Probability Analysis of Cyber Attack Paths

against Business and Commercial Enterprise Systems

Dmitry Dudorov, David Stupples, Martin Newby
School of Engineering and Mathematical Sciences
City University London
London, United Kingdom
{dimitry.dudorov.1, d.w.stupples, m.j.newby}@city.ac.uk

Abstract — **The level of risk of attack from new cyber-crime related malware is difficult to quantify as standard risk analysis models often take an incomplete view of the overall system. In order to understand the full malware risk faced by organisations any model developed to support the analysis must be able to address a statistical combination of all feasible attack scenarios. Moreover, since all parametric aspects of a sophisticated cyber-attack cannot be quantified, a degree of expert judgement needs to be applied. We develop a modeling approach that will facilitate risk assessment of common cyber attack scenarios together with likely probabilities of successful attack for each scenario. The paper demonstrates through use cases how a combined attack can be assessed.**

Keywords – cyber-security; cyber-terrorism; risk analysis; probability of cyber-attack ; malware

I. Paper Overview

The last five years has seen a significant increase in malware sophistication with the latest being Stuxnet, Flame and Gauss. Stuxnet famously attacked Iranian uranium enrichment industrial processes in 2010 [1]. The worm specifically targeted the Siemens supervisory control and data acquisition (SCADA) [2]. In 2012, Flame infected governmental organizations, educational institutions and private individuals, mainly based in Middle Eastern countries, with the aim of stealing intelligence, banking and intellectual property data, and sensitive private information. Gauss, also from 2012, is thought to be a development of Flame with a primary objective of collecting online banking credentials [3]. With the advent of these new sophisticated worms we must be cognisant that economically important systems such as those that support financial services, critical infrastructure and industry are now substantially at risk from cyber attack [4].

The level of risk of attack from this new sophisticated malware is difficult to quantify as standard risk analysis models often take an incomplete view the overall security of the system involved. In order to understand the true malware risk faced by organisations any model developed to support the analysis must be able to address a statistical combination of all feasible attack scenarios. Moreover, since all parametric aspects of a sophisticated cyber-attack cannot be quantified, a degree of expert judgement needs to be included.

In this paper we propose a comprehensive risk analysis model that can be used by institutions and organisations to quantify the cyber-threat risk where the attack is being orchestrated simultaneously through several parallel paths. This identification will be made possible as the model has realistic attack scenarios at its foundation. Using results from the model effective countermeasures can be assessed for cost-effectiveness, as the combinational effect will be carried through to a composite result. The proposed modeling approach is widely applicable and is illustrated in the paper using data from multiple sources. We use indicative data to show how a quantitative risk analysis can be performed. In a real case appropriate empirical data should be used.

II. Cyber-Attack Diagram

Critical systems connected to public networks are at risk of cyber-attacks. Attack scenarios with elements of social engineering, phishing or planting are becoming popular [5]. A cyber-attack scenario diagram (Fig. 1) provides a visual representation of attack routes to facilitate detailed risk analysis. Five feasible attack routes are identified and associated with real malware infections. The “mule” is an employee working for the target organisation. We assume that the cyber-attack can be carried out by stealth (via the “mule”), by subverting a “mule”, or directly by penetrating the organisation’s infrastructure. Infection of home computers or mobile devices is achieved through traditional malware routes, such as email, drive-by-download, file-sharing websites etc. These are represented as Source groups 1 and 2 according to the target and the intermediate infection state.

III. Probability Estimations

3.1 Route 1 - paid “mule” either bribed or coerced

There are a number of important papers associated with bribery risk assessment [6]; however, most focus on limited population groups that are inappropriate for this research. However, Hunt [7] provides a comprehensive bribery risk assessment that involves a broad number of factors influencing the probability of bribe including: age, culture, religion, occupation, marital status, geopolitical location, household size, salary, annual salary change, experience etc. Hunt calculates that the probability of bribing an employee as 12%. Assuming that the bribed “mule” will deliver the malicious payload to the target organisation with the probability of 100%, we get the final probability $P_{br} = P_{Hunt} / J_s$ where J_s is the job satisfaction level [8] from 0 to 100% and is assumed as 15%:

$P_{br} = 80\%$.

