Data leakage triage : quantitative measure of effectiveness of containment plan

Pan, Jonathan

2015

Pan, J. (2015). Data leakage triage : quantitative measure of effectiveness of containment plan. International Journal of Information and Computer Security, 7(2/3/4), 240-254.
doi:10.1504/IJICS.2015.073032

<https://hdl.handle.net/10356/144714>

<https://doi.org/10.1504/IJICS.2015.073032>

© 2015 Inderscience. All rights reserved. This paper was published in International Journal of Information and Computer Security and is made available with permission of Inderscience.

Downloaded on 26 Apr 2023 20:22:26 SGT

Data Leakage Triage: Quantitative Measure of Effectiveness of Containment Plan

---

Jonathan Pan*

Wee Kim Wee School of Communication and Information
Nanyang Technological University

Singapore

E-mail: jonathan.pan.jy@gmail.com

*Corresponding author

Abstract: Increasingly organizations' deployed security defences are challenged by persistent malware attacks. There are many risks induced by successful malicious assaults against an organization. One of the significant risks is *data leakage*. This risk has significant impacts to organizations and nations. To mitigate the likely occurrence of such risk, adequate preparation and response are needed. Epidemiological modelling of such Malware outbreaks can aid in the development of mitigation plans. Most of these models used the number of infected hosts or time duration to analyse Malware behaviours or the effectiveness of response plan. This paper proposes an extension of the Agent Based Malware Containment Model (AMCM) to provide a quantitative measurement of the amount of data leakage from Malware infected hosts in order to evaluate the effectiveness of the response containment plan. It is believed that this form of measurement will enhance the preparation of incident response plans to deal with Malware outbreak.

Keywords: Malware Containment; Agent Based Modelling; Malware Incident Response Plan.

Biographical Notes:

Jonathan Pan is currently a Part-Time Lecturer at the Wee Kim Wee School of Communication and Information, Nanyang Technological University, Singapore. He received his doctoral degree in Information Technology from Murdoch University and Master of Technology Management from University of Queensland and Bachelor of Computer Science from University of Newcastle. His research interests are in Cyber Security Analytics and Simulation Modeling using Agent Based Modeling. He authored over 10 papers in journals and proceedings of international conferences.

---

1 Introduction

Malware have advanced in their ability to infiltrate an organization's computing resources and to keep these resources infected for prolonged durations. This poses a significant challenge to the organization's Information Technology (IT) security defences. Possible risks induced by a successful Malware infection and subsequent outbreaks include the loss of sensitive and important data, the loss of IT services used to support the organization's business operations and tarnished organizational image. The effects of data leakage and loss of control to preserve and protect sensitive data have significant impacts to organizations and in some instance, this has greater bearings to affect the trust relationships between countries. Data leakage, in the context of a Malware intrusion, occurs with the exfiltration of sensitive data from the organization's storage and network into the hands of untrusted entities or the public domain.

In order to deal with such risks from the threat of Malware, formally formed incident responders like the Computer Security Incident Response Team (CSIRT), or ad-hoc IT individuals or teams, are deployed as incident responders. Their primary objective would be to "*minimize the impact of an incident to a company and allow it to get back to work as quickly as possible*", according to Wiik et al. (Wiik, Gonzalez and Kossakowski, 2006)). Containment is generally a resource intensive activity requiring the deployment of incident responders to deal with infected computers. There are many limitations to overcome when executing a containment plan. This includes environmental constraints like dispersed physical locale and heterogeneous configurations, high demands of skills and

