Magdeburger Journal zur Sicherheitsforschung

Gegründet 2011 | ISSN: 2192-4260
Herausgegeben von Stefan Schumacher und Jörg Sambleben
Erschienen im Magdeburger Institut für Sicherheitsforschung

This article appears in the special edition »In Depth Security – Proceedings of the DeepSec Conferences«.
Edited by Stefan Schumacher and René Pfeiffer

An innovative and comprehensive framework for Social Vulnerability Assessment

Enrico Frumento & Roberto Puricelli

Nowadays security attacks greatly rely on the human vulnerabilities, hence is fundamental to include the human factor into corporate risk analysis. However, is it possible to evaluate this risk through a specific type of vulnerability assessments? Since 2010, we have been working on the extension of traditional security assessment, going beyond the technology and including the “Social” context. In these years, we assessed several big European enterprises, understanding the impact of these activities on the relations among employees and employer, both from ethical and legal points of view. We developed a innovative methodology for Social Driven Vulnerability Assessments (SDVAs) that we present in this paper beside the early results. As part of their Advanced Threat Protection (ATP) programs, we performed more than 15 SDVAs in big enterprises with a gross number of 12.000 employees; this gave us a first-hand sight on the real vulnerabilities against modern non-conventional security threats.

Keywords: Social Driven Vulnerability Assessment, Social Engineering, Risk Analysis

Citation: Frumento, E. and Puricelli, R. (2014). An innovative and comprehensive framework for Social Driven Vulnerability Assessment. *Magdeburger Journal zur Sicherheitsforschung*, 2, 493–505. Retrieved December 16, 2014, from http://www.sicherheitsforschung-magdeburg.de/publikationen.html Version 2014/12/17 21:51

1 Introduction

Latest insights into security breaches reveal that most of the security incidents include the human element as a major component of their attacks: about 90% of them include several enabling steps belonging to the area of Social Engineering (SE) [1]. Current approaches to IT security and risk management tend to underestimate, or even ignore, the human element in their calculation due to a lack of assessment models, tools, processes and legal backing. However recent statistics [1] provide additional insights on the actual concerns of SE: (1) 1 year is the average time to discover an attack performed via SE; (2) 5 emails are the average number of emails needed to create an entry point in a company; (3) attacks are typically discovered by third parties. These points show that something important is happening in the way attackers perform their actions against citizens and enterprises.

SE is not new; it has been actively used in specific attacks since the 1980 and 90s [2][3], but lately evolved into a new model, which we conveniently call Social Engineering 2.0, characterized by several new aspects:

- higher level of complexity;
- heavy usage of open available information;
- extended scope of (potential) attackers;
- nearly full automation of SE attacks;
- focus on making money [4].

The transition from SE to SE 2.0 follows the societal evolution of the last few decades which saw an increasing exposure over the network of people's personal details and information, across the whole society [5]. This exposure is already extensively reshaping the concepts of identities, privacy and even the perception of the ego [6] with profound impacts on the way people work [7] and are attacked.

In this context a quotation of Bruce Schneier [10] helps to understand the situation in just a few words: »Good old days of (in)security are back«. This quotation builds on important trends in SE: (1) main stream entities demonstrated to be incredibly weak against SE based attacks (e.g., [8][9]); (2) crushing attacks can be launched even by a single attacker; (3) awareness programs are incredibly inefficient [11]; (4) Classical protection technologies (e.g., antivirus, firewall, etc.) are inefficient against these attacks [12].

This is the scenario where the new cybercrime activities prosper and the most critical for enterprises because require new protection mechanisms, supported by proper risk assessment methodologies.

The paper in Section II presents the Advanced Persistent Threat (APT) model and its relations with SE 2.0. Section III introduces the SDVA concept, while Section IV introduces our framework. Section V shows some insights related our experience in SDVA. Section VI concludes with an analysis of the cumulative results collected with our assessments.

2 Advanced Persistent Threats

Advanced Persistent Threats (APT) commonly refers to a new kind of cyber threats, targeted against a specific entity, with the purpose to obtain control over the internal perimeter, using a combination of multiple attack vectors and techniques. APT often begins with sophisticated social engineering attacks, in order to break the perimeter, and uses »advanced« malware to avoid detection. Indeed, evolution of the infection and Open Source INTtelligence (OSINT) technologies allows for a wider range of attackers to hit normal victims. The following phenomenon allowed the change: (1) the evolution of the social media (SM) even through mobile platforms and the corresponding new people's habits; (2) the possibility to automate SE against a large number of people/victims; (3) the possibility to automate most of the attack steps using low cost homemade tools; (4) the use of a »glocalization« approach to precisely select the victims[1].

As a consequence, modern attacks follows the steps of Fig. 1 that describes the whole process involved in APTs: (1) preparatory OSINT phase, especially on the Social Media (in this case we refer to Social Media Analysis –SMA); (2) selection of the most vulnerable human targets, followed by contextualized SE attacks (3). The victims of the SE attacks are hence hit with ad-hoc infections (4), followed by an expansion of the attack inside the perimeter (5), whose aim is the automated evaluation of which assets the victims accesses (typically using a digital shoulder spy approach) and a seek-and-infect phase. The last step being the data exfiltration (6), once the attacker find an interesting target [29].

2.1 The Social Engineering 2.0

Modern SE evolved in the last years into something more complex that we call Social Engineering 2.0. This evolution is probably the relevant reason behind the spread of APT-like attacks. Fig. 2 sums the most relevant trends of SE 2.0 that we identified:

- Malware Ecosystem 2.0: SE became an important part of the malware 2.0 and its main infection strategy [13]. The inclusion of SE shaped the malware and the infections strategies [40]. For example, the need of privilege escalation is greatly reduced since the probability of infecting the right victim (which already owns the asset the attack is searching for) is higher.
- Automatic Social Engineering Attacks (ASE): automation of SE attacks through information collection and data mining and through the sentiment analysis from Social Media has been already anticipated [14], but only nowadays be-

[1] The term is a crush of globalization and localization and in this paper, we use it in the area of phishing to identify the essence of the modern phishing techniques that use »globally« automated social media scanning and spamming technologies to »locally« customize the hook emails for each victim's context.

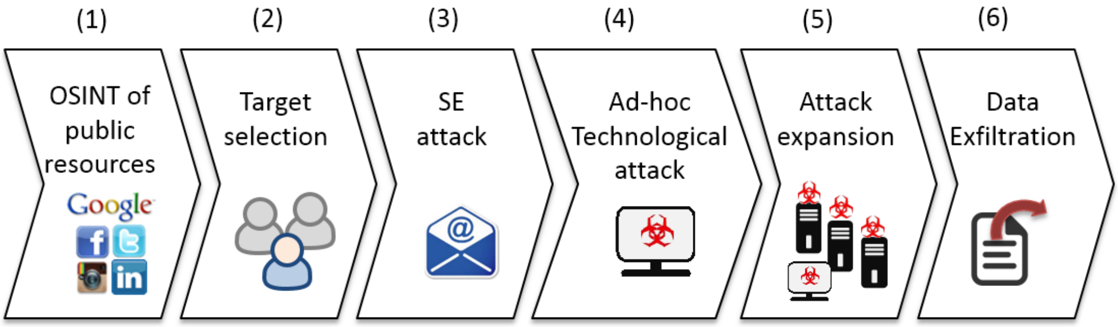


Figure 1: Steps involved in a typical APT attack.

came mainstream. The advantage of social media, by the attackers« point of view, is that they return machine processable data, validated by other peers (e.g., classmates, friends).