978-0-7695-5062-6/13 $26.00 © 2013 IEEE
DOI 10.1109/EISIC.2013.13

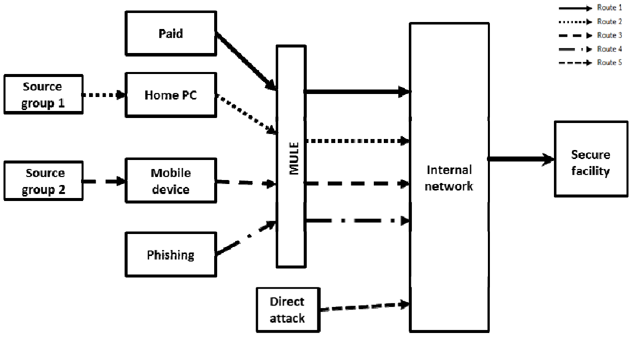

Figure 1. Cyber attack scenario diagram

3.2 Route 2 – via Home PC

3.2.1 Source group 1 – direct attack

A direct attack on a home PC can be performed in ways described by Wilson [9]. Our assumption is that an attacker requires full control of a compromised home PC in order to remotely control virus propagation. We use expert judgement [10] to assess the likelihood of attack as security applications and technology change rapidly; any purely quantitative method could prove difficult to formulate.

Judgments made by 15 cyber security experts participating in an international cyber defence exercise [11] estimated the probability of a successful remote exploitation of software vulnerabilities given different scenarios. Variables that influence the probability of a remote code execution attack were:

1) *Non-executable memory (NX) {Yes, No}.* This is a paging-based countermeasure that marks the data segment of a computer memory as non-executable making it more difficult for an attacker to execute code injected into the data segment without costs in computational power or memory usage [12].

2) *Access {No Access, Access}.* In all scenarios considered in this study it is assumed that the attacker can connect to the vulnerable service. Some vulnerabilities would require access as an authorized user to the service to be exploited.

3) *Severity of the exploited vulnerability {Medium, High}.* In all scenarios considered there is an exploit available to the attacker corresponding to vulnerability on the targeted machine. In this investigation, exploits for High and Medium severity vulnerabilities as defined by the Common Vulnerability Scoring System (CVSS) [13] are highlighted. NX access and the severity of the vulnerability exploited are all believed to affect the probability that successful arbitrary code execution is in the state yes.

The probability that successful arbitrary code execution is in the state "yes" was assessed through a discrete scale of 0-100 percent [11]. All possible combinations (2^3) between the studied variables were evaluated and the result of expert judgement is presented in Table 1.

The percentages in the Table 1 are the conditional probabilities, $P(access \mid scenario)$ so that the overall probability of access $P_{da} = \sum_{i=1}^{n} P(access \mid scenario_i) \cdot P(scenario_i)$. Since the scenario of the direct attack on a home PC will not be known in advance, we take $P(scenario_i) = 1/n$, where n is the number of scenarios.

TABLE I. STUDIED SCENARIOS AND THEIR RESULTS

Scenario	NX	Access	Exploit	Mean success percentage	Samples
1	Yes	Yes	High	85.6	15
2	Yes	Yes	Medium	74.6	15
3	No	Yes	High	81.2	15
4	No	Yes	Medium	65.7	15
5	Yes	No	High	54.7	15
6	Yes	No	Medium	42.9	15
7	No	No	High	52.3	15
8	No	No	Medium	43.7	15

Thus, the estimated expected value for the probability of attack success is:

$$P_{da} = \frac{1}{n}\sum_{i=1}^{n} a_i \approx 63\% , \qquad (1)$$

where n is the number of scenarios and $a_i = P(access \mid scenario_i)$.

3.2.2 Source group 1 – software-sharing websites

This source of infection is believed to be the most dangerous. The probability of infecting a home computer with such a virus is equal to the probability of an epidemic outbreak on a file sharing network, assuming that a hacker is exploiting 0-day vulnerability. To simulate realistic cyber epidemics caused by a torrent worm, we analyse a popular torrent tracker "ThePirateBay" (TPB). A Bit Torrent worm based on a 0-day vulnerability exploitation may propagate without any problems only until the vulnerability is patched. Having analysed TPB we propose a possible scenario for worm propagation which provides the necessary input data for a botnet growth simulation. We assume that for a successful infection, a hacker needs to hit popular files on the tracker. According to [14], the Stuxnet worm, which purportedly propagates via memory sticks, infected 14000 hosts in 72 hours. This rate is probably not possible given there is only one epicentre for the infection. We assume that the cyber criminals involved needed to constantly upload malicious files to a torrent tracker, creating more infection sources. We analysed the top 50 files by popularity for three months. Each file on the list contains; name, upload date, size, seeds, leeches, and downloads. We calculated the download rate of the typical file as the average number of downloads divided by the average number of hours since the upload, which gives 102.19 times/hour. However, the hacker tends to create more infection sources before the vulnerability is plugged. Therefore there is a need to upload more popular files to a BitTorrent tracker. Creation and camouflaging of worm-carrying files takes time and the content needs to be divided into small parts for the seeding process. Internet upload speed is also taken into consideration (see Table 2). According to [15], the average upload speed in the UK is 0.9257 Mb/s.