---

knowledge like Malware analysis and digital forensics, and complex steps involved in handling of advanced Malware. This in turn will limit the available staff resources to deploy and respond to the situation. Hence the incident response team will need to devise an optimal Malware containment plan. Such optimization may be done prior to or during the Malware outbreak. In order to facilitate the ability to assess the effectiveness of plan options, simulation modelling can be used. Such simulation allows the incident responders or decision makers to assess the effectiveness of containment plans and to minimize the risk exposure from the Malware outbreak. Currently, most simulation models use the number of infected computers or the duration of containment as the measurement of the effectiveness of containment plans. In practice, incident responders would focus more with dealing with infected hosts instead of considering

the impact of data leakage from these hosts. This paper proposes the use of quantitative assessment of data leakage to measure the extent of the risk induced by Malware outbreak and effectiveness of containment plans. This is an extension to the Agent Based Malware Containment simulation model (AMCM) (Pan and Fung, 2012). However, there are qualifications to the use of this approach in measuring containment plans. Firstly, the infection is assumed to occur within the environment where the incident responders have the means to mitigate the impact of the outbreak. Secondly, the infection may be wide spread, that is, the outbreak spreads to more than just a limited or isolated group of computing devices.

The next section looks into the need for an effective Malware containment. The proposed solution and a survey of existing research are covered next. A description of the modelling configuration and analysis then follows. The last section discusses the findings, future research options and conclusion.

2 Need to Re-Enforce New Defence Line

While a successful containment is the desired outcome for incident responders, there are many constraints and challenges that have to be dealt with. Malware are not considerate about the situation and will pursue relentlessly to achieve its malicious objectives. On the other hand, the demand on incident responders and decision makers is to devise an optimal containment to minimize the risk impact and to restore computing continuity and reliability.

A. The Losing Battle

Malware has been advancing in their ability to infiltrate, to evade detection or eradication, and to launch attacks against their targets. Advancement that entails the use of Artificial Intelligence (Pan and Fung, 2008) and ability to deceive their victims through human-like interactivity capability is becoming pervasive (Pan and Fung, 2009). Custom developed Malware are also increasingly being used against targeted organizations or individuals through means to circumvent deployed defences (Pan and Fung, 2010).

Security companies are working hard to develop preventive measures and countermeasures to fend off these threats. However, AUSAERT reckons that 80% of the Anti-Virus solutions are ineffective in detecting and removing Malware (Kotadia, 2006). In a report by MessageLab, it noted that out of 31 Anti-Virus companies, only 6 recognized malicious

files that contain a Malware (Shipp, 2007). Finally, people are the weakest link in the whole security defence posture. Hence, successful Malware penetration is a scenario that an organization cannot afford to ignore, or unprepared against such eventuality.

B. Containment Challenges

Beyond the sophistication of the adversary, incident responders have added problems to their limitations and constraints to deal with. This includes the lack of sufficient staff with appropriate skills and knowledge, lack of comprehensive policies and procedures in the organisation to mitigate such operations, and the lack of required tools needed for incident response work. A report by Kossakowski (Kossakowski, 2005) exposed that Incident Responders are overstretched. One of the noted reasons for this was the lack of coordination mechanisms in place. This is understandably so when the environment in which the responders operate in vary widely in several dimensions like secondary logistical support differences and heterogeneous computing environment.

Also it is noted that containment consumes a considerable amount of time before a Malware infiltration is dealt with. According to Verizon's 2010 Data Breach Investigation Report (Verizon, 2010), containment takes days or even months to complete while a Malware infection occurs within seconds to minutes. There is a significant risk exposure incurred from the time of infiltration to its containment.

C. Data Leakage

When Malware successfully infiltrates the organization's defences and starts spreading, the Malware gains control over the organization's computing assets. However the most significant computing asset lost from such occurrence is the organization's data. Data loss can be in the form of data leakage or data destruction. Such occurrences can incur significant damages to the organization. On 20 April 2011, Sony's Playstation Network (Messmer, 2011) was down for a prolonged duration. However the greatest impact to Sony from this incident was not the outage although ingress revenue was affected, the greatest impact was the likely compensation lawsuit from many communities due to such incident and that could amount to millions of dollars. This reinforces the importance of this research on data leakage.