- (ab)use of linked-data: Public Administration are moving to the Web 3.0 paradigm based on Linked open Data (LoD) [15]. An huge opportunity to improve the efficiency of an SE attack, automatically increasing the accuracy of the phishing hooks, comes from the correlation of this mass of information with the victim's context and social media [20]. Despite still relatively low exploited it is a trending tactic [16].
- Chat-bot: diffused use of chat-bots, as in ASE attacks, to start and maintain conversations with other social media users and to balance the lack of a human social engineers (i.e. mass SE attacks) [17]. The average communications on the social media are quite simple (e.g., twitter messages are very short) and this help to overcome the known limitations of chat-bots.
- (Ab)use of psychology, personality profiling systems and cognitive science models: professional use of memetics [18] and personality models [19] of the attacked users, especially of models coming from theories of cognitive psychology [21]. A fundamental evolution is the application of cognitive sciences and semantics technologies in order to automatically profile personalities and find potential victims on large mass of online persons.
- Email attack vector: the massive use of mails, if compared to other attack vectors (e.g., presence, voice, chat), increased a lot in sophistication, since it does need less talented hackers (e.g., the ability to use the voice attack vector –i.e. on the phone– is more complex, because requires control of non-verbal messages, voice, tones, cadence etc.) and it can reach lot of victims at a time [22].
- Economic Drivers: as for malware now, SE 2.0 is an investment and all the attacks using it are prepared only to make money. It makes no sense to use SE 2.0 for non-professionals attacks since it is an instrument whose aim is just to make money [23]. This is an important aspect that creates a methodological »connection« between SE 2.0 and the modern marketing tactics, like viral marketing or social advertising [28].

All the characteristics listed are rooted on technologies that are also used for a proper meaning, but at the same time could be abused by SE to perform attacks and collect information, which are exploited for highly contextualized attacks (e.g., linked open data). Summing up, the real essence of SE 2.0 is the abuse versus the use.

2.2 Spear Phishing

SE 2.0 most used attack vector is the email. Unlike in traditional phishing, modern spear phishing attacks are sophisticated and contextualized, hence they are effective. Research in this field showed that the effectiveness is related to the fact that users mainly discriminate the legitimacy of an email or a website based on the look and feel [22]. Using public available information, cybercriminal are able to replicate credible templates in order to deceive users to perform risky behaviours. This could be related to the fact that, although the problem is well known, users are not completely aware about the severity of their action [37] and useful methods for identifying phishing attacks [38].

3 The Social Driven Vulnerability assessment model

Sections I, II explained where research in the security area is lagging behind, fully operational solutions that address this problem at an integrated level are still not present on the market. In practice, companies currently face a major challenge due to the lack of established countermeasures [24]. Employees usually have knowledge on critical company data and are fully integrated in the company security system, whilst often are the »weak points«. Testing the resilience of the human factor inside companies is an important element of any ATP program and brings important benefits, but as Fig. 3 shows, it is a multifaceted problem.

Our first and most important objective was to make SE attacks a known and properly evaluated risk for

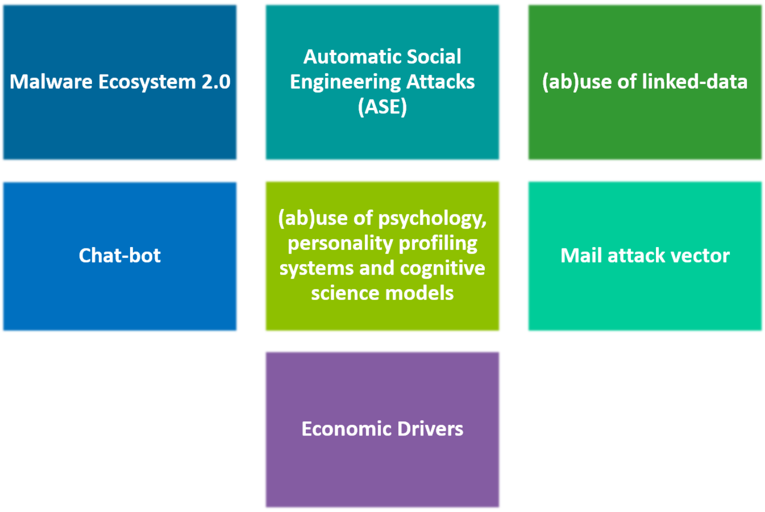

Malware Ecosystem 2.0
Automatic Social Engineering Attacks (ASE)
(ab)use of linked-data
Chat-bot
(ab)use of psychology, personality profiling systems and cognitive science models
Mail attack vector
Economic Drivers

Figure 2: Overview of the main SE 2.0 characteristics.

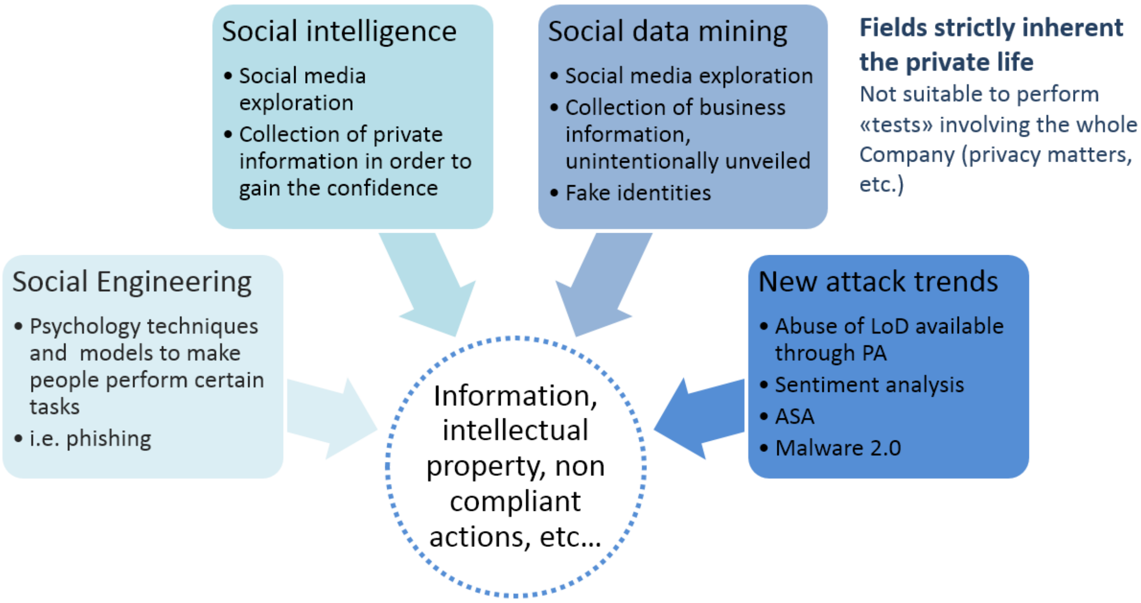

Social intelligence
• Social media exploration
• Collection of private information in order to gain the confidence
Social data mining
• Social media exploration
• Collection of business information, unintentionally unveiled
• Fake identities
Fields strictly inherent the private life
Not suitable to perform «tests» involving the whole Company (privacy matters, etc.)
Social Engineering
• Psychology techniques and models to make people perform certain tasks
• i.e. phishing
Information, intellectual property, non compliant actions, etc...
New attack trends
• Abuse of LoD available through PA
• Sentiment analysis
• ASA
• Malware 2.0