TABLE II. UPLOAD AND DOWNLOAD SPEEDS OF UK ISPS

ISP	Average download speed	Average upload speed
AOL	3.919	0.433
BE and O2	5.945	0.871
BT	6.981	1.325
Eclipse	9.266	1.415
Orange	3.321	0.598
Plusnet	3.122	0.394
Sky	4.571	0.573
Talk Talk	5.276	0.622
Tiscali	3.989	0.552
Virgin Media	15.18	2.474
Av.value(Mb/s)	6.157	0.926

The creation and uploading time for an average size file is 0.4 hours. Therefore, until the 0-day vulnerability is patched, 951.14 malicious files can be uploaded. In normal working conditions a hacker may create 0.3 infection epicentres per hour. Using the expression:

$$I_h = \frac{\alpha \bar{c} \varphi (1+\varphi)}{2}, \tag{2}$$

where I_h is total number of involved hosts, $\bar{c}$ is the average download rate, α is the rate for new infection source formation, and φ is a number of infected files uploaded. This gives epidemic numbers as 13,913,279. A stochastic differential equation model from [16] provides a propagation scenario. The critical variables that affect the dynamics are the epidemiological contact and removal rates. While the removal process is usually never affected by any external conditions, contact rate is more complicated and depends on infection source nature. For Internet worm modeling, [17] proposes an alternative formulation for pair-wise infection rate. Contact rate for the uniform scanning worms is defined as $\beta = \eta / \Omega$, where η is the scan rate, and Ω is the size of the IP space scanned by the worm. The number of scanned computers represents the number of contacts made during some time period. In a torrent propagation scenario this is represented the total number of downloads of infected files from different sources, or I_h. Ω is the total number of unique users visiting an average BitTorrent website. Audience statistics from TPB over a 3 month period identify 16,865,321 users [18]. The epidemiological contact rate is given as:

$$\beta = \frac{\alpha \bar{c} \varphi (1+\varphi)}{2\Omega} = 0.82496 \tag{3}$$

One thousand iterations of the simulation show the average value for the infected population to be very close to the value that we calculated by analysing torrent tracker statistics. According to [16], the probability of an epidemic outbreak is given by:

$$P_{outbreak} = 1 - \left(\frac{b+\gamma}{\beta} \right)^{i_0}, \tag{4}$$

where b is the birth rate, γ is the infection removal rate and i_0 is the initial number of infected hosts. Applying this to our previous calculations we get the probability of getting an infection from torrent environment as:

$$P_t \approx 78\% .$$

3.2.3 Source group 1 – E-mail

Many researchers have investigated the topological effect on the propagation of email worms using graph theory [19]. Other researchers use epidemiological equations [20] and branching theory [21]. Although being theoretically correct, these models do not match the statistical data found in annual reports of different security companies [22], [23], [24]. Cyber criminals tend to use databases containing victims' email addresses and social engineering techniques to improve their success [25]. We propose a new probability formula for email malware:

$$P_e = \frac{P_s c (1-\bar{r})}{\omega} \tag{5}$$

where P_s is the probability of a malicious message getting through mail server spam filters, c represents the number of clicks on a malicious message, $\bar{r}$ is the average detection rate of an antivirus program installed on user's PC (1-$\bar{r}$ is the probability of non-detection), ω is the total number of views of a malicious message. The probability of penetrating a spam filter is not trivial. When using social engineering techniques, hackers tend to create a realistic message so that filtering on word frequency or entropy method will fail. Also, the content of the message may change, but the payload remains the same, making spam filters return false probabilities. According to one study into scam message filtering [26], the probability of a single fraudulent e-mail treated as legitimate is $P_s \approx$ 12%. The number of clicks on the malicious message/total number of views of that message represents the probability of opening an infected email. This can only be quantified by running an experiment, since the human action is required and the content of the message is initially unknown. We assume that a fraudulent message emanates from an unknown source and hacker does not use any social network. According to [27], 16% of the tested group responded to an email containing social engineering techniques and sent from an unknown address. Therefore, we assume that c/ω= 0.16. Even when clicked, an email attachment can remain harmless because of antivirus software. The probability of detecting a malicious attachment is estimated by analysing detection rates of different antivirus products [28]. For viruses exploiting 0-day vulnerabilities the detection rates are different. [29] conducted an experiment where 124 malware instances which did not have antivirus definitions were tested – approximately 56% were not detected. The average rate for detection of 0-day viruses is $\bar{r} = 0.44$. Therefore, the final probability of being infected via email is:

$P_e = 0.118 \cdot 0.56 \cdot 0.16 \approx 1\%$.

