Optimal Security Protection for Sensitive Data

George O. M. Yee
Computer Research Lab, Aptusinnova Inc., Ottawa, Canada
Dept. of Systems and Computer Engineering, Carleton University, Ottawa, Canada
email: george@aptusinnova.com, gmyee@sce.carleton.ca

Abstract—**The growth of the Internet has unfortunately been accompanied by an increasing number of attacks against an organization's computing infrastructure, leading to the theft of sensitive data. In response to such incursions, the organization installs security measures (e.g., intrusion detection system) for protecting its sensitive data. However, this installation is often done haphazardly, without any objective guidance regarding how many vulnerabilities must be secured in order to achieve an acceptable level of protection. This paper shows how an organization can calculate estimates of security protection, and objectively use them to adjust the number of security measures installed, until an optimal level of protection is achieved, subject to certain constraints. This work extends the paper "Assessing Security Protection for Sensitive Data" published in SECURWARE 2017. Additional explanations, application examples, and related works have been included.**

Keywords-optimal; security protection; assessment; sensitive data; vulnerability.

I. INTRODUCTION

This work extends Yee [1] by adding explanations, application examples, and related works.

Recent attacks against computing infrastructure, resulting in the theft of sensitive data, have grabbed the headlines, and have devastated the victim organizations. The losses have not only been financial (e.g., theft of credit card information), but more importantly the damage to the organization's reputation. Consider the following data breaches that happened in 2016 [2] and 2017 [3]:

- February, 2016, University of Central Florida: Data breach affected approximately 63,000 current and former students, faculty, and staff, with the theft of information including social security numbers, first and last names, and student/employee ID numbers.
- February, 2016, U.S. Department of Justice: Hackers released data on 10,000 Department of Homeland Security employees one day, and the next day released data on 20,000 FBI employees. Stolen information included names, titles, phone numbers, and email addresses.
- March, 2016, Premier Healthcare: Theft of a laptop containing sensitive data pertaining to more than 200,000 patients, including names, dates of birth, and possibly social security numbers or financial information.
- March, 2016, Verizon Enterprise Solutions: Hackers stole information for about 1.5 million customers; the information was found for sale in an underground cybercrime forum by cyber security journalist Brain Krebs.
- September, 2016, Yahoo!: The company announced that a hacker had stolen information from 500 million accounts in 2014. The hacker, believed to be working for a foreign government, stole email addresses, passwords, full user names, dates of birth, telephone numbers, and in some cases, security questions and answers.
- February and April, 2017, InterContinental Hotels Group (IHG): The company that owns popular hotel chains like Crowne Plaza, Holiday Inn, and Kimpton Hotels, announced in February a data breach that affected 12 of its properties. This number was enlarged to 1,200 properties in April. Malware was found on payment processing servers. The stolen data included cardholder names, card numbers, expiration dates, and internal verification codes.
- March, 2017, Dun & Bradstreet: This business services company found its marketing database with over 33 million corporate contacts shared across the web. The company claimed that the breach occurred to businesses, numbering in the thousands, that had bought its 52 GB database. The leak may have included full names, work email addresses, phone numbers, and other business-related data from millions of employees of organizations such as the US Department of Defense, the US Postal Service, AT&T, Walmart, and CVS Health.
- July, 2017, Verizon: 14 million Verizon subscribers may have been affected by a data breach simply by having contacted Verizon customer service in the past 6 months. The customer service records were kept on a server controlled by Israel based Nice Systems. The leaked data consisted of log files generated when Verizon customers contacted the company by phone.
- September, 2017, Equifax: This is one of the three largest credit agencies in the US. It announced a breach that may have affected 143 million customers, one of the worst breaches ever due to the sensitivity of the data stolen. The compromised data included social security numbers, driver's license numbers, full names, addresses, birth dates, credit card numbers, and other personal information. Hackers had access to the company's system from mid-May to July by exploiting

2018, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

a vulnerability in website software. Equifax discovered the breach on July 29, 2017.

- November, 2017, Uber: Uber revealed that it became aware of a data breach in late 2016 that potentially exposed the personal information of 57 million Uber users and drivers. However, Uber chose to pay the hackers $100,000 to keep the breach a secret instead of immediately alerting the affected victims. The hackers gained access to the data stored on GitHub, which was used by Uber engineers for collaboration, and included names, email addresses, and phone numbers of Uber users worldwide.

This is only a sampling, as there were many more breaches in 2016 and 2017, and in fact, no year can be said to have been breach-free. Moreover, the problem appears to be getting worst, as 2017 has been mentioned [4] as a "record-breaking year for the numbers of publicly reported data breaches and exposed records in 2017 worldwide: a total of 5,207 breaches and 7.89 billion information records compromised."

To protect themselves from attacks, such as the ones described above, organizations determine their vulnerabilities to attack, and then secure the vulnerabilities with security measures. Common measures include firewalls, intrusion detection systems, two-factor authentication, encryption, and training for employees on identifying and resisting social engineering. However, today's organizations install security measures without any way of calculating the overall level of protection that will result. They proceed based on recommendations from consultants, in reaction to attacks that have been observed, or worst, as a result of having suffered an attack themselves. And in many cases, they are forced to stop this deployment once their security budget runs out. It would be far better if an organization can follow a top-down approach, by setting a target level of protection and then install security measures to achieve the target. The target would be set according to the expected threat situation, the nature of the business, the sensitivity of information kept, and an estimated financial budget. Before this can be done, it would be useful to have quantitative estimates of the level of protection based on the number of vulnerabilities secured. This work derives such estimates and shows how to apply them to not only set a protection target, but also how security measures can be installed to achieve the target.

The objectives of this work are i) derive estimates of the resultant protection level obtained by an organization through the installation of security measures to secure vulnerabilities, ii) show how these estimates can be calculated, iii) show how the estimates can be applied in a structured, objective, quantitative approach to secure an organization, and finally iv) illustrate ii) and iii) using examples.

The rest of this paper is organized as follows. Section II discusses the nature of sensitive data and derives the estimates. Section III explains how the estimates are calculated and applied in a structured, objective, quantitative approach to secure an organization. Additional application areas are also included. Section IV presents two application examples. Section V discusses related work. Finally, Section VI gives conclusions and future research.

II. Estimating Security Protection Levels

Before deriving estimates of security protection levels, it is useful to examine the nature of sensitive data.

