. There is a growing need for new tools and methods for detecting such threats, and we believe that it is incumbent on cyber security researchers in academia to address this challenge.

B. Ethical Considerations

Our main goal is to help those targeted by malicious attackers through social networks, while also respecting those individuals. We accomplish this by considering the ethical

issues involved with our research and those we are trying to help, and by doing everything in our power to minimize an invasion of their privacy.

As part of our effort to respond to ethical considerations, we carried out the following actions (based on the main guidelines in [59]) in order to protect the privacy of the social network profiles, we contacted as follows.

1) The creation and maintenance of the profiles relied upon the cooperation of the organization, and each honeypot profile was created only with the strict approval of the organization.
2) The data used for profile generation were publically available online information.
3) The profiles' privacy settings were defined such that an external entity could not obtain information about the profiles we contacted.
4) We did not access or store any information about the profiles we contacted, and we only recorded the fact of accepting the friend request.
5) The profiles' identifiers were stored securely on a password protected server during the study.
6) Only incoming friend requests and incoming emails were processed and analyzed for the research.
7) The profiles were deleted from the two social networks at the end of the study.
8) We used the OSNs' APIs for the monitoring and crawling process.
9) Identifiers of the profiles that contacted our profiles or were contacted by our profiles were deleted, and only aggregated statistical data was retained.
10) Data processing and analysis are as follows.
    a) Email communication with the profiles was inspected and cleared by a security officer of Organization *A* in order to avoid unintentional leakage of information.
    b) Suspicious emails that did not contain any personal or confidential information were forwarded to the research team and were inspected by automatic tools for determining spam, newsletters, and detecting malicious content/code.

The issue of ethical considerations is a subject that needs to be addressed quickly, due to the evidence of increased use of social networks by attackers and the inability of current security solutions to address the problem. Therefore, we believe security solutions like ours are necessary, and furthermore we believe that the insight gathered from our experiment will serve to assist the cyber security community in their efforts to create defense mechanisms to protect private individuals and organizations.

VII. Conclusion

In this paper, we propose a method that is based on social network honeypots to detect APT attacks at early phases of the APT life cycle. We implemented the method and conducted a field trial to demonstrate the effectiveness of the suggested method with the cooperation of a European organization. The artificial profiles that we created were able to assimilate into OSNs and appeared genuine and attractive to other users. We can conclude that the wiring method proved successful by having more than 70% average acceptance rate when sending friend requests to members of the organization from the artificial profiles. The artificial profiles received suspicious friend requests and emails.

We were unable to completely validate the indications of potential forthcoming attacks during our case study. This could be attributed to: the short period of time in which the case study was performed, the small number of profiles we created for the purpose of demonstrating the framework, and/or stopping the wiring process before it finished due to investigation by organization employees.

In the future, we plan to continue and scale up the case study for an additional time period in order to further examine the effectiveness and attractiveness of the generated profiles, generating a more diverse pool of profiles, and significantly increasing the number of created profiles.

Furthermore, since the number of profiles was small, we created the profile information manually so there was no need to use the complete framework. For the next study, we plan to use and examine the complete framework.

We plan to upgrade the automatic generation of the artificial profiles component to support not only the automatic generation of basic profile information, but also the automatic generation of more advanced profile information such as employment and education history.

Acknowledgment

The authors would like to thank Deutsche Telekom AG for its support in this research. The field trial was conducted with the help of a large European organization (not Deutsche Telekom), and in order to protect the organization's anonymity, they presented the organization as Organization *A*. They would also like to thank the cyber security department from Organization *A* for their assistance during the field trial.

References

[1] C. Technologies. *Defending Against Advanced Persistent Threats: Strategies for a New Era of Attacks*, accessed on Jun. 2012. [Online]. Available: https://www.necam.com/docs/?id=759a3111-a395-4c40-ae0a-43e299159c81

[2] N. Virvilis, B. Vanautgaerden, and O. S. Serrano, "Changing the game: The art of deceiving sophisticated attackers," in *Proc. 6th Int. Conf. IEEE Cyber Conflict (CyCon)*, Jun. 2014, pp. 87–97.