Figure 3: The role of the human factor in SE 2.0 and problematic areas.

companies. Unfortunately, the assessment and calculation of the risk connected to social driven vulnerabilities are extremely complex. Indeed, it requires a mix of expertise:

- psychological, to find a possible vulnerability to exploit and to find an effective way to exploit it;
- technical, to create an effective attack that can exploit the vulnerability;
- legal and strategic, since the measurement of the risk might lead to the exposure of sensitive information;
- societal, since the consequences of an attack can have major consequences, ranging from the resignation of a CEO to the bankruptcy of the enterprise [25].

Despite a number of software tools (both to gather information and to exploit the vulnerability) and well known practices that can be used for this purpose are available [26], a comprehensive framework is still lacking.

A security assessment aims at simulating, in a way as realistic as possible, attack patterns, before they really happen, in order to measure the real vulnerability of an enterprise. A SDVA is a new type of assessment, a crucial element of holistic risk management, which actively uses SE 2.0 techniques to simulate an attack against the enterprises. SDVAs might also foresee infection of the victims« terminals or, more realistically, infections of their clones, using ad-hoc malware.

The most important elements of an SDVA are:

1. realistically simulate the SE 2.0 based attacks;
2. evaluate the technology-enabled breaks opened as a consequence of the SE based vulnerabilities;
3. ethically respect the employee and comply with

the existing legislations (i.e. at the Italian level for what concerns this paper);

4. contextualize the attacks at either enterprise, teams or single employee levels;
5. involve the strictly required departments, with only the required details;
6. analyse and interpret findings correctly, in order to create a report of results;
7. use the results to find long-term lasting solutions (e.g., through innovative awareness methodologies).

3.1 Legal and Ethical perspective

The essential aim of a SE attack is to trick the employee and force him to violate a policy. By doing this cybercriminals do not have scruples, using whatever information they can retrieve. Despite a SDVA has the same purpose, companies have to observe severe moral and legal limitations. In particular, from a moral perspective, the assessment should be executed guaranteeing the respect of the relationship between employer and employee, avoiding to invade the personal sphere. For example, impersonation using fake identities is a common attack strategy that cannot be simulated in SDVAs. Moreover, it is necessary to consider the labour legislation that particularly in Europe protects employees from any interference of the employer. For example in Italy, a law prohibits the employer to monitor the behaviour of employees or interfere with their private lives (i.e., how they behave on the SMs); hence in an assessment it is not possible to reveal the details of single users involved. This impacts the SDVA at the technical level because the information of employees must be inaccessible either from the security testers or the employer, only the system knows them. However US and Europe have very different legal frameworks and these activities are easier in the US market (e.g. [34]). Despite these limitations the interest on this topic is increasing even in EU (e.g., [35]), and it is important to consider that to realistically simulate an SE attacks for a SDVA implies some legal and ethical risks [27], hence the overall legal compliance is a strong requirement.

4 A Framework for SDVAs

During the last five years we had the opportunity to work on this topic with several European big enterprises, allowing us to face the difficulties related to the impact of this kind of activities on the relational issues between employees and employer both from the ethical and legal points of view.

This experience allowed us to develop a specific methodology for performing SDVAs, ensuring ethical respect for employees and legal compliance with European work regulations and standards.

This Section explains the methodology and all the phases of an assessment alongside the main activities.

4.1 Setup

A SDVA is a relatively new type of security testing in the enterprises and often a risky one on its own. Hence, the first operation is a setup phase, whose purpose is to involve only the strictly required stakeholders, explain the threat, share the objectives, define the scope of the assessment, obtain agreement and retrieve the needed information.

Although this step might seem obvious it is of paramount importance for SDVAs, because is the earliest moment where the stakeholders face the security problem and raise ethical and legal concerns that must be immediately addressed, as reported in III.A. Consider also that these activities can be presented either as a risk reduction strategy or as a part of the corporate responsibility program.

The stakeholders usually come from different company's departments and in our experience the required ones belong to:

- IT, to define/configure the IT services, for assessment and to solve any possible technological constraints that could invalidate the test (e.g., tweak spam filter, warn security helpdesk responsible);
- HR, to define the characteristics of the users sample (i.e. how the sample is composed) target of the SDVA;
- Legal, to share the precautions taken to not violate the laws and gain his placet;
- Communication, to properly design the hooks used in the SDVA[2], with a coherent style and to avoid collisions with existing company activities. While an attacker does not care about consequences of his attacks, in the enterprise avoidance of the internal conflicts is mandatory.

The most important output of this phase is to share the objectives and the scope of the activity, in particular the boundaries of the assessment:

- for the social media information mining phase, the Social Media included and the type of scannings (i.e. how deep);
- the spear phishing attack simulation phase, usually performed by email, the level of contextualization of the hooks and the definition of the employees sample;

4.2 Passive Social Information mining

In this phase, we simulate an attacker seeking information about the employees of a company, published mainly on Social Media in order to gain knowledge of potential victims for creating an effective attack.

[2] A hook is in general the trick used to catch the user, either using a drive-by-download or drive-by-infection strategy. Possible hooks are baiting, phishing emails, malevolous sites or forms, phone calls, etc.

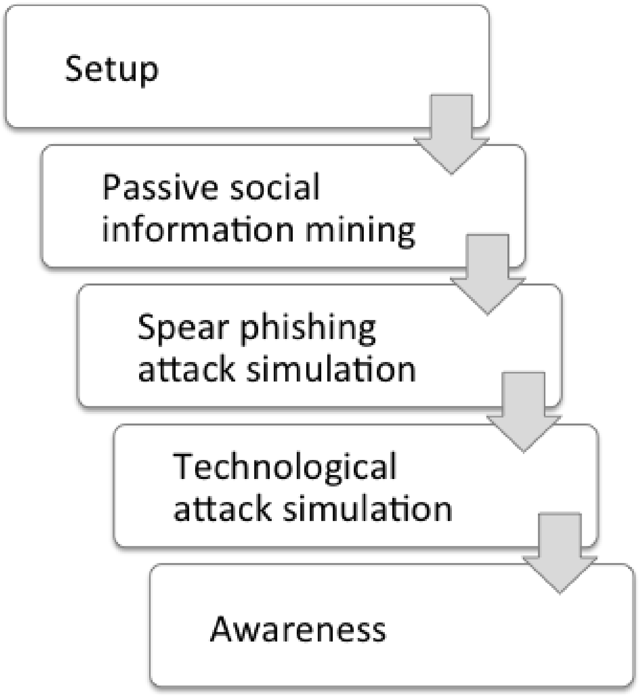

Setup
Passive social information mining
Spear phishing attack simulation
Technological attack simulation
Awareness

Figure 4: Social Driven Vulnerability Assessment Framework.