A. *Sensitive Data*

We all have some sense of what is meant by sensitive data: first and foremost it is data that must be safeguarded from falling into the wrong hands, the consequence of which would be damaging to an individual or an organization.

For an individual, sensitive data usually means private information, which is information about the individual and is owned by that individual. The individual's privacy then refers to his or her ability to control the collection, purpose of collection, retention, and distribution of that information by another party. Private information is also called personal information or personally identifiable information because it can be used to identify the individual. For example, an individual's height, weight, or credit card number can all be used to identify the individual and are therefore considered as personal information. Continuing this example, the extent to which the individual has control over who collects this information, the purpose for which the collector will use this information, how long the collector will retain this information, and to which other parties the collector will disclose this information, determines the individual's degree of privacy. The nature of private information will not be explored further here but the reader is encouraged to consult [5] for more details.

For an organization, sensitive data may encompass private information, but may additionally include information that may compromise the competitiveness of the organization if divulged, such as trade secrets or proprietary algorithms and secret formulas. For government organizations, sensitive data may include information that is vital for the security of the country for which the government organization is responsible. For this work, sensitive data is defined as follows:

DEFINITION 1: *Sensitive data* is information that must be protected from unauthorized access in order to safeguard the privacy of an individual, the well being of an organization, or the well being of an entity for which the organization has responsibility.

This work considers losses arising from sensitive data or sensitive information being in the possession of unintended malicious parties or entities. This covers theft and any unintended exposure of sensitive information such as accidental leakage or posting. Per Definition 1, "sensitive data" and "sensitive information" are used interchangeably in this work. Some researchers make a distinction between these terms but the popular usage calls for no distinction.

A. *Attacks on Organizations*

Attacks carried out against sensitive information residing with organizations may be categorized as "outside attacks" and "inside attacks". We define these as follows.

DEFINITION 2: An *attack* is any action carried out against sensitive information held by an organization that, if successful, results in that information being in the hands of the attacker. An *outside attack* (A_o) is an attack that is carried out by an outsider of the organization (i.e., the attacker is not associated with the organization in a way that gives her special access privileges to sensitive data, e.g., a regular member of the public). An *inside attack* (A_i) is an attack that is carried out by an insider of the organization (i.e., someone who has special access privileges to sensitive data by virtue of her association with the organization, e.g., employee).

DEFINITION 3: A *vulnerability* of an organization is any weakness in the organization's infrastructure, platform, or business processes that can be targeted by an attack with some expectation of success. A *secured-vulnerability* was originally a vulnerability that has had protective security measures put in place so that it is no longer a vulnerability. For example, a vulnerability is private information stored in the clear. This becomes a secured vulnerability if the private information is encrypted.

Outside attacks target a range of security vulnerabilities, from software systems that can be breached to access the sensitive information to simple theft of laptops and other devices used to store sensitive information. An example of an outside attack is the use of a Trojan horse planted inside the organization's computer system to steal sensitive information.

Inside attacks arise from the attacker making use of her privileged position (e.g., as an employee) to cause a loss of sensitive data. In this case, the attack is often difficult to detect, since it would appear as part of the normal duties of the insider attacker. An example of an inside attack is where a disgruntled employee secretly posts the organization's sensitive information on the Internet to try to harm the organization. An inside attack can also be unintentional (e.g., an employee casually providing client names for a survey).

Both outside and inside attacks target the organization's vulnerabilities. Vulnerabilities that invite outside attacks include the use of badly provisioned firewalls, the failure to encrypt data, and simple carelessness (e.g., leaving a laptop containing sensitive information in a car). Vulnerabilities that attract inside attacks include a) poor business processes that lack mechanisms to track which data is used where, used for what purpose, and accessed by whom, b) poor working conditions that give rise to employees feeling unfairly treated by management which can lead to employees seeking revenge, and c) poor education and enforcement of company policies regarding the proper care and handling of sensitive information (e.g., the above survey example).

The location of an attacker carrying out an attack does not determine whether the attack is an inside attack or an outside attack. An inside attack can be carried out outside the organization's premises; similarly, an outside attack can be carried out inside the premises.

We have so far used the expressions "level of protection" and "protection level" informally relying on their everyday meaning. We now formalize this meaning in terms of vulnerabilities, introducing the idea of "security protection level".

DEFINITION 4: An organization's security protection level (SPL) is the degree of security protection from attacks that results from the organization having secured q vulnerabilities, leaving p vulnerabilities unsecured, where the organization has a total of $p+q$ vulnerabilities. Each pair of values *(p, q)* corresponds to a different SPL.

Suppose an organization has a total of N vulnerabilities, where $p + q = N$. Then each organization has a value of N that corresponds to a set of SPL points lying on a straight line in the *(p, q)* plane, where the higher values of q correspond to higher or greater security protection levels. Figure 1 shows this relationship for two organizations having $N=50$ and $N=100$.

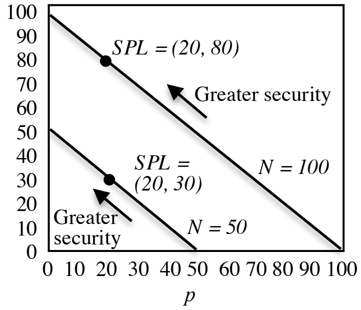

Figure 1. SPL points on lines corresponding to two organizations, one with N=50 and the other with N=100.

B. Deriving the Estimates

Intuitively, for the same organization, SPL A is more capable of protecting from sensitive information loss than SPL B if A is composed of more secured vulnerabilities than B, where all vulnerabilities have roughly the same level of loss risk. This is the idea behind the derivation below.

We seek the capability C of an organization's SPL to protect sensitive data. Suppose that an organization's SPL has p vulnerabilities and q secured-vulnerabilities, where no distinction is made between outside and inside attacks. The number of original vulnerabilities before any vulnerabilities were secured is $p+q$. Let $P(e)$ represent the probability of event e. For convenience, "data" is understood to be "sensitive data". We have

$$C = P(\text{no data losses}) = 1\text{-}P(\text{data losses}) \tag{1}$$

Since a data loss is the result of a successful attack on a vulnerability,

$$P(\text{data losses}) \approx p/(p+q) \tag{2}$$

where we have applied the additive rule for the union of probabilities of attacks on the p vulnerabilities, assuming that 2 or more attacks do not occur simultaneously. This is a fair assumption confirmed by experience. Substituting (2) into (1) and adjusting for a possible zero denominator gives

$$C \approx 1-[p/(p+q)] = q/(p+q) \quad \text{if } p+q > 0 \qquad (3)$$
$$= 1 \quad \text{if } p+q = 0 \qquad (4)$$

Since C is a probability, its value is between 0 and 1, attaining 0 if the organization has no secured vulnerabilities ($q=0$, (3)) and 1 if either all of its vulnerabilities are secured ($p=0$, (3)) or if the organization has no vulnerabilities ($p+q=0$, (4)). Since an organization having no vulnerabilities is highly improbable, (4) is unlikely to apply.