D. Modelling of Containment

According to a recommendation made by the Department of Homeland Security (DHS) in the handling of Malware outbreak, it advocated stages of Planning and Triage to prioritize and assess the situation in order to devise an optimal containment plan (Killcrece, 2013). A means to develop an understanding of the situation based on the condition of the outbreak and the effects of mitigation options will aid decision makers and incident responders in their conduct of Triage. According to Boase and Wellman (Boase and Welman, 2001), the propagation characteristics of Malware are likened to that of biological equivalence. Epidemiological models have been developed to study the characteristics of Malware outbreak. Hence a quantitative measure of the modelling of the Malware epidemic can aid responders to develop an optimal remediation containment plan.

A consideration while developing this research hypothesis is whether data leakage could be immediately rectified through the breakage of connectivity to the Internet. However there are conditions whereby such a 'kill switch' to stop data leakage cannot be applied. Examples of such conditions include the need to ensure business continuity using connectivity to the Internet, or due to constraints of the operating environments such as remote computing.

3 Research Proposition

This research proposition entails the assessment of data leakage as a measurement output from the simulation model for Malware containment analysis. The research is based on an Agent based Malware Containment Model (AMCM) (Pan and Fung, 2012) which was developed against the popular and generic SIR (Susceptible-Infected-Recovered) epidemiological model (Kermack and McKendrick, 1927). The latter compartmental-based model is frequently used to study Malware epidemics. The AMCM similarly uses the compartmental model that has three states – 'Susceptible', 'Infected' and 'Recovered'. It was developed using the Netlogo agent based modelling development toolkit (Netlogo).

The extended AMCM model has a number of input parameters with new additions included described later in the model configuration section. The model has three type of agents interacting with each other in the simulated world. The model parameters and agents have been coded into the AMCM simulation model to simulate the epidemiological characteristics of a Malware outbreak among computing hosts to support various 'what-if' scenarios.

The first type of agents is the computing hosts whose population size is configurable and is defined at the start of the simulation. The second is the Malware. The third is the Incident Responder. Each host agents will only have one of the three epidemiological status. The initial state of all hosts will be assigned the 'Susceptible' status. When a host agent becomes infected with a Malware agent, it will have the 'Infected' status. When an Incident Responder agent 'treats' the 'Infected' host, its status changes to 'Recovered'. In this simulation, the status of host agent will not revert to 'Infected' status. The host agents primarily role is to hold their own infection status. For the model extension, the host agents will be initially assigned a prioritization value that represents the amount of important data the host carries. The Malware agents' population size starts with one at the start of the simulation. At each simulation cycle, the population of

Malware agents will seek to find other 'Susceptible' host and infect it with a cloned copy of itself. Hence the size of the Malware agent population will match the size of host agent population when the infection epidemic reaches its peak. In the simulation extension of AMCM, the Malware agent contributes to the measure of the epidemiological condition in the form of data leakage that will be described in greater details later. The Incident Responder agents' population size, like the computing hosts, is configurable. However the size of this type of agents is considerably smaller to simulate the real-world constraints of the team size limitations whose is responsible to deal with all Malware incidents. The Incident Responder agents' execution rule is to find infected hosts and 'treat' them by removing the Malware agent and setting the treated hosts' statuses to 'Recovered'. Depending on the setting prior to the simulation run, each Incident Responder agent will either 'treat' the next nearest 'Infected' host or the next highest prioritized selected host based on the extent of important data the host carries. Details of the prioritization will be covered later. The following pseudo code illustrates the rules coded into the two active agents in extended AMCM simulation which are the Malware and Incident Responder agents.

```
Malware_Agent () {  
  For each Agent {  
    If status is 'Infected' {  
      Find next nearest 'Susceptible' Host Agent  
      If any host found {  
        Clone a copy of Malware to new Host Agent  
        Change status of that Host Agent to 'Infected' Status  
      }  
    }  
  }  
}
```

```
        Contribute to the output of simulated world condition
        // Or Data Leakage in this research
    }
}
}

Incident_Responder_Agent() {
    For each Agent {
        If Priorization setting is 'OFF' {
            Find nearest 'Infected' Host Agent
        }
        Else if 'ON' {
            Find 'Infected' Host Agent with highest importance
        }
        If any 'Infected' Host Agent found {
            Remove Malware Agent from Host Agent
            Change 'Infected' Host Agent to 'Recovered' status
        }
    }
}
```

