MODELLING THE EFFECTS OF MALWARE PROPAGATION ON MILITARY OPERATIONS BY USING BAYESIAN NETWORK FRAMEWORK

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS OF
THE MIDDLE EAST TECHNICAL UNIVERSITY
BY

ZAFER ŞENGÜL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
THE DEPARTMENT OF CYBER SECURITY

AUGUST 2019

Approval of the thesis:

MODELLING THE EFFECTS OF MALWARE PROPAGATION ON MILITARY OPERATIONS BY USING BAYESIAN NETWORK FRAMEWORK

Submitted by Zafer ŞENGÜL in partial fulfillment of the requirements for the degree of **Master of Science in Cyber Security Department, Middle East Technical University** by,

Prof. Dr. Deniz Zeyrek BOZŞAHİN
Dean, **Graduate School of Informatics** _____

Assoc. Prof. Dr. Aysu Betin CAN
Head of Department, **Cyber Security** _____

Assoc. Prof. Dr. Cengiz ACARTÜRK
Supervisor, **Cyber Security Dept., METU** _____

Examining Committee Members:

Asst. Prof. Dr. Aybar Can ACAR
Health Informatics Dept., METU _____

Assoc. Prof. Dr. Cengiz ACARTÜRK
Cognitive Science Dept., METU _____

Assoc. Prof. Dr. Hacer KARACAN
Computer Engineering Dept., Gazi University _____

Date: 07.08.2019

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name : Zafer ŞENGÜL

Signature :

ABSTRACT

MODELLING THE EFFECTS OF MALWARE PROPAGATION ON MILITARY OPERATIONS BY USING BAYESIAN NETWORK FRAMEWORK

ŞENGÜL, Zafer
MSc., Department of Cyber Security
Supervisor: Assoc. Prof. Dr. Cengiz ACARTÜRK

August 2019, 80 pages

Malware are malicious programs that cause unwanted system behavior and usually result in damage to IT systems or its users. These effects can also be seen during military operations because high-tech military weapons, command, control and communication systems are also interconnected IT systems. This thesis employs conventional models that have been used for modeling the propagation of biological diseases to investigate the spread of malware in connected systems. In particular, it proposes a probabilistic learning approach, namely Bayesian Network analysis, for developing a framework for the investigation of mixed epidemic model and combat models to characterize the propagation of malware. Compared to the classical models, which have employed formula-based representations, the results of this thesis reveal more enriched representations of the superiority of one military force over the other in probabilistic terms.

Keywords: Combat and Epidemic Models, Cyber Warfare, Bayesian Network Framework, Artificial Intelligence, Machine Learning.

ÖZ

ZARARLI YAZILIMLARIN YAYILMALARININ ASKERİ OPERASYONLAR ÜZERİNDEKİ ETKİSİNİN BAYES AĞI YAPISI KULLANILARAK MODELLENMESİ

ŞENGÜL, Zafer

Yüksek Lisans, Siber Güvenlik Bölümü

Tez Yöneticisi: Doç. Dr. Cengiz ACARTÜRK

Ağustos 2019, 80 sayfa

Kötü amaçlı yazılımlar, istenmeyen sistem davranışlarına neden olan ve genellikle BT sistemlerine veya kullanıcılarına zarar veren kötü amaçlı programlardır. Bu etkiler askeri operasyonlar sırasında da görülebilir, çünkü yüksek teknoloji ürünü askeri silahlar, komuta, kontrol ve haberleşme sistemleri de birbirine bağlı BT sistemleridir. Bu tezde, kötü amaçlı yazılımların bağlı sistemlerdeki yayılmasını araştırmak için biyolojik hastalıkların yayılımını modellemek için kullanılan geleneksel modeller kullanılmıştır. Özellikle, kötü amaçlı yazılımların yayılmasını karakterize etmek kullanılan karma salgın modeli ve savaş modellerinin araştırılmasında bir çerçeve geliştirmek için Bayes Ağı analizi gibi olasılıksal bir öğrenme yaklaşımı önermektedir. Formüle dayalı temsiller kullanan klasik modellerle karşılaştırıldığında, bu tezin sonuçları, bir askeri gücün diğerine göre üstünlüğünün olasılıksal açıdan daha zenginleştirilmiş temsillerini ortaya koymaktadır.

Anahtar Sözcükler: Savaş ve Salgın Modelleri, Siber Savaş, Bayes Ağı Yapısı, Yapay Zekâ, Makine Öğrenmesi.

To My Wife Burcu and

To My Sons Çağatay and Kıvanç

ACKNOWLEDGEMENTS

First and foremost, I would like to express my gratitude to my supervisor Assoc. Prof. Dr. Cengiz ACARTÜRK for the useful comments, remarks and engagement through the learning process of this master thesis. I would like to thank him for introducing me to the topic, for his support on the way and his steering me in the right direction whenever I needed.

I would like to thank my committee members, Asst. Prof. Dr. Aybar Can ACAR and Assoc. Prof. Dr. Hacer KARACAN for their technical guidance, encouragement and support.

I would also like to thank Informatics Institute's faculty members and staff for their help during my MSc period.

Finally, I must express my very profound gratitude to my wife Burcu ŞENGÜL for providing me with inexhaustible support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. She changed the way I live, warm my soul and made my life meaningful. This accomplishment would not have been possible without her.

I would like to thank my precious children Çağatay and Kıvanç who have always made me smile from the moment I learned about their existence. I am sorry to have stolen their time while studying and writing my thesis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

DDoS	Distributed Denial of Service
DoS	Denial of Service
FN	False Negative
FP	False Positive
ROC	Receiver Operating Characteristic
PRC	Precision Recall Curve
SEIQR	Susceptible-Exposed-Infected-Quarantine-Recovered
SEIR	Susceptible-Exposed-Infected-Recovered
SIR	Susceptible-Infected-Recovered
TN	True Negative
TP	True Positive
WEKA	Waikato Environment for Knowledge Analysis

CHAPTER 1

INTRODUCTION

1.1. Purpose and Scope

War is defined as *"a conflict among political groups involving hostilities of considerable duration and magnitude"* in Britannica Encyclopedia[1]. It is a violent movement that aims at rivals accepting the will of its enemy (Clausewitz, 1982). It has a multi-disciplinary structure which includes biological, psychological, economic and political concepts. With the advent of technology, one more concept has been added recently: the cyberspace. According to the US Department of Defense, cyberspace is an information domain which comprises of technology infrastructures such as internet, telecommunications networks and computer systems. Cyber warfare is an additional ability that both prevents and destroys the enemy's technological attacks and helps us protect our own networks, systems, information from malicious cyber actions (DoD, 2018). It can be carried out as a force multiplier during contemporary wars.

Warfare can be modelled. There are various computational models such as the Lanchester linear model, the aimed fire model, and the ambush model. These are mathematical models that are based on differential equations.

The main goal of this study is to develop a probabilistic approach to investigate the results of malware propagation in a network of computers, possibly used by military units during operations. The more the skills of the weapons increase, the more they are dependent on technology. However robust a system is developed; it is susceptible to malware, such as a computer virus. In modern day warfare, a novel type of malware may change all the expected outcomes of the war. For instance, country with fewer number of weapons and soldiers may menace a stronger country by using its cyber superiority. Even in closed networks, malware can cause damage by affecting ongoing processes as seen in Stuxnet (Lindsay, 2013). As of the time when computers began to spread rapidly, connected

[1] Britannica Encyclopedia. Retrieved on July 23, 2019 from https://www.britannica.com/topic/war

devices have apparent cyber effects on military operations and this situation should be taken into account by military decision makers.

In this study, I analyzed several types of malware and classified them into three models according to their characteristics. Then, I used Bayesian Network Framework to show the probabilities of the effects of malware propagation on military operations.

1.2. Research Question

The models of propagation of infectious diseases gave inspiration for researchers to use them for the simulation of the computer viruses. However, there is limited research to apply a probabilistic approach to study likely outcomes of malware propagation. The research question of this thesis is how to model the effects of malware propagation by using Bayesian Network Framework to enrich the representation of the model outcomes.

1.3. Research Method

Firstly, I analyzed *combat models* (Clausen, 2003) and *epidemic models* (Kermack & McKendrick, 1927*)*. These models are combined and obtained *mixed models* (Schramm & Gaver, 2013). Based on the distinctive characteristics of specific types of malware, I employed three mixed models (namely SIR, SEIQR, SEIR) to model malware. Secondly, I generated sample data (in accordance with the ones used in the relevant literature) to simulate the differential equations of the models. The outcome of the models is basically two categories (in this context, army forces) which show the winning side of the war; "Blue" for the friendly forces, and "Red" for the enemy forces. Finally, I applied the Bayesian Network approach to compute the winning probabilities of the warring sides. The goal of this thesis is to propose a mapping between epidemic models and cyber-attacks, in addition to analyze cyber-attacks in a Bayesian Network framework for probabilistic analysis of the model outcomes.

1.4. Thesis Layout

Chapter 2 presents background and relevant work, chapter 3 reports Bayesian Network and its implementation on the models. Finally, chapter 4 reports conclusion and future work.

CHAPTER 2

BACKGROUND AND RELEVANT WORK

2.1. Background

2.1.1. Combat Models

Combat modelling is a sort of mathematical modelling. Its aim is to find out the result of the combat (Lanchester, 1916). Frederick William Lanchester established a mathematical analysis of air combat and paved the way for combat modelling and calculation of attrition of the fighting sides in military operations.

In the course of cold war, these combat modellings were used as a main tool to guide NATO military decision makers. In fact, real life combat is affected by a lot of factors such as personnel, leadership, moral, training and education, weapon and sensor systems, command, control and information systems, strategies and tactics, terrain, weather and light conditions. However, bearing these in mind, in order to compute the warfare, only some quantitative factors can be taken into account (Clausen, 2003).

Although most of the combat simulations are stochastic, heterogeneous, complex and presents better forecasting (if the initial countless assumptions are correct), Lanchester models have excellence in simplicity. These models make good simplifying acceptances at the beginning of the battle. Moreover, they are close to real-time situations and bring marvelous results (MacKay, 2006).

2.1.1.1. Lanchester Linear Model (Unaimed Fire Model)

This model is based on ancient combat conditions where one soldier is fighting against one enemy soldier, assuming that they have the same fighting value and all other conditions are equal.

At the beginning of the battle, if Blue and Red forces both have 1000 soldiers, all of them will be dead in the end. If Blue forces have 1000 soldiers and Red forces have 700 soldiers initially, Blue forces will win the war and have 300 soldiers in the end as shown in Figure 1.

So, the number of soldiers remaining at the end of the battle is simply the difference between the larger army and the smaller one.