The above derivation can be done within each of the categories of outside attacks and inside attacks (we did not distinguish between outside and inside attacks above). Let C_o, C_i represent the capabilities of an organization's SPL to protect sensitive information from outside attacks and inside attacks, respectively. Let p_o, p_i represent the number of vulnerabilities to outside attacks and inside attacks, respectively. Let q_o, q_i represent the number of secured vulnerabilities to outside attacks and inside attacks, respectively. Then, repeating the above derivation for outside attacks and inside attacks gives

$$C_o \approx q_o/(p_o+q_o) \quad \text{if } p_o+q_o > 0 \qquad (5)$$
$$\approx 1 \quad \text{if } p_o+q_o = 0 \qquad (6)$$
$$C_i \approx q_i/(p_i+q_i) \quad \text{if } p_i+q_i > 0 \qquad (7)$$
$$\approx 1 \quad \text{if } p_i+q_i = 0 \qquad (8)$$

As above, C_o (C_i) have values between 0 and 1, attaining 0 if the organization has no secured vulnerabilities to outside (inside) attacks ((5) and (7)) and 1 if either all of the vulnerabilities are secured ((5) and (7)) or if the organization has no vulnerabilities ((6) and (8)). Since an organization having no vulnerabilities to outside and inside attacks is highly improbable, (6) and (8) are unlikely to apply.

The estimates of data protection capability are now assigned as follows for a given SPL (no distinction between inside and outside attacks), SPL_o (for outside attacks), and SPL_i (for inside attacks). Let E be an estimate of data protection capability, where no distinction is made between outside and inside attacks. Let E_o be an estimate of data protection capability against outside attacks. Let E_i be an estimate of data protection capability against inside attacks. Then for the SPL, SPL_o, and SPL_i

$$E = q/(p+q) \quad \text{if } p+q > 0 \qquad (9)$$
$$= 1 \quad \text{if } p+q = 0 \qquad (10)$$
$$E_o = q_o/(p_o+q_o) \quad \text{if } p_o+q_o > 0 \qquad (11)$$
$$= 1 \quad \text{if } p_o+q_o = 0 \qquad (12)$$
$$E_i = q_i/(p_i+q_i) \quad \text{if } p_i+q_i > 0 \qquad (13)$$
$$= 1 \quad \text{if } p_i+q_i = 0 \qquad (14)$$

E has the advantage of providing a single number for ease of comparison between different SPLs within an organization. A threshold T for E may be pre-determined such that for E above T, the security measures installed by the organization to secure vulnerabilities against both outside and inside attacks (corresponding to a SPL) are deemed adequate. For given SPL_o and SPL_i, E_o *and* E_i have the advantage of focusing in separately on where an organization stands in terms of its security measures against outside and inside attacks. Thresholds T_o and T_i may be pre-determined for E_o and E_i respectively, such that for both estimates above their respective thresholds, the corresponding installed security measures against outside and inside attacks are deemed adequate. If this is the case, we call the corresponding SPL an *adequate SPL*. In practice, E_o and E_i may be expressed as percentages that define a region in a 100 x 100 plane in which an organization's capability to protect data is adequate (acceptable), as represented by the shaded region in Figure 2. Each point in this shaded region corresponds to an adequate SPL. An organization strives to have the "best" adequate SPL (one which has highest number of security measures possible against both outside and inside attacks) as allowed by its financial budget for adding security measures (see Section III).

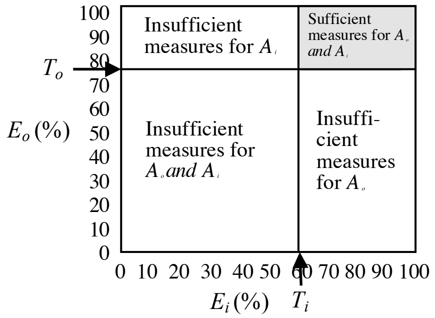
100
90
80
70
60
50
40
30
20
10
0
T_o
E_o (%)
Insufficient measures for A_i
Sufficient measures for A_o and A_i
Insufficient measures for A_o and A_i
Insuffi-cient measures for A_o
0 10 20 30 40 50 60 70 80 90 100
E_i (%)
T_i

Figure 2. Sufficiency of Security Measures Against Outside Attacks (A_o) and Inside Attacks (A_i)

III. Applying the Estimates

This section shows how an organization may use the estimates to establish "best" adequate SDLs as permitted by its financial budget. The description below separates outside attacks from inside attacks since organizations would need to account for them separately.

A. *Determining the Vulnerabilities*

For outside attacks, we recommend a threat analysis of security vulnerabilities in the organization's systems that could allow outside attacks to occur. Threat analysis or threat modeling is a method for systematically assessing and documenting the security risks associated with a system (Salter et al. [6]). Threat modeling involves understanding the adversary's goals in attacking the system based on the system's assets of interest. It is predicated on that fact that an adversary cannot attack a system without a way of supplying it with data or otherwise accessing it. In addition, an adversary will only attack a system if it has some assets of interest. The method of threat analysis given in [6] or any other method of threat analysis will yield $N_o = p_o + q_o$, which is the total number of vulnerabilities to outside attacks. The method presented here for threat modeling is based on [6], and consists of the following steps:

1. Identify threats: examine all available details of the system and enumerate possible threats.

2. Create attack trees for the system: for each threat, take the attacker's view and find the weak points in the system and the paths that can lead to realizing the threat.
3. Apply weights to the leaves: for each leaf, assign qualitative values for risk, access, and cost to the attacker.
4. Prune the tree so that only exploitable leaves remain: prune leaves that represent objectives that are beyond the attacker's capabilities or that offer an inadequate return.
5. Generate corresponding countermeasures: identify security measures for rendering the threat non-realizable.