Information mining could be performed in two different ways: active which includes creation of fake identities in order to get in touch with the victim actively, and passive where it is included only the gathering of information publicly released by the victim (i.e. not properly protected). To respect the employees involved in the SDVA and to avoid legal problems, we only do passive scanning at the level of the company's brand.

We develop a toolchain, combining Open Source Intelligence (OSINT) tools that seeks most of the information almost automatically.

At the end of this phase, we obtain three different outputs:

- insight related to company initiatives, templates, or any other information that allow to craft a contextualized emails that could be used during the test;
- a list of employees email address, publicly available, or inferred from names and format of the company email, that could be the potential target of the assessment;
- evidence of specific content shared from user that can constitute a risk itself, such as picture in high quality of offices, badges, post-it with passwords, or internal documents containing critical information.

The most critical part of this phase is reporting: due to legal constraints, the employer cannot know the identities of whom illicitly shared information on the Social Media hence we properly anonymize the results collected.

4.3 Spear Phishing attack simulation

The central core of an SDVA is to test the personnel behaviour against a customized hook, which tries to trick the user to perform an action that could put at risk the company's assets. Possible hooks are baiting, tailgating, but the most requested is contextualized phishing using drive-by-infection [31] and/or drive-by-download [30]. The email is properly crafted and contains links to a controlled website that asks to insert a critical information, typically enterprise credentials.

Fig. 5 shows where the phishing tests we usually run in an SDVA are placed in a Contextualization (Volumes) space. The Volumes axis refers to the number of identical email sent in a phishing campaign and has three values: (1) mails sent to few selected victims; (2) mail sent to a subset of the whole company (e.g., a department); (3) mail spread to all the employees. The Contextualization axis refers to the degree of contextualization the email has (e.g., custom graphic, real argument) and has three values: (1) generic, thus not customized at all; (2) company, thus properly customized for the specific enterprise (e.g., use of the official look or logo); (3) person, thus contextualized to a single person's interests. To help understanding this classification we placed the classic and the RSA phishing samples [29][31] as reference. This graphical taxonomy helps to immediately spot four important areas:

- Today unfeasible attacks: phishing customized at personal level, but spread to a large number of victims. This area will become popular with the improvement of the semantic and sentiment analysis technologies.
- Anti-economic attacks: phishing attacks targeted

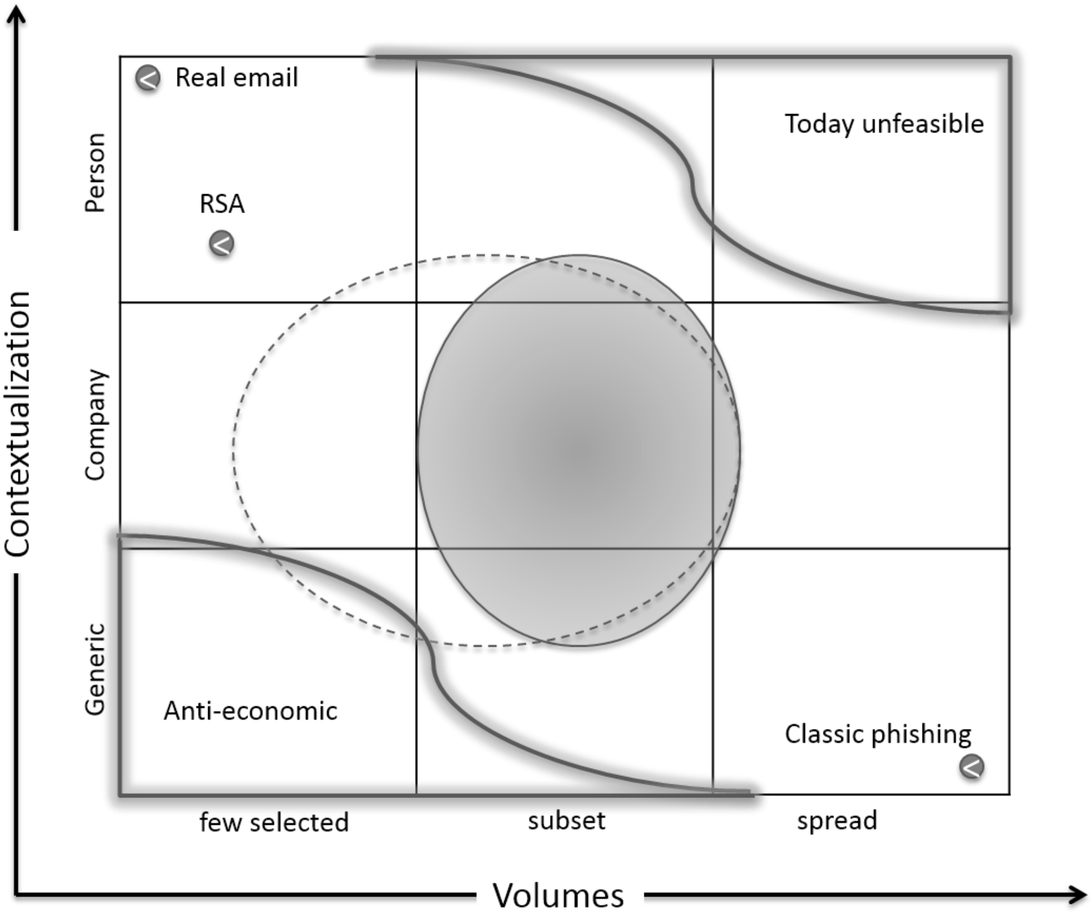


Figure 5: A conceptualization of the mail space useful to identify the type of phishing tests performed during an SDVA.

at few selected people, but not customized at all, are not economically sustainable nor convenient.

- The upper left corner, where is the RSA sample, is unfeasible in the SDVAs due to legal reasons (i.e. this test would require active SM scanning).
- The lower right corner, where is the classic phishing sample, is not useful to be tested with a SDVAs (i.e. companies already have lot of samples of this type).

Hence, the most convenient place for the SDVAs email tests is the grey area in the centre. The dotted circle reports a legally possible, but nowadays still unexplored, extension of SDVAs on restricted groups for very mission critical employees (e.g., only directors or restricted project teams).

Specifically the assessment allows evaluating two different types of risks:

1. The user clicks on the link inside the email and visits the website, exposing himself to a drive-by-infection attack;
2. The user provides also the requested information into the website form, providing critical information, such as enterprise credentials[3].