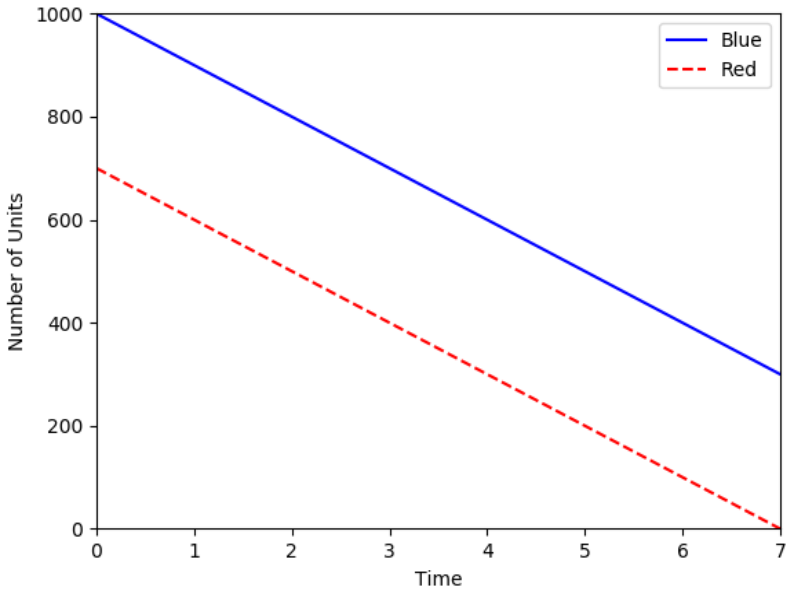


Figure 1: Lanchester Linear Model (Unaimed Fire Model)

2.1.1.2. Lanchester Square Model (Aimed Fire Model)

In this model, it is assumed that an operational unit is able to notice many enemy operational units at any time and has the ability to kill them. Each unit knows the location and the condition of all the enemy forces, so that its fire is directed only to live units or running weapons. When a target is killed, the fighting unit starts to search for a new target.

$$\frac{db(t)}{dt} = -kr \text{ x } r(t)$$
$$\frac{dr(t)}{dt} = -kb \text{ x } b(t) \qquad (2.1)$$

b(t) : Number of Blue units at time t
r(t) : Number of Red units at time t
kr : Kill rate of one Red unit
kb : Kill rate of one Blue unit

Combatants are not equally trained and they fight under different morale situations. Furthermore, the efficiency of their weapons is not the same. So, the kill rate of one Blue unit and that of the Red unit is different.

At the beginning of the battle, if Blue has 720 units and Red has 900 units and kill rate of Blue is 0.07 and kill rate of Red is 0.04, which side wins the battle?

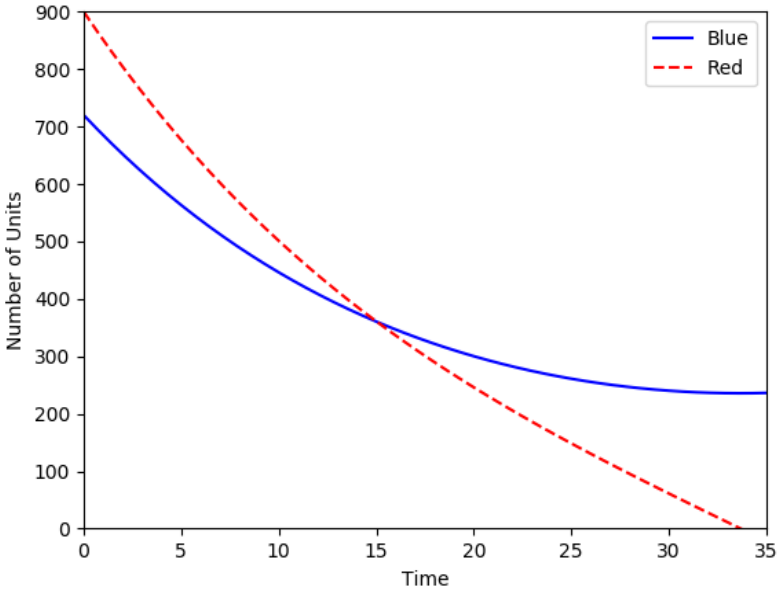


Figure 2: Lanchester Square Model (Aimed Fire Model)

By using the differential equations 2.1, we can obtain the result as shown in Figure 2. Even if the initial number of Red forces is higher than the Blue forces, Blue side wins the battle because the kill rate of one operational Blue unit is bigger and this affects the ongoing battle.

2.1.1.3. Lanchester Guerrilla Model (Area Fire Model)

According to the guerrilla-counterguerrilla warfare, weaker side tries to use intelligence better, takes advantage of terrain and uses special tactics in order not to be detected by the enemy. In this model, each operational unit is able to kill all enemy operational units that have been detected. So, this model can be described as "if you are seen, you are dead". This situation might be logical for battle between two guerrilla units trying to find each other (Deitchman, 1962). This model is also used to compute the result of the forces after indirect fires such as artillery or mortar fire.

$$\frac{db(t)}{dt} = -dr \text{ x } b(t) \text{ x } r(t)$$
$$\frac{dr(t)}{dt} = -db \text{ x } b(t) \text{ x } r(t) \qquad (2.2)$$

db : Detection rate of one Blue unit
dr : Detection rate of one Red unit

In order to compute the result of the battle, the differential equations 2.2 are used for this model. If the initial number of Blue forces is 720 units and Red forces is 900 units and detection rate for one operational Blue unit against an operational Red unit is 0.0006 and

detection rate for one operational Red unit against an operational Blue unit is 0.0004, Blue side will win the battle as shown in Figure 3. Because detection rate of Blue forces is higher. Blue forces can detect and kill more than Red side does.

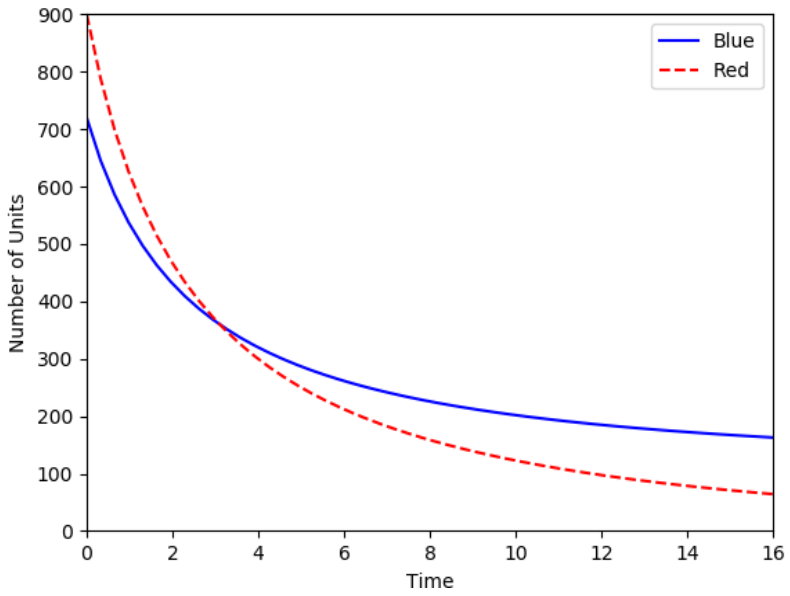


Figure 3: Lanchester Guerrilla Model (Area Fire Model)

2.1.1.4. Lanchester Mixed Combat Model

This model is the mixture of the Lanchester square model and guerrilla model. One side is forwarding over an unconcealed area (square model) and the other side is fighting from a hidden position (guerrilla model).

$$\frac{db(t)}{dt} = -dr \text{ x } b(t) \text{ x } r(t)$$
$$\frac{dr(t)}{dt} = -kb \text{ x } b(t) \qquad (2.3)$$

By using the differential equations 2.3, we can make calculation in order to find the outcome of the battle. Red forces are advancing over an open area searching for Blue forces to destroy. Besides, Blue forces are fighting from a secret place to destroy the enemy. Both sides use different type of combat models.

If Blue has 720 units and Red has 900 units at the beginning of the battle and detection rate for one operational Red unit against an operational Blue unit is 0.0004 and kill rate of Blue is 0.25, then Blue forces will win the battle as shown in Figure 4. Blue forces who use guerrilla tactics take advantage of the terrain and although initially fewer in number, they achieve the victory.

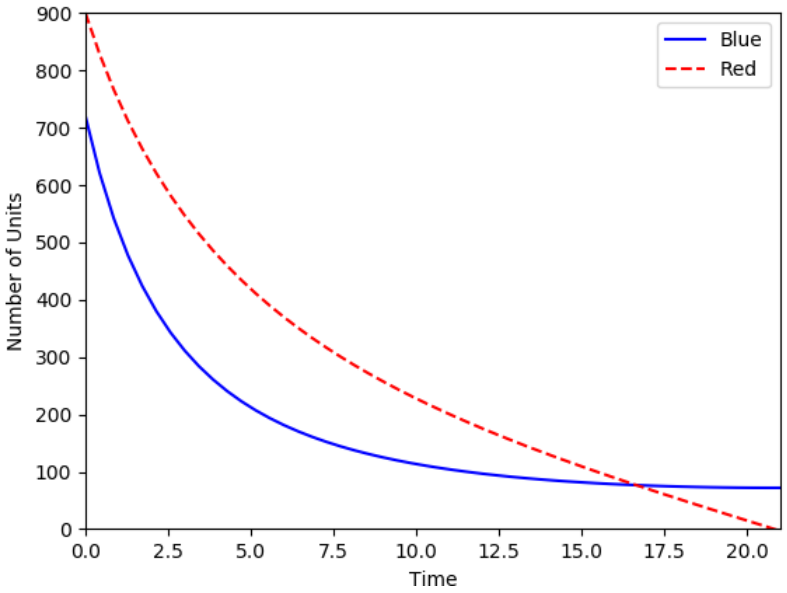


Figure 4: Lanchester Mixed Combat Model

In summary, I analyzed four combat models. I used Lanchester aimed fire model in my thesis because of its simplicity and fitting well for the mixed epidemic model. In the following section I presented epidemic models.

2.1.2. Epidemic Models

The Epidemic Model is a configuration that was firstly developed in order to investigate the contagion of biological diseases (Kermack & McKendrick, 1927). The generally used mathematical model for the spread of biological pathogens is SIR model. It is also known as Kermack-McKendrick Infection Model. In this model, total population is separated into three different groups. The first group is *Susceptible* (S) who are vulnerable to catch the disease. The second group is *Infected* (I) who have the disease and are able to spread it to others. The third group is *Recovered* (R) who are immune to the infection or isolated. The total number of the population is sum of the *Susceptible*, *Infected* and *Recovered*.

$$\frac{dS}{dt} = -r \text{ x } S \text{ x } I$$

$$\frac{dI}{dt} = r \text{ x } S \text{ x } I - a \text{ x } I \qquad (2.4)$$

$$\frac{dR}{dt} = a \text{ x } I$$

r : Infection rate
a : Removal rate

The usage of this model can be seen in real life. For example, in 1978 there was a flu epidemic in a boarding school in Britain. 512 of 763 boys were infected and had to stay in bed for two weeks. It came out that one boy started to spread the epidemic (Murray, 2002). The continuous curves of *Susceptible* and *Infected* group have close similarity with the dotted real values taken from British medical journal.