As an illustration of the above method for threat analysis, consider the hypothetical software system of an online seller of merchandise (e.g., Amazon.com). Figure 3 shows the essential components of this system, using solid arrows to represent sensitive data flow, dashed arrows to depict non-sensitive data flow, circles to represent processing modules, squares to represent data storage, and a dashed square to enclose components that execute on the same computing platform. Additionally, data items are identified using numbers; all other components are identified with letters.

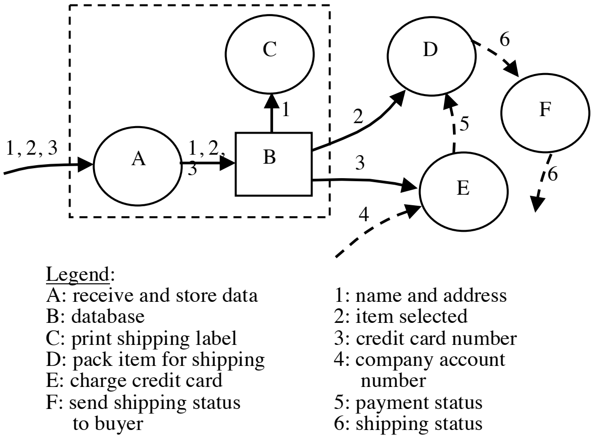

Figure 3. Software system of an online seller of merchandise.

Suppose the goal of an attacker is to steal sensitive data from this system. The above steps are applied as follows:

1. Identify threats: An examination of the system in Fig. 3 found the following threats: a) theft of sensitive data flowing into A, D, and E, b) theft of sensitive data from A, C, D, and E, and c) theft of sensitive data from B.
2. Create attack trees: the weak points in the system that can lead to realizing the threats found in step 1 are: i) the paths for data flowing into A, D, and E can be exploited using man-in-the-middle attacks, ii) the processing modules A, C, D, and E can be exploited using Trojan horse or hacker attacks, and iii) the database B can be exploited using SQL attacks. These locations correspond to vulnerabilities in the system. Note that the paths of data flow into B and C are excluded as weak points because they are not considered externally accessible, due to the fact that A, B, and C all run on the same computing platform. This attack tree can be represented using hierarchical numbering as follows:

0 Theft of sensitive data from system
- 1.1 Theft of sensitive data flowing into A, D, E
  - 2.1 Man-in-the-middle attack on data paths Into A, D, E
- 1.2 Theft of sensitive data from A, C, D, E
  - 2.2 Trojan horse or hacker attack on A, C, D, E
- 1.3 Theft of sensitive data from B
  - 2.3 SQL attacks on B

This can also be depicted graphically as the attack tree in Figure 4, using the same number labels as above.

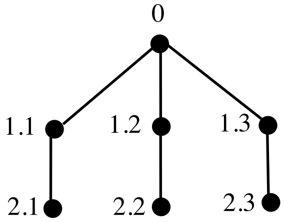

Figure 4. Attack tree for the online seller system in Fig. 3

3. Apply weights to the leaves: the leaves are 2.1, 2.2, and 2.3. Assigning weights L (low), M (medium), and H (high) to each of risk, access, and cost in turn yields (L, H, L) for 2.1, (L, M, L) for 2.2, and (L, M, L) for 2.3. This is read as low risk, high access, and low cost for 2.1, for example, meaning that there is low risk to the safety of the attacker (he is attacking from a remote location), high access to the path, and low cost to the attacker in carrying out the attack. The attacker's access for 2.2 and 2.3 are rated as medium because he or she has to actually get into the system.
4. Prune the tree so that only exploitable leaves remain: in this case, no pruning is necessary because all the leaves are within the capability of the attacker to exploit. If there was a leaf with a weighting of say, (H, L, M), i.e., high risk to the attacker's safety, low access, and medium cost, then this leaf may be pruned, in that the attacker would be unlikely to exploit the leaf.
5. Generate corresponding countermeasures: the countermeasures or security measures for the

2018, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

vulnerabilities identified in step 2 are: i) encrypt the data that flow along paths into A, D, and E, ii) use firewalls to protect against Trojan and hacker attacks on A, C, D, and E, iii) use a combination of firewall and database hardening to defend against SQL attacks on B.

The threat analysis may be carried out by a project team consisting of the system's design manager, a security and privacy analyst, and a project leader acting as facilitator. In addition to having expertise on privacy and security, the analyst must also be very familiar with the organization's systems.

For inside attacks, we recommend that the above project team carry out a special insider threat analysis, to identify vulnerabilities to inside attacks and identify measures to secure these vulnerabilities. The team would accomplish this by brainstorming answers to the questions in Table I, or other questions from experience, identifying the vulnerabilities and measures to secure the vulnerabilities in the process. In Table I, questions 1 to 6 address motivational or environmental vulnerabilities, which may also be "secured" by applying mitigating measures. Questions 7 and 8 address security vulnerabilities. In identifying vulnerabilities to inside attack, the project team may weigh the vulnerabilities in terms of how likely they are to lead to attacks, and eliminate the unlikely ones. The weighing process may consider such factors as risk to the attacker that she could be caught as well as her motivation for the attack. The value of $N_i = p_i + q_i$ would be determined at the end of this process.

B. Determining the Thresholds T_o and T_i

The values of T_o and T_i should be determined by the same threat analysis team mentioned above. The values would depend on the following:

- The potential value of the sensitive data – the more valuable the data is to a thief, a malicious entity, or a competitor, the higher the thresholds should be.
- The damages to the organization that would result, if the sensitive data were compromised – of course, the higher the damages, the higher the thresholds.
- The current and likely future attack climate – consider the volume of attacks and the nature of the victims, say over the last 6 months; if the organization's sector or industry has sustained a large number of recent attacks, then these thresholds need to be higher.
- Consider also potential attacks by nation states as a result of the political climate; attacks by individual hacktivist groups such as Anonymous or WikiLeaks may also warrant attention.

In general, an organization would like to be as secure as possible and establish a "best" adequate SPL. Therefore, values above 80% would not be uncommon. However, whatever the thresholds, the organization must find them acceptable after considering the above factors. The financial budget available for securing vulnerabilities also plays an important role here, since higher thresholds call for securing more vulnerabilities, which means more financial resources will be needed.