Figure 1 – Pseudo Code of extended AMCM At Each Simulation Cycle

There are assumptions incorporated into the original AMCM model. One key assumption is that the lead time, which is a global variable parameter, is sufficient to identify the required remediation solution to contain Malware and to turn an 'Infected' host to 'Recovered' status. The Incident Responders are only deployed after the remediation solution identification lead time has lapsed. The effectiveness of the remediation is not considered in this model. An assumption that the remediation solution works with the Malware is included. Another assumption is travelling latency between hosts for the incident responder is negligible and not included into the model though the model can be modified to include it.

The output of the original model is the measured simulation duration in terms of time ticks used in Netlogo and the number of hosts with their infectious statuses. The proposition is to extend the model to include another output variable that is the measure of data leakage from infected hosts. This is illustrated in the diagram below.

Figure 2 – Agent Based Malware Containment Model (AMCM).

The original validation done with AMCM was done using model alignment that involved aligning the AMCM's output to match the output of the mature epidemiological model using available published information about the Malware spread characteristics. The mature model used was SIR and the Malware spread characteristic used was the infection rate of the notorious Conficker Malware.

As with most epidemiological models, the rate of infection and disinfection is included in the model. Each infected host contributes to the extent of data leakage. The rate of data leakage is a percentage of the maximum output an infected host can generate onto the network. The following formulae define the epidemiological characteristics of the extended AMCM model with the data leakage attribute.

$$\frac{ds}{dt} = -\beta s(t)i(t) \quad (1)$$

$$\frac{di}{dt} = \beta s(t)i(t) - ki(t) \quad (2)$$

$$\frac{dr}{dt} = ki(t) \quad (3)$$

$$\frac{dz}{dt} = \xi i(t) \quad (4)$$

β = Infection parameter or rate of infection

k = Disinfection parameter or rate of disinfection

ξ = The measure of data leakage per infected host

Data leakage in this research is only computed for infected hosts. The unit measurement of data leakage is not in actual bytes or megabytes per second but a representative value of significant volume of data leaked per unit time. In this research, the higher the value of data leakage, the more

significant it is to the organization. Insignificant data from hosts will not contribute to the measure of data leaked.

There are new assumptions included into the extended model. Firstly, duplicity of the data is not considered. Hence the same data may be leaked into the Internet more than once hence the extent of data loss may be lower. Secondly, the connectivity to the Internet from the Infected hosts will always exist.

4 Related Work

There are very few models that have considered data leakage as an output measurement. One such is a model (Papadimitriou and Garcia-Monlina, 2010) that computed the probability of data leakage by trusted data recipients with sensitive data from data distributor. Another model (Clark, Hunt and Malacaria, 2002) used basic information theory to analyse the confidential information that may be leaked by programs written in a simple imperative language. However there is none that models the effects of data leakage from Malware intrusion with containment efforts.

Most simulation models were developed to simulate the behaviour of Malware. Garetto et al. (Garetto, Gong and Towsley, 2003) used interactive Markov chains to model the propagation behaviour of Malware on a network, Staniford et al. (Staniford, Paxson and Weaver, 2002), used empirical data derived from the outbreak of the CodeRed worm to model its propagation, Chen et al. (Chen, Gao and Kwiat, 2003) used mathematically based discrete time model called Analytical Active Worm Propagation (AAWP) to characterize worm propagation. Many models used epidemiology models to simulate Malware. Vlachos et al. (Vlachos, Vouzi, Chatziantoniou and Spinellis, 2005) used network graph to model computer epidemiology. Zou et al. (Zou, Gong, Towsley and Gao, 2005) used simple epidemic model and exponential models to analyse their monitoring and early detection system that studies trend of malicious activities. Kim et al. (Kim, Park, Lee, Choi and Seo, 2010) extended the basic Susceptible-Infected-Recovered (SIR) Model to support the analysis of the dynamics of incubation period of epidemic Malware code. These models use the number of infected hosts or the percentage of the host population that is infected against time to study the epidemiological effects of Malware outbreak. However none considered data leakage as a measurement. Hence the proposition in this paper is to use the measure of data leakage to assess the output of the simulation.