For both steps it is necessary to track the user behaviour, thus each email contains a unique link, managed by the system to be completely anonymous: the most important requirement of the assessment methodology is that the system must prevent the identification of the employees who fall victim of the hook. Only statistically anonymized results are allowed.

In order to correlate the technological risk (e.g., unpatched systems) with the hook effectiveness, the system also fingerprints the terminal. The information collected helps us to understand the level of exposure to ad-hoc technological attacks (e.g., [32]).

A tricky activity required by Step B is to check the information that the users supplied, against their real credentials. Due to its extreme value of this information for the company, we shaped our methodology to comply with confidentiality requirements, hence we check under a closed cryptographic system the credentials match.

What we developed is also a dashboard for execut-

[3] The enterprise's credentials are not useful from outside the company's network without a VPN access, but we usually ask them since a release of this valuable asset is a strong sign of an exploitability. The enterprises insist about the extreme value of this information since the first working day.

ive, that graphically allows to monitor real time statistics about the ongoing and past (as a reference) tests. Checking against past tests is also important because helps to understand if the risk is reduced.

Our experience says that the adherence to the real attacks is of paramount importance to not bias the results. The most important point of attention is the contextualization level of the hooks: too much adherence to the real emails (i.e. too contextualized) is usually a con, because people in companies are often security-trained at different levels and it is interesting to pair with the SDVA a verification of the training effectiveness. We usually do this adding tiny inconsistencies in both the emails and the website, which could potentially be detected by people. Another attention we usually pay is to spoof the emails and offshore host the website, in order to hide the identity of the penetration tester and prevent the association of these emails/websites with the enterprise (e.g., this is useful if the enterprise wants to keep the SDVA secret).

As also reported in [21] we also test the knowledge of the notification policies to the IT help centre, tracking the warnings flows inside the company (e.g., a common situation is, a victim recognizes the threat but handles it badly, forwarding it too late and/or to the wrong contact, thus delaying the company's reaction). This step also offers the opportunity to test the reaction of the internal CERT.

This phase output is an anonymized analysis of the collected data during, also during time (See Section V) whose main results are:

- help understanding the overall risk exposure to spear phishing attacks,
- estimate the rapidity of an attack during time.

The correlation of the characteristics of the used hook with the overall impact and the response time provides a deep analysis of the most critical employees profiles (we use the personas approach [33]) and gives insights for the awareness programs. In addition, it is interesting to compare the click-rate of the credential theft with the notifications, in order to evaluate the users »readiness«.

4.4 Technological Attack Simulation

The attacks simulated/assessed by SDVAs usually ends with a digital shoulder spying activity: a backdoor inside the private network from which starts the silent expansion of the infection, while searching for valuable assets to exfiltrate. This kind of attacks are thought to run undercover and exploit systematic vulnerabilities or incorrect prevention solutions.

Our SDVAs reports usually include a Proof-of-Concepts (PoC) that shows how to compromise the typical terminals of the company, for example after a visit to a malicious website. The PoC is tied to the results of the phases B and C: the phished data are used to create a custom ad-hoc malicious program.

The main PoC requirement is to not add risks or create problems on its own, hence this step is executed on a specific isolated installation, cloned from the company setup for the victims profiles identified in Step C.

The great advantage of performing such a PoC at this stage, after phases from A to C, is that it is extremely easy to create an ad-hoc malware that deeply exploits the weaknesses between the defence systems. Without mentioning those companies which have badly configured defence solutions, modern countermeasures follows a »defence in depth« approach; an ad-hoc malware must be properly studied to avoid at the right moment the controls provided by antivirus, inline anti-malware, firewall, etc. Implementation of such »surgical« malware is easier using the information gathered with an SDVA.

A common follow-up of this activity is a normal vulnerability assessment/penetration test of the internal network from the terminals perspective and the integration of its results with those coming from the SDVA.

4.5 Awareness

Companies usually provide training programs to their employees about the phishing risks and warnings to be vigilant using the Social Media. The extension of these programs is shaped by legal constraints: as already shown in Fig. 3 the most sensitive areas where awareness plays an important role belong to the private lives. What companies usually do is to »convince« the employees, writing Social Media Guidelines, to better control their own SM lives. It is known that SMs are a potential source of attacks, however these training activities are blindly submitted to all the employees and there is no way to measure their effectiveness. As a matter of facts, according to our results, the tactics used in the SDVAs are effective.

The awareness actions after an SDVA are often of two types:

- Sensibilization of the management: the type of threats exploited by an SDVA are easily understandable by non-technical people.
- Sensibilitazion of the employees: usually the results of an SDVA are not published because the vulnerabilities found are not easily patchable and could last for long times. Hence, the publication of these reports is often strictly confidential. Anyway, these results are used to shape the global awareness program or to train specific groups or profiles.

Not all the awareness programs work for all the employees [21][26], SDVAs offer a tool to correctly address the efforts. The general problem of any security related awareness program is anyway how to create long-lasting training programs [11]. This is still an open issue in security since there are no best practices and a large space for experimentations is still open.

5 Results Highlights

Our work and research on SDVAs is ongoing since several years. Thanks to the assessments performed, we collected many data on user behaviour facing spear phishing attacks that gave us a first-hand measurement of the risk.

This section presents the main facts and insights. Our aim is to make companies aware of the actual risks of this threat.

5.1 Results comparison

In these years, we performed about 15 SDVA in big enterprises with thousands of employees (a gross number of 12.000 people). Note that what we present here is a selection of the aggregated results, that have been elaborated to comply the privacy and non-disclosure agreements we signed. The results focuses on the spear phishing attacks phase, because it is the most representative for comparison among different companies and does not reveal anything about possible real vulnerabilities of the assessed companies. Furthermore, to make these results consistent, we selected only those assessments that are similar in terms of threat and type of enterprise. The phishing tests presented have all the same characteristics:

- the target is a sample of the employees that represents the entire population;
- the campaign is someway contextualized for the company, using at least colours, logos, template or name of the company and a proper style of communication;
- the campaign is related to general arguments, such as promotions or discounts for the employees, but not related to specific initiatives concerning directly the company.

An example of a possible contextualized hook is shown in Fig. 6: note that the argument is generic, but still potentially interesting and the contextualization is related to the company logo and colours. Nevertheless it is interesting to underline that variations on the proposed template in some cases do not upsets the results of the SDVA. However the characteristics presented are quite common in our SDVAs also due to legal constraints; using the schematization shown in Fig. 5, our hooks are in the central sector.

5.2 Benchmarking

Spear phishing is known to be one of the most dangerous risks for companies [36]. This is also evident from our on field tests. Fig. 7 shows a comparison between the overall results of the two steps, which the assessment is composed of (see Section IV.C). Each circle represents one company and its colour is the corresponding industrial sector, the x-axis is the percentage of employees of the sample who clicked on the link, the y-axis is the number of them who also inserted credentials. The radius of the circle is logarithmically proportional to the dimension of the company.