TABLE I. QUESTIONNAIRE TO IDENTIFY VULNERABILITIES TO INSIDE ATTACK

	Question	Rationale
1.	Is the sensitive information of high value to outside agencies or a competitor?	The higher the value, the more an inside attacker will be tempted to steal and sell the information.
2.	Does the organization have an employee assistance program that includes counselling and help with financial difficulties?	Such a program may eliminate some financial motivation for an inside attack.
3.	Does the organization have an ombudsman or other impartial agent to assist employees with their grievances?	Such an impartial agent may eliminate or reduce the motivation to seek revenge by committing an inside attack.
4.	Does the organization have a history of perceived injustices to employees?	If the answer is 'yes', employees may be motivated by revenge to commit an inside attack.
5.	Does the organization conduct a stringent background and reliability check on a candidate for employment prior to hiring the candidate?	While a background and reliability check is not guaranteed to weed out potential inside attackers, it should eliminate those with criminal pasts.
6.	Does the organization require candidates for employment to disclose any potential conflicts of interest they may have with respect to their new employment and any outside interests prior to hire? Does the organization require ongoing disclosure of conflicts of interest after hire?	Eliminating conflicts of interest should reduce related motivations for malicious inside attacks. For example, an inside attacker may secretly compromise private information in favour of an outside interest, believing that the compromise is undetected.
7.	What are some possible ways for an insider to gain access to sensitive information she should not be accessing? How to secure?	This question will identify security weaknesses.
8.	What are some possible ways for an insider to transmit sensitive information outside the organization undetected? How to secure?	This question will identify additional security weaknesses.

C. Applying the Estimates to Determine Optimal or "Best" Adequate SPLs

We now have values for the following: $N_o = p_o + q_o$, $N_i = p_i + q_i$ (Section IIIA), and T_o, T_i (Section IIIB). Rewriting (11) and (13) and using the ceiling function to avoid fractional numbers of secured vulnerabilities gives:

$$q_o = \lceil N_o E_o \rceil \qquad \text{where } T_o \le E_o \le 1 \qquad (15)$$

$$q_i = \lceil N_i E_i \rceil \qquad \text{where } T_i \le E_i \le 1 \qquad (16)$$

Equations (15) and (16) give all possible values of q_o and q_i such that the associated E_o and E_i (with $p_o = N_o - q_o$ and $p_i = N_i - q_i$) fall within the shaded region of Figure 1. In other words, these equations give all possible values of q_o and q_i

for adequate SPLs. The ceiling function biases the security level upward by taking the number of secured vulnerabilities to the next higher integer where applicable, which should be fine since more security should be better than less security. The quantities $q_o = \lceil N_oT_o \rceil$ and $q_i = \lceil N_iT_i \rceil$ from (15) and (16), termed respectively the threshold q_o and the threshold q_i, will be useful below.

To obtain an optimal "best" adequate SPL$_\text{o}$ and an optimal "best" adequate SPL$_\text{i}$ from among the adequate SPLs generated by (15) and (16), the organization applies the constraint that the total cost of implementing the $(q_o + q_i)$ security measures from (15) and (16) must be less than or equal to the financial budget for security measures. The organization separately prioritizes its outside attack and inside attack vulnerabilities in terms of urgency, and then selects them for securing in order of high priority to low priority, until both the financial budget is exhausted and the number of secured vulnerabilities are at least as great as the threshold q_o and the threshold q_i. In this way, the organization determines the q_o and q_i, as well as the p_o and p_i (which are just N_o - q_o and $N_i – q_i$ respectively) that define its "best" adequate SPL$_\text{o}$ and "best" adequate SPL$_\text{i}$, respectively. This procedure may be precisely described in the form of a computer algorithm as follows. Let $u_1, u_2, ... u_{No}$ and $v_1, v_2, ... v_{Ni}$ be the organization's outside attack and inside attack vulnerabilities prioritized in terms of urgency, respectively, such that u_1 has higher or equal priority (urgency) than u_2, u_2 has higher or equal priority (urgency) than u_3, and so on. Similarly, v_1 has higher or equal priority (urgency) than v_2, v_2 has higher or equal priority (urgency) than v_3, and so on. Figure 5 illustrates these relationships in which equal priority does not occur. Let B_o and B_i represent the budgets for securing against outside and inside attacks, respectively. Let C_o and C_i be the costs of securing the vulnerabilities to outside and inside attacks respectively. Let k be a counter variable. Then the pseudo code in Figure 6 comprises a computer algorithm for obtaining "best" adequate SPLs. Running this algorithm will produce the following: a) q_o and q_i, defining the "best" adequate SPL$_\text{o}$ and "best" adequate SPL$_\text{i}$, or b) one or two "insufficient budget" messages, in which case the organization has to increase the corresponding budgets and re-run the algorithm. Only result a) would be acceptable. If result b) is obtained, decreasing the thresholds T_o and T_i may result in fewer vulnerabilities needing to be secured, and may therefore generate result a). However, decreasing T_o and T_i is not recommended at this point, since these values were determined only after thorough analysis and consideration (see Section IIIB).

Prioritizing the vulnerabilities may be based on four aspects of an attack, namely "risk", "access", "cost", and the resulting damages from the attack, where "risk" is risk to the safety of the attacker, "access" is the ease with which the attacker can access the system under attack, "cost" is the monetary cost to the attacker to mount the attack, and resulting damages is self evident. A full explanation of this prioritization procedure is given in Yee [5].

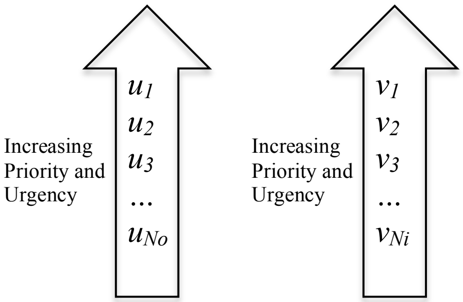
Increasing Priority and Urgency
u_1
u_2
u_3
...
u_{No}
Increasing Priority and Urgency
v_1
v_2
v_3
...
v_{Ni}

Figure 5. Vulnerabilities with increasing priorities and urgencies.

```
Begin;
    $C_o = 0; C_i = 0; k = 0;$
    While $k \le N_o$ and $C_o \le B_o$;
        $k = k + 1;$
        $C_o = C_o +$ cost of securing $u_k$;
    EndWhile;
    If ($k \ge$ threshold $q_o$) $q_o = k$;
    Else Print "$q_o$ unavailable -insufficient budget";
    $k = 0;$
    While $k \le N_i$ and $C_i \le B_i$;
        $k = k + 1;$
        $C_i = C_i +$ cost of securing $v_k$;
    EndWhile;
    If ($k \ge$ threshold $q_i$) $q_i = k$;
    Else Print "$q_i$ unavailable – insufficient budget";
End;
```

Figure 6. Algorithm for obtaining a "best" adequate SPL.