5 Methodology

The following details how the quantitative measure of data leakage can be used to assess the effectiveness of containment plans of Malware infected computing hosts in simulation modelling.

A. Evaluation Approach

The evaluation entails two stages of tests. The first stage is to validate the proposed use of data leakage measurement to quantitatively assess containment scenarios. The tests involve the study of the characteristics of the measure of data leakage of infected hosts against other existing measurement. If the characteristic of the measure of data leakage mimics the characteristics of other proven measure of Malware containment that is the number of infected hosts, then the model is ready for the next stage of tests.

The second stage of tests is to conduct two scenarios to measure the effectiveness of certain containment strategy using data leakage measurement. Each scenario will have five tests. The first scenario involves no prioritization hence incident responders will attend to the nearest infected host regardless of the extent of data leakage the remaining infected hosts contributes. This involves setting the parameter of “Find Next Host based on Prioritization” parameter of extended AMCM to “Off”. The next scenario test involves prioritizing incident responders to attend to infected hosts that contribute more significantly to data leakage. In order to activate this condition, the parameter of “Find Next Host based on Prioritization” in AMCM is set “On”. Finally, the results of the two scenarios of tests in this stage are compared.

B. Configuration of the Simulation Model

The following describes the parameter configuration used in this research with the extended AMCM model in order to simulate the Malware outbreak and containment effort. The model has been extended to support the measure of data leakage. The model, as with the original one, has been configured to mimic the infection characteristics of the notorious Conficker Malware that infected millions of computers in the Internet since Nov 2008. The infection rate for Conficker was 4% according to a technical report by SRI International for Office of Naval Research, USA (Porrás, Saidi and Yegneswaran, 2010). The measure of data leakage is in megabytes of important data per unit time. Unit time is

measured in ticks with Netlogo. The extent of data leakage for an infected host is computed to be 2 megabyte per unit time (20% of 10 Megabytes per unit time). The following are the global constants and variables.

Parameter	Default Value
Population size of hosts	3721
Random Seed	0
Lead Time for incident responders to identify remediation	400 time steps
Infection Rate	0.04 (or 4%)
Data transmission rate per host	10 Megabytes per unit time
Data Leakage rate per host	20%

Table 1 – Global Constants

Parameter	Initial Value
Number of Susceptible hosts	3720
Number of Infected hosts	1
Number of Recovered hosts	0

Table 2 – Global Variables

At the start of simulation scenario, one of a randomly selected host will be tagged as the first infected host. Subsequently, more susceptible hosts will be infected based on the infection rate of Conficker. A lead-time, which is a constant parameter, is incorporated into the model as the time needed by the incident responders to identify an effective eradication approach to the Malware involved. The assumption here is that the Malware eradication approach can be acquired by that lead-time. Incident response only starts after the lead-time. When an Incident Response agent attends to an infected host, it will turn the ‘Infected’ host to ‘Recovered’ status. Recovered host do not get re-infected by the same Malware.

Parameter	Range Values
Number of incident responders	Integer (0 to 100) – Set to 20
Prioritization Assignment	Boolean (On or Off) – Set to ‘On’
Number of Prioritization Levels	Integer (Max of ‘3’) – Set to ‘3’
Find Next Host based on Prioritization	Boolean (On or Off)

Table 3 – Scenario Variables

Prioritization Assignment is set ‘On’ in all tests conducted in this research where each and every host will be assigned a random zero based value (between values of 0 and 2) for its prioritization value assignment to each and every host. The higher the number of the value, the host has access to more important data. If that host gets infected, it contributes

more to the extent of data leakage. If that host is not infected or is disinfected, it does not contribute to data leakage. With the model set to 3 levels of prioritization, the least important hosts contribute zero amount of importance to the measure of data leakage. The moderately important infected hosts contribute the 2 megabytes per unit time. The most important infected hosts contribute twice the value of the moderately important hosts.