[3] HACKING. *Incident Response*, accessed on Feb. 18, 2015. [Online]. Available: http://resources.infosecinstitute.com/current-trends-apt-world/

[4] A. T. Micro. *Custom Defense Against Targeted Attacks*, accessed on Dec. 2014. [Online]. Available: http://www.trendmicro.com/cloud-content/us/pdfs/business/white-papers/wp_custom-defense-against-targeted-attacks.pdf

[5] Websense. *Advanced Persistent Threats and Other Advanced Attacks*, accessed on 2011. [Online]. Available: https://www.websense.com/assets/white-papers/whitepaper-websense-advanced-persistent-threats-and-other-advanced-attacks-en.pdf

[6] N. Villeneuve and J. Bennett. *Detecting Apt Activity With Network Traffic Analysis. Trend Micro Incorporated*, accessed on 2012. [Online]. Available: http://www.trendmicro.it/media/wp/detecting-apt-activity-with-network-traffic-analysis-whitepaper-en.pdf

[7] R. Jasek, M. Kolarik, and T. Vymola, "APT detection system using honeypots," in *Proc. 13th Int. Conf. Appl. Informat. Commun. (AIC)*, 2013, pp. 25–29.

[8] S. Kemp. *Digital in 2017: Global Overview*, accessed on Jan. 24, 2017. [Online]. Available: https://wearesocial.com/special-reports/digital-in-2017-global-overview

[9] ISACA. *Advanced Persistent Threat Awareness*, accessed on 2013. [Online]. Available: http://www.trendmicro.com/cloud-content/us/pdfs/business/datasheets/wp_apt-survey-report.pdf

[10] I. Ahmad. *How Many Internet and #SocialMedia Users are Fake?* accessed on Apr. 2, 2015. [Online]. Available: http://www.digitalinformationworld.com/2015/04/infographic-how-many-internets-users-are-fake.html

[11] Section 9 lab. *Automated LinkedIn Social Engineering Attacks*, accessed on Sep. 1, 2014. [Online]. Available: https://medium.com/section-9-lab/automated-linkedin-social-engineering-attacks-1c88573c577e

[12] G. Torston. *The Next Big Cybercrime Vector: Social Media*, accessed on Dec. 1, 2014. [Online]. Available: http://www.securityweek.com/next-big-cybercrime-vector-social-media

[13] S. Webb, J. Caverlee, and C. Pu, "Social honeypots: Making friends with a spammer near you," presented at the CEAS, California, CA, USA, 2008.

[14] K. Lee, J. Caverlee, and S. Webb, "Uncovering social spammers: Social honeypots+ machine learning," in *Proc. 33rd Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 2010, pp. 435–442.

[15] G. Stringhini, C. Kruegel, and G. Vigna, "Detecting spammers on social networks," in *Proc. 26th Annu. Comput. Secur. Appl. Conf.*, 2010, pp. 1–9.

[16] K. Lee, B. D. Eoff, and J. Caverlee, "Seven months with the devils: A long-term study of content polluters on twitter," in *Proc. ICWSM*, Barcelona, Spain, Jul. 2011, pp. 1–8.

[17] Q. Zhu, A. Clark, R. Poovendran, and T. Basar, "Deployment and exploitation of deceptive honeybots in social networks," in *Proc. IEEE 52nd Annu. Conf. Decision Control (CDC)*, Dec. 2013, pp. 212–219.

[18] Y. Boshmaf, I. Muslukhov, K. Beznosov, and M. Ripeanu, "The socialbot network: When bots socialize for fame and money," in *Proc. 27th Annu. Comput. Secur. Appl. Conf.*, 2011, pp. 93–102.

[19] A. Elyashar, M. Fire, D. Kagan, and Y. Elovici, "Homing Socialbots: Intrusion on a specific organization's employee using Socialbots," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining*, Aug. 2013, pp. 1358–1365.

[20] A. Elyashar, M. Fire, D. Kagan, and Y. Elovici, "Guided Socialbots: Infiltrating the social networks of specific organizations' employees," *AI Commun.*, vol. 29, no. 1, pp. 87–106, 2016.

[21] L. Bilge, T. Strufe, D. Balzarotti, and E. Kirda, "All your contacts are belong to us: Automated identity theft attacks on social networks," in *Proc. 18th Int. Conf. World Wide Web*, 2009, pp. 551–560.

[22] L. M. Aiello, M. Deplano, R. Schifanella, and G. Ruffo, "People are strange when you're a stranger: Impact and influence of bots on social networks," *Links*, vol. 697, no. 483, pp. 1–566, 2012.