D. *Application Areas*

The main application area for the approach proposed in this paper is to secure an organization from attacks that target the sensitive data in the organization's computer system. An organization would implement security using the following steps:

1. Identify vulnerabilities to inside and outside attacks using threat analysis and prioritize them in terms of urgency. Identify the costs of security measures required to secure the vulnerabilities.
2. Identify "how secure" the organization needs to be given its nature and the existing attack environment, i.e., determine the values of T_o and T_i. Identify the organization's security budget. Identifying how secure it needs to be is necessary, since securing all vulnerabilities is probably not feasible due to a finite financial budget.
3. Run the algorithm in Fig. 6 to obtain q_o and q_i. Secure the prioritized outside and inside vulnerabilities up to and including q_o and q_i respectively.

The above steps may be carried out by a security consulting firm or by the organization's security department, depending on the latter's level of confidence. A good compromise may be to hire a consultant for steps 1 and 3, since step 2 involves careful consideration that may be better done internally within the organization. The above steps can also be performed for an organization that already has some vulnerabilities secured. In this case, the already secured vulnerabilities would simply not be part of the above steps and the security budget and q_o, q_i found would be for the unsecured vulnerabilities.

Another application area is in marketing and gaining consumer confidence. Organizations that hold sensitive private information and provide services to the public may wish to advertise the fact that they have high SPLs in order to gain consumer confidence and gain competitive advantage.

A third application area lies in standardization. Studies could be undertaken by standards bodies to determine recommended "best" adequate SPL values for organizations, based on organization type, size, activity, the quantity and nature of the sensitive data held, history of information breaches, and so on. These values could be published and made available as targets or guidance for organizations seeking to implement security using this approach.

One comment received while presenting the original paper at SECURWARE 2017 was that the structured approach of securing vulnerabilities against an objective security target is not how "things are done" in industry and therefore this approach would be useless. Well, of course its not how security is implemented today – that's the whole point of this paper, to suggest a better way, one that is more quantitative, and structured, as opposed to guessing and reacting emotionally when an attack has occurred. How implementing security is currently done was described at the beginning of this paper.

IV. Application Examples

This section presents two examples of applying the proposed approach. The first example is that of an online seller of merchandise implementing security for the first time. The second example looks again at the online seller in the first example, but 5 years later than in the first example, when the online seller decides to replace its computer system for a new more high performance model. Naturally, this introduces new vulnerabilities that need to be secured.

A. *Implementing Security for the First Time*

Alice Inc., an online seller of goods (e.g., Amazon.com), has an objective to secure its vulnerabilities to outside and inside attacks and to establish corresponding "best" adequate SPLs using the approach in this work. The company hires a security consulting firm to perform threat analyses of its systems, resulting in a report of vulnerabilities found that could be targeted by outside and inside attackers. The report also provides values for the number of vulnerabilities as $N_o = 10$ and $N_i = 8$, and includes prioritizations of outside and inside vulnerabilities. For each type of vulnerability (i.e., outside or inside) the prioritizations identified which vulnerability required securing first, which one second, and so on, in declining order of urgency. Based on the consultant's recommendations, as well as its own internal deliberations, Alice Inc. assigned the following values:

$$T_o = 0.80,\ T_i = 0.90,\ B_o = \$100,000,\ B_i = \$150,000$$

Therefore

$$\text{threshold } q_o = \lceil N_oT_o \rceil = \lceil 10 \text{ x } 0.80 \rceil = 8$$
$$\text{threshold } q_i = \lceil N_iT_i \rceil = \lceil 8 \text{ x } 0.85 \rceil = 7$$

meaning that at least 8 vulnerabilities to outside attacks and 7 vulnerabilities to inside attacks must be secured in order to have "best" adequate SPL$_o$ and "best" adequate SPL$_i$. Table II identifies the costs of securing the prioritized vulnerabilities where vulnerability 1 has the highest priority (urgency), vulnerability 2 has the next highest priority (urgency), and so on.

TABLE II. COSTS OF SECURING OUTSIDE AND INSIDE VULNERABILITIES

u_k	Cost of Securing	v_k	Cost of Securing
1	$7,000	1	$10,000
2	$15,000	2	$40,000
3	$5,000	3	$5,000
4	$10,000	4	$20,000
5	$8,000	5	$40,000
6	$20,000	6	$5,000
7	$10,000	7	$30,000
8	$5,000	8	$5,000
9	$3,000		
10	$2,000		

As in Section III, outside and inside vulnerabilities are denoted as u_k *and* v_k respectively. Running the algorithm in Figure 6 yields $C_o = \$85,000$ at $q_o = 10$ and $C_i = \$150,000$ a*t* $q_i = 7$. The budget for securing outside vulnerabilities was more than enough to secure all outside vulnerabilities. The budget for securing inside vulnerabilities was only enough to secure 7 inside vulnerabilities. Given the existing budgets, Alice Inc.'s "best" adequate SPL$_o$ is realized with $q_o = 10$, $p_o = 0$ and its "best" adequate SPL$_i$ has $q_i = 7$, $p_i = 1$. Any additional security measure against inside attacks would require an increase in the budget.