The 'Find Next Host based on Prioritization' parameter is used to control the incident responder agents in their selection of hosts to respond to. When it is set to 'On', incident responder agent will attend first to the highest prioritized host nearest to it before others. When it is set to 'Off', the incident responder attends to the nearest computing assets to relative to its current position regardless of the prioritization.

6 Analysis of Results

This section analyses the results of the staged tests outlined in the previous section. The first set of tests validated the extent of data leakage against the duration of the AMCM. The second set of tests involved conducting two test scenarios with the intention of validating the premise that emphasis to deal first with infected hosts with more important data results in lower data leakage outcome.

A. Stage 1 – Validation Test

The test objective for this stage is to assess whether the data leakage measurement matches closely to the original output of the AMCM model that is total duration to disinfect all infected hosts disinfected. As there are random characteristics included in the model, like the assignment of prioritization value to each and every host, five tests were carried out. The mean values were then computed from the tests.

Figure 3 – Epidemiological Characteristics of Containment Effort Against Malware Outbreak

Figure 4 – Rate of Data Leakage

This test shows that the characteristic of data leakage follows the behaviour of other form of measurement used in AMCM that is the number of infected hosts that has yet to be disinfected. It is noted that as more 'Susceptible' hosts become 'Infected', the rate of data leakage rises. As 'Infected' hosts become 'Recovered' or is disinfected, the rate of data leakage declines at the same proportional rate. Under both conditions, the rates of increase and decrease of data leakage are proportional to rates of increase and decrease in the number of infected hosts. Also when the all

‘Infected’ hosts were dealt with, or when the Malware is completely eradicated, the rate of data leakage drops to zero. Pearson product-moment correlation coefficient was computed to assess the relationship between both data sets. There was a positive correlation between the two variables with $r = 1.000$, $n = 876$ with $p < 0.05$. Hence this stage of tests validated the mathematical equation for data leakage (4) and that it can be used to carry out further modelling scenarios with data leakage as its measurement.

B. Stage 2 – Two Scenario Tests

This stage involves two scenarios of tests. The following table shows the data leakage measurements over five tests when incident responders respond to infected hosts based on distance proximity between incident responders and infected hosts. It does not consider the measure of data leakage that each remaining infected host contributes. Such condition will be largely applicable when any incident response triage excludes the effects of data leakage by individual involved computing hosts.

	Test 1	Test 2	Test 3	Test 4	Test 5	Average
Duration (ticks)	868	861	876	867	859	866
Total Data Leakage	3,269,670	3,200,364	3,275,098	3,353,796	3,160,364	3,251,858

Table 4 – Stage One Measurements (5 tests)

The second scenario set of tests involves having the incident responders respond first to the highest prioritized hosts before proceeding to the next highest. Higher prioritization assignment was to the hosts that contribute more to the quantitative measure of the risks of data leakage when infected. In this simulation test condition, the incident response triage is to disinfect the higher prioritized hosts before attending to the lesser prioritized hosts. The expected effects of such a triage decision is that incident responders will go to the higher prioritized hosts even when there are infected hosts nearer to the incident responders requiring less travelling time (or simulation tick step). Such condition may occur when the organization intends to deal first with hosts that pose a greater risk to the organization in terms of data leakage. As with the previous stage of testing, five tests were carried out and the mean result values were taken. The following is the output for this stage of testing.

Test 1	Test 2	Test 3	Test 4	Test 5	Average

Data Leakage Triage: Quantitative Measure of Effectiveness of Containment Plan

Duration (ticks)	1,146	1,125	1,120	1,144	1,139	1,135
Total Data Loss	3,110,650	2,997,188	3,097,624	3,174,682	3,080,480	3,092,125

Table 5 – Stage Two Measurements (5 Tests)

Figure 5 – Epidemiological Characteristics of Conficker Malware Outbreak With Prioritized Containment Response

It is noted from this stage of two scenario tests that the later scenario in which the incident responders disinfect the higher prioritized hosts first, the full extent of the measure of data leakage after the event is lower than the former scenario. The following chart compares the extent of data leaked between non-prioritized and prioritized incident response plans.