3.2.4 Source group 1 – rogue websites

Drive-by downloads are particularly common form of such attacks [30]. The probability of this infection type cannot be estimated analytically as there is; diversity in web users' population, disparate content viewing, variable frequency of web surfing, different levels of security awareness, variation in security settings, etc. Kaspersky Labs has calculated the frequency of web antivirus detections in different countries for each quarter [31]. These figures are based on the raw number of web antivirus alerts from computers and are not adjusted to reflect the number of Kaspersky Security Network [32] users in each country. On average in Q3 2012, 36.7% of KSN users were attacked at least once while surfing online. The UK figures give us the probability of being attacked while surfing the Internet as:

$P_{db} \approx 30\%$.

3.2.5 Source group 1 – Instant messaging

Malicious applications use IM clients to send spam and links to applications to steal users' account numbers and passwords. The probability is given by $P_{im} = 1 - e^{-vm/k}$ [35]. The number of vulnerabilities is calculated as $v = (V_d \cdot V_a)/365 =$ 0.3589, where V_a is the annual number of 0-day vulnerabilities detected in instant messaging software (1 for 2011) [33] and V_d is the average duration of exploitable 0-day vulnerability [34]. The attacker is assumed to be an expert, and therefore we can use m = 450 exploits [35]. Thus, probability is:

$P_{im} \approx 0.8\%$.

3.2.6 Concluding Route 2 Probability of transerring an infection from a home computer

Transfer to a target organisation's internal network depends on statistical data for external device usage. File transfer is only possible when personal external devices are used at work. According to [36], 81% of employees personally owned electronic device for work-related functions. We note, 52.65% of employees use either a personal desktop or laptop. Assuming that there is no transfer from a home PC to a MULE state (see Fig.1) we estimate the final probability for Route 2 according to Bayes' theorem, where a priori probability of hypothesis is 53%, and the probability of an event is the probability of individual infection sources from Source group 1.

3.3 Route 3 – via Mobile device

3.3.1 Source group 2 infecting via Bluetooth

Smart phones offer Internet connectivity as well as external data drives. Its mobility provides a perfect environment for disease spread and therefore it is very attractive to cyber criminals. Malware spreads mainly via Bluetooth connectivity and the Multimedia Messaging Service (MMS). [37] proposes a propagation model for smart phone viruses to calculate the probability of a smart-phone epidemic outbreak; a Bluetooth epidemic model is defined by:

$$\frac{dI(t)}{dt} = \beta\bar{k}S(t)I(t) - \gamma I(t), \qquad (6)$$

where β is the infection rate, γ is the removal rate from infectious nodes and $\bar{k}$ is the node average degree in unit time. Coverage area of a mobile node is presented in the figure 2 below:

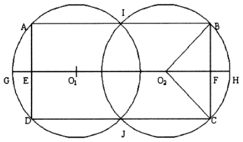

Figure 2. Mobile node coverage area

The authors assume that mobile nodes are uniformly distributed and defined by the following parameters; r is the coverage radius of Bluetooth signal, σ is the distribution density of smart phones, Δt is the time for worm to replicate itself. Not all nodes in the coverage area have enough time to be infected as devices move out of range within the replication period. Suppose the node was at position O_1 and after Δt, the node moves to O_2, i.e. $O_1O_2 = v\Delta t$. The area between straight-line AB and straight line CD is defined as the infection effective region and h is the width of that region $h = AD = \sqrt{4r^2 - (\Delta t)^2 v^2}$. The area of valid coverage region of Bluetooth signal in one second is defined as $S = vh + \pi \ \frac{h}{2}^{2}$ and the average degree formula is defined as:

$$\bar{k} = \sigma S - 1 = \sigma \ \ v\sqrt{4r^2 - (\Delta t)^2 v^2} + \pi r^2 - \frac{\pi}{4}(\Delta t)^2 v^2 \ \ -1, \qquad (7)$$

