

Master Thesis
Computer Science
Thesis no: MCS-2009-13
Month Year: June, 2009



Modeling the Spread of Malware in Computer Networks

Patlolla Pradeep Reddy

Pasam Raghava Reddy

School of Computing
Blekinge Institute of Technology
Soft Center
SE - 372 25 Ronneby
SWEDEN

This thesis is submitted to the Department of Computer Science, School of Computing at Blekinge Institute of Technology in partial fulfillment of the requirements for the degree of Master of Science in Computer Science. The thesis is equivalent to 20 weeks of full time studies.

Contact Information:

Author(s):

Patlolla Pradeep Reddy,

E-mail: ppradeep88@gmail.com

Pasam Raghava Reddy

E-mail: praghavreddy77@gmail.com

University advisor(s):

Stefan J.Johansson

E-Mail: sja@bth.se

Blekinge Institute of Technology

SE – 371 79, Karlskrona, Sweden

Phone: +46 455 38 50 00/ Fax: +46 455 38 50 57

School of Computing
Blekinge Institute of Technology
Soft Center
SE - 372 25 Ronneby
SWEDEN

Internet : www.bth.se/tek
Phone : +46 457 38 50 00
Fax : + 46 457 102 45

ABSTRACT

Malware in the form of viruses, rootkits, trojans and email worms are a menace to personal computers as well as corporate networks. They not only cause obstruction to work, but significantly reduce productivity. Understanding how these malware spread and their impact on costs is important while designing prevention mechanisms for organizations. We have created a model to simulate the spread of malware in Wi-Fi networks and have used the simulator to understand the impact of various parameters which come into play during a viral attack.

Our experiments with parameters such as replication rate of email worms, mail checking frequency, effect of firewalls have led to insightful and non-intuitive results. Most importantly, the cost analysis has shown how beyond a point having expensive anti-virus systems might not make economical sense

Keywords: Malware, Spread Model, Wi-Fi Network, Cost Analysis.

ACKNOWLEDGEMENT

We would like to thank Mr. Stefan J.Johansson and Mr. Guohua Bai for their contribution in one or another ways to our thesis.

We would also like to take this opportunity to thank our parents and friends, for their blessings and support without which we could not have achieved this.

TABLE OF CONTENTS

Abstract

Acknowledgement

Table of Contents

Introduction.	7
Chapter 1: Literature Survey.	8
1.1. Information Security and viruses.	8
1.2. Virus Propagation.	8
1.3. Malware in Mobile Networks.	8
1.4. Malware Spread Models.	9
1.5. Economics of Anti-Malware Software.	10
Chapter 2: Problem Definition/Goals.	11
2.1. Challenge/Problem Focus.	11
2.2. Goals/Results.	11
2.3. Research Questions.	11
2.4. Hypothesis.	12
Chapter 3: Methodology.	13
Chapter 4: Model for Spread of E-mail Worms in a Network of Computers.	15
4.1. System Description.	15
4.2. Assumptions.	16
4.3. States in the System.	16
Chapter 5: Model Description.	17
5.1. Initialization.	17
5.2. Continuation.	17
5.3. Model of Estimation of Losses During a Virus Attack in a Network. . .	18
5.4. Additional Cost.	19
5.5. Cost Benefit Analysis for Anti-Virus.	19
5.6. Simulator.	19

Chapter 6: Results.	21
6.1. Impact of % of Terminals Fire-Walled.	21
6.2. Impact of Frequency of Checking E-mail.	22
6.3. Impact of Standard Deviation in Frequency of Checking E-mail.	24
6.4. Impact of Replication Rate of Malware.	25
6.5. Impact of Rate of Death on the System.	26
6.6. Impact of Dormant System's Efficiency.	26
Chapter 7: Cost Benefit Analysis of the Firewalls.	28
Chapter 8: Discussion.	30
8.1. Impact of various factors on the spread of malware and losses.	30
8.2. Monetary Losses due to spread of malware and cost benefit analysis.	31
Conclusions.	32
Reference.	33

INTRODUCTION

Computer systems and networks infested with malware are commonplace today and are a menace to workplace efficiency. “Malware is a broad term including computer viruses, worms, Trojan horses, rootkits, spyware, adware and other forms of unwanted software which intend to harm the efficiency of systems” [2]. Today malware are increasingly designed to impact networks of computers and propagate through networks to inflict maximum damage to the network of computers. Therefore one computer infested with malware implies that within a short span of time, the entire network can get affected, increasing the downtime. Every minute of downtime in turn means a loss in terms of revenue and efficiency. In other cases these could lead to severe crisis such as security concerns. Hence, understanding how malware spreads and infests networks is very important.

“Malware attacks against computer networks constitute a growing area of research in order to prepare network administrators to prevent potentially disastrous attacks” [6]. Park et al. (2007) state that “email worms not only affect the productivity leading to loss of time and money, but also affect intangible assets of companies such as brand and customer loyalty” [11].

Our research is an exploratory study on how various parameters in the attack, ranging from that of the worm (replication rate), to those of the network (number of nodes, % fire-walled computers) as well as user behaviour (frequency of checking mail) impact the spread of malware. Through the development of a simulator we have created various experiments and have studied the impact of all possible parameters.

CHAPTER 1: LITERATURE SURVEY

1.1. Information Security and Viruses

Information security deals with the integrity, confidentiality and availability (service disruption) aspects of a network. “Viral attacks vary in their effects. While some attacks harm the integrity of information, others disclose confidential information. Some other viruses affect the system availability” [7]. The denial of service (DoS) attacks degrades system availability.

“There have been various viruses which have caused havoc in the information technology world. For example, the Slammer worm, also known as Sapphire and SQL Hell paralyzed a number of hosts immediately after it was released” [9]. Its speed to infect the host as well as reproduce is extraordinary thus producing massive levels of network traffic as it scans random IP addresses looking for other vulnerable SQL servers. “Rootkits are used for the purpose of system administration or to protect licensed systems. However, some rootkits are used by hackers to hide or protect malicious codes” [10].

Symantec, Klez.A and Kama Sutra worms cause damage to file systems of compromised systems. According to a PWC report, almost 60% of all companies in the United Kingdom have faced different kinds of information security breaches in 2006, almost 50% of which were caused by malware, i.e. viruses, worms and Trojans. Therefore, there is a need to understand what security systems should be used, and how much of it must be used to protect our networks.

1.2. Virus Propagation

One major way in which viruses propagate is through mail user-agents, such as Netscape Mail, Microsoft Outlook Express, and Eudora Mail. Systems that run or host mail user-agents are exploited by a variety of malicious attacks. Some worms are able to replicate themselves through vulnerable nodes after they are introduced into a system. In general, there are hostile codes which are non-self replicating while others replicate themselves. By virtue of their mobility, viruses and worms can cause a vast number of incidents.

1.3. Malware in Mobile Networks

Self-propagating malware is well understood in the Internet however malware spread in mobile phone networks are increasingly become common place. “These have very different characteristics in terms of topologies, services, provisioning and capacity, devices, and communication patterns” [15]. Mobile phones are therefore the next frontier for malware. The potential effects of virulent malware propagation on consumers and mobile phone providers are severe, including excessive charges to customers, deterioration of mobile phone services, public relations disasters, and ultimately loss of revenue for mobile phone providers.

1.4. Malware Spread Models

Several models exist in the literature which analyze spread of malware in computer networks. Some of the major ones are discussed below:

Stochastic Epidemic Model:

Siewiorek and Swarz [15] have developed a model based on a finite-Markovian process. In these epidemiology based models, a node can have two states – one in which a node is susceptible to become infected and the other in which it becomes “infected forever”. The authors have also considered the fact that an “unhealthy node” can actually recover and become healthy again. This makes the model very realistic and resembles how an unhealthy human being can recover again. The model presented by the authors is useful in determining state transition dynamics for estimating infection and recovery rates of susceptible systems. The equations below are a common setup used:

$$\begin{aligned}\frac{d}{dt}S(t) &= -\beta S(t)I(t), \\ \frac{d}{dt}I(t) &= \beta S(t)I(t) - \gamma I(t), \\ \frac{d}{dt}R(t) &= \gamma I(t),\end{aligned}$$

Here the differential equations for $S(t)$, $I(t)$ and $R(t)$ represent the “continuous time functions of susceptible, infected, and recovered systems.” The authors state that Markov models can be used to analyze “time-dependent reliability of systems.”