Figure 6 – Comparison of Extent of Data Leakage Between Prioritized and Non Prioritized Strategies

he difference of data leakage averaged around 5% between prioritized and non-prioritized containment strategy. It is also noted that it took relatively longer time to contain a Malware outbreak when prioritization was applied. It took 30% more time than non-prioritized. This is because the incident responder agents focused on the higher priority hosts before attending to the others even when the other infected hosts were nearer to the agent. The longer duration can be reduced if an optimal placement of hosts to support prioritized containment is in place.

7 Future Research

The extended AMCM model with data leakage measurement can be used to define optimal incident response plans to minimize the impact caused by Malware outbreak. There are many research opportunities that can be extended from this research work. Firstly, regarding the way the simulation model is used, aside from the prioritization containment strategy, other reactionary strategies may be evaluated. Also besides adjusting the reactionary strategies, proactive planning of the placements of critical data sources may be evaluated to assess possible reduction remediation duration through centralization of storage of critical data resulting in reduction of traveling time required of incident responders. This could affect the way network designs are developed, that is to have an optimal network design to improve Malware related or cyber security incident response.

Model adjustments can be done as part of future research development. The AMCM model with data leakage measurement can be extended to include other forms of data loss. This includes data destruction that

implies that the data in the infected hosts are destroyed hence not recoverable by the incident responders. Data destruction may occur with certain types of Malware that detects attempt to contain it or is instructed by its Command and Control entity to unilaterally destroy the data that Malware has hostage of. Also the AMCM's output measurement dimensions can be further extended beyond the number of infected hosts, duration of disinfection, measurement of data leakage to include computation of human resource involvement required to contain an outbreak. Another future research consideration is to enhance the AMCM simulation model to closer mimic the environmental conditions of a Malware outbreak like the environmental limits to the extent of data leakage due to the egress bandwidth limit to the Internet or from the Infected hosts. Other forms of 'what-if' test conditions may be codified and applied to the model to test game-changing strategies, for example if remote administration of remediation on the infected hosts can be done hence removing the need to compute physical traveling duration of incident responders. Another direction is to simulate reactive Malware when containment is being applied hence extending the possibility of incorporate game theory into the simulation.

8 Conclusion

This reported research illustrates the possible use of the measure of data leakage in assessing the Malware containment strategies. Based on the simulation experiments, it is noted that the strategy to focus on more important computing assets will result in lower extent of data leakage despite longer containment duration. Hence the measure of data leakage can be used to facilitate the planning of a reactive containment plan during a Malware outbreak or pre-incident response preparation planning. Data leakage also allows the security personnel to assess the containment strategies from a different perspective, beyond the measure of time and number of infected hosts to deal with, and allows the consideration of alternative approaches to the containment plan. Finally, this model extension of AMCM opens opportunities to further research in the development of containment strategies and approaches. This proposed model is relatively simple and it is believed that it can contribute significantly to improve decision-making and definitions of response plans to deal with the threat of Malware and data leakage risks.