[23] S. Mitter, C. Wagner, and M. Strohmaier. (Feb. 2014). "A categorization scheme for socialbot attacks in online social networks." [Online]. Available: https://arxiv.org/abs/1402.6288

[24] K. Krombholz, D. Merkl, and E. Weippl, "Fake identities in social media: A case study on the sustainability of the Facebook business model," *J. Ser. Sci. Res.*, vol. 4, no. 2, pp. 175–212, 2012.

[25] A. Paradise, R. Puzis, and A. Shabtai, "Anti-reconnaissance tools: Detecting targeted socialbots," *IEEE Internet Comput.*, vol. 18, no. 5, pp. 11–19, Oct./Oct. 2014.

[26] A. Paradise, A. Shabtai, and R. Puzis, "Hunting organization-targeted socialbots," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining*, Aug. 2015, pp. 537–540.

[27] C. A. Freitas, F. Benevenuto, S. Ghosh, and A. Veloso, "Reverse engineering socialbot infiltration strategies in twitter." [Online]. Available: https://arxiv.org/abs/1405.4927

[28] N. Abokhodair, D. Yoo, and D. W. McDonald, "Dissecting a Social Botnet: Growth, Content and Influence in Twitter," in *Proc. 18th ACM Conf. Comput. Supported Cooperat. Work Soc. Comput.*, 2015, pp. 839–851.

[29] J. Messias, L. Schmidt, R. Oliveira, and F. Benevenuto, "You followed my bot! Transforming robots into influential users in Twitter," *First Monday*, vol. 18, no. 1, Jul. 2013.

[30] E. Ferrara, O. Varol, C. Davis, F. Menczer, and A. Flammini. (Jul. 2014). "The rise of social bots." [Online]. Available: https://arxiv.org/abs/1407.5225

[31] H. Yu, P. B. Gibbons, M. Kaminsky, and F. Xiao, "SybilLimit: A near-optimal social network defense against sybil attacks," in *Proc. IEEE Symp. Secur. Privacy*, May 2008, pp. 3–17.

[32] G. Danezis and P. Mittal, "SybilInfer: Detecting sybil nodes using social networks," presented at the NDSS, California, CA, USA, Feb. 2009.

[33] D. Wang, D. Irani, and C. Pu, "A social-spam detection framework," in *Proc. 8th Annu. Collaboration, Electron. Messaging, Anti-Abuse Spam Conf.*, 2011, pp. 46–54.

[34] Z. Yang, C. Wilson, X. Wang, T. Gao, B. Y. Zhao, and Y. Dai. (Jun. 2011). "Uncovering social network sybils in the wild." [Online]. Available: https://arxiv.org/abs/1106.5321

[35] Y. Xie *et al.*, "Innocent by association: Early recognition of legitimate users," in *Proc. ACM Conf. Comput. Commun. Secur.*, 2012, pp. 353–364.

[36] Q. Cao, M. Sirivianos, X. Yang, and T. Pregueiro, "Aiding the detection of fake accounts in large scale social online services," in *Proc. 9th USENIX Conf. Netw. Syst. Des. Implement.*, 2012, p. 15.

[37] Q. Cao, X. Yang, J. Yu, and C. Palow, "Uncovering large groups of active malicious accounts in online social networks," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, 2014, pp. 477–488.

[38] O. Lesser, L. Tenenboim-Chekina, L. Rokach, and Y. Elovici, "Intruder or welcome friend: Inferring group membership in online social networks," in *Proc. Int. Conf. Social Comput., Behavioral-Cultural Modeling, Predict.*, 2013, pp. 368–376.

[39] Y. Kim, I. Kim, and N. Park, "Analysis of cyber attacks and security intelligence," in *Mobile, Ubiquitous, and Intelligent Computing*. Berlin, Garmany: Springer, 2014, pp. 489–494.

[40] M. Ask, P. Bondarenko, J. E. Rekdal, A. Nordbø, Ruthven, and P. B. Nordbo "Advanced persistent threat (APT) beyond the hype," Presented at the IMT4582 Netw. Secur. GjoviN Univ. College, 2013.