The epidemic dynamics are sensitive to the epidemiological contact rate parameter β. We propose a statistical approach, and use statistical data from [38] and [39]. Although, their approaches were different and experiments were conducted in different environments and at different times, there is good correlation in their results. The first experiment discovered 1269 Bluetooth devices over a period of four days, giving $\beta_1 = 0.110156$ (assuming that there are 8 working hours per day). The second experiment results are provided in the table below:

TABLE III. EXPERIMENT RESULTS

Traces	Site 1	Site 2	Site 3
Duration(min)	107	204	801, intermittent trace
Total devices found	90	100	106
Cell phones	87	84	96
Computers	3	15	9
Headsets	0	1	0
Unknown Device	0	0	1

We assume that Bluetooth connectivity in a subway was successful only 15% of the time. Therefore 296 devices were found in 437 minutes, which gives β_2 = 0.011289. We decided to use the average between two experiments and our

$\beta_{bt} = (\beta_1 + \beta_2)/2 = 0.0111523$. We define the probability of a Bluetooth epidemic outbreak as $P_{bt} = 1 - \left((b+\gamma)/(\beta_{bt}\bar{k})\right)^{i_0}$. We assume that there is one unique source of infection and the birth rate is equal to zero. According to [37] it takes users an average of 50 minutes to get technical support, among all users 1 out of 10 cannot get immunity, therefore γ =0.018. In order to estimate the distribution density of smart phones we make two assumptions; every person has a mobile phone, but not every mobile phone is a smart phone, and hence we can apply population density and mobile phones statistical data, and the epidemic outbreak can only happen in large cities where the chances of successful attack are higher. According to [40] urban population in the UK is approximately 54 million, and the area of urban territories is 2500 km^2, giving the urban population density of 0.022 per m^2. [41] states that 56% of UK consumers own a smart phone. Therefore, smart phone density is estimated as $\sigma \approx$ 0.0123 devices per m^2. Applying all parameters we get:

$$P_{bt} \approx 52\%.$$

3.3.2 Source group 2 –direct attack on a mobile device

This can be performed in a limited number of ways; eavesdropping, impersonations of a user, impersonation of the network, man-in-the-middle, and compromising authentication vectors. The man-in-the-middle attack is the most dangerous as an intruder can force the device to accept unwanted code and hence carry the disease further. A fake broadcast Base Transceiver Station signal dupes mobile phones in its range area to communicate with it. According to ASMONIA research studies [42], direct attack on a mobile device can be performed with the probability of:

$$P_{dma} \approx 33\%.$$

3.3.3 Source group 2 –mobile network messaging

These viruses are adequately described with epidemiological equations [39], [37], [43]. While contact list quantity influences the final size of an epidemic, user contact rate is very useful in the calculation of the main parameter for an epidemic outbreak probability – the basic reproduction number. The probability of an outbreak in a mobile messaging environment is calculated as in part 3.2.1, where the birth rate is equal to zero and the initial number of infected phones is equal to one.

The contact rate is calculated as $\beta_{mms} = u * <k>$ according to [44], where u is the inverse of time that a virus takes to infect a susceptible phone and $<k>$ is the average number of phones that can contact each other and estimated as $<k> = T_i \cdot n_{tp} / (2 * T_i \cdot r)^2$, where r is the radius of a cell tower, n_{tp} is total number of cell phones in the service area.

The recovery rate γ depends on user's own security awareness that can be quantified by $\gamma = \upsilon_i \cdot p_{click}$ and the recovery proportion that depends on the public security awareness. Recovery rate is calculated as in [37], giving $\gamma = \upsilon_i \cdot p_{click}$ = 0.018. We estimate u =50 minutes according to [45]. The range of a cell tower is assumed to be approximately 400 m, and $n_{tp} = 10^4$ because the infection is started in urban area where the density of mobile users is very high. Therefore, from equation (4), the probability of an outbreak in a mobile environment can be estimated as:

$$P_{mms} = 1 - \left(\frac{\upsilon_i \cdot p_{click}}{u * T_i \cdot n_{tp} / (2 * T_i \cdot r)^2} \right)^{i_0} \text{ giving } P_{mms} \approx 76\% \qquad (8)$$

3.3.4 Concluding Route 3 probability of transerring an infection from a mobile device

In a similar manner to section 3.2.6, we assume that the mobile device is owned by an employee and therefore there is no transfer from state Mobile device to state MULE. According to [36], the percentage of employees who use smart phones or tablets for work is 32%. The final probability for Route 3 is estimated using Bayes' theorem, where a priori probability of hypothesis is 32% and the probability of an event is the probability of individual infection sources from Source group 2.