B. *Securing Additional Vulnerabilities*

We return to Alice Inc., 5 years after implementing security in the first example. The company has grown rapidly and decides to replace its aging computer system with a new more high-performing one. However, a new system brings new vulnerabilities, so Alice Inc. decides to re-apply the proposed approach to generate new "best" adequate SPLs to make sure that the new vulnerabilities are secured. The company hires the same security consulting firm as before (the firm does good work) to perform threat analyses of its systems, resulting in a report of vulnerabilities found that could be targeted by outside and inside attackers. The report also provides values for the number of vulnerabilities as $N_o = 6$ and $N_i = 3$, and includes prioritizations of outside and inside vulnerabilities. The vulnerability numbers are lower than 5 years ago because i) the new system did not introduce very many additional vulnerabilities, ii) the company's internal work processes have not changed very much, so that there are not many additional vulnerabilities to inside attacks, and iii) vulnerabilities secured 5 years ago are still secured. For each type of vulnerability (i.e., outside or inside) the prioritizations identified which vulnerability required securing first, which one second, and so on, in declining order of urgency. Based on the consultant's recommendations, as well as its own internal deliberations, Alice Inc. assigned the following values:

$T_o = 0.80$, $T_i = 0.90$, $B_o = \$60{,}000$, $B_i = \$30{,}000$

Therefore

threshold $q_o = \lceil N_o T_o \rceil = \lceil 6 \text{ x } 0.80 \rceil = 5$
threshold $q_i = \lceil N_i T_i \rceil \ = \lceil 3 \text{ x } 0.90 \rceil = 3$

meaning that at least 5 vulnerabilities to outside attacks and 3 vulnerabilities to inside attacks must be secured in order to have "best" adequate SPL$_o$ and "best" adequate SPL$_i$. Table III identifies the costs of securing the additional vulnerabilities, where vulnerability 1 has the highest priority (urgency), vulnerability 2 has the next highest priority (urgency), and so on, as in example 1.

TABLE III. COSTS OF SECURING ADDITIONAL OUTSIDE AND INSIDE VULNERABILITIES

u_k	Cost of Securing	v_k	Cost of Securing
1	\$9,000	1	\$7,000
2	\$10,000	2	\$10,000
3	\$7,000	3	\$8,000
4	\$8,000		
5	\$16,000		
6	\$10,000		

Running the algorithm in Figure 6 yields $C_o = \$60{,}000$ at $q_o = 6$ and $C_i = \$25{,}000$ at $q_i = 3$. The budget for securing outside vulnerabilities was just enough to secure all outside vulnerabilities. The budget for securing inside vulnerabilities was more than enough to secure all 3 inside vulnerabilities. Given the existing budgets, Alice Inc.'s "best" adequate SPL$_o$ is realized with $q_o = 6$, $p_o = 0$ and its "best" adequate SPL$_i$ has $q_i = 3$, $p_i = 0$. Alice Inc. has budgeted enough funds to obtain "best" adequate SPLs that secured all vulnerabilities.

V. RELATED WORK

Related work found in the literature includes risk and threat analysis applied to various domains as well as research on vulnerabilities and countermeasures. No work was found that is similar to this work. One work, Duffany [7], is related in that it looks at protecting an enterprise's information infrastructure. This author develops an economic model for optimal resource allocation in terms of countermeasures to protect an enterprise information infrastructure. The model is solved as a linear program to determine the optimal resource allocation. However, the author does not distinguish between sensitive and non-sensitive data, but considers the organization's overall information infrastructure, including its computing devices. In addition, the author employs an economic model for optimization whereas this work optimizes based on the increase in security obtained through the addition of security measures.

In terms of risk analysis, Jing et al. [8] present an approach that uses machine learning to continuously and automatically assess privacy risks incurred by users of mobile applications. Aditya et al. [9] catalog privacy threats introduced by new, sophisticated mobile devices and applications. Their work emphasizes how these new threats are fundamentally different and inherently more dangerous than prior systems, and present a new protocol for secure communications between mobile devices. Islam et al. [10] present a risk assessment framework specifically tailored for the automotive industry. The framework starts with a threat analysis followed by a risk assessment to estimate the threat level and the impact level. This leads to an estimate of a security level, which is used to formulate high level security requirements. It is interesting that these authors also consider security levels, although the levels they use are only descriptive, such as "low", "medium", and "high".

In terms of threat analysis, Schaad and Borozdin [11] present an approach for automated threat analysis of software architecture diagrams. Their work shows that automated threat analysis is feasible. Shi et al. [12] describe a hybrid static-dynamic approach for mobile security threat analysis, where the dynamic part executes the program in a limited way by following the critical path identified in the static part. Sokolowski and Banks [13] describe the implementation of an agent-based simulation model designed to capture insider threat behavior, given a set of assumptions governing agent behavior that pre-disposes an agent to becoming a threat. Panou et al. [14] propose a cyber investment management framework named RiSKi that detects and continuously monitors insiders' societal behavior, as permitted by law, to proactively treat implied anomalies, threats, and their potential business impacts and risks. RiSKi also provides access to security incidents data to enable businesses to advance their understanding of cyber security and breaches. Sanzgiri and Dasgupta [15] present a taxonomy and classification of insider threat detection techniques based on strategies used for detection. Their classification should assist researchers and readers of this work to better understand the insider threat landscape. Baluta et al. [16] propose a discrete-event simulation model to investigate the effect of insider threats on system vulnerabilities. Their model considers both users and computer systems along with their interactions. The authors claim that the model is useful for "what-if" analysis and for gaining insights into anti-cyber intrusion strategies. Kul et al. [17] present two attack models that pose high risks for sensitive data stored in an organization's database. They discuss the complexities of both models and the defense mechanisms available in the literature.

With regard to vulnerabilities, Gawron et al. [18] investigate the detection of vulnerabilities in computer systems and computer networks. They use a logical representation of preconditions and postconditions of vulnerabilities, with the aim of providing security advisories and enhanced diagnostics for the system. Spanos et al. [19] look at ways to improve the open standard to score and rank vulnerabilities, known as the Common Vulnerability Scoring System (CVSS). They propose a new vulnerability scoring system called the Weighted Impact Vulnerability Scoring System (WIVSS) that incorporates the different impact of vulnerability characteristics. In addition, the MITRE Corporation maintains the Common Vulnerability and Exposures (CVE) list of vulnerabilities and exposures [20], standardized to facilitate information sharing. In terms of vulnerability mitigation, Oladimeji et al. [21] present a goal-centric and policy-driven framework for obtaining security and privacy risk mitigation strategies for health information interchange. They use scenario analysis and other techniques to model security and privacy objectives, threats, and mitigation strategies. Alqahtani et al. [22] propose a security vulnerability analysis framework that establishes bi-directional traceability links between security vulnerability databases and traditional software repositories. Their framework allows researchers to take advantage of semantic inference services to determine both direct and transitive dependencies between reported vulnerabilities and potentially affected software projects.

VI. Conclusion and Future Work

Organizations need to protect their sensitive data from outside and inside attacks against their computer systems that store the data. This protection is achieved by adding security measures to secure vulnerabilities to attack. However, organizations have been implementing security measures without any way of setting security protection level targets, or knowing how an added security measure contributes to the protection target. Organizations also did not have a way of selecting which security measures to implement in order to stay within the financial budget. This work proposes a structured, objective, quantitative approach to estimate, set, and achieve safe, acceptable security protection levels in terms of securing outside and inside vulnerabilities. In addition, the work proposes an algorithm for selecting which security measures to implement in order to achieve optimal adequate protection levels against outside and inside attacks, within the allowable financial budget.