The immediate result is that a spear phishing attack, slightly contextualized to the company, generates the following risky behaviours:

- An average of 34% of employees follow the link included in the email;
- An average of 21% of employees also inserted their own credentials.

Results of Fig. 7 are higher than one would expect, looking at the worldwide incidence statistics of phishing (approx. 10% [39]), but similar to the results recently presented by McAfee (approx. 80% [36]), collected using a simulated web-quiz environment (our SDVAs do not simulate the working environment and the emails are delivered to real inboxes). Looking at the results it is evident how few emails are enough to potentially break into a company.

5.3 Temporal analysis

One of the most used psychological tricks with phishing is to put the users in an urgent situation in order to shortcut their decision processes and forcing them to commit errors. We study our hooks using cognitive psychology or even memetics [18], but the effectiveness of a hook is measured looking at how fast victims fall in the trap. This not only gives a measure of the used »meme«, but also helps the companies to understand if their reactions procedures are fast enough. The effectiveness of spear phishing in the early minutes is also important to have an idea of the rapidness of the attacks: when the attacker collects enough victims, the hooks are quickly dismantled and all the afterward investigation efforts are doomed to fail.

All the data are normalized to the success rate of each assessment, in order to allow a direct comparison.

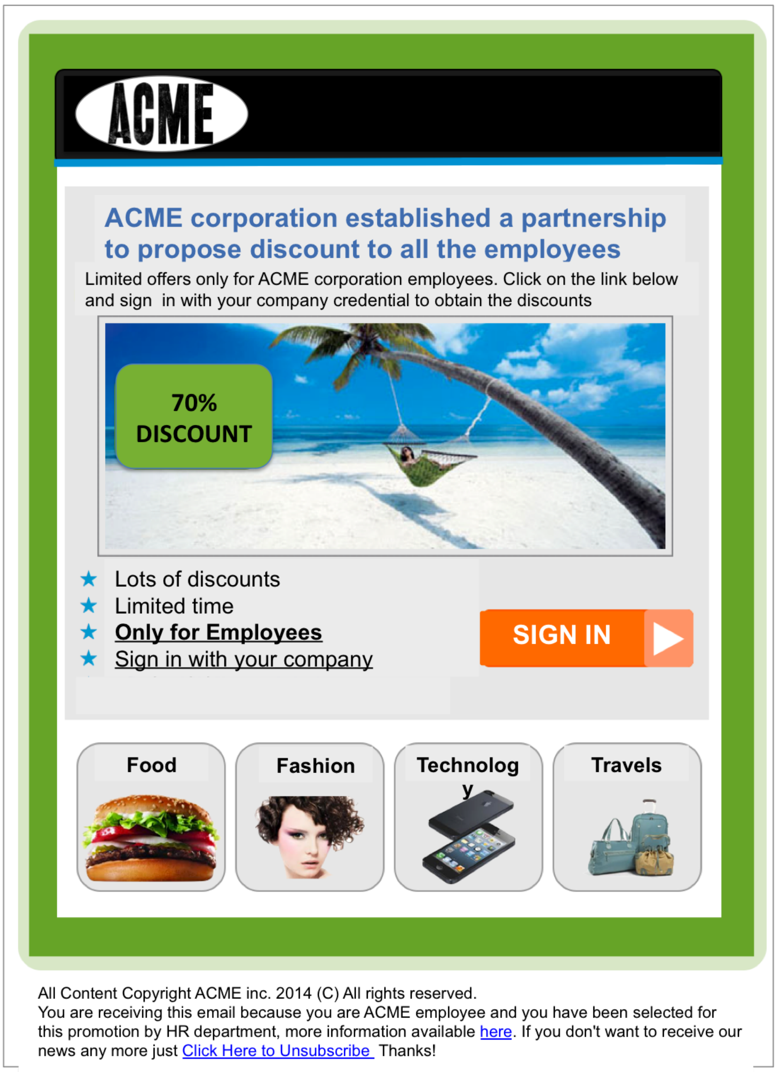
The chart of Fig. 8 maps the success rate of the campaign, meant as the ratio between the number of employees who performed an action that could introduce a risk for the company in a certain time (either simple visit or credential insertion), and the overall result of the campaign, all this on the first two hours of the assessment.

Although there are differences between the curves, most of them show a rapid growth in the early moments of the assessment and afterward a slower increase, until a plateau, where the »hook power« can be considered exhausted.

Considering the averaged results, it is interesting to observe that:

- 41% of the effectiveness of the hook is reached in the first 10 minutes;
- 50% of the global effectiveness is reached after around 20 minutes.

Combining this analysis with the results reported in the previous section we must underline that, even in the best case when the campaign is successfully



ACME

ACME corporation established a partnership to propose discount to all the employees

Limited offers only for ACME corporation employees. Click on the link below and sign in with your company credential to obtain the discounts

70% DISCOUNT

★ Lots of discounts
★ Limited time
★ **Only for Employees**
★ Sign in with your company

SIGN IN ▶

Food | Fashion | Technology | Travels

All Content Copyright ACME inc. 2014 (C) All rights reserved.
You are receiving this email because you are ACME employee and you have been selected for this promotion by HR department, more information available here. If you don't want to receive our news any more just Click Here to Unsubscribe Thanks!

Figure 6: An anonymized example of phishing hook used in a SDVA.

blocked after 20 minutes (this is a fast reaction time in an enterprise), about 14% of employees included in the assessment visited the website and 9% also inserted the credentials. This result poses a strong warning on the effectiveness of the automated contrast methods: this interval is short, also considering the user reaction described in the following section.

5.4 User reaction

During assessments, with the collaboration of the IT department, we track the users reactions, meant as any warning sent out that reveals some suspect (e.g., alerts/request for clarification).

What we saw is that most of the companies do not have formal procedures describing how to behave in case of a suspicious email, or at least employees are not aware of them.

Indeed, only an average 1% of tested users started some type of warnings, and performed it in different ways (contact ICT friends, lawyer's office, their boss, etc. through email, phone, voice . . .). Despite the lack of coordinated reactions (which anyhow slows down the enterprise reaction), the average time between the start of the attack and the first alert is about 6 minutes, and, according to Section V.C, at this stage the attack has already collected enough information.

After this first alert the average reaction time of a medium sized enterprise is usually above 20 minutes, considering the best scenario where the warning directly reaches the right person, who understands the problem and proactively acts. Comparing these values with the success rate described in Section V.C, it is evident that the attack and reaction times are not matching.