3.4 Route 4 – via Phishing or Planting

According to the latest report [5] Phishing remains the most successful type of attack. We propose an expression for estimation of its probability as $P_{fp} = P_{pick} \cdot P_{plug} \cdot P_{av}$. Here, $P_{av} = (1 - \bar{r})$ is the probability for a malicious payload to get through the antivirus, and $\bar{r}$ is the average detection rate from [28]. $P_{pick} \cdot P_{plug}$ is calculated with respect to this probability tree:

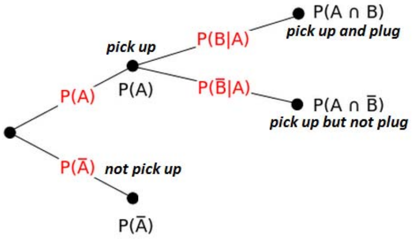

Figure 3. Survey probability tree

Here, $P(A)$ is the probability of picking up a previously planted infected flash drive and $P(B)$ is the probability of plugging it in order to check the data. This part of our research is based on the survey where a number of respondents were asked two questions reflecting $P(A)$ and $P(B)$. Respondents represented different age groups, education background, occupation and income levels [46].

The result gives probability of $P(A \cap B)$ = 60%. The probability of phishing/planting attack is calculated as:

$$P_{fp} \approx 60\%.$$

3.5 Route 5 – Direct attack on internal network

A direct attack on a target organisation's internal network is dependent on network's topology and infrastructure. Simple enterprise architecture is considered to model the attack. Paper [47] proposes a model for estimating the time to compromise a system component that is visible to an attacker. This model provides an estimate of the expected value of the time-to-compromise as a function of known and visible vulnerabilities and attacker skill level. The time-to-compromise a random process model is a composite of three sub processes associated with attacker actions aimed at the exploitation of vulnerabilities.

The probability of such attack is composed as $P_{attack} = P_t \cdot P_a \cdot P_b \cdot P_s \cdot P_c$, where P_t is the probability the system is on an attacker target list, P_a is the probability of being attacked given that the system is targeted, P_b is the probability of a perimeter breach given that the system was attacked, P_s is the probability of a successful attack given that there was a perimeter breach and P_c is the probability of damage given the system was successfully attacked. We set P_t=P_a=P_c=100% since we assume that attacker has already decided to attack the target organisation and is not focused on damage.

Attack on the final system is considered as another state for this model. The probability that the attacker has at least one exploit readily available that can be used against one of the known vulnerabilities is calculated by using 'search theory' in a similar fashion as has been applied to physical security systems by [35].

The following expression uses the simplifying assumption that the available exploits are uniformly distributed over all vulnerabilities $P_{dna} = 1 - e^{-vm/k}$ where P_b is the probability that the attacker has an exploit readily available, v is number of vulnerabilities on the component of interest, m is the number of exploits readily available to the attacker, and k is the total number of vulnerabilities. The value of k is 19871 and is defined to be the total number of non-duplicate known vulnerabilities found in [48].

The value of m is a function of the attacker skill level. The *Novice* skill level is defined as m = 50 because there is a Metasploit web site [49] that has 50 exploits that are trivial to use. The higher skill levels are defined by increasing the value of m for each increase in skill level.

For the *Beginner* it is 150, *Intermediate* has 250 and for *Expert* it is 450. The specific choices are based on a postulated exponential growth in readily available exploits as a function of skill level. The number of vulnerabilities on the component of interest is calculated as $v = (V_d \cdot V_a)/365$ =2.87, where V_a = 8 is the annual number of 0-day vulnerabilities detected [50] and V_d =131 is the average duration of exploitable 0-day vulnerability [34].

The attacker is assumed to be an expert, and therefore we can use m = 450 exploits. Thus, the probability of a direct attack in organisation's internal network is:

$P_{dna} \approx$ 6%.

IV. Conclusion

Table 4 provides a probability representation of Figure 1. The figures presented are from the concluding arguments of the sections above. The order of presentation reflects the model at Figure 1.

Use cases demonstrate this model (Fig. 1). A target organisation is attacked via an unaware employee bringing an infected laptop computer into the workplace. We assume that the attack on the laptop computer is undertaken by a lone-wolf hacker via luring people to an infected website and spreading malware via software sharing websites.