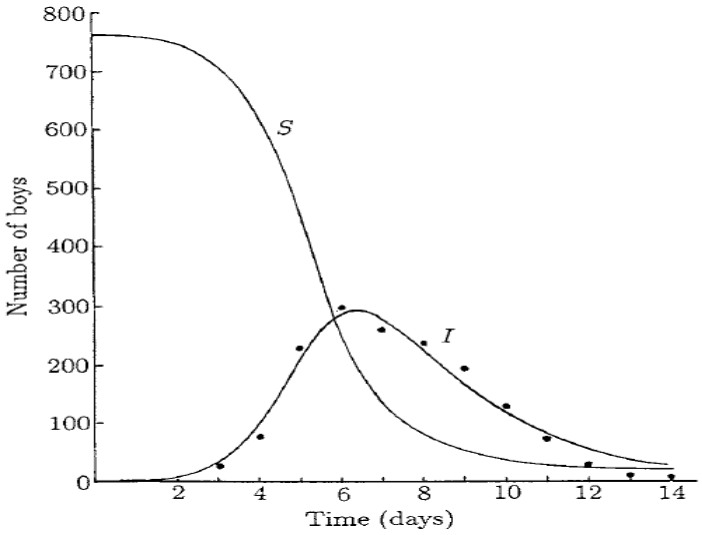


Figure 5: Influenza Epidemic in an English Boarding School in 1978

In the third chapter, I will use this epidemic model and derivation of it to combine them with the combat models. In the next section, I presented what Bayesian Network is and how it is computed.

2.1.3. Bayesian Network

Bayesian Network is a probabilistic model. This model uses graphically shown provisional probabilities between disparate variables (Fenton & Neil, 2012). Bayesian Network is generally used to make judgement under uncertain conditions. The probability of realization of the elements in the classification problem are obtained by using the probabilities of the individual components and the probable effects on the output (Nielsen & Jensen, 2009).

Bayesian Network can be practiced in lots of fields such as biology, bioinformatics, document classification, risk analysis and engineering.

Given the data, the probability of the hypothesis can be calculated with the formula 2.5.

$$\mathrm{P}(\mathrm{hypothesis}|\mathrm{data}) = \frac{\mathrm{P}(\mathrm{data}|\mathrm{hypothesis}) * \mathrm{P}(\mathrm{hypothesis})}{\mathrm{P}(\mathrm{data})} \qquad (2.5)$$

For instance, in a classification problem we want to determine the susceptibility of the system given the user profile is novice or aware on security and operating system of the

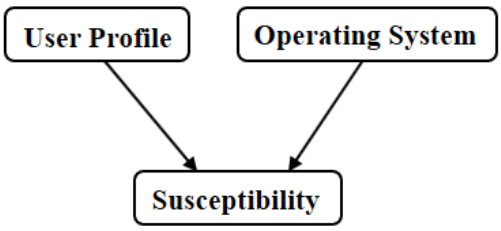
computers is either Windows or Mac OS. In the end, we will obtain the probabilistic value of how susceptible the system is.

In Figure 6, rectangles represent nodes and links between parent and child nodes show the conditional relationship.

$$P(S = \text{True}|OS = W \text{ and } UP = N) = \frac{P(OS = W|S = \text{True}) * P(UP = N|S = \text{True}) * P(S = \text{True})}{P(OS = W) * P(UP = N)}$$

$$P(S = \text{True}|OS = W \text{ and } UP = N) = \frac{\frac{28}{34} * \frac{23}{34} * \frac{34}{80}}{\frac{40}{80} * \frac{40}{80}} = \%94.7$$

As a result, given the operating system is windows and user profile is novice, the probability of the system's susceptibility is %94.7 according to the sample data.



Operating System (OS)	User Profile (UP)	Susceptible (S)	
		True (T)	False (F)
Windows (W)	Aware on Security (A)	10/34	10/46
Windows (W)	Novice (N)	18/34	2/46
Mac (M)	Aware on Security (A)	1/34	19/46
Mac (M)	Novice (N)	5/34	15/46

Figure 6: Bayesian Network Example

In this thesis, Bayesian Network is performed to display the probabilistic result of combat outcome between two forces which have cyber capabilities. There are binary outcomes; either Blue wins (0) or Red wins (1). The rate of transition among cyber-relevant compartments are included to compute the effects of them on the battle outcome. In the following section, I presented mathematical models for simulating malware propagation.

2.2. Mathematical Models for Simulating Malware Propagation

There are mathematical models designed to compute the spread of biological virus. Epidemic models consist of a few phases connected to each other according to the typical feature of the diseases. Total population is split into different groups as shown in Figure 7 (del Rey, 2015).

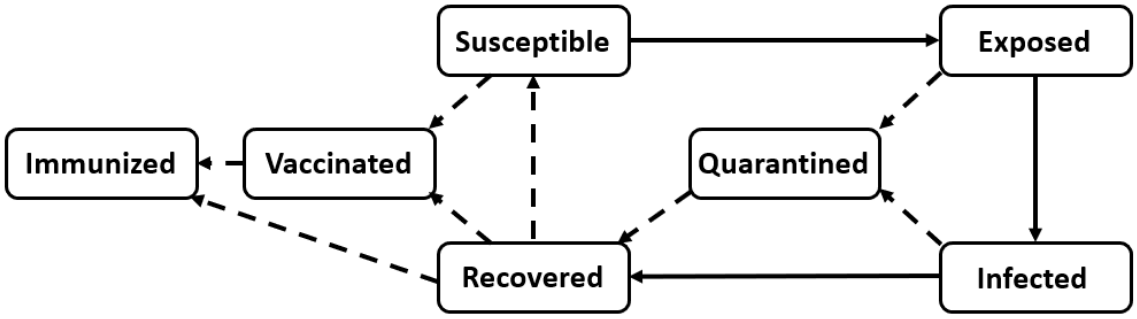


Figure 7: Diagram of Transition Among Different Compartments

Inspired by these models, similar models were used to compute the spread of computer viruses. Group definitions are made as follows: *Susceptible* is for the computers in which there is not an infected file or process; *Exposed* is for the computers in which there is a virus but that virus in inactive state until user's execution. *Infected* is for the computers in which there is an active virus and this virus has the capacity to spread throughout the network, thus damages other computers. *Quarantined* is for the computers in which there is a virus, but these computers are isolated from the network. *Recovered* is for the computers in which all the viruses are deleted and damaged parts are completely repaired. When a computer is in *Recovered* state, it means that all new patches have been made and all updates have been installed.

Lanchester aimed fire model and specific malware infection model are combined to display the outcome of the cyber effects during a military operation. The compartments used in mixed models are *Susceptible*, *Exposed*, *Infected*, *Quarantined* and *Recovered*. The parameters used in the models are the rate of transition from one state to another, the number of computers removed or added to the system.

2.2.1. SIR Model

The changes in the number of units can be calculated with the differential equations 2.6 and the transition of the compartments can be seen in Figure 8.

$$\frac{dS_B}{dt} = (-\varepsilon_B S_B I_B - \eta_B S_B R_B) - [\rho_U (S_Z + R_Z) + \rho_D I_Z] \frac{S_B}{S_B + I_B + R_B}$$

$$\frac{dI_B}{dt} = (\varepsilon_B S_B I_B - \eta_B I_B R_B) - [\rho_U (S_Z + R_Z) + \rho_D I_Z] \frac{I_B}{S_B + I_B + R_B}$$

$$\frac{dR_B}{dt} = (\eta_B S_B R_B + \eta_B I_B R_B) - [\rho_U (S_Z + R_Z) + \rho_D I_Z] \frac{R_B}{S_B + I_B + R_B} \quad (2.6)$$

$$\frac{dS_Z}{dt} = (-\varepsilon_Z S_Z I_Z - \eta_Z S_Z R_Z) - [\beta_U (S_B + R_B) + \beta_D I_B] \frac{S_Z}{S_Z + I_Z + R_Z}$$

$$\frac{dI_Z}{dt} = (\varepsilon_B S_Z I_Z - \eta_Z I_Z R_Z) - [\beta_U (S_B + R_B) + \beta_D I_B] \frac{I_Z}{S_Z + I_Z + R_Z}$$

$$\frac{dR_Z}{dt} = (\eta_Z S_Z R_Z + \eta_Z I_Z R_Z) - [\beta_U (S_B + R_B) + \beta_D I_B] \frac{R_Z}{S_Z + I_Z + R_Z}$$

The inside of the brackets at beginning of each equation demonstrates the cyber effects on the overall changes in the number of units. The inside of the square brackets in the middle of each equation demonstrates the kinetic effects of the warring forces. The fraction at the end of each equation is for the kinetic effect of the forces on the specific parts of the enemy according to their *Susceptible*, *Infected* or *Recovered* state.

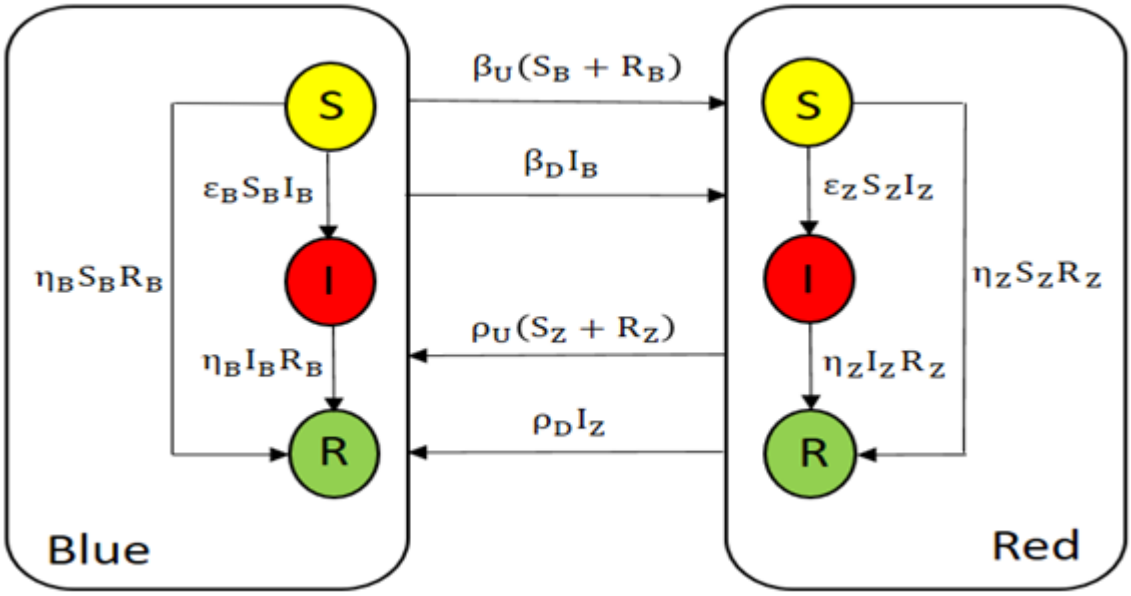


Figure 8: SIR Model

B : Blue forces
Z : Red forces
S : The number of *Susceptible*
I : The number of *Infected*
R : The number of *Recovered*
ε : Infection spread rate
η : Patch rate
ρ_U : Normal kinetic attack rate of Red forces
ρ_D : Diminished kinetic attack rate (because of infection) of Red forces
β_U : Normal kinetic attack rate of Blue forces
β_D : Diminished kinetic attack rate (because of infection) of Blue forces