5.5 User Characterization

During SDVA, we anonymously correlate the results to the characteristics of the employees tested, in order to better shape for example the awareness initiatives. Despite a general statistic is not possible, we identified some common patterns:

- younger employees are more exposed: this could be probably related to the habits of new generations, used to online sharing services, combined

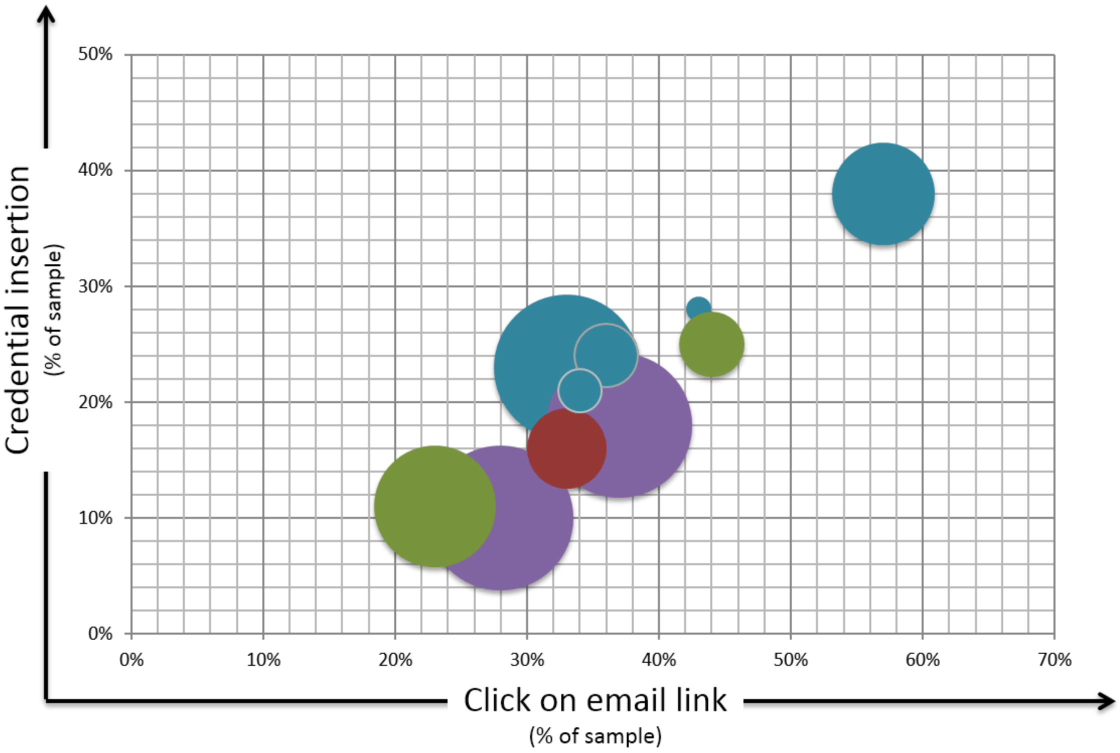


Figure 7: Overall incidence statistic of SDVAs performed.

with less perception about the online risks;

- management is often quite vulnerable: in general, what we observed is that the higher the role in the company the lesser is the exposure. Nevertheless, the percentage of managers who click on the link, or insert credentials is not low. Being the assets managed by these figures relevant, their incidence is high.
- awareness and education mitigates the risk: in some cases, we assessed the same people before and after attendance to specific training tracks (see [21]). Some awareness methods performed better than others did and this opened the road for our future researches.

6 Conclusions and Future Works

The SDVA Framework we presented [26] is a holistic approach to measure the risk related to the »human factor« inside companies and a test for the overall enterprise reactivity.

What we found is that, even only doing passive OSINT, it is possible to find a lot of relevant information on both company initiatives and employees that can be used to contextualize the attacks. Furthermore, according to our results the users have two different levels of perception of the threats: the awareness that their credentials must not be inserted in a generic web site is relatively higher than the awareness that just a click on a web-page/link could infect a computer. The drive-by-download infection schema seems to be better known than the drive-by-infection one. Our tests also report that most of the companies are heavily exposed to these new risks and often, before performing the first SDVA, there is no perception of how extended the exposure is.

A possible solution, according our experience with the follow-ups of SDVAs is an integrated approach represented by a set of actions defined through a model that is shared by all the company's functions and in synergy also with the allocated budgets for structures different from IT. In particular, the SDVAs have also a beneficial impact on the enterprise internal dynamics:

- an increased internal collaboration among the involved departments and a better understanding of the security risks by whom are less used to security (e.g., communication department are less used to think about security consequences of their actions);
- a sharing of internal budgets (not only IT) on security related activities;
- a renewed attention to the internal security problems versus the perimetral-only defence.

This approach also fosters the diffusion of people awareness and increases their knowledge of the new Social-driven Vulnerability's dynamics. The collected results are so interesting that on the one hand we are expanding our SDVA approach/tools and on the other hand we are investigating new research directions studying innovative ways to do awareness, experimenting new methods to trigger alerts to improve the overall incident response readiness.

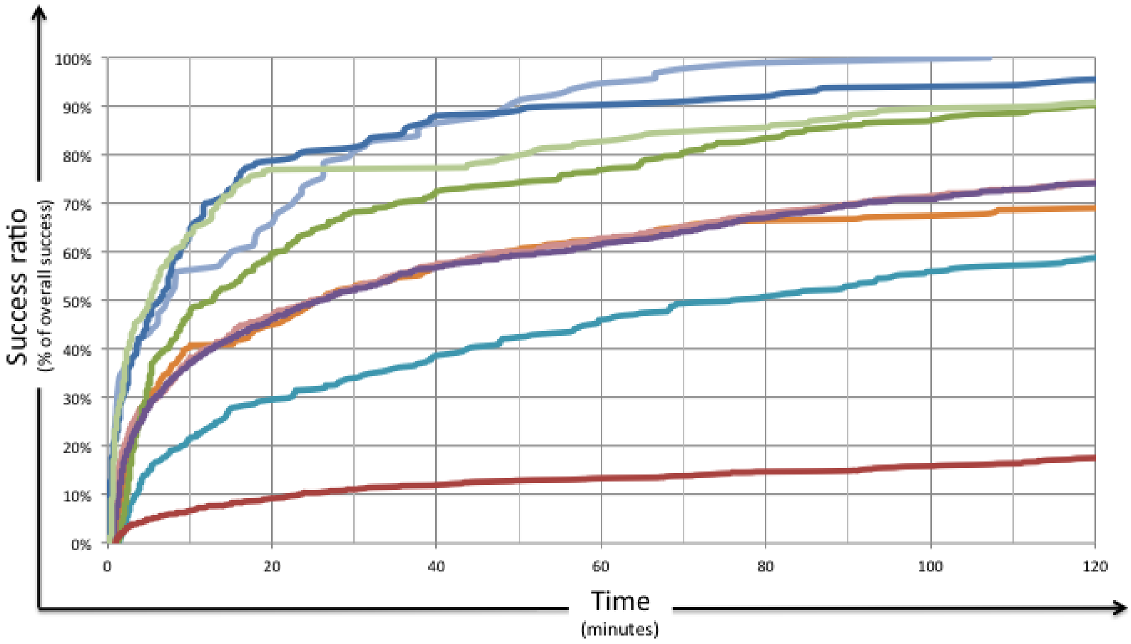


Figure 8: Normalized click-through trends for each SDVA, in the first 2 hours of testing.