TABLE IV. Probability Values

Variable	Infection source	Probability value, %
P_{br}	Dissatisfied and paid employee	80
P_{pc}	Home computer (transfer)	53
P_{md}	Mobile device (transfer)	32
P_{fp}	Planting technique	60
P_e	E-mail	1
P_{db}	Drive-by-download	30
P_t	Software sharing websites	78
P_{im}	Instant messaging software	0.8
P_{da}	Direct attack	63
P_{dma}	Fake BTS	33
P_{mms}	Mobile messaging	76
P_{bt}	Bluetooth	52

The probability of infecting a laptop and successfully transmitting the infection to its final target can then be calculated using the following expression:

$$P(case\_1) = \frac{P_{pc} P_{da} (s_t P_t + s_{db} P_{db})}{n \sum_{i=1}^{5} P_i}, \tag{9}$$

where P_i is the probability value from Source Group 1, s_t and s_{db} are the number of attacks from torrent and rogue website respectfully, and n is the total number of attacks on a laptop. We assume that the target secure facility is no more than a computer inside the organisation's internal network, and therefore the probability of attack on it can be estimated as in section 3.2.1, and hacker sends s_{db} =250 malicious links and launches s_t =1 torrent attack. Therefore the probability of a successful penetration is:

$P(case\_1) \approx 6\% \cdot$

A further use case represents an attack from a group of 3 hackers using more trivial methods such as bribing an employee, dropping a number of infected flash drives on the target organisation's car park and walking around the building with an infected phone. The probability of a successful infection attack can be estimated by:

$$P(case\_2) = P_{da} \frac{\left. P_{br} s_{br} + P_{fp} s_{fp} + P_{bt} s_{bt} \middle/ \sum_{i=1}^{3} P_q \right.}{k \sum_{j=1}^{5} P_j} \tag{10}$$

Where P_q is the probability value from Source Group 2, P_j is the probability value from MULE state inputs, s_{br} is the number of bribe trials, s_{fp} is the number of flash drives dropped, s_{bt} is the number of Bluetooth attacks, k is the total number of attacks on target organisation. We assume that

Bluetooth attacks can be carried out only three times a day (morning, lunchtime, end of the day – to maximise the smart phone population density) and only 10 days during the 131 day "stealth" period to remain unsuspicious, so that s_{bt}=30; hacker group drops only s_{fp} = 50 flash drives and bribes s_{br} =5 people. Therefore the probability for this use case is:

$$P(case\_2) \approx 19\% .$$

Combination of attacks based on the routes from Figure 1 can be calculated using values from Table 7 and constructing an appropriate probability expression. When setting up specific experiment for a designated location it would be necessary to revisit the data used. More specifically, it may be necessary to undertake further data collection or scale existing results to tailor the model to the environment being considered.

References

[1] Stuxnet worm hits Iran nuclear plant staff computers. (2010, September 26). Retrieved from BBC News Middle East: http://www.bbc.co.uk/news/world-middle-east-11414483

[2] Matrosov Aleksandr, R. E. (2012). Stuxnet Under the Microscope. ESET.

[3] Global Research & Analysis Team (GReAT), Kaspersky Lab. (2012). Gauss: Abnormal Distribution.

[4] Moteff, J. (2005). Risk Management and Critical Infrastructure Protection: Assessing, Integrating, and Managing.

[5] Emerging Cyber threats report 2013. (2012). Georgia Tech Cyber Security Summit.

[6] Avoiding Corruption Risks in the City: The Bribery Act 2010. (2010). London: Transparency International.

[7] Hunt, J. (2004). Trust and Bribery: The Role of the Quid Pro Quo and the Link with Crime. IZA Discussion Paper No.1179.

[8] Herzberg, Frederick. (1959). The Motivation to Work, New York: John Wiley and Sons.

[9] Wilson, C. (2008). Botnets, Cybercrime, and Cyberterrorism: Vulnerabilities and Policy Issues for Congress

[10] T. Bedford, R. Cooke. Probabilistic risk analysis: Foundations and methods, (2001). p.191.

[11] Holm H., S. T. (2012). Success Rate of Remote Code Execution Attacks Expert Assessments and Observations.

[12] Younan, Y. (2008). Efficient countermeasures for software vulnerabilities due to memory management errors.

[13] Mell, P. S. (2007). A complete guide to the common vulnerability scoring system version 2.0.