[41] C. Wuest. *The Risks of Social Networking*, accessed on 2010. [Online]. Available: https://www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/the_risks_of_social_networking.pdf

[42] P. Chen, L. Desmet, and C. Huygens, "A study on advanced persistent threats," in *Communications and Multimedia Security*. 2014, pp. 63–72.

[43] N. N. A. Molok, S. Chang, and A. Ahmad, "Information leakage through online social networking: Opening the doorway for advanced persistence threats," *J. Austral. Inst. Professional Intell. Officers*, vol. 19, no. 1, pp. 70–80, 2011.

[44] P. Giura and W. Wang, "A context-based detection framework for advanced persistent threats," in *Proc. Int. Conf. Cyber Secur. (CyberSecurity)*, Dec. 2012, pp. 69–74.

[45] Federal Bureau of Investigation. *Internet Social Networking Risks*, accessed on 2014. [Online]. Available: https://www.fbi.gov/file-repository/internet-social-networking-risks-1.pdf/view

[46] A. Banerjee, C. Banerjee, and A. Poonia, "Security threats of social networking sites: An analytical approach," *Network*, vol. 13, no. 12, p. 16, Dec. 2014.

[47] Trend Micro. *Social Media Malware on the Rise*, accessed on Feb. 24, 2015. [Online]. Available: http://blog.trendmicro.com/social-media-malware-on-the-rise/

[48] Proofpoint, *Cybersecurity Predictions for 2015*, accessed on Dec. 17, 2014. [Online]. Available: https://www.proofpoint.com/us/threat-insight/post/Cybersecurity-Predictions-2015

[49] Zerofox. *Top 9 Social Media Threats of 2015*, accessed on Jan. 20, 2015. [Online]. Available: https://www.zerofox.com/blog/top-9-social-media-threats-2015/

[50] P. Paganini. *LinkedIn—How to Exploit Social Media for Targeted Attacks*, accessed on Nov. 5, 2013. [Online]. Available: http://securityaffairs.co/wordpress/19446/cyber-crime/linkedin-targeted-attacks.html

[51] C. Pernet. *Reconnaissance Via Professional Social Networks*, accessed on Jun. 2, 2015. [Online]. Available: http://blog.trendmicro.com/trendlabs-security-intelligence/reconnaissance-via-professional-social-networks/

[52] B. Leiba. *Security and Social Networking*, accessed on 2012. [Online]. Available: https://www.pwc.com/us/en/it-risk-security/assets/social-networking-final.pdf

[53] P. Rashid. *Hackers Create Fictional People on LinkedIn to Engage in Industrial Espionage and Social Engineering Attacks*, accessed on Oct. 12, 2015. [Online]. Available: http://www.infoworld.com/article/2991532/security/fake-linkedin-profiles-lure-unsuspecting-users.html

[54] R. Stern, L. Smama, R. Puzis, T. Beja, Z. Bnaya, and A. Felner, "TONIC: Target oriented network intelligence collection for the social Web," presented at the AAAI, 2013.

[55] M. Berkovich, M. Renford, L. Hansson, A. Shabtai, L. Rokach, and Y. Elovici, "HoneyGen: An automated honeytokens generator," in *Proc. IEEE Intell. Secur. Informat.*, Jul. 2011, pp. 131–136.

[56] S. Patil, "Will you be my friend?: Responses to friendship requests from strangers," in *Proc. iConference*, 2012, pp. 634–635.

[57] P. Gupta, B. Srinivasan, V. Balasubramaniyan, and M. Ahamad, "Phoneypot: Data-driven understanding of telephony threats," presented at the NDSS, California, CA, USA, Feb. 2015.
[58] S. J. Bell, "Building a honeypot to research cyber-attack techniques interim report," Univ. Sussex, Brighton, U.K., Tech. Rep., 2013.
[59] Y. Elovici, M. Fire, A. Herzberg, and H. Shulman, "Ethical considerations when employing fake identities in online social networks for research," *Sci. Eng. Ethics*, vol. 20, no. 4, pp. 1027–1043, 2014.
[60] D. Dittrich, "The ethics of social honeypots," *Res. Ethics*, vol. 11, no. 2, pp. 192–210, 2015.

Abigail Paradise received the B.Sc. and M.Sc. (Hons.) degrees in information systems engineering from the Ben-Gurion University of the Negev, Beersheba, Israel, where she is currently pursuing the Ph.D. degree with the Department of Software and Information Systems Engineering.