The setup used in such models is shown in the figure 1 below:

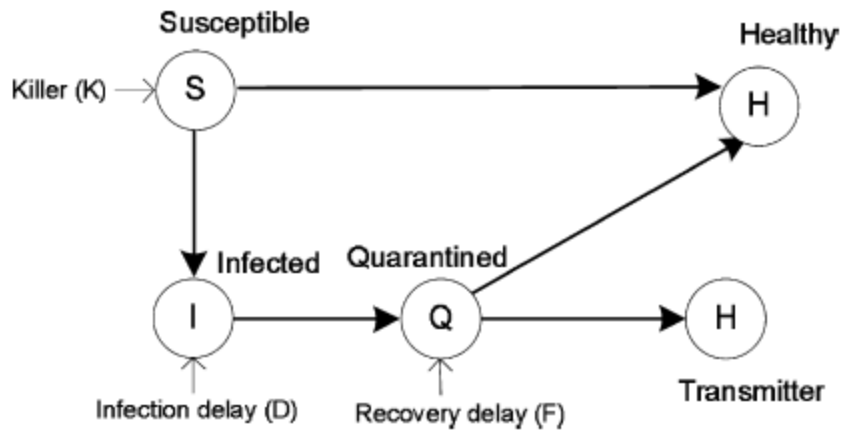


Fig. 1: “A Belief Network (graph) representation of the virus infection model for the computation of prior distribution” [7].

As shown in the schematic represented in Fig. 1, any node can be either of the following: susceptible (S), infected (I), quarantined (Q), healthy (H) or transmitter (T).

A susceptible node is one that can detect the viruses it is vulnerable to. As soon as the susceptible node catches a given virus, it is exposed and enters as the authors have named, the “latent period.” Note that the nodes in the latent period are infected but are not infectious yet. Following the latent period, the susceptible node becomes an “active transmitter of the virus it had caught.” Hence, the transmitter node further distributes the virus to the user-agent addresses in the address book of the transmitter. There is an additional time period during which a transmitter node could become healthy (H) if the required virus removal mechanism is available. The removal of the viruses is done using the quarantine (Q) process.

Genetic Algorithm Based Model

Goranin and Čenys [5] have proposed a genetic algorithm based model for estimating the propagation rates of known and perspective Internet worms after their propagation reaches the satiation phase. Estimation algorithm is based on the known worms’ propagation strategies with correlated propagation rates analysis and is presented as a decision tree.

Yua et al. [14] have developed a malware propagation model for P2P networks. As the surge of peer-to-peer (P2P) systems continues with large numbers of users and rich connectivity, P2P systems are becoming a potential vehicle for the attacker to achieve rapid worm propagation in the Internet.

The authors argue that in general, there are two stages in an active worm attack: (1) scanning the network to select victim hosts; (2) infecting the victim after discovering its vulnerability. Infected hosts further propagate the worm to other vulnerable victims and so on. The three key factors that decide worm propagation speed are

- (1) how fast the worm can scan other hosts in the network
- (2) the probability of the worm to scan a real host; and
- (3) Vulnerability of the scanned host.

The first factor has been modeled as the scan rate S , which is the number of hosts per unit time that a worm infected host can scan. The second and third factors are victim independent. The third factor, namely vulnerability of victim hosts is quite high in the case of P2P systems as most P2P hosts are un-trusted and un-validated during the entry into the P2P system. The authors have concluded that P2P size, topology degree, host vulnerability, etc. have important impacts on attack effects. They observe that attack effects are more pronounced in the case of unstructured P2P systems compared to structured P2P systems.

1.5. Economics of Anti-Malware Software

Lelarge [8] show the economic considerations while investing in anti malware software on computer networks. Negative externalities exist if the anti-malware is strong leading to the free-rider problem. This means that everyone waits for the others to invest in anti-malware technologies and hence nobody eventually invests. Positive externalities exist if the malware technologies are weak. The authors hypothesize that anti-malware manufacturers could deliberately keep their products weaker to create positive externalities in the market.

CHAPTER 2: PROBLEM DEFINITION GOALS

2.1. Challenge/Problem Focus

Modeling of malware spread through simulations will be important in understanding how systems could be configured in order to prevent malware spread or at least to minimize the speed of propagation of malware to the extent possible. Using the simulator developed based on the model developed, we will present different kinds of networks and how they rank in terms of vulnerability to malware propagation. This analysis should be useful for organizations making decisions on network infrastructure and design and can help them create network designs which are optimally configured.

2.2. Goals/Results

The results that are to be extracted from the simulation are:

1. Studying the Speed of spread, i.e., rate of computers being infected (infected computers/day),
2. Estimating the Total downtime (computer days) for the entire duration of the attack,
3. Calculating the % Loss in efficiency taking into account downtime as well as CPU usage by the anti malware program activity due to the attack.

Modelling monetary loss

A second simplistic model will be developed to connect the system downtime to monetary loss. For this each computer will be allocated a productivity rate, i.e., how is the revenue of the organization dependant on each computer (Euros/system hour). Using this rate and the downtime results from the simulator we will be able to estimate the monetary loss.

The monetary loss estimates will provide insights on two fronts: a. which computers should be more protected than the others, and b. cost benefit analysis of installing anti-malware software on each computer. This should provide organizations valuable insight on how much to spend on protecting systems and which systems to protect first.

2.3. Research Questions

Based on the results and the analysis, following information should potentially be gathered from the study:

1. How does factors such as email checking frequency, installation of anti-malware on a % of systems, worm reproduction rate impact speed of propagation of a malware worm?

2. How does worm propagation result in monetary losses? What factors determine the extent of monetary losses?
3. What are the benefits vs. costs of installing anti-malware in computer systems and how can we prioritize installation?

2.4. Hypothesis

There are a few hypothesis which we would prove disprove through our work and model for malware propagation. These are listed below:

1. The email checking frequency should intuitively be positively correlated with the losses that result, because this facilitates the spread of malware from one computer into the other
2. Worm reproduction rate should also be directly proportional to the losses inflicted by the malware because a higher reproduction rate implies a faster rate of infection spread in the system
3. A straight forward hypothesis is that the extent of losses will be inversely proportional to the % of terminals fire-walled as they kill the virus spread
4. There should be some cost of anti-malware software, for which the incremental monetary benefit brought by an additional firewall is less than the incremental cost of installing the software. Therefore a cost-benefit analysis should be possible.

We attempt to prove or disprove these hypotheses in the work and analysis in subsequent sections of this report.

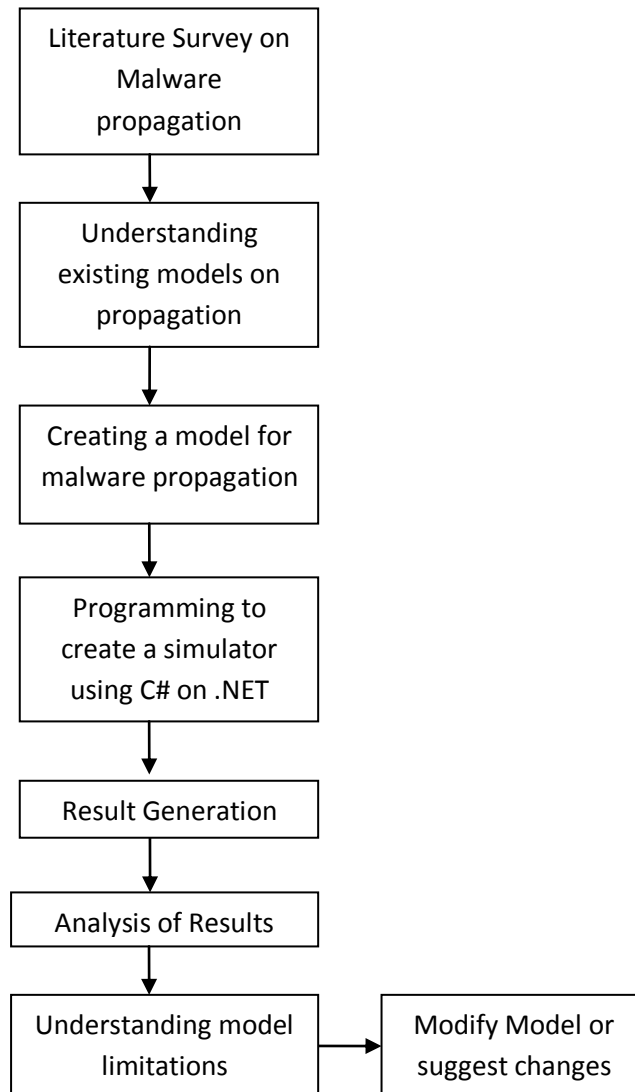
CHAPTER 3: METHODOLOGY

We look at the propagation malware in different forms of computer networks. We attempt to model the spread of malware in computer networks to understand the speed of spread as a function of the malware type, number of computers in the network and the number of nodes in the network. For our specific case we intend to use self propagating worms with payloads to understand malware propagation.

Overall Research Approach

The overall research approach would be quantitative rather than qualitative because modeling monetary losses in qualitative terms would be meaningless. However, the impact of various parameters on the spread of malware will be discussed qualitatively based on the simulator results seen.

The work flow of methodology followed is shown below:



We will define a network of N computers with i nodes, each node being allocated the same number of computers and connected in different configurations (ex. In series, in a closed ring configuration and so on).