References

- Boase J., Wellman B.: A Plague of Viruses: Biological, Computer and Marketing. *Current Sociology*, vol. 49, no. 6, 39-55 (2001)
- Chen Z., Gao L., Kwiat K.: Modeling the Spread of Active Worms. In: INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies, Vol 3, Pg 1890 (2003)
- Clark D., Hunt S., Malacaria P., "Quantitative Analysis of the Leakage of Confidential Data", *Electronic Notes in Theoretical Computer Science*, Vol 59, Issue 3, Pg 238-251, 2002.
- Garetto M., Gong W., Towsley D.: Modeling Malware Spreading Dynamics. In: INFOCOM 2003, Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies, Vol 3, Pg 1869-1879 (2003)
- Kermack W. O., McKendrick A. G.: A Contribution to the Mathematical Theory of Epidemics. *Proceedings of the Royal Society of London. Series A*, Vol. 115, Pg 700–721 (1927)
- Killcrece G.: Incident Management. Software Engineering Institute, Carnegie Mellon University & Department of Homeland Security National Cyber Security Division (2013). Available via Internet. <https://buildsecurityin.us-cert.gov/articles/best-practices/incident-management/incident-management>. Online. [Accessed Jun. 7, 2014].
- Kim S.W., Park J. H., Lee E D., Choi M. E. and Seo S. W., "Threat Analysis of Incubation Period in Malware Epidemics", IEEE 71st Vehicular Technology Conference (VTC'10), Taipei, Taiwan (2010)
- Kossakowski K., "Dynamics of Incident Response", 17th Annual FIRST Conference on Computer Security Incident Handling, FIRST.Org (2005)
- Kotadia M., "Eighty percent of new malware defeats antivirus", ZDNet Australia, 2006. Available via Internet. <http://www.zdnet.com.au/eighty-percent-of-new-malware-defeats-antivirus-139263949.htm>. Online. [Accessed Jun. 7, 2014].
- Messmer E., "Sony creates CISO position to clean up mess after PSN hack", *Computer World UK*, May 3, 2011. [Online]. Available: <http://www.computerworlduk.com/in-depth/careers/3277005/sony-creates-ciso-position-to-clean-up-mess-after-psn-hack/>. Online. [Accessed Jun. 7, 2014].
- NetLogo. <http://ccl.northwestern.edu/netlogo/>.
- Pan J., Fung C.C., "An agent-based model to simulate coordinated response to malware outbreak within an organization", *International Journal of Information and Computer Security*, 5 (2). Pg 115-131, 2012.
- Pan J. and Fung C.C., "Artificial Intelligence in Malware – Cop or Culprit?", The Ninth Postgraduate Electrical Engineering & Computing Symposium PEECS 2008, the University of Western Australia, Perth, Australia, 2008.
- Pan J., Fung C. C., Koh W. W., "Devious Chatbots – Interactive Malware with a Plot", In: FIRA 2009, CCIS 44, pp. 110–118 (2009)
- Pan J., Fung C. C., "Boutique Malware – Custom made for e-business", In: The 9th International Conference on e-Business iNCEB, Bangkok, Thailand (2010).
- Paradimitriou P., Garcia-Monlina H., "Data Leakage Detection", *IEEE Transactions on Knowledge and Data Engineering*, Vol 23, Issue 1, Pg 51-63, 2010.
- Porras P., Saidi H. and Yegneswaran V., "Malware Pandemics", Technical Report 1, SRI International Menlo Park CA Computer Science Lab, Sep. 1, 2010.
- Shipp A.: MessageLabs Intelligence Special Report: Targeted Attacks April 2007, MessageLabs (2007)

Data Leakage Triage: Quantitative Measure of Effectiveness of Containment Plan

- Staniford S., Paxson V., Weaver N.: How to Own the Internet in Your Spare Time. In: Proceedings of the 11th USENIX Security Symposium (2002)
- Verizon, “Data Breach Investigations Report”, Verizon Business, (2010). Available via Internet. <http://www.verizonbusiness.com/go/2010databreachreport/>. Online. [Accessed Jun. 7, 2014].
- Vlachos V., Vouzi V., Chatziantoniou D., Spinellis D.: NGCE – Network Graphs for Computer Epidemiologists. In: Advances in Informatics: 10th Panhellenic Conference on Informatics, PCI 2005, Pg 672–683, Berlin (2005)
- Wiik J., Gonzalez J. J. and Kossakowski K., “Effectiveness of Proactive CSIRT Services”, 18th Annual First Conference, Baltimore, Maryland USA, 2006.
- Zou C. C., Gong W., Towsley D., Gao L.: The monitoring and early detection of internet worms. IEEE/ACM Transactions on Networking, Vol 13, Issue 5 (2005)