7 About the Authors

Enrico Frumento (twitter: enricoff) and Roberto Puricelli (twitter: robywankenoby) work in the security practice at CEFRIEL, a center for innovation of the Politecnico di Milano. Since years they do research on the security of the human factor, through measurable risk assessment and new awareness methods.

8 References

1. G. Mann, Forget the horse, this is the year of the F[ph]ish and the RAT, The Future of Cybersecurity, London, March 2014
2. K.D. Mitnick, The art of deception: Controlling the human element of security, John Wiley & Sons, 2002.
3. K.D. Mitnick, Ghost in the wires, Little Brown & Co, 2011.
4. »Data Brokers: A Call For Transparency and Accountability: A Report of the Federal Trade Commission«. May 2014. [Online]. http://goo.gl/CTqPLc [Accessed: Sep-2014].
5. »We are data«. WatchDogs. [Online]. http://wearedata.watchdogs.com
6. AA.VV., Task Force 1 - Personal information space. Talk in the Tower. [Online]. http://goo.gl/IlfAvN [Accessed: Sep-2014].
7. E. Frumento, Redefinition of the digital identity through the evolution of modern workforces, Talk in the Tower. [Online]. http://goo.gl/AN9043 and http://goo.gl/mRf5HV [Accessed: Sep-2014].
8. »Target CEO resigns as fallout from data breach continues«, LA Times. April 2014. [Online]. http://goo.gl/C8oOuL [Accessed: Sep-2014].
9. »DHS: Spear Phishing Campaign Targeted 11 Energy Sector Firms«. SecurityWeek.Com. April 2013. [Online] http://goo.gl/bJpve [Accessed: Sep-2014].
10. B. Schneier, »The Human side of HeartBleed«. [Online]. http://goo.gl/6hF9It [Accessed: Sep-2014].
11. R. Abrams, D. Harley. People Patching, is user education of any use at all?. ESET. [Online]. http://www.eset.com/us/resources/white-papers/People_Patching.pdf [Accessed: Sep-2014].
12. »Symantec Develops New Attack on Cyberhacking. Declaring Antivirus Software Dead, Firm Turns to Minimizing Damage From Breaches« , WSJ, May 2014, http://goo.gl/CssQYF [Accessed: Sep-2014].
13. »Social Engineering, Hacking The Human OS«. Kaspersky Labs. [Online]. https://blog.kaspersky.com/social-engineering-hacking-the-human-os [Accessed: Sep-2014]
14. M. Huber, S. Kowalsky, Towards Automating Social Engineering Using Social Networking Sites, Int. Conf. on Computational Science and Engineering, 2009
15. T. Berners-Lee, »The year open data went worldwide«, TEDTalk Videos, 2010. [Online]. http://goo.gl/n1wWJ4 [Accessed: Sep-2014].
16. J. Mahmud, J. Nichols et al., Home Location Identification of Twitter Users, ACM Transactions on Intelligent Systems and Technology, Vol. 5, No. 3, Article 47, July 2014.
17. D. Shounak, G. Debojyoti et al., A Method for

Bypassing Keystroke Recognition Based Security System Using Social Engineering, OSR-JCE, Volume 16, Issue 2, PP 87-93, Mar-Apr. 2014.

18. S. Blackmore, The Meme Machine, Oxford University Press, 1999. ISBN 0198503652.
19. C. Hadnagy, Social Engineering: The Art of Human Hacking, Wiley, 2010, ISBN 0470639539.
20. I. Danesh, M. Balduzzi et al. Reverse Social Engineering Attacks in Online Social Networks, Proc. of 8th DIMVA, 2011.
21. E. Frumento, C. Lucchiari et al., Cognitive Approach for Social Engineering, DeepSec Conference 2010. Wien. Nov 2010.
22. R. Dhamija, J. Tygar et al., Why Phishing Works, Proceedings of the SIGCHI conference on Human Factors in computing systems - CHI «06, 2006.
23. S. Li, X.Yun et al., A Propagation Model for Social Engineering Botnets in Social Networks, Proc. of 12th PDCAT, PP 423-426, Oct 2011.
24. »Cisco 2014 Annual Security Report«. Cisco. [Online]. http://www.cisco.com/web/offer/gist_ty2_asset/Cisco_2014_ASR.pdf [Accessed: Sep-2014].
25. »DigiNotar Files for Bankruptcy in Wake of Devastating Hack«, Wired, 2011. [Online]. http://www.wired.com/2011/09/diginotar-bankruptcy [Accessed: Sep-2014].
26. R. Brenna, E. Frumento et al., Social driven vulnerability. Facing and managing vulnerabilities driven by Social Media. 2014. [Online]. http://www.slideshare.net/CEFRIEL/social-driven-vulnerability-english-version [Accessed: Sep-2014].
27. F. Mouton, M.M. Malan et al., »Social engineering from a normative ethics perspective«. Information Security for South Africa, PP 1-8. Aug 2013.
28. »Systems 2014 Mobile Malware Report«, BlueCoat. [Online]. Available: http://goo.gl/VUhwVV. [Accessed: Sep-2014].
29. »Anatomy of an Attack«, RSA Blog, 01-Apr-2011. [Online]. Available: http://goo.gl/2a0QD. [Accessed: Sep-2014].
30. »Drive-by download«, Wikipedia. 09-Jul-2014.
31. »Drive-by infections«, eBanking but secure. [Online]. Available: http://goo.gl/D9wmO5. [Accessed: Sep-2014].
32. »Large-Scale Water Holing Attack Campaigns Hitting Key Targets«, threatpost, 25-Sep-2012. [Online]. Available: http://goo.gl/bLLLe2. [Accessed: Sep-2014].
33. T. Adlin and J. Pruitt, The Persona Lifecycle: Keeping People In Mind Throughout Product Design. United States: Morgan Kaufmann Publishers In Interactive Technologies, 2006.
34. Wombat Security. [Online]. Available: http://www.wombatsecurity.com. [Accessed: Sep-2014].
35. Digital Shadows, 25-Jun-2014. [Online]. Available: http://www.digitalshadows.com. [Accessed: Sep-2014].
36. »McAfee Labs Threats Report, August 2014 | Phishing lures the unsuspecting: business users easily hooked«, in Threats Report, August 2014. McAfee Labs, 2014. [Online]. Available: http://goo.gl/ucei9R [Accessed: Sep-2014].
37. J. Downs, M. Holbrook et al., Behavioral Response To Phishing Risk, Proceedings of the anti-phishing working groups 2nd annual eCrime researchers summit on - eCrime «07, 2007.
38. J. Downs, M. Holbrook et al., Decision Strategies And Susceptibility To Phishing, Proceedings of the second symposium on Usable privacy and security - SOUPS «06, 2006.
39. M. Sparshott, »The psychology of phishing«, Help Net Security, 23-Jul-2014. [Online]. Available: http://www.net-security.org/article.php?id=2078. [Accessed: Sep-2014].
40. E. Frumento, »Security in mobile work environments«, MUSES Project, 20-Sep-2014. [Online]. Available: https://www.musesproject.eu/security-in-mobile-work-environments. [Accessed: Sep-2014].