2.2.2. SEIQR Model

The changes in the number of units can be calculated with the differential equations 2.7 and the transition of the compartments can be seen in Figure 9.

$$\frac{dS_B}{dt} = (A_B - c_B S_B I_B - d_B S_B + \eta_B R_B) - [\rho_U(S_Z + E_Z + R_Z) + \rho_D(Q_Z + I_Z)]\frac{S_B}{S_B + E_B + I_B + Q_B + R_B}$$

$$\frac{dE_B}{dt} = (c_B S_B I_B - (d_B + \mu_B)E_B) - [\rho_U(S_Z + E_Z + R_Z) + \rho_D(Q_Z + I_Z)]\frac{E_B}{S_B + E_B + I_B + Q_B + R_B}$$

$$\frac{dI_B}{dt} = (\mu_B E_B - (d_B + \alpha_B + \Upsilon_B + \delta_B)I_B) - [\rho_U(S_Z + E_Z + R_Z) + \rho_D(Q_Z + I_Z)]\frac{I_B}{S_B + E_B + I_B + Q_B + R_B}$$

$$\frac{dQ_B}{dt} = (\delta_B I_B - (d_B + \alpha_B + \varepsilon_B)Q_B) - [\rho_U(S_Z + E_Z + R_Z) + \rho_D(Q_Z + I_Z)]\frac{Q_B}{S_B + E_B + I_B + Q_B + R_B}$$

$$\frac{dR_B}{dt} = (\Upsilon_B I_B + \varepsilon_B Q_B - (d_B + \eta_B)R_B) - [\rho_U(S_Z + E_Z + R_Z) + \rho_D(Q_Z + I_Z)]\frac{R_B}{S_B + E_B + I_B + Q_B + R_B} \quad (2.7)$$

$$\frac{dS_Z}{dt} = (A_Z - c_Z S_Z I_Z - d_Z S_Z + \eta_Z R_Z) - [\beta_U(S_B + E_B + R_B) + \beta_D(Q_B + I_B)]\frac{S_Z}{S_Z + E_Z + I_Z + Q_Z + R_Z}$$

$$\frac{dE_Z}{dt} = (c_Z S_Z I_Z - (d_Z + \mu_Z)E_Z) - [\beta_U(S_B + E_B + R_B) + \beta_D(Q_B + I_B)]\frac{E_Z}{S_Z + E_Z + I_Z + Q_Z + R_Z}$$

$$\frac{dI_Z}{dt} = (\mu_Z E_Z - (d_Z + \alpha_Z + \Upsilon_Z + \delta_Z)I_Z) - [\beta_U(S_B + E_B + R_B) + \beta_D(Q_B + I_B)]\frac{I_Z}{S_Z + E_Z + I_Z + Q_Z + R_Z}$$

$$\frac{dQ_Z}{dt} = (\delta_Z I_Z - (d_Z + \alpha_Z + \varepsilon_Z)Q_Z) - [\beta_U(S_B + E_B + R_B) + \beta_D(Q_B + I_B)]\frac{Q_Z}{S_Z + E_Z + I_Z + Q_Z + R_Z}$$

$$\frac{dR_Z}{dt} = (\Upsilon_Z I_Z + \varepsilon_Z Q_Z - (d_Z + \eta_Z)R_Z) - [\beta_U(S_B + E_B + R_B) + \beta_D(Q_B + I_B)]\frac{R_Z}{S_Z + E_Z + I_Z + Q_Z + R_Z}$$

E : The number of *Exposed*
Q : The number of *Quarantine*
A : The number of new computers added to the network at each moment in time.
d : The rate of removal of computers from the network owing to causes not due to malware
μ : The rate of passage from the *Exposed* state to the *Infected* state
δ : The rate of passage from the *Infected* state to the *Quarantine* state
c : The rate of removal of a computer from the network owing to the action of malware
α : The rate of removal of a computer which is in the *Infected* or *Quarantine* state from the network due to the action of malware
Υ : The rate of recovery due to the action of the antivirus software
ε : The rate of passage from *Infected* or *Quarantined* state to the *Recovered* state

η : The rate of the loss of immunity

The inside of the brackets at beginning of each equation demonstrates the cyber effects on the overall changes in the number of units. The inside of the square brackets in the middle of each equation demonstrates the kinetic effects of the warring forces. The fraction at the end of each equation is for the kinetic effect of the forces on the specific parts of the enemy according to their *Susceptible*, *Exposed*, *Infected*, *Quarantine* or *Recovered* state.

2.2.3. SEIR Model

The changes in the number of units can be calculated with the differential equations 2.8 and the transition of the compartments can be seen in Figure 10.

$$\frac{dS_B}{dt} = (-\frac{\beta_B \alpha_B}{N_B} E_B S_B - \psi 1_B S_B + \phi_B R_B) - [\beta_U (S_Z + E_Z + R_Z) + \beta_D I_Z] \frac{S_B}{S_B + E_B + I_B + R_B}$$

$$\frac{dE_B}{dt} = (\frac{\beta_B \alpha_B}{N_B} E_B S_B - (\alpha_B + \psi 2_B) E_B) - [\beta_U (S_Z + E_Z + R_Z) + \beta_D I_Z] \frac{E_B}{S_B + E_B + I_B + R_B}$$

$$\frac{dI_B}{dt} = (\alpha_B E_B - (Y_B + \Theta_B) I_B) - [\beta_U (S_Z + E_Z + R_Z) + \beta_D I_Z] \frac{I_B}{S_B + E_B + I_B + R_B}$$

$$\frac{dR_B}{dt} = (\mu_B N_B + \psi 1_B S_B + \psi 2_B E_B + Y_B I_B - \phi_B R_B) - [\beta_U (S_Z + E_Z + R_Z) + \beta_D I_Z] \frac{R_B}{S_B + E_B + I_B + R_B} \quad (2.8)$$

$$\frac{dS_Z}{dt} = (-\frac{\beta_Z \alpha_Z}{N_Z} E_Z S_Z - \psi 1_Z S_Z + \phi_Z R_Z) - [\rho_U (S_B + E_B + R_B) + \rho_D I_B] \frac{S_Z}{S_Z + E_Z + I_Z + R_Z}$$

$$\frac{dE_Z}{dt} = (\frac{\beta_Z \alpha_Z}{N_Z} E_Z S_Z - (\alpha_Z + \psi 2_Z) E_Z) - [\rho_U (S_B + E_B + R_B) + \rho_D I_B] \frac{E_Z}{S_Z + E_Z + I_Z + R_Z}$$

$$\frac{dI_Z}{dt} = (\alpha_Z E_Z - (Y_Z + \Theta_Z) I_Z) - [\rho_U (S_B + E_B + R_B) + \rho_D I_B] \frac{I_Z}{S_Z + E_Z + I_Z + R_Z}$$

$$\frac{dR_Z}{dt} = (\mu_Z N_Z + \psi 1_Z S_Z + \psi 2_Z E_Z + Y_Z I_Z - \phi_Z R_Z) - [\rho_U (S_B + E_B + R_B) + \rho_D I_B] \frac{R_Z}{S_Z + E_Z + I_Z + R_Z}$$

N : The total number (S+E+I+R)
β : The contact rate
α : The rate of passage from the *Exposed* state to the *Infected* state
ψ1 : The rate of passage from the *Susceptible* state to the *Recovered* state
ψ2 : The rate of passage from the *Exposed* state to the *Recovered* state
Y : The rate of passage from the *Infected* state to the *Recovered* state
ϕ : The rate of passage from the *Recovered* state to the *Susceptible* state
Θ : The dysfunctional rate
μ : The replacement rate