Parameters to be defined for computers in the network will include: email checking frequency, installation of malware (yes/no), mailbox configuration. The single parameter we would use for the worm is the reproduction rate. This is because there are worms that propagate as quickly as possible to inflict maximum damage. On the other hand there are slow reproducing worms which avoid detection and can be even more dangerous.

We also assume that computers without protection through anti-malware software suffer “death” or shutdown because of the attack.

We attempt to model the propagation through a simple sequence of steps:

1. Introduce malware in one of the computer system (assume comes in through an external email link)
2. The worm randomly scans all computers in the node and probabilistically chooses the one computer which is most vulnerable. The probability is calculated using the parameters defined above. We can make the choice probabilistically random by comparing a random number generated to the probability of attack on each computer in the vicinity.
3. Based on the choice made in step 2, the worm propagates, reproduces into another live worm and again spreads to another computer in the network.
4. The propagation stops when there are no more unaffected computers left
5. A protected computer (having anti-malware software installed) is continuously bombarded with the worms causing no direct damage but causing processor time and RAM usage.

CHAPTER 4: MODEL FOR SPREAD OF E-MAIL WORMS IN A NETWORK OF COMPUTERS

4.1. System Description

In line with the infrastructure being used in modern corporations, we have assumed a modern office setup with all computers connected with each other through a Wi-Fi internet network. This means that each computer can interact with another computer without issues. Hence in our case, nodes and terminals are used interchangeably. Taking the example of a medium – large office, as a base case we assume 1000 terminals in the network. All of the computers use common email clients such as Microsoft Outlook and access and send emails regularly.

We assume that a certain % of computers are fire-walled and that the fire-walled systems are immune to any virus attack.

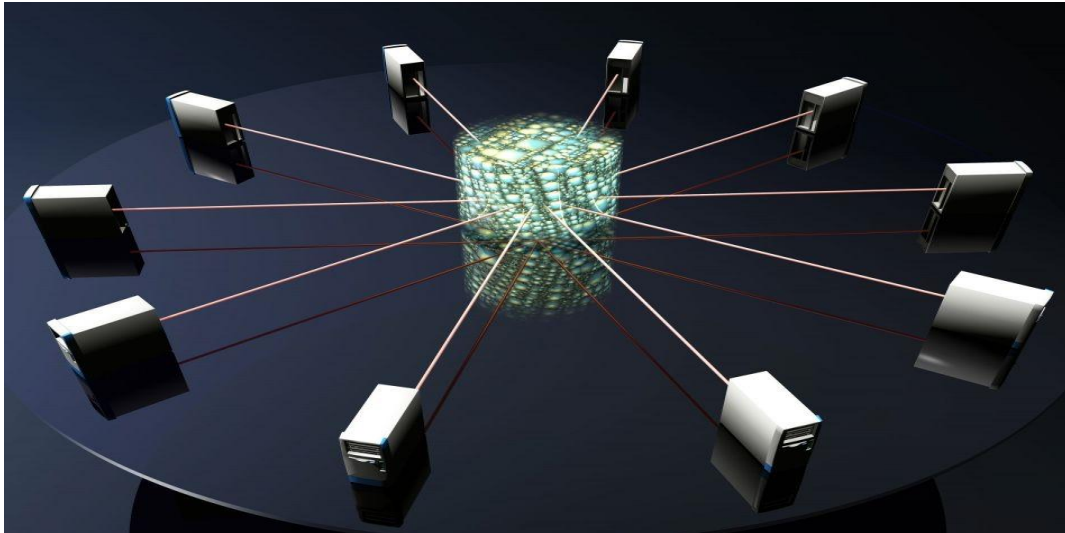


Fig 2: Schematic representation of a network in which each computer can communicate with another directly via a central hub.

The parameters used are described below:

- N = Total Number of nodes/terminals in the system
- F = % of nodes/terminals fire-walled
- μ_f = mean frequency of checking emails (expressed in time steps)
- SD_f = Standard deviation in frequency of checking mails (expressed in time steps)
- μ_e = Mean number of people emailed node or terminal (expressed as %)
- SD_e = Standard Deviation of number of people emailed (expressed as %)
- R_d = Rate at which affected computers die (expressed in time steps)
- Loss (dormant) = % loss in efficiency for dormant nodes
- Loss (active) = % loss in efficiency for active nodes
- I = % cost of installing firewall per node
- L = Total cumulative loss

4.2. Assumptions

1. All computers are used to check emails using common clients like Microsoft Outlook and the internet based mail services (Ex. Yahoo Mail)
2. People interact with each other using emails amongst other forms of communication – this leads to the spread of email worms
3. The usage of email (checking) has an average and standard deviation which has already been measured
5. On any given day or instant, a user can access emails any number of times, i.e., a random number which is drawn from a normal distribution curve with the mean and standard deviation mentioned in assumption 3.
6. Users send emails to each another and we can measure the average and standard deviation of the number of emails one terminal on average sends
7. On any given day or instant, a user may send emails to any number of terminals, i.e., a random number which is drawn from a normal distribution curve with the mean and standard deviation mentioned in assumption 5.

4.3. States in the System

Every terminal under a viral attack can be in one of the following 4 states:

1. Clean: This terminal has not been infected by viruses nor does it have unopened emails which have viruses in them (C)
2. Dormant: This is the terminal which has a virus in its mailbox, however the virus containing email has not be accessed yet, hence the virus is dormant (Do)
3. Active: As soon as the email containing the virus is accessed, it becomes active and hence we label the terminal or node as active (A)
4. Dead: The virus-active terminal gradually loses its efficiency and dies, thereby becoming irrelevant and disconnected with the network

Note that when the virus is active it multiples via emails. Hence, when a user sends mails from his/her terminal (which is in active state), it affects all those terminals who receive the mails making them initially dormant and then active. Similarly, each active terminal dies eventually because of the viral attack.

CHAPTER 5: MODEL DESCRIPTION

The model has been described in the algorithm steps below. The terminals are represented in the form of an array. Hence, if the number of computers is 1000, the terminals will be represented in the form of array elements with each element containing its identity number (say, i) and its state (C, Do, A or D). We describe the steps in two parts, A: Initialization, and B: Continuation.

5.1. Initialization:

1. Accept user inputs for $N, F, \mu_f, SD_f, \mu_e, SD_e, R_d, Loss (dormant), Loss (active), I, L$
2. Create an array of N elements. Initialize all elements with the state C
3. Initialize Time step with 0
5. Randomly chose one array element and initiate the viral attack from here. Change state to Do
6. Calculate the number of time steps for which this terminal will be dormant, depending on the values of μ_f, SD_f . In order to do so, we chose a random number from $N(\mu_f, SD_f)$, i.e., we draw a random number from the normal distribution curve with mean and standard deviation of μ_f, SD_f respectively. This makes the system resemble an actual attack as we do not use a fixed frequency of checking mails but a random value determined from the normal distribution of the average mail checking frequency of various users in the system. Using this value we compute the time steps until which time the system remains dormant. Suppose this value is T_{rand} . Hence we wait until T_{rand} time steps (This is like a count down).
7. When the time steps from Step 4 are over, the state of the terminal is changed from Do to A , thereby making this an active spreader of viruses
8. We compute the number of terminals emailed by each cell again using a random number (as in reality this is not fixed). We draw this random number from $N(\mu_e, SD_e)$. Suppose this random number is N_{infect} . Hence, this terminal communicates (by sending emails) to N_{infect} terminals and makes them Dormant, Do .

5.2. Continuation

Increment Time step by 1 unit

1. Check for the condition: Are terminals either clean or dead?
IF Yes go to Step 3
ELSE Goto Step 6
2. Assess the entire array of elements (from $i=1$ to N)

2.1. If an array element is in the *Do* State:

$T_{rand, i} = T_{rand, i} - 1.$

IF $T_{rand} = 0$

 Compute N_{infect} .

 Select N_{infect} terminals randomly

 Check whether any of the selected terminals are fire-walled or Active or Dormant

 IF Yes: Do Nothing

 ELSE

 Change state of these from *C* to *Do*

 Update the count of Clean and Dormant terminals

 Calculate $T_{rand, i} = N(\mu_f, SD_f)$

 Assign the element *i* the state *Do* and T_{rand}

2.2. IF an array element is in the *A* State:

$T_{dead} = T_{dead} - 1$

IF $T_{dead} = 0$ Change State from *A* to *D*

 Update the number of terminals in State *A* and *D*

3. Print Time step, Number of Clean, Dormant, Active and Dead terminals

4. Go to Step 2

6. END

Therefore, the algorithm scans the array and continues to change their states from *C* to *Do* to *A* if there is no fire-wall installed, increments time steps and prints the state of the system, until the system has only dead and fire-walled systems.