Her master's and doctoral research focused on protecting organizations from attacks through social networks, and her current research interests include security and social networks.

Asaf Shabtai received the Ph.D. degree in information systems engineering from the Ben-Gurion University of the Negev (BGU), Beersheba, Israel.

He is currently a Senior Lecturer (Assistant Professor) with the Department of Software and Information Systems Engineering, BGU, where he is also a Senior Researcher with the Telekom Innovation Laboratories. He is a recognized expert in information systems security and has led several large-scale research projects in this field. He has authored over 60 refereed papers in leading journals and conferences and has co-authored a book on information leakage detection and prevention. His current research interests include computer and network security, machine learning, security awareness, smart mobile security, user profiling, social network security, Internet of Things security, and security of avionic systems.

Rami Puzis received the B.Sc. (Hons.) degree in software engineering and the M.Sc. and Ph.D. (Hons.) degrees in information systems engineering from the Ben-Gurion University of the Negev (BGU), Beersheba, Israel.

He was a Post-Doctoral Research Associate with the Laboratory for Computational Cultural Dynamics, University of Maryland, College Park, MD, USA. He is currently a Faculty Member with the Department of Software and Information Systems Engineering, BGU. Over the past few years, he has managed several research projects funded by Deutsche Telekom AG, the Israeli Ministry of Defense, the Israeli Ministry of Trade and Commerce, and leading cyber security companies in Israel. His recent research projects have focused on web intelligence, security awareness in mobile environments, and protecting organizations from attacks through social networks. His current research interests include network analysis with applications for security, social networks, computer communication, and simulations.

Aviad Elyashar received the B.Sc. and M.Sc. degrees in software and information systems engineering from the Ben-Gurion University of the Negev (BGU), Beersheba, Israel.

He is currently a Data Scientist with the Telekom Innovation Laboratories, BGU. His master's degree research focused on protecting organizations from attacks through social networks. His current research interests include network analysis with applications for detecting abusers within social networks, and fake news detection.

Yuval Elovici received the B.Sc. and M.Sc. degrees in computer and electrical engineering from the Ben-Gurion University of the Negev's (BGU), Beersheba, Israel, and the Ph.D. degree in information systems from Tel Aviv University, Tel Aviv, Israel.

For the past 14 years, he has led the cooperation between BGU and Deutsche Telekom, Israel. He is currently the Director of the Telekom Innovation Laboratories, BGU, the Head of BGU Cyber Security Research Center, the Research Director of iTrust at SUTD, Singapore, the Lab Director of ST Electronics-SUTD Cyber Security Laboratory, Singapore, and a Professor with the Department of Software and Information Systems Engineering, BGU. He also consults professionally in the area of cyber security. He is the Co-Founder of Morphisec, Israel, a startup company that develops innovative cyber security mechanisms that relate to moving target defense. He has authored articles in leading peer-reviewed journals and in various peer-reviewed conferences and has co-authored a book on social network security and a book on information leakage detection and prevention. His current research interests include computer and network security, cyber security, web intelligence, information warfare, social network analysis, and machine learning.

Mehran Roshandel received the master's degree in computer science from the Technical University of Berlin, Berlin, Germany, in 1994.

From 1991 to 1997, he was a Programmer, Software Engineer, and Developer Team Leader with the field of distributed platforms and AAA, Fraunhofer Fokus, Berlin. Since 1997, he has been with different units of Deutsche Telekom AG (T-Systems and Telekom Innovation Laboratories), Germany, where he was involved in the fields of platforms, security, usability, and data analytics with machine learning methods. He has led several large international research projects in these fields. Furthermore, he is responsible for productization of systems in the fields of security, risk management, fraud detection, and monitoring at Deutsche Telekom AG, Germany.

Christoph Peylo received the Ph.D. degree in artificial intelligence from the University of Osnabrück, Osnabrück, Germany. He has studied computer science, computational linguistics, and artificial intelligence.

He held various positions ranging from Software Engineer to Managing Director of a software company. In 2006, he joined the Deutsche Telekom Laboratories, Berlin, Germany, as a Vice President, where he was involved in the area of artificial intelligence, cyber security, Industrie 4.0, and Internet of Things. He is currently the Global Head of the Bosch Center for Artificial Intelligence, Palo Alto, CA, USA; Bangalore, India; and Renningen, Germany.