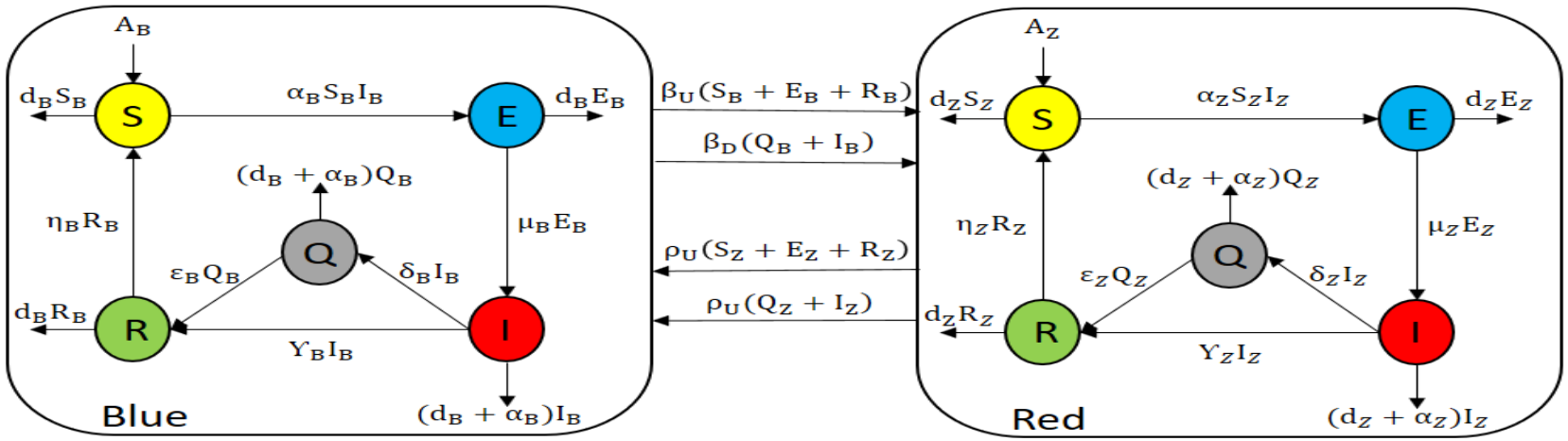


Figure 9: SEIQR Model

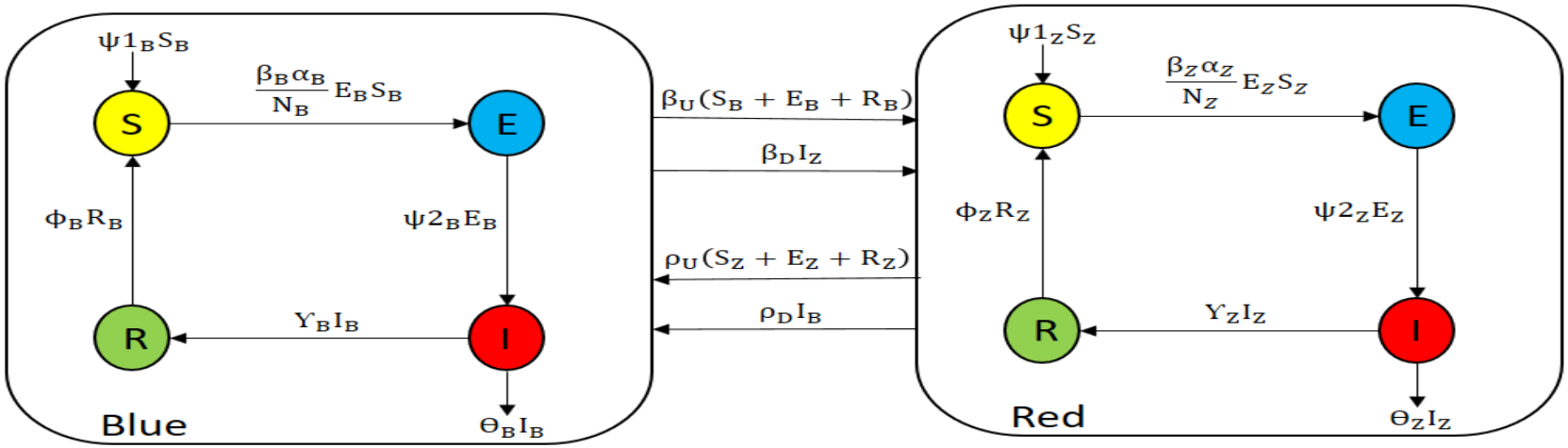


Figure 10: SEIR Model

The inside of the brackets at beginning of each equation demonstrates the cyber effects on the overall changes in the number of units. The inside of the square brackets in the middle of each equation demonstrates the kinetic effects of the warring forces. The fraction at the end of each equation is for the kinetic effect of the forces on the specific parts of the enemy according to their *Susceptible*, *Exposed*, *Infected* or *Recovered* state.

I wrote my python code in order to execute the differential equations used in SIR, SEIQR and SEIR models and used the results throughout my thesis.

I used the diagram in Figure 7 as the main framework and analyzed mostly encountered nine types of malware, their propagation in a computer network and obtained three forms of models, namely SIR, SEIQR and SEIR. In the next section the characteristics of these malware and their compatibility with the models are shown.

2.3. Types of Malware and Their Compatibility with the Models

The most common types of cyber-attacks seen throughout the world are denial of service, malicious codes, viruses, worms and trojans, malicious insiders, stolen devices, phishing and social engineering, web-based attacks. The aim of these attacks is cyber-crime, cyber espionage, cyber war and hacktivism (Bendovschi, 2015).

It is difficult to make a concise taxonomy of malware. When an attack happens, there is always a probability that it uses several attack vectors in order to accomplish that cyber-attack. According to the classification by attack vector, we can enumerate misconfiguration, kernel flaws, buffer overflow, insufficient authentication validation, insufficient input validation, symbolic links, file descriptor, race condition, incorrect file/directory permission and social engineering. When we try to make the classification by operational impact, we can enumerate misuse of resources, user compromise, root compromise, web compromise, installed malware (virus, spyware, trojan, worm, arbitrary code execution) and denial of service (Simmons, Ellis, Shiva, Dasgupta & Wu, 2015).

Bearing the features of malware such as functionality, spread type and their damage in mind, we can categorize malware in various types: computer viruses, worms, trojans, rootkits, spywares, ransomwares and so on (del Rey, 2015).

I chose the most popular types of malware considered above and analyzed their propagation over the network. Then I placed them in a particular model as shown in Table 1.

Virus, worm, trojan horse and ransomware can be categorized in SEIQR Model. Initially all the computers are in *Susceptible* group. When attacker tries to infect computers in the system via removable devices, SMS, email or links, those computers turn into *Exposed* group. If users do the attacker's directions and trigger the malicious file or execute the malicious codes, then those computers turn into *Infected* group. In the meantime, antivirus

programs already installed in computers (if correctly patched and updated) may activate and take those malware to *Quarantine*. After that, system admins may detect the malware in the network and try to delete them and save the damaged parts of the devices. In that case, those computers start to take part in the *Recovered* group.

Similar scenario happens in rootkit, DoS and replay attacks. These attack types can be categorized in SIR Model. Taken into account the characteristics of these attacks, there is no *Exposed* or *Quarantine* group. Because users do not trigger anything to start the malicious codes and antivirus software does nothing to take those methods to *Quarantine*.

Finally, impersonation and social engineering attacks can be categorized in SEIR Model. Same conditions are also valid for these types of attacks. By using the phishing email for example, attacker can deceive the user to click on a fake link and try to compromise the computer. In this case those computers are in *Exposed* group. If user is fooled by attacker, then his computer turns into *Infected* group. Having cleaned from the malware, that computer will be in the *Recovered* group.

Table 1: Types of Malware and Their Compatibility with the Models

No	Type of Malware	Model
1	Virus	SEIQR
2	Worm	SEIQR
3	Trojan Horse	SEIQR
4	Ransomware	SEIQR
5	Rootkit	SIR
6	DOS	SIR
7	Replay	SIR
8	Impersonation	SEIR
9	Social Engineering	SEIR

The table above is an assumption list. I assumed that these malware types fit into the models as shown. However, different types of malware can be categorized in various models and can be studied in future work.

2.4. Relevant Work

Lanchester models have been applied to compute the result of the wars since 1916. The Battle of Iwo Jima took place towards the end of the World War II on a small island in the Pacific Ocean. It was between Japanese and American forces. American forces landed 54000 troops on the first day, 6000 on the third day and 13000 on the sixth day. The kill rate of Americans was 0.0113 and the kill rate of Japanese was 0.0088. 21000 Japanese soldiers were killed during the battle. These values were used in Lanchester aimed fire model and the results were very close to the real results as shown in Figure 11 (Engel, 1954). When analyzing mixed epidemic combat models in chapter 3, I used aimed fire model as combat model.

In the Battle of Kursk, Germans attacked Soviets in July and August 1943 during the WWII. The number of the units, as well as attrition rate values were applied to the Lanchester combat models. It is really hard to put complex variables into simple Lanchester model. Many significant elements such as equipment, leadership, training, morale and weather cannot be added to the model. However, it was found that the fighting combat units data fitted to the Lanchester equations. Even so, more present-day and complex values should be added to calculation in order to make better conclusions about the outcome of the Lanchester models (Lucas & Dinges, 2004). It is difficult to add the values mentioned here such as equipment, leadership, training, morale and weather. I also couldn't take into account the effects of these elements in this thesis.

Fighting forces change their strategy and tactics according to the phases of the war. For that reason, Lanchester models need to be revised. The results obtained from the Lanchester models are only accurate for some specific situations because the outcome on each time period during the war changes dynamically and exponentially. This situation clarifies why Lanchester models sometimes fit well, but sometimes not fit to the real results. Therefore, arbitrary parameter adaptation should be applied to the combat model. A novel theory which depends on dynamic strategic learning provides exponential loss and exchange rates and presents better results (Duffey, 2017). I performed Lanchester aimed fire model in this thesis. Dynamic strategic learning model can be studied in the future work.

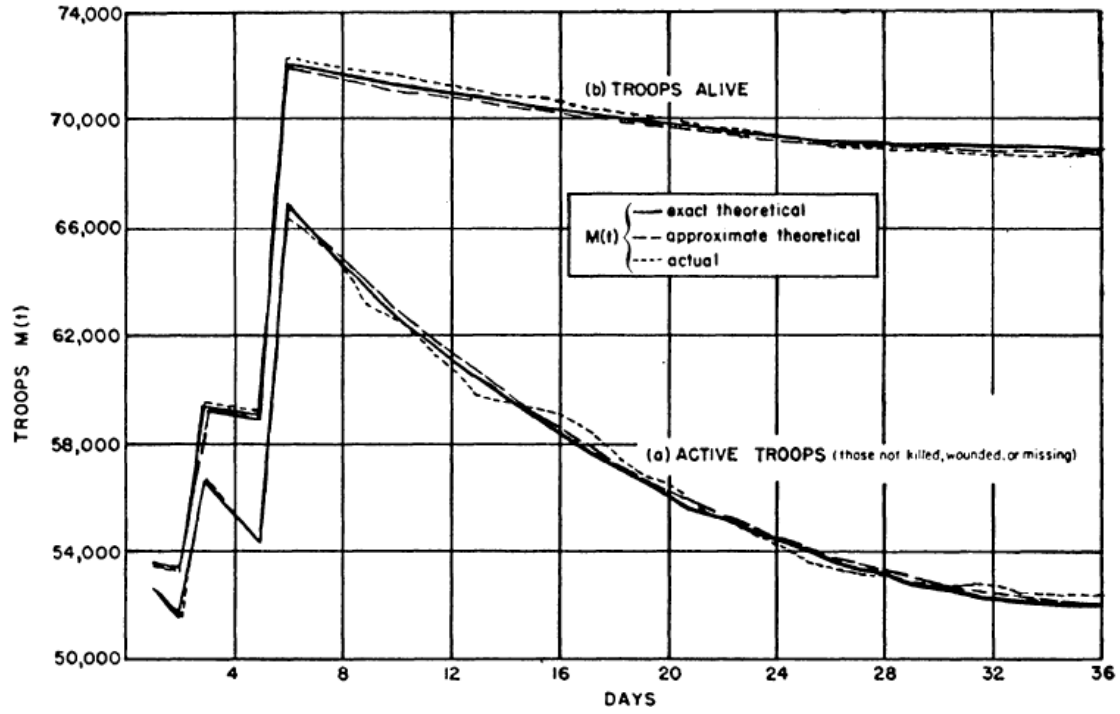


Figure 11: The Battle of Iwo Jima Example

Intelligence is an important factor for two sides of the fighters. However, its effect hasn't been taken into account during the calculations of the war outcomes. Lanchester models provide significant clues about the results of the war but lack from intelligence values. When intelligence values are added to the Lanchester equations, it presents that the intelligence performed as a force multiplier (Coulson, 2018). In this thesis, intelligence values didn't take into account when calculating the model outcome. The effects of the intelligence on the combat outcome can be studied in the future work.

Lanchester models were used in order to compute the outcome of the battles in 21st century. Some of them fit the model output, but some of them did not. Because Lanchester models make some assumptions at the beginning, which considers the forces as homogeneous. In addition, model does not give permission for changes in unit type. But it was emerged that the outcome of these models brings good result. The Battle of Britain was the first battle in which only aircrafts fought with each other. It was between England's royal air forces and Germany's air force in 1940. The model provided good result. In fact, Lanchester's ideas were well accepted by military decision makers and constituted the significant part of the mental background calculations of the WWII tactics (Johnson, & MacKay 2011). The models used in this thesis assume the fighting forces as they are homogeneous.

The Battle of Kursk was used very much in the research papers on Lanchester models. Because the losses of each side, that is Germans and Soviet Russians, had been recorded and known. Until now, the Lanchester model were used to compute the homogenous

losses (i.e. tanks to tanks or artillery to artillery). However, in Das's study both tank and artillery losses were taken into account while using Lanchester model. In order to validate the heterogenous model, he used R^2, sum-of-square-residuals, root-mean-square error, Kolmogorov-Smirnov and chi-square statistics methods. He found out rational enough and efficient estimates (Das, 2019). Heterogeneous fighting forces can be modeled to use in the future work.