By drawing graphs of the results one is able to visualize how the system would change with time. We have done this in our results and discussion section

5.3. Model for Estimation of Losses During a Virus Attack in a Network

The cost model is based on the revenue loss per terminal basis. In order to estimate the losses due to an email worm attack, we use the following parameters:

- 1. Revenue per terminal**
- 2. % Inefficiency for a dormant terminal**
- 3. % Inefficiency for a active terminal**

The loss for the organization is actually due to the loss in efficiency due to the viral activity on the computers. Hence, after the viral attack begins:

Cumulative Losses for the organization at any time *t*
 $= \sum (Revenue\ of\ terminal, i) * (State\ of\ terminal\ i) * (\% \ Inefficiency\ due\ to\ viral\ attack\ on\ i)$

This assumes that an unaffected terminal performs with 100% efficiency. Hence, by summing over all terminals, we calculate the cumulative losses for the organization. Obviously, the dead state terminals are assumed to have 0% efficiency.

5.4. Additional Cost

Resurrection of infected terminals can be a significant cost. Investments needed to resurrect infected terminals should also be counted in the cumulative losses. This additional cost would be: (# of hours of system engineers to resurrect PCs)*(hourly rate of engineers). However, being subjective we have not included this consideration in our current analysis.

5.5. Cost-Benefit Analysis for Anti-Virus

This can be calculated by assuming an investment cost for anti-virus software for all systems. Assuming, this investment to be I , the probability of a viral attack in a year to be $p\%$, and the annual losses as calculated in the earlier step as L , we need to compare I with $p*L$. If $I < p*L$, then investments in an anti-virus system is justified.

Simplifying our assumptions further, let us assume that it is 100% probable that there will be a viral attack each year (going by present day situation), hence we can just understand the trade off by comparing I and L .

As part of the algorithm, for each time step the loss in efficiency (cumulative) is calculated. In order to make the analysis independent of the number of computers and the revenue, we chose I relative to the revenue potential of any computer. For example, if the revenue of each terminal is \$100 per hour and the firewall cost is \$5000 (on an annual subscription basis). We express the fire-wall investment as 5000%. This makes it easy to analyze results. By comparing the cumulative result as a result of a viral attack with the annual investment in firewall we understand the utility of installation of firewalls.

5.6. Simulator

A simulator has been developed on the .NET Platform based on the algorithm described above. The simulator generates a text output which can be used to plot graphs using standard packages such as Microsoft Excel. The user interface is shown in the figure 3 below:

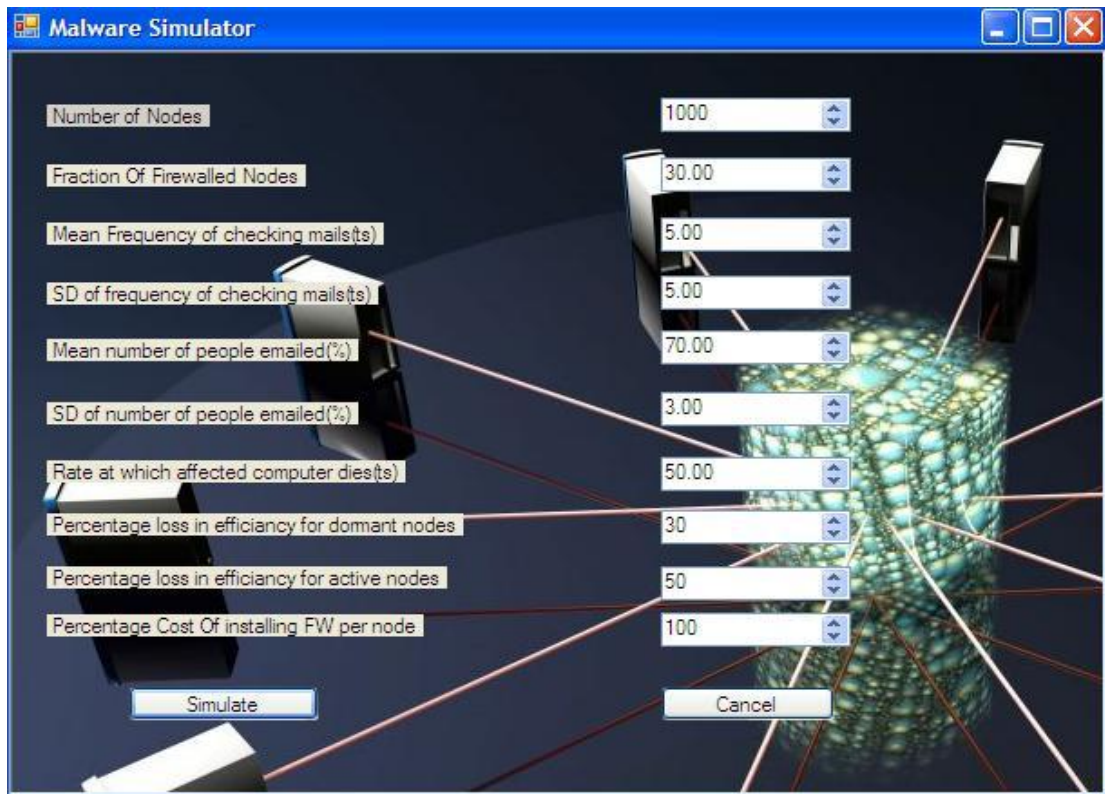


Fig 3a: Simulator Developed

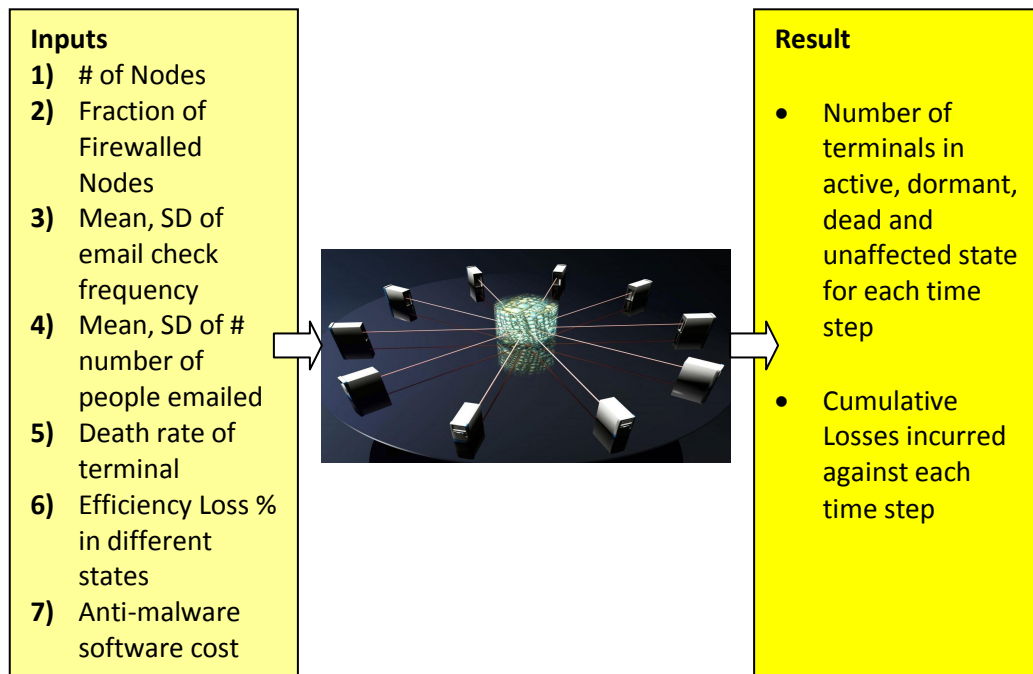


Figure 3b: Schematic representation of the simulator inputs and output

CHAPTER 6: RESULTS

As is intuitively clear, the spread of malware is dependent on all the factors chosen, for example, the frequency of checking emails, the % of employees with whom one employee interacts with on email, kind of worm which implies, its rate of death, % inefficiency it causes etc. We take this one by one and show the impact of each parameter graphically drawing insights from the same.

6.1. Impact of % of Terminals Fire-Walled

We plot the cumulative loss, number of clean, active, dormant and dead terminals at each time step. The value of N has been taken as 1000 and the values of the other parameters are

- $\mu_f = 5$ time steps; $SD_f = 5$; $\mu_e = 1\%$; $SD_e = 3\%$; $R_d = 50$ time steps; Loss (dormant) = 30%; Loss (active) = 50%

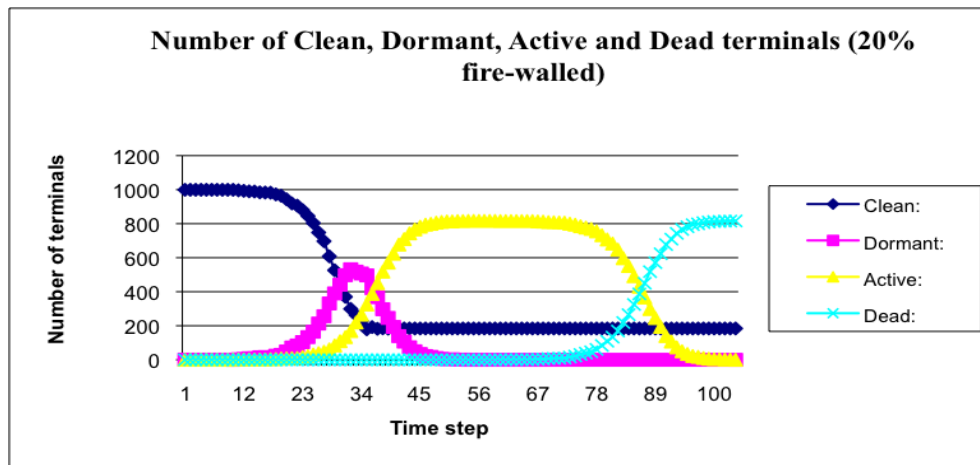


Fig 4a.: Impact of % fire-walled terminals

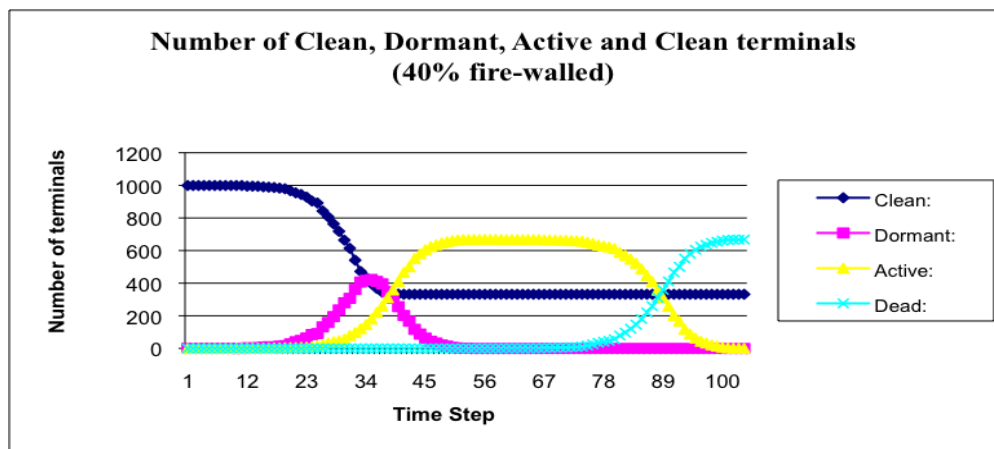


Fig 4b.: Impact of % fire-walled terminals

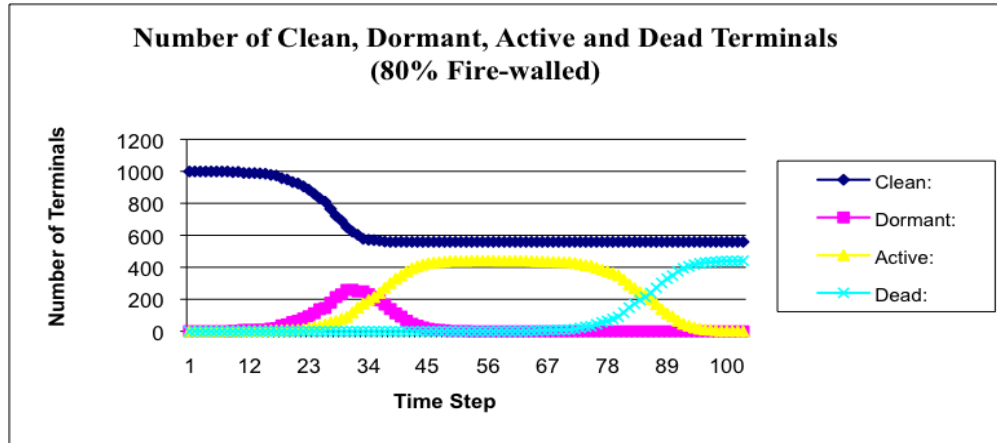


Fig 4c.: Impact of % fire-walled terminals

The change in the number of clean, dormant, active and dead terminals in the system as a function of time steps and % of terminals fire-walled

The above illustration in Figure 4a, 4b and 4c shows the following:

1. The characteristic curves of number of terminals in each state is identical in all graphs, independent of the % of fire-walled computers:
 - a. The number of dead terminals rises like an S curve and plateaus
 - b. The number of dormant terminals is a bell curve
 - c. The number of active terminals is a plateau shaped curve and falls to 0
 - d. The number of clean terminals falls and then becomes flat
2. At the end as designed, there are only clean and dead terminals. The cleans are clean by virtue of the fact that they are fire-walled

These characteristic curves should change with other parameters as we will illustrate in the sections below.

6.2. Impact of Frequency of Checking E-mail

The implication of the frequency of checking emails is that higher the frequency, the faster the systems become dormant. As shown in the graphs below generated using results from the simulator, the dormant terminal distribution shifts left as the frequency of mail checking increases (i.e., the time steps decreases).

The results are illustrated in the graphs shown below. The characteristic curves in Figure 5a, 5b and 5c show how the dormant curve shifts leftward and the active curve becomes larger with increase in email checking frequency.

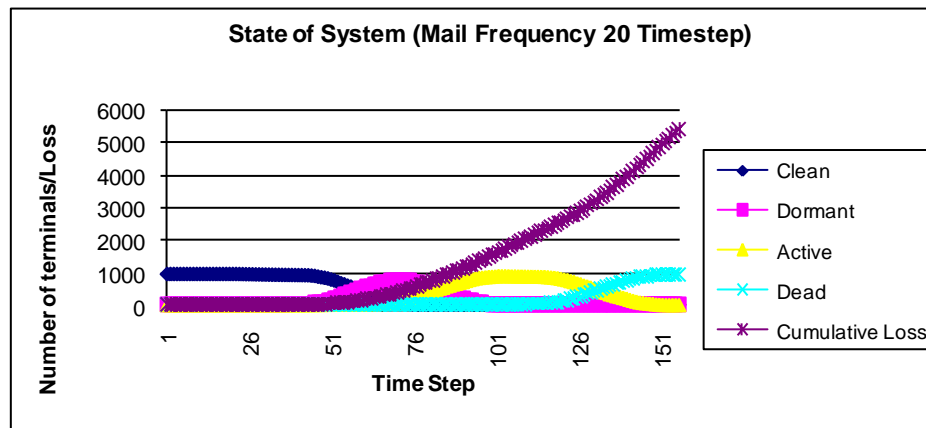


Fig 5a: Impact of mail frequency

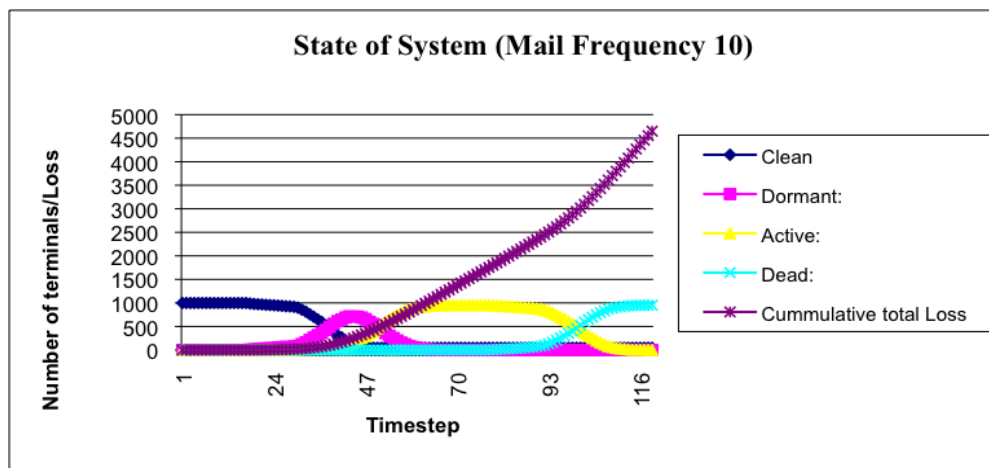


Fig 5b.: Impact of mail check frequency

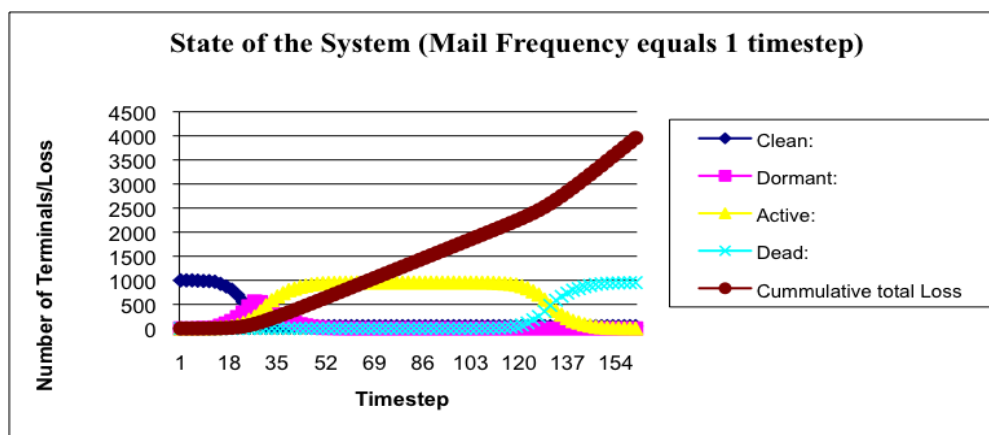


Fig 5c.: Impact of mail check frequency

The impact of mail checking frequency: Higher the frequency, lower the losses as the dormant stage vanishes quickly whereas in the case of lower frequency the virus lingers on for a longer time inflicting losses.

6.3. Impact of Standard Deviation in Frequency of Checking E-mail

The impact of standard deviation on checking email is shown in the graphs generated below. As is clear, the graphs show increasing losses as the standard deviation in frequency of email checks go higher.

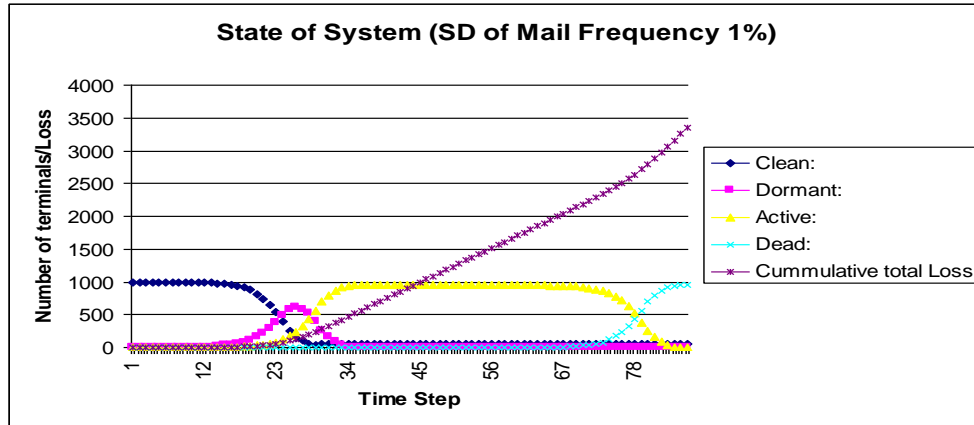


Fig 6a.: Impact of SD of mail check frequency

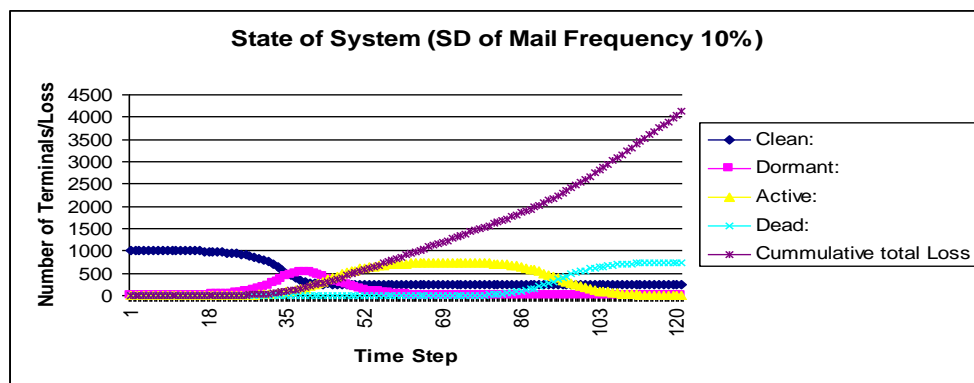


Fig 6b.: Impact of SD of mail check frequency

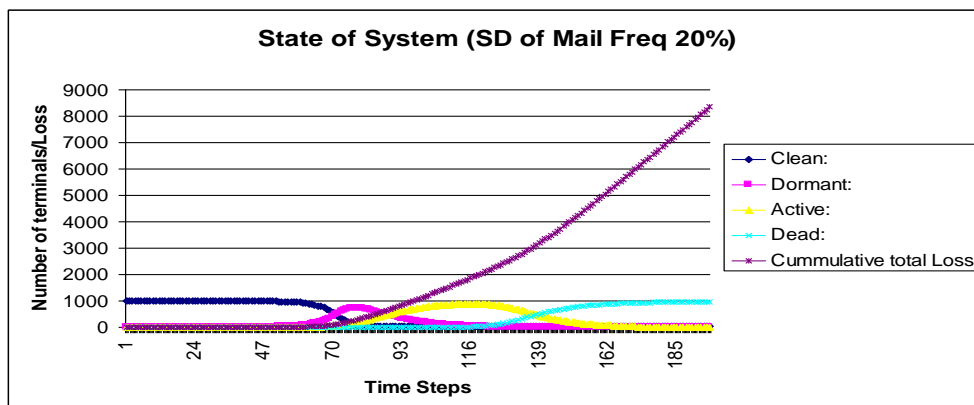


Fig 6c.: Impact of SD of mail check frequency

Impact of SD of mail checking frequency on the characteristic curves of the system infected by virus. Higher the standard deviation, more losses as virus linger in the system for a longer time.

6.4. Impact of Replication Rate of Malware

In the system under consideration, the worms replicate through emails. Hence if the average number of emails interchanged between employees in the system is high, the virus replication rate is high. We see that the nature of the spread does not change and hence there must be some dual effects canceling each other. We discuss this further in the next section. The results are show in Fig 7a and 7b.

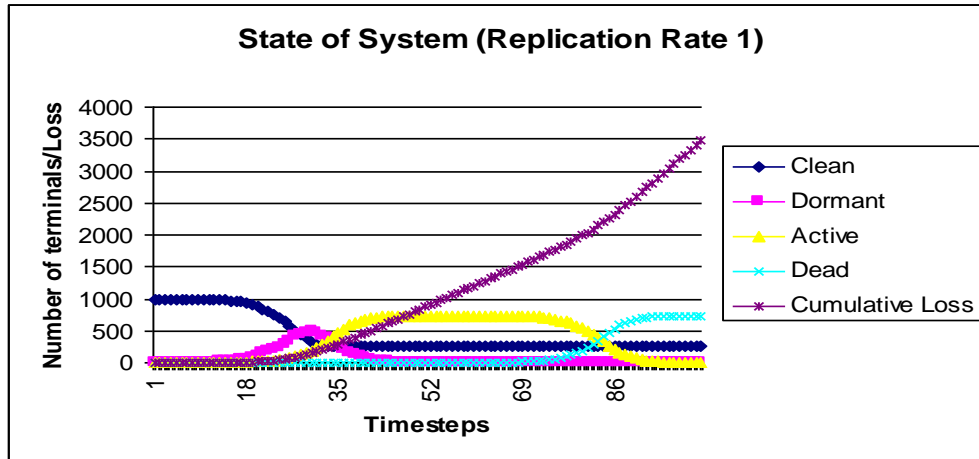


Fig 7a.: Impact of Replication rate

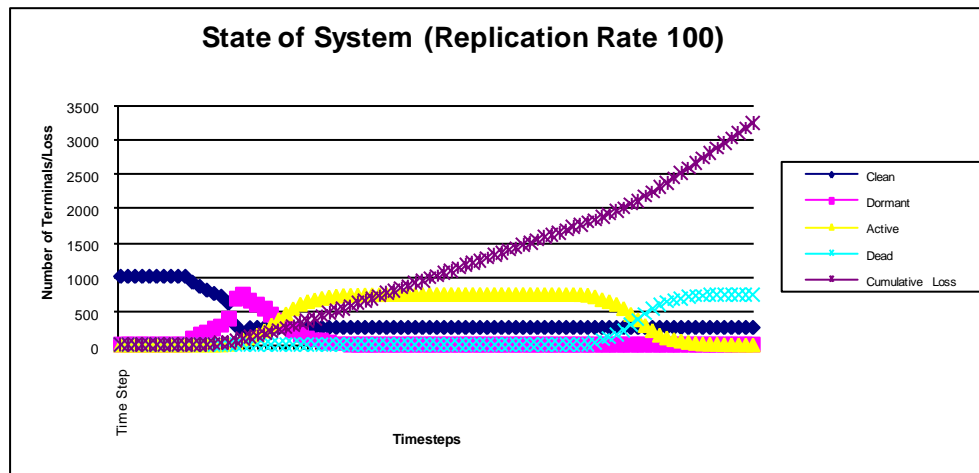


Fig 7b.: Impact of Replication rate

A similar affect is expected for the impact of standard deviation of the number of people emailed.

6.5. Impact of Rate of Death on the System

The results are shown in the graph with Rate of Death = 1 ts and 100 ts.

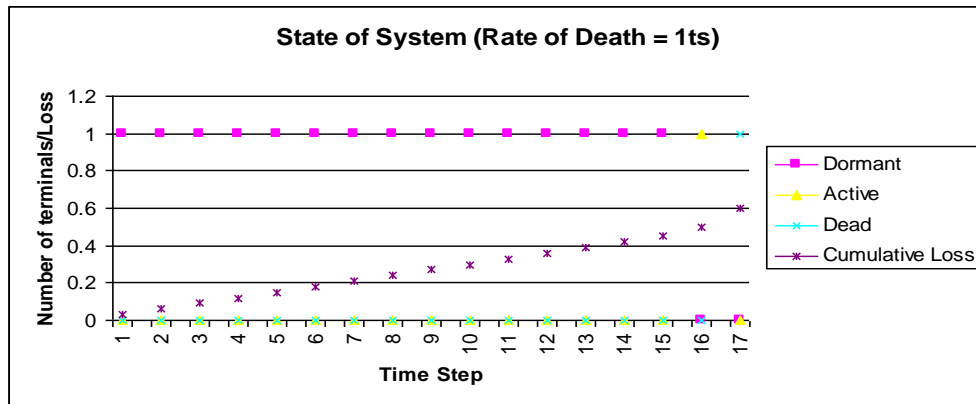


Fig 8a.: Impact of Death Rate

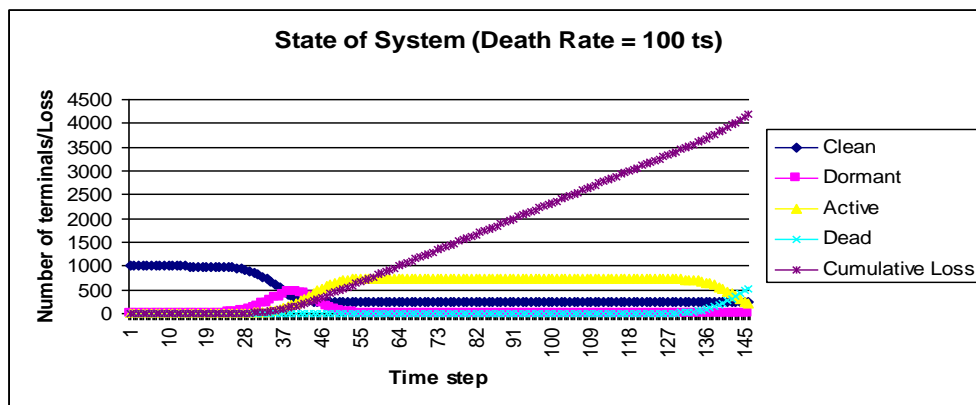


Fig 8b.: Impact of Death Rate

6.6. Impact of Dormant System's Efficiency

Fixing the efficiency of the active system as 50% of normal, we have tried to see the impact on losses if the loss in efficiency of dormant terminals is 10% vs. 40%. The losses increase, however, the losses do not increase proportionately (i.e. not by 40%). This is because the time for which the computers remain in the dormant state is low and hence the impact is not felt to a great extent.

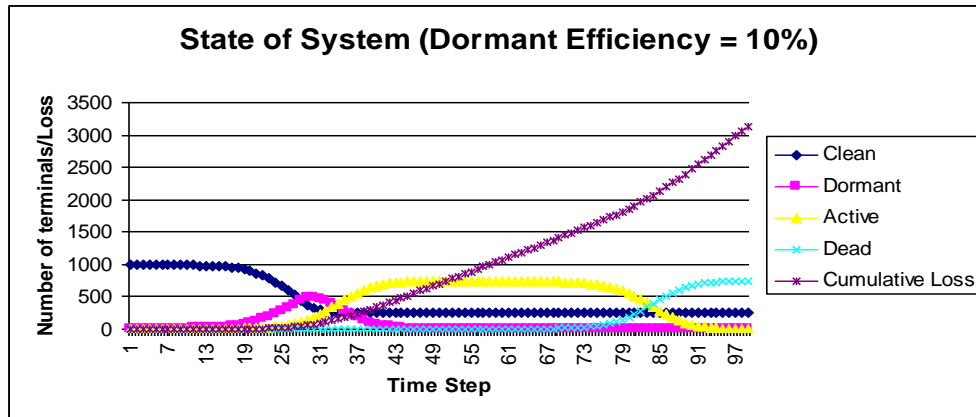


Fig 9a: Impact of Dormant Computer Efficiency on Losses

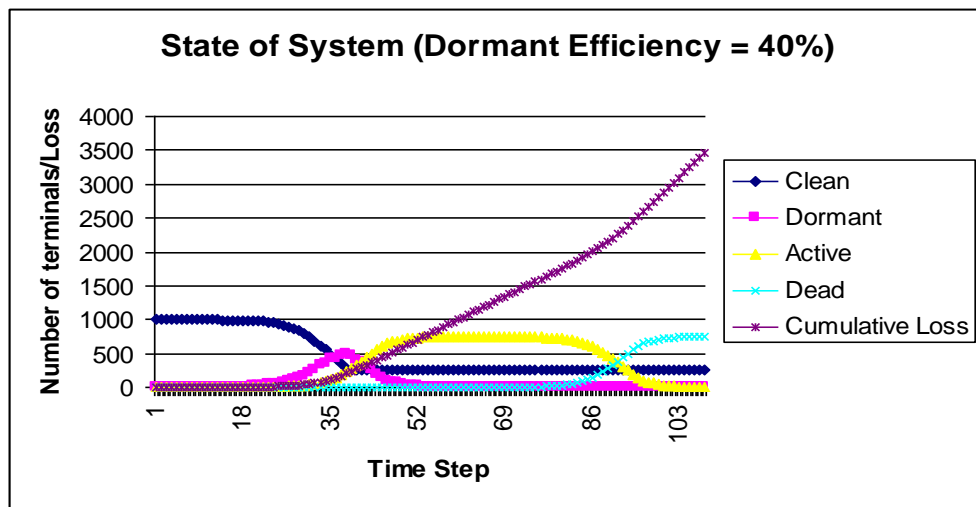


Fig 9b.: Impact of Dormant Computer Efficiency on Losses

It is obvious that increasing or decreasing the efficiency of active terminals will have an impact on the losses because of the longer duration (as shown in figures 8a and 8b) for which terminals remain active.

CHAPTER 7: COST BENEFIT ANALYSIS OF THE FIREWALLS

The cost benefit analysis of installing fire-walled computers can be done by comparing the revenue losses avoided vs. the installation cost of the fire-wall. Let us assume the revenue per terminal is R. Using different % of fire-wall usage we compute the efficiency lost. By comparing the marginal loss avoided by putting in more firewalls as compared to cost of the additional firewalls, we can understand the tradeoff.

The table below compares the cumulative loss due to the viral attack with the cost of installing a firewall at different firewall cost levels. Everything is relative to the Revenue per time step.

Fire-wall %	Cumulative Loss	Savings with firewalls	Fire-wall one time cost						
			2000%	3000%	5000%	8000%	10000%	20000%	30000%
0%	4740%		0%	0%	0%	0%	0%	0%	0%
10%	3935%	805%	200%	300%	500%	800%	1000%	2000%	3000%
20%	3686%	1054%	400%	600%	1000%	1600%	2000%	4000%	6000%
30%	3289%	1451%	600%	900%	1500%	2400%	3000%	6000%	9000%
40%	3078%	1662%	800%	1200%	2000%	3200%	4000%	8000%	12000%
50%	2849%	1891%	1000%	1500%	2500%	4000%	5000%	10000%	15000%
60%	2567%	2173%	1200%	1800%	3000%	4800%	6000%	12000%	18000%
70%	2321%	2419%	1400%	2100%	3500%	5600%	7000%	14000%	21000%
80%	1949%	2791%	1600%	2400%	4000%	6400%	8000%	16000%	24000%
90%	1885%	2855%	1800%	2700%	4500%	7200%	9000%	18000%	27000%

Table 1: Cost – Benefit Analysis Table. Understanding the trade-offs

As shown in the highlighted cells in the table 1, the cost of the fire-wall cost becomes more than the savings which result from having firewalls and at these levels it is unadvisable to install fire-walls. Let us take an example: Let us suppose that the time step is 1 hour and the revenue per hour per terminal is \$100. Hence the total loss per terminal, if no firewall installed is $\$100 \times 4740\% = \4740 . If we install 10% firewalls, we could save $\$100 \times 805\% = \805 per terminal. Hence it would be advisable to buy a fire-wall which costs lower than the savings resulting from it. Similarly, for different firewall levels we can calculate the per terminal fire-wall cost which makes sense.

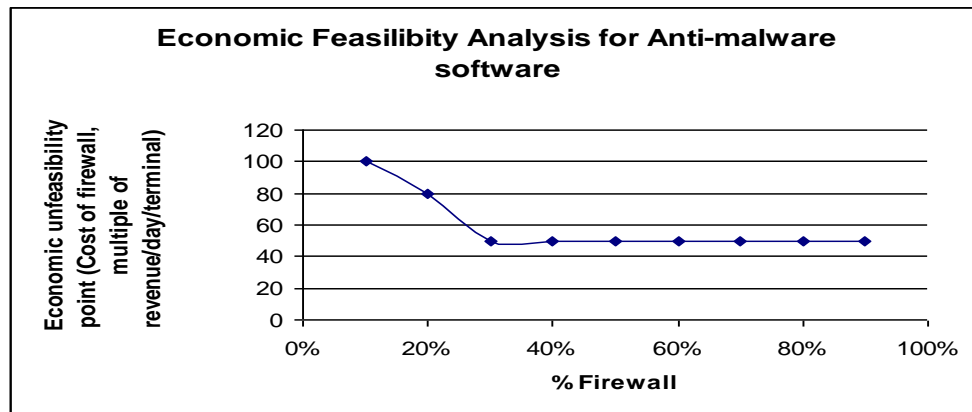


Figure 10: Economic Feasibility Analysis for Anti-Malware software

Figure 10 above shows the % of computers fire-walled against the economic unfeasibility point, i.e., at what price of software would it not make economic sense to install it on more terminals. As shown, if the price of the software is around 50 times the daily revenue generated by each terminal, it does not make economic sense to install the software on more than 40% of terminals

This is a useful tool for people involved in decision making regarding installation of fire-walls and other security investments.

CHAPTER 8: DISCUSSION

8.1. Impact of various factors on the spread of malware and losses

Impact of % of terminals fire-walled: The results as shown in Fig. 4a, 4b and 4c support what we could guess intuitively: the more the % of terminals fire-walled the better. However, the % changes the spread as well. For example, if the % of fire-walled computers is high, the system reaches steady state (only clean and dead terminals) sooner. The characteristic curves for different levels of fire-walling remains similar showing that these are fundamental to the system.

Impact of mail checking frequency and standard deviation: In a result which is counter intuitive, the losses are more if the mail checking frequency is less. This is because the more period of time that the system is in dormant or active state, the more the losses. If the mail checking frequency is high, the dormant state vanishes quickly thereby reducing the losses, whereas if the mail checking frequency is low, the virus lingers on and reduces the overall efficiency of the system. These are shown by Fig 5a, 5b and 5c.

As expected the standard deviation also leads to similar effect on the spread of malware as shown in Fig. 6a, 6b, 6c. We find that the more regularity in the patter of email checking, the lower the losses. This is because if the mails are checked irregularly, the systems stay in the dormant state for longer periods of time, inflicting losses. These are clear from the graphs in figure 6a, 6b, 6c. If the dormant state vanished quickly, the losses are limited.

Impact of replication rate of the worm: The results are not completely inline with the hypothesis we started with. To explain the results we have a new hypothesis which is that the impact of a high rate can be two fold:

1. With a high replication rate, the virus spreads faster and hence the system is impacted early and more terminals stay active and dormant longer thereby increasing losses
2. On the other hand, if the replication rate is high, it also implies that the system reaches steady state faster and hence the exposure time of the virus is shorter.

Hence we expect to see the opposing effects of 1 and 2 in the results. This is exactly what the results show. When the replication frequency is low, the virus stays in the system for a long time and causes losses as shown above. However, when the replication rate is very high (100), the virus spreads fast and makes many terminals active causing losses. Hence for both the extremes, because of the opposing effects (more dormant vs. more active), the losses turn out to be the same.

Impact of Rate of Death on the System: The rate of death turns out be a very interesting factor in the simulations. As shown in the Fig 8a and Fig 8b, if the Rate of death is too fast, i.e., the infected computer just dies instantaneously; the loss to the system is negligible. This is because the infected system does not get a chance to

replicate the virus through emails and hence the first computer in the network which acquires the virus is the last one as well. This is very counter intuitive at first glance as we expect a higher death rate to result in higher losses. Similarly, taking the extreme case if the death comes at a slow rate for the active terminals, the losses could be high as the virus replicates and results in many terminals becoming dormant and active for a longer time period.

Impact of Efficiency: The losses will increase with lower efficiency in the dormant or the active state and therefore the obtained values are in line with expectations.

8.2. Monetary Losses due to spread of malware and cost benefit analysis

As with every tool there is a trade-off between their costs and the benefits. The fire-wall cost and benefits have been brought out in the results on the incremental loss avoidance vs. the cost of the fire-wall. As has been shown in Table 1, for each price of fire-wall we can estimate the % of computers which need to be fire-walled while making economic sense. Vice versa, if we have a certain % of fire-walling in mind, we can understand what the threshold price for software is, while selecting between different options of fire-walls.

CONCLUSIONS

Using the model developed we have been able to measure the impact of various parameters which define a viral attack in a computer network, namely, replication rate of the virus, the death rate of the terminal, mail checking frequency etc. Importantly, we have been able to measure the economics of using fire-walls and how we could compare the costs and benefits in a scientific manner.

The overall conclusions have been:

- Increasing the % of fire-walled terminals can help limit monetary losses
- A very high email checking frequency can in fact limit monetary losses as the “active” and “dormant” states remain for limited time
- The reproduction rate of the virus does not have a direct impact on losses. This is because a high rate means a more rapid spread whereas a low rate implies longer time duration for “dormant” and “active” state. Both these cases result in higher losses
- A very high death rate of the virus can actually lead to very low monetary loss as the chance of replication is not there. The reverse is also true.
- A trade off exists for incremental investments in installing anti-malware software and the incremental savings it brings during a malware attack.

The model could have been more robust had we been able to validate against actual empirical data to understand how the results and conclusion compare. This could be taken up as content for future work.

REFERENCES

- [1] Chantler, A. N. and Broadhurst, R. (2006) Critical Information Infrastructure Protection (ciip). *Draft Technical Report for the Australian Institute of Criminology*. <http://eprints.qut.edu.au> (accessed May 6th 2009).
- [2] Chen, Z., Ji, C. (2005) Spatial-Temporal Modeling of Malware Propagation in Networks. *IEEE Transactions on Neural Networks*. 16. pp. 5.
- [3] Daley, D., Gani, J. (1999) Epidemic Modeling, an Introduction, Cambridge University Press, Cambridge, UK.
- [4] Fleizachy, C., Liljenstamz, M., Johanssony, P., Voelkery, G. M., and András Méhesz. (2007) *WORM'07*. Alexandria, Virginia, USA.
- [5] Goranin, N., Čenys, A. (2008) Malware Propagation Modeling by the Means of Genetic Algorithms. *Electronics and Electrical Engineering*. Kaunas Technologija. No. 6(86). pp. 23–26.
- [6] Katsikas, S., Spyroua, T., Gritzalisa, D., Darzentasa, J. (1996) Model for network behaviour under viral attack. *Computer Communications*. 19. pp. 124-132.
- [7] Kondakci, S. (2008) Epidemic state analysis of computers under malware attacks. *Simulation Modelling Practice and Theory*. 16. pp. 571–584.
- [8] Lelarge, M. Epidemic Risks Model, Network Externalities and Incentives. *Economics of Malware*. INRIA-ENS.
- [9] Litchfield, D. (2002) Threat profiling microsoft sql server (*a guide to security auditing*) <http://www.nextgenss.com/papers/tp-SQL2000.pdf> (accessed May 10th 2009).
- [10] McMillan, R. (2006) IDG News Service, Computerworld, IDG Inc. <http://www.computerworld.com/>. (accessed April 2nd 2009).
- [11] Park, I., Sharman, R., Rao, H. R., Upadhyaya, S. (2007) Short Term and Total Life Impact analysis of email worms in computer systems. *Decision Support System*. 43. pp. 827–841.
- [12] Pricewaterhouse Coopers. (2006) Information security breaches survey 2006 Technical report. UK Department of Trade and Industry.
- [13] Siewiorek, D. P. and Swarz, R. S. (1998) *Reliable Computer Systems*, 3rd ed. *Design and Evaluation*. A.K. Peters Ltd., Natick, MA, USA.

- [14] Yu, W., Chellappan, S., Wang, X., Xuan, D. (2008) Peer-to-peer system-based active worm attacks: Modeling, analysis and defense, *Computer Communications*. 31. pp. 4005–4017.
 - [15] Zyba, G., Voelker, G., Liljenstam, M., Mehes, A., and Johansson, P. (2009) Defending mobile phones from proximity malware. In *Proceedings of INFOCOM*
-