This work has extended [1] in terms of updating the definition of sensitive data and adding a) more examples of attacks on sensitive data, b) more explanation and Fig. 1 on the nature of SPLs, c) an explanation and example of how to do threat analysis, d) a second application example, and e) additional references.

Future work includes investigating other formulations of security protection levels, such as incorporating the effectiveness of security measures, as well as improving the methods for threat analysis and prioritization. In addition, it would be interesting to explore how this work complements existing work in the standardization community.

References

[1] G. Yee, “Assessing Security Protection for Sensitive Data,” Proc. The Eleventh International Conference on Emerging Security Information, Systems and Technologies (SECURWARE 2017), pp. 111-116, 2017.

[2] Identity Force, “The Biggest Data Breaches in 2016,” retrieved: July, 2017, https://www.identityforce.com/blog/2016-data-breaches

[3] Identity Force, “2017 Data Breaches – The Worst So Far,” retrieved: February, 2018, https://www.identityforce.com/blog/2017-data-breaches

[4] Dark Reading, “2017 Smashed World’s Records for Most Data Breaches, Exposed Information,” retrieved: February, 2018, https://www.darkreading.com/attacks-breaches/2017-smashed-worlds-records-for-most-data-breaches-exposed-information/d/d-id/1330987?elq_mid=83109&elq_cid=1734282&_mc=NL_DR_EDT_DR_weekly_20180208&cid=NL_DR_EDT_DR_weekly_20180208&elqTrackId=700ff20d23ce4d3f984a1cfd31cb11f6&elq=5c10e9117ca04ba0ad984c11a7dfa14b&elqaid=83109&elqat=1&elqCampaignId=29666

[5] G. Yee, “Visualization and Prioritization of Privacy Risks in Software Systems,” International Journal on Advances in Security, issn 1942-2636, vol. 10, no. 1&2, pp. 14-25, 2017, http://www.iariajournals.org/security/

[6] C. Salter, O. Saydjari, B. Schneier, and J. Wallner, “Towards a Secure System Engineering Methodology,” Proc. New Security Paradigms Workshop, pp. 2-10, 1998.

[7] J. Duffany, “Optimal Resource Allocation for Securing an Enterprise Information Infrastructure,” Proc. 4th International IFIP/ACM Latin American Conference on Networking (LANC ’07), pp. 35-42, 2007.

[8] Y. Jing, G.-J. Ahn, Z. Zhao, and H. Hu, “RiskMon: Continuous and Automated Risk Assessment of Mobile Applications,” Proc. 4th ACM Conference on Data and Application Security and Privacy (CODASPY '14), pp. 99-110, 2014.

[9] P. Aditya, B. Bhattacharjee, P. Druschel, V. Erdélyi, and M. Lentz, “Brave New World: Privacy Risks for Mobile Users,” Proc. ACM MobiCom Workshop on Security and Privacy in Mobile Environments (SPME '14), pp. 7-12, 2014.

[10] M. Islam, A. Lautenbach, C. Sandberg, T. Olovsson, “A Risk Assessment Framework for Automotive Embedded Systems,” Proc. 2nd ACM International Workshop on Cyber-Physical System Security (CPSS ’16), pp. 3-14, 2016.

[11] A. Schaad and M. Borozdin, “TAM2: Automated Threat Analysis,” Proc. 27th Annual ACM Symposium on Applied Computing (SAC ’12), pp. 1103-1108, 2012.

[12] Y. Shi, W. You, K. Qian, P. Bhattacharya, and Y. Qian, “A Hybrid Analysis for Mobile Security Threat Detection,” Proc. IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), pp. 1-7, 2016.

[13] J. Sokolowski and C. Banks, “An Agent-Based Approach to Modeling Insider Threat,” Proc. Symposium on Agent-Directed Simulation (ADS '15), pp. 36-41, 2015.

[14] A. Panou, C. Ntantogian, and C. Xenakis, “RiSKi: A Framework for Modeling Cyber Threats to Estimate Risk for Data Breach Insurance,” Proc. 21st Pan-Hellenic Conference on Informatics (PCI 2017), article no. 32, pp. 1-6, 2017.

[15] A. Sanzgiri and D. Dasgupta, “Classification of Insider Threat Detection Techniques,” Proc. 11th Annual Cyber and Information Security Research Conference (CISRC ’16), article no. 25, pp. 1-4, 2016.

[16] T. Baluta, L. Ramapantulu, Y. Teo, and E. Chang, “Modeling the Effects of Insider Threats on Cybersecurity of Complex Systems,” Proc. 2017 Winter Simulation Conference (WSC), pp. 4360-4371, 2017.

2018, © Copyright by authors, Published under agreement with IARIA - www.iaria.org

[17] G. Kul, S. Upadhyaya, and A. Hughes, “Complexity of Insider Attacks to Databases,” Proc. 2017 International Workshop on Managing Insider Security Threats (MIST ’17), pp. 25-32, 2017.

[18] M. Gawron, A. Amirkhanyan, F. Cheng, and C. Meinel, “Automatic Vulnerability Detection for Weakness Visualization and Advisory Creation,” Proc. 8th International Conference on Security of Information and Networks (SIN ’15), pp. 229-236, 2015.

[19] G. Spanos, A. Sioziou, and L. Angelis, “WIVSS: A New Methodology for Scoring Information System Vulnerabilities,” Proc. 17th Panhellenic Conference on Informatics, pp. 83-90, 2013.

[20] MITRE, “Common Vulnerabilities and Exposures”, retrieved: July, 2017, https://cve.mitre.org/

[21] E. Oladimeji, L. Chung, H. Jung, and J. Kim, “Managing Security and Privacy in Ubiquitous eHealth Information Interchange,” Proc. 5th International Conference on Ubiquitous Information Management and Communication (ICUIMC ’11), article no. 26, pp. 1-10, 2011.

[22] S. Alqahtani, E. Eghan, and J. Rilling, “SV-AF – A Security Vulnerability Analysis Framework,” Proc. 2016 IEEE 27th International Symposium on Software Reliability Engineering (ISSRE), pp. 219-229, 2016.