Malware attacks along with the kinetic attacks in a battle may provide triumph for the forces smaller in number. Malware attacks can change the combat outcome. Cyber capabilities can damage the kinetic superiority of the enemy. Malware usage during the kinetic attacks and employing cyber warfare are needed little attempt but provide important destruction. Lanchester model was used to analyze a sample fictitious system which describes a conventional homogenous battle between Red and Blue forces. SIR model was used to represent malware propagation. When Blue starts malware attack, Red side needs time to understand the attack and begin to take precaution to stop and recover it. The model has an extension to represent this period of time in order to fit the model more likely to real-time scenarios. Above all, it should be noted that real-life comparison data are not reachable (Draeger & Öttl, 2018). In this thesis, it is assumed that the cyber-attack starts at the same time with the kinetic attack. In the future work, the situation that the cyber-attack starts earlier or later than the kinetic attack can be studied.

The effects of cyber war can be felt and these effects can be computed with mathematical models. There are three participants in the model. X and Y states are fighting sides and Z state is the peacemaker side. In the first phase, X employs cyber attack to Y. In the second phase, X diminishes its cyber attack because of the negotiations done with Z. In the third phase, all cyber attacks are excluded. The output of the system with respect to time was evaluated. According to the simulated results, a stability analysis of the model was presented (Mishra & Prajapati, 2013). In this thesis, cyber-attack starts at the same time with the kinetic attack and lasts until the end of the war.

Susceptible-Infectious-Removed-Susceptible (SIRS) model was used to simulate the virus and worm propagation. Once the virus infiltrates the computer network, the *Susceptible* computers turn into Infectious. Until the antivirus software is run, virus goes around in the network. This period is taken into consideration as a latent period and computed as well. In addition, a clear formula was derived to indicate the infection-free equilibrium (R). If $R<0$, it means that the virus or worm does not spread in the network, because recovery is faster than infection. On the other hand, if $R>0$, then it means that the virus or worm spread and damage the network (Mishra & Ansari, 2012). One of the three models analyzed in this thesis is SIR model and explained in detail.

The Battle of Bulge (also known as Ardennes Campaign) was the last major German attack against Soviet Russians in 1944. In order to compute the output of the battle, Lanchester models were tried. In fact, it is difficult to validate the Lanchester model, because there is no available historical data to use. The model is homogenous. The casualties in the number of tanks, armored personal carrier, artillery and manpower are

weighted and added to the model. This model used five parameters: Russian individual effectiveness, German individual effectiveness, exponent of shooting force, exponent of target force, and a tactical parameter reflecting which side is defending and attacking. The effectiveness of the Russians was better than that of Germans. According to the data of 10-days of the battle, the result obtained from the research is that Lanchester linear model fits the historical data (Bracken, 1995). I performed Lanchester aimed fire model in this thesis.

Bracken's study on the Battle of Bulga was improved to better fit Lanchester model to the historical data by using linear regression. All the available data throughout the campaign were used. Moreover, air sortie data were added. The result of the study indicates that neither Lanchester linear model nor square model are proper to the historical data. Instead, a revised Lanchester equations emerged. In order to estimate the parameters in the model, linear regression was applied to the logarithmically converted Lanchester equations. Furthermore, not 10-days, but full 32-days data of the battle were used. Lanchester logarithmic model indicates that the attrition parameter which was used to calculate the kill rate of the forces is valid. Boosted fire capability of the forces in time was computed in the model (Fricker, 1998). The assumption done in the models studied in this thesis at the beginning of the battle does not change during the war.

In Bracken's study, it was figured out that the Lanchester linear model was fit well for the Battle of Ardennes. But, in Fricker's analysis, it was shown that the Lanchester linear model did not fit. He indicated that Lanchester logarithmic laws were fit to that battle. Bayesian framework can be used to make predictions about the outcome of the battle. If previous data from the battle are taken into account, predictions about mortality can be made. Bayesian framework helps us make predictions about future losses. Evaluating the one day losses of the forces in Ardennes campaign, predictions for the next days can be done according to the Bayesian model. In addition, this model can be used to predict the casualties in real battles (Wiper, Pettit & Young, 2000). I studied Bayesian Network Framework to make predictions on the outcome of the battle between two forces which has cyber-attack capabilities.

In order to compute the surviving number of units when three homogenous forces fight with each other, deterministic mathematical models are used. These models use the data of previous wars and are theoretical, sometimes not successful to fit the contemporary wars as well. Attrition rate indicates the power of the weapons. When three different homogenous forces fight with same attrition rates, in the end all the units of all warring forces will be dead. On the other hand, when the attrition rate of two forces are equal, but the attrition rate of other state is different, then in the end of the war, remaining units of the winning or losing side will be change according to that specific attrition value (Sfikas, 2017). I presented both the kinetic and cyber-attack attrition rates in the models in this thesis, but as for the Bayesian Network framework, I analyzed cyber-attack attrition rates in detail.

There are various models to show malware propagation, actually based on Kerman-McKendrick's SIR epidemic model. One of them is SIRA (*Susceptible-Infected-Recovered-Antidotal*) model. The novel term *Antidotal* is related to the robustness of the antivirus program installed in the computer network. *Antidotal* group carries out damage and vaccinates bordering computers. When a node is *Susceptible*, it either turns into *Infected* or *Antidotal*. Continuously when a node is *Infected*, it either turns into *Recovered* or *Antidotal*. By analyzing the results obtained from the executed formulas, designer of the network can determine defense strategies (Batistela & Piqueira, 2018). One of the three models analyzed in this thesis is SIR model, but I didn't use a compartment as *Antidotal*.

Bayesian Network approach was used to predict the combat results which was obtained from Lanchester linear and square laws. Given the data about battle, model parameters and winner of the battle can be estimated with Bayesian Network Framework. This can also be used to predict the initial number of the fighting states. Furthermore, in order to make a conclusion which Lanchester model fits well to the data about the battle, Bayesian Network approach can be used (Pettit, Wiper & Young, 2003). I studied Bayesian Network approach to make predictions on the outcome of the battle.

Situation assessment is critical, which helps decision-makers decide about anything in the continuing battle in a military perspective. It should be adaptive, that is, it changes according to the new emerging events during the battle. Bayesian Network approach was used to provide probabilistic solution on the flexible situation assessment. By using the full or partial data, we can obtain a reasonable probabilistic result about the changes in the battle surroundings (Mirmoeini & Krishnamurthy, 2005). Bayesian Network approach is performed in this thesis and it provides probabilistic estimations on the combat outcome.

In the next chapter, I studied Bayesian Network with its implementation on the SIR, SEIQR and SEIR models. I used Naïve Bayes and BayesNet classifiers and presented the probabilistic results of the sample situations.

CHAPTER 3

BAYESIAN NETWORK AND ITS IMPLEMENTATION ON THE MODELS

In this chapter, I develop and implement a Bayesian Network model for the analysis of specific malware types as represented by specific mixed epidemic models. This analysis is based on a set of assumptions for the purpose of simplicity of modeling. In particular, I assume that the kinetic effect values are constant for both sides at the beginning of the war. There are further assumptions as listed below.

- The kinetic attack rates are equal for both sides.
- Both sides have same vulnerability in their communications systems. There is only one type of infection.
- When a unit is patched, no other damage remains in that unit. A patched unit has the strength just like before the infection.
- The cyber-attack starts at the same time with the kinetic battle.

3.1. SIR Model

Rootkit, DOS and replay attacks are compatible with the SIR model.

The initial assumptions for the number of forces and their kinetic attack rates are equal because I want to analyze the cyber effects during the battle.

$S_B = S_R = 90$ $\beta_U = \rho_U = 0.1$
$I_B = I_R = 5$ $\beta_D = \rho_D = 0.01$
$R_B = R_R = 5$

The cyber effect rates are epsilon (infection spread rate) and eta (patch rate) as shown in Figure 12.

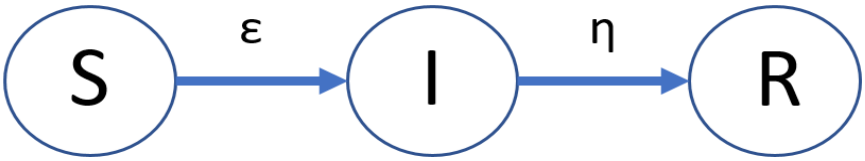


Figure 12: Cyber Effect Parameters in SIR Model

Table 2: Sample Cyber Effect Values in SIR Model

ε_B	ε_R	η_B	η_R
0.002	0.003	0.0008	0.0003

The initial number of Blue and Red forces and both kinetic attack and cyber effect rates are given above. When I execute the differential equations 2.6, I obtain a graph as seen in Figure 13. The changes in number of units among compartments for Blue side are shown in Figure 14 and the changes in number of units among compartments for Red side are shown in Figure 15. We can conclude that under such circumstances, Blue side has superiority over Red side and in the end of the war, Blue wins. I would like to state specifically that these values are in accordance with the ones used in the relevant literature on this subject.

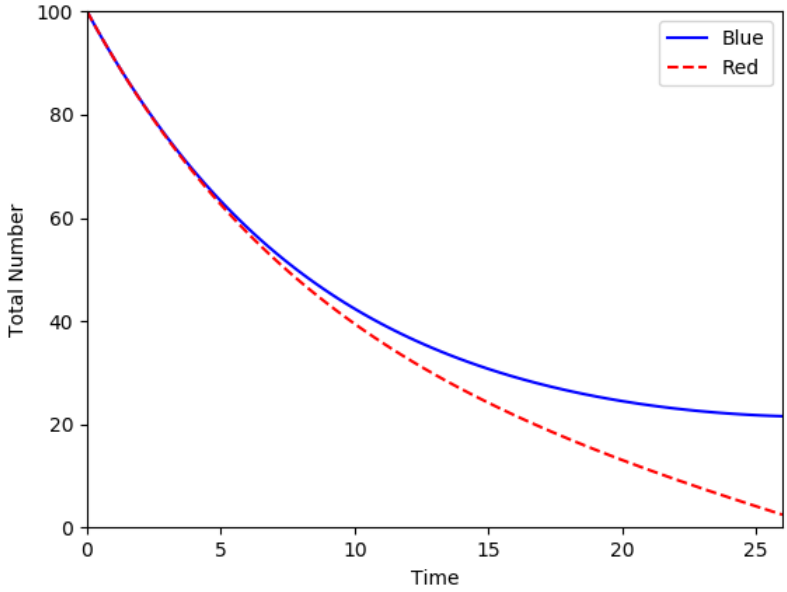


Figure 13: SIR Model Implementation

According to the values above, the malware spreads among Red units faster than Blue units. Moreover, the rate of installing patches and recovering the damaged parts is faster among Blue units than Red units.

By doing minor changes in the values of cyber effect rates, I created a set of samples by executing my python code. My set consists of 75 sample values as shown in Appendix A. According to the graph I draw in line with the output after the execution of the equations,

I can see which side has superiority over the other side. According to my classification problem there are two classes. 0 (zero) means Blue side's superiority and 1 (one) means Red side's superiority.

Conforming to the values in the dataset, I obtained a likelihood table of compartment transition rates as shown in Table 3. For example, in ε_B table (upper left corner) 0.001 value is used for 9 times. Blue side won 8 times, while Red side won 1 time. In addition, in total number of 75 trials, Blue won 38 times, while Red won 37 times. Keeping these values in mind, the likelihood that the Blue won when $\varepsilon_B = 0.001$ is 8/38, whereas the likelihood that the Red won when $\varepsilon_B = 0.001$ is 1/37.

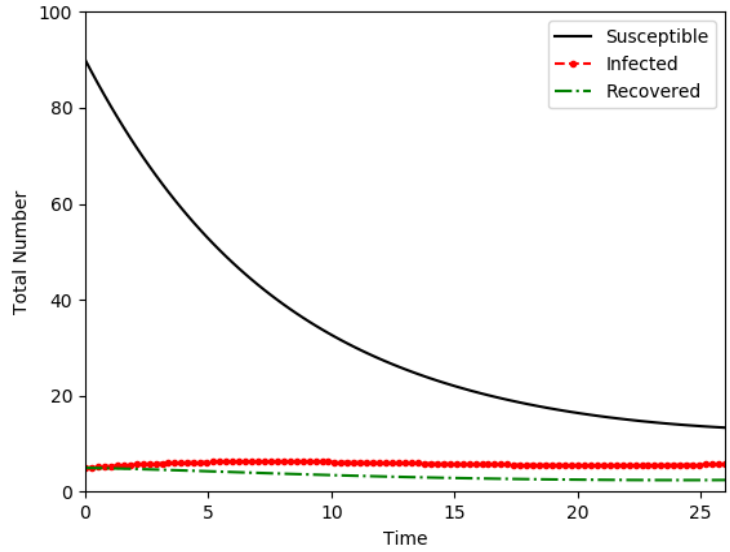


Figure 14: Compartmental Changes of Blue Units in SIR Model

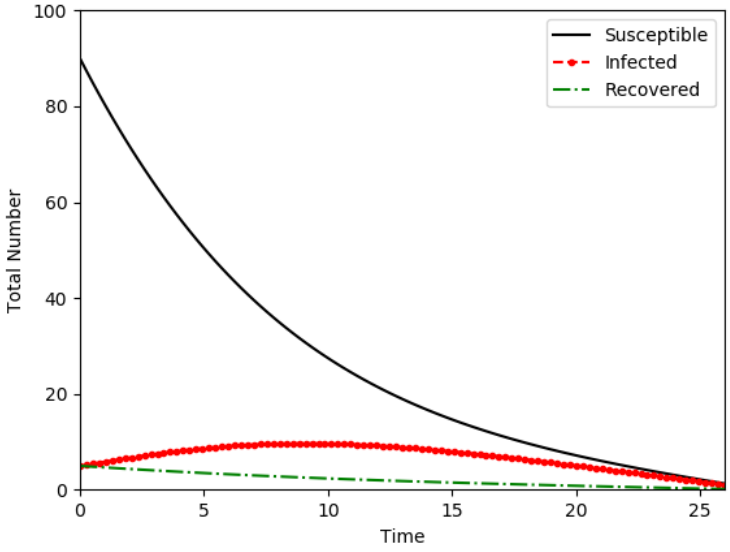


Figure 15: Compartmental Changes of Red Units in SIR Model

Although there are many values which can be assigned to the cyber effect parameters, in order to execute the model in Bayesian Network approach, I specified only a certain amount of values.

Table 3: Likelihood Table of Cyber Effect Parameters in SIR Model

Likelihood Table	Winner	
ε_B	**Blue**	**Red**
0.001	8/38	1/37
0.002	4/38	1/37
0.003	7/38	3/37
0.004	5/38	1/37
0.005	10/38	11/37
0.006	1/38	3/37
0.007	1/38	5/37
0.008	1/38	7/37
0.009	1/38	5/37

Likelihood Table	Winner	
ε_R	**Blue**	**Red**
0.001	0/38	11/37
0.002	1/38	3/37
0.003	1/38	4/37
0.004	3/38	5/37
0.005	11/38	10/37
0.006	5/38	1/37
0.007	7/38	1/37
0.008	5/38	1/37
0.009	5/38	1/37

Likelihood Table	Winner	
η_B	**Blue**	**Red**
0.001	4/38	1/37
0.002	2/38	2/37
0.003	2/38	2/37
0.004	2/38	2/37
0.005	20/38	21/37
0.006	1/38	3/37
0.007	1/38	1/37
0.008	4/38	1/37
0.009	2/38	4/37

Likelihood Table	Winner	
η_R	**Blue**	**Red**
0.001	6/38	4/37
0.002	2/38	3/37
0.003	1/38	1/37
0.004	1/38	1/37
0.005	20/38	22/37
0.006	2/38	2/37
0.007	4/38	2/37
0.008	1/38	1/37
0.009	1/38	1/37

A major contribution of this thesis is the proposal for enriching the Lanchester models and the epidemic models by embedding cyber-attacks by means of probabilistic modeling. This approach is of importance since the set of differential equations is limited to binary outcomes (i.e., the winning and losing forces as model outputs), whereas a probabilistic model provides the probability of winning and losing as the model output.

By using the values in Table 3 and Bayesian Network formula 2.5, I obtained the probability of the winning side as described below.

Hypothesis = Blue wins (BW)
Data = (ε_B=0.002, ε_R=0.003, η_B=0.0008 and η_R=0.0003)

$$P(BW|Data) = \frac{P(\varepsilon_B \mid BW) * P(\varepsilon_R \mid BW) * P(\eta_B \mid BW) * P(\eta_R \mid BW) * P(BW)}{P(Data)}$$

Given the data (ε_B=0.002, ε_R=0.003, η_B=0.0008 and η_R=0.0003) the probability of Blue Units' winning is the probability of ε_B=0.002 given BW multiplied by the probability of ε_R=0.003 given BW and multiplied by the probability of η_B=0.0008 given BW and multiplied by the probability of η_R=0.0003 given BW and multiplied by the probability of BW divided by the probability of Data.

$$P(BW|Data) = \frac{\frac{4}{38} * \frac{1}{38} * \frac{4}{38} * \frac{1}{38} * \frac{38}{75}}{\frac{5}{75} * \frac{5}{75} * \frac{5}{75} * \frac{2}{75}} = 0.492$$

When I change the hypothesis in favor of Red's winning, similar calculations can be done as follows.

Hypothesis = Red wins (RW)
Data = (ε_B=0.002, ε_R=0.003, η_B=0.0008 and η_R=0.0003)

$$P(RW|Data) = \frac{P(\varepsilon_B \mid RW) * P(\varepsilon_R \mid RW) * P(\eta_B \mid RW) * P(\eta_R \mid RW) * P(RW)}{P(Data)}$$

Given the data (ε_B=0.002, ε_R=0.003, η_B=0.0008 and η_R=0.0003) the probability of Red Units' winning is the probability of ε_B=0.002 given RW multiplied by the probability of ε_R=0.003 given RW and multiplied by the probability of η_B=0.0008 given RW and multiplied by the probability of η_R=0.0003 given RW and multiplied by the probability of RW divided by the probability of Data.

$$P(RW|Data) = \frac{\frac{1}{37} * \frac{4}{37} * \frac{1}{37} * \frac{1}{37} * \frac{37}{75}}{\frac{5}{75} * \frac{5}{75} * \frac{5}{75} * \frac{2}{75}} = 0.133$$

Probability of Blue wins = 0.492
Probability of Red wins = 0.133
Sum of Probabilities = 0.625

Likelihood of Blue wins = 0.492/0.625 = %78.69
Likelihood of Red wins = 0.133/0.625 = %21.31

Finally, we can conclude that given the values ε_B=0.002, ε_R=0.003, η_B=0.0008 and η_R=0.0003, Blue side wins the battle with %78.69 probability.

In order to analyze my sample data set, I used WEKA 3.8 software. WEKA (Waikato Environment for Knowledge Analysis) is an open source and publicly available tool, which provides a great number of classification methods for data mining (Kabakchieva, 2013). I used two Bayesian classifiers in WEKA, which are NaïveBayes and BayesNet. Both classifiers are evaluated for 10-fold cross validation. 10-fold cross validation means that the algorithm runs 10 times and each time 9/10 of dataset is used for training and remaining 1/10 of dataset is used for testing.

The basic difference between NaïveBayes and BayesNet is that in NaïveBayes algorithm an individual attribute has effects on the overall result independently, whereas in BayesNet algorithm attributes which depend on another attributes all together have effects on the overall result.

I executed WEKA with my dataset in Appendix A and obtained the confusion matrix as shown in Table 4 and results in Table 5. Confusion matrix is a table, which shows how many instances in the dataset are predicted as True or False. It is actually the summary of the classification problem and provides significant information about the confusion the classification model made.

Table 4: Confusion Matrix in SIR Model

Confusion Matrix		NaïveBayes		BayesNet	
		Real Values		Real Values	
		0	1	0	1
Predicted Values	0	31	7	24	14
	1	7	30	11	26
Total Number of Instances		75		75	
Correctly Classified Instances		61		50	
Accuracy		% 81.3		% 66.6	

According to the confusion matrix done with NaïveBayes, when the actual class is Blue (0), the predicted class is also Blue for 31 times (True Positive), but assessed wrongly for 7 times (False Negative). In addition, when the actual class is Red (1), the predicted class is also Red for 30 times (True Negative), but assessed wrongly for 7 times (False Positive). Furthermore, according to the confusion matrix done with BayesNet, when the actual class is Blue (0), the predicted class is also Blue for 24 times (True Positive), but assessed wrongly for 11 times (False Negative). In addition, when the actual class is Red (1), the

predicted class is also Red for 26 times (True Negative), but assessed wrongly for 14 times (False Positive).

61 out of 75 instances are classified correctly in NaïveBayes and the accuracy is very high (% 81.3). On the other hand, 50 out of 75 instances are classified correctly in BayesNet and the accuracy is good enough (% 66.6).

TP rate (also known as sensitivity or recall) and Precision (also known as positive predictive value) statistically indicates the performance of binary classification tests.

Table 5: Results for the Bayesian Classifiers in SIR Model

Class	NaïveBayes		BayesNet	
	TP Rate	Precision	TP Rate	Precision
0	0.816	0.816	0.632	0.686
1	0.811	0.811	0.703	0.650
Weighted Average	0.813	0.813	0.667	0.668

TP rate shows the ratio of how many instances were predicted positive when they are actually positive. It is calculated by dividing TPs to the sum of the TPs and FNs. On the other hand, Precision shows the ratio of how many instances were actually predicted positive when they are positive. It describes how good a model is at predicting the positive class and is calculated by dividing TPs to the sum of the TPs and FPs.

In NaïveBayes classifier TP rate and Precision for class 0 and 1 are 0.816 and 0.811 respectively. These ratios indicate that the model performs very well. Besides, in BayesNet classifier TP rate for class 0 and 1 is 0.632 and 0.703, while Precision for class 0 and 1 are 0.686 and 0.650 respectively. Not high as much as the ratio in NaïveBayes, but still these ratios indicate that the model performs well enough.

ROC (Receiver Operating Characteristic) Curve is used to demonsrate the performance of a classifier. It is a graph which X-axis shows FP, while Y-axis shows TP. ROC Curve which is close to the upper left corner indicates that the classifier is good and performs well enough to separate the classes. If that curve is far to the upper left corner, it means that the classifier is not good at prediction.

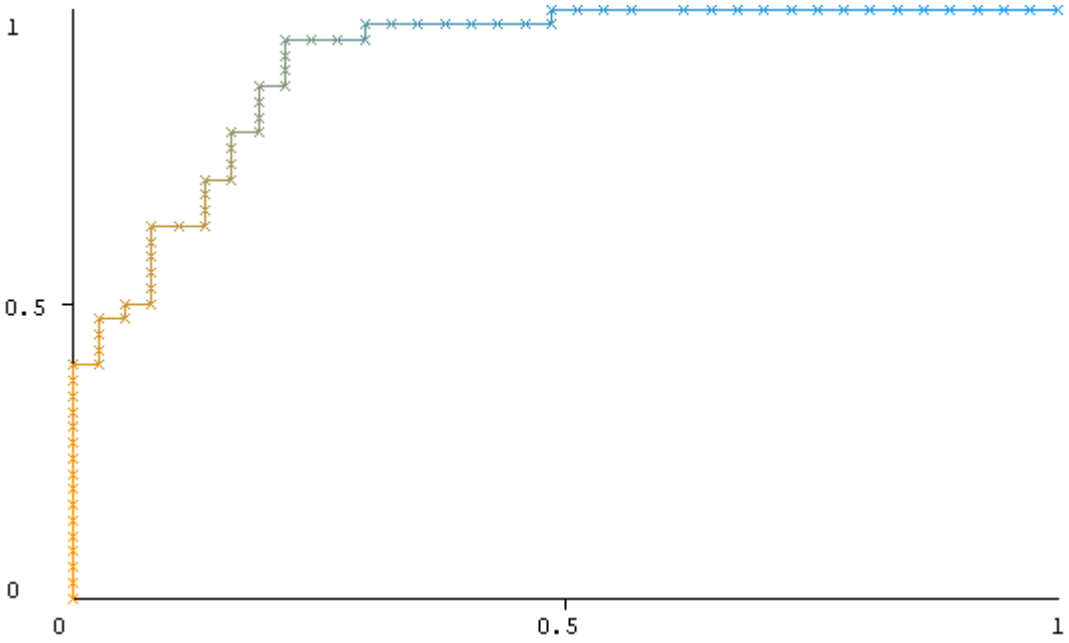


Figure 16: ROC Curve for NaïveBayes in SIR Model

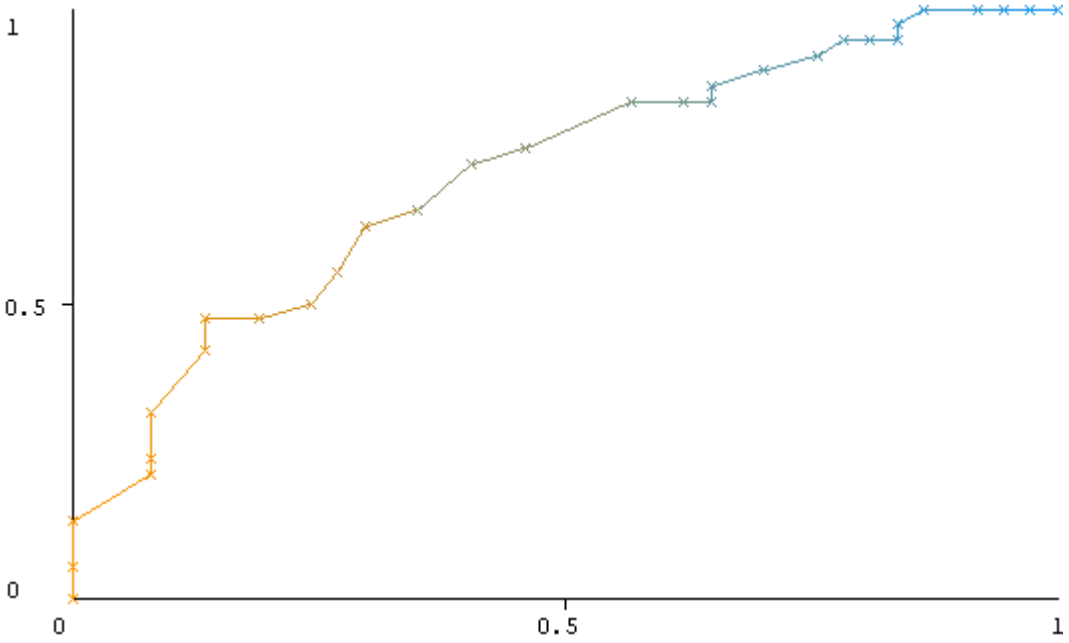


Figure 17: ROC Curve for BayesNet in SIR Model

ROC Curve for NaïveBayes is presented in Figure 16 and ROC Curve for BayesNet is presented in Figure 17. We can analyze that NaïveBayes classifier classifies our dataset better than BayesNet classifier. Furthermore, we can figure out that our model performs well.

Precision-Recall Curve (PRC) is another graph used also to show the performance of a classifier. There exists recall values in X-axis and precision values in Y-axis. The difference is that there is a no-skill line on that graph. It is calculated by dividing total number of positive cases to the sum of total number of positive and negative cases. The curve above this line indicates the skill of the model. The more to the upper right corner the PRC is, the better the classifier is.

PRC for NaïveBayes is presented in Figure 18 and PRC for BayesNet is presented in Figure 19. No-skill line is 0.51. We can analyze that both NaïveBayes and BayesNet classifier performs above the no-skill line. In addition, we can declare that NaïveBayes classifier performs better than BayesNet classifier.

By evaluating the results above, we can make a conclusion that SIR model performs well enough and can be used for further probabilistic predictions.

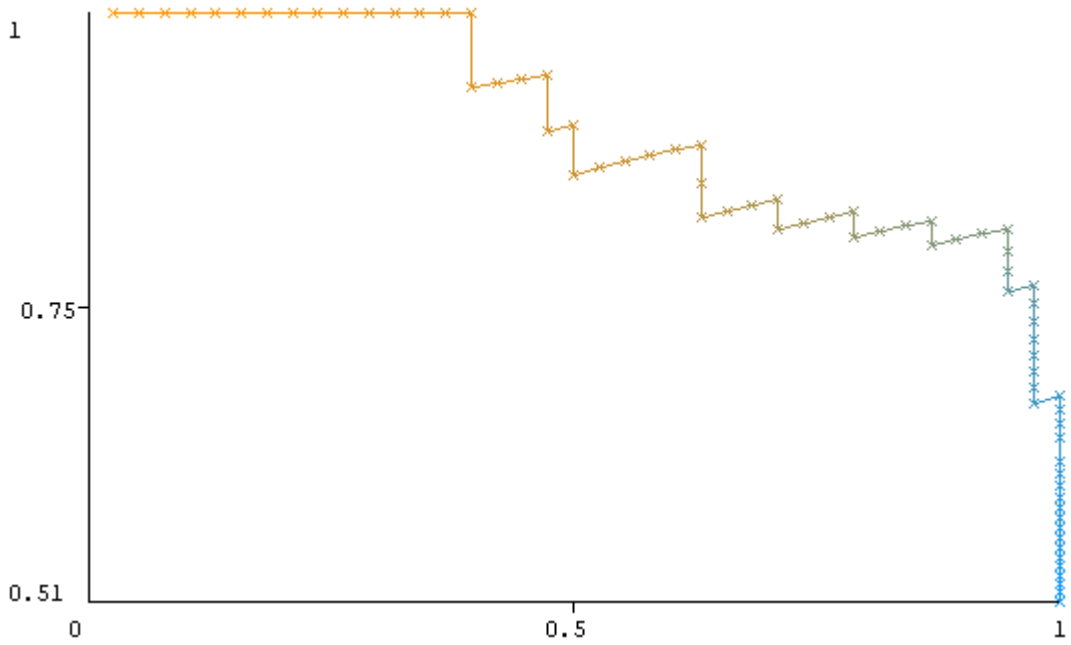


Figure 18: PRC for NaïveBayes in SIR Model

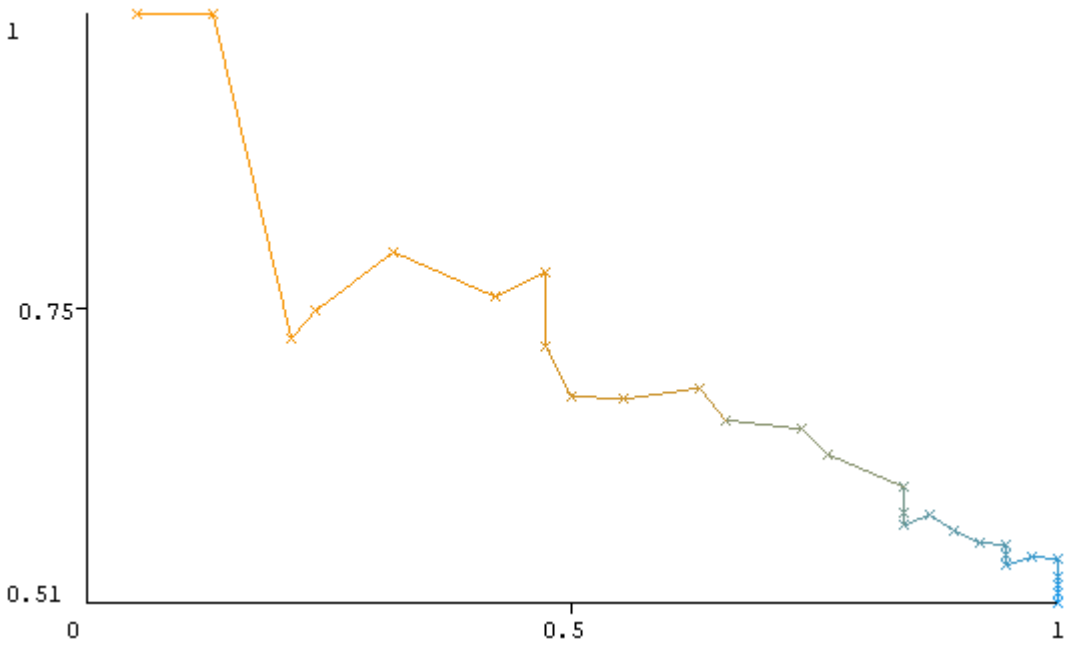


Figure 19: PRC for BayesNet in SIR Model

3.2. SEIQR Model