[14] W32.Stuxnet Dossier whitepaper. (2011, February). Retrieved from Symantec official website: http://www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/w32_stuxnet_dossier.pdf

[15] UK fixed-line broadband performance: The performance of fixed-line broadband delivered to UK residential customers. (2011, July 27). Retrieved from OFCOM official website: http://stakeholders.ofcom.org.uk/binaries/research/telecoms-research/bbspeeds2011/bb-speeds-may2011.pdf

[16] F. Brauer, P. van den Driessche, J. Wu. (2008). Mathematical Epidemiology, Springer-Verlag Berlin Heidelberg.

[17] Dagon, D., Zou, C., & Lee, W. (2006). Modeling Botnet Propagation Using Time Zones. In Proceedings of the 13 th Network and Distributed System Security Symposium NDSS.

[18] The Pirate Bay official website. (2012). http:/thepiratebay.se/tag/Traffic

[19] Yin Ke-xin, Z. J.-q. (2011). Simulation on Email Worms Propagation. IEEE, International Conference on Mechatronic Science, Electric Engineering and Computer Science.

[20] Zou, C., Towsley, D., & Gong, W. (2004). Email Worm Modeling and Defense. Computer Communications and Networks, 2004. ICCCN 2004., (pp. 409 - 414). Chicago, IL, USA.

[21] Yang Xiang, X. F. (2009). Propagation of Active Worms. Centre for Intelligent and Networked Systems, Central Queensland University.

[22] Symantec. (2012). Internet Security Threat Report.

[23] Kaspersky Lab. (2011). Threat Evolution Report.

[24] Cisco Corp. (2011). Cisco Annual Security Report.

[25] MAAWG Email Security Awareness and Usage Report. Ipsos Public Affairs. (2010).

[26] E. Airoldi, B. M. (2005). Technologies to Defeat Fraudulent Schemes Related to Email Requests. AAAI Spring.

[27] Rachna Dhamija, J. T. (2006). Why Phishing Works. Conference on Human Factors in Computing Systems.

[28] Anti-Virus Comparative report 2011. (n.d.). Retrieved from www.av-comparatives.org.

[29] Finn Michael Halvorsen, R. W. (2008). Zero-day Malware.

[30] Marco Cova, C. K. (2010). Detection and Analysis of Drive-by-Download Attacks and Malicious JavaScript Code.

[31] Yury Namestnikov, Kaspersky Lab. (2012). IT Threat Evolution: Q3 2012.

[32] Retrieved from Kaspersky Security Network official website: http://ksn.kaspersky.com/en. (2012).

[33] Symantec. (2011). Internet Security Threat Report, part 2. Total number of vulnerabilities.

[34] L. Bilge, T. Dumitras. (2012). Symantec Research Labs. An Empirical Study of Zero-Day Attacks In The Real World.

[35] Major, J. A. (2002). Advanced Techniques for Modeling Terrorism Risk. Journal of Risk Finance.

[36] Camp, Cameron, ESET. (2012). Retrieved from ESET Threat Blog.

[37] XIA Wei, L. Z. (2007). Dynamic Epidemic Model of Smart Phone Virus Propagated through Bluetooth and MMS.

[38] Bluetsnarfing, M. H. (2004). Detecting and Attacking Bluetooth-enabled Cellphones at the Hanover Fairground. CeBIT 2004.

[39] Jing Su, K. K. (2006). A Preliminary Investigation of Worm Infections in a Bluetooth Environment. WORM'06.

[40] 2011 Census, Population and Household Estimates for England and Wales. (2011).

[41] Gavin Sudgen, IPSOS. (2012). NewMedia TrendWatch report.

[42] ASMONIA project. (2012). Threat and Risk Analysis for Mobile Communication Networks and Mobile Terminals, p.175

[43] Liu, C. G. (2011). Modeling and Predicting the Dynamics of Mobile Virus Spread Affected by Human Behavior. World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2011 IEEE International Symposium.

[44] Pu Wang, M. C.-L. (2009). Understanding the spreading patterns of mobile phone viruses. Science, 324, 1071-1076.

[45] Chao Gao, J. L. (2011). Modeling and Restraining Mobile Virus Propagation. IEEEXplore.

[46] Centre for Cyber Security Sciences, City University London. Lost flash drives and User Behaviour. Survey. February 2013.

[47] Miles A. McQueen, W. F. (2005). Time-To-Compromise Model For Cyber Risk Reduction Estimation. Quality of Protection Workshop, ESORICS.

[48] Retrieved from Inj3ct0r, the ultimate database of exploits and vulnerabilities: http://1337day.com/. (2012).

[49] Retrieved from Metasploit: http://www.metasploit.com/. (2012).

[50] Positive Technologies. (2012). Vulnerability Statistics for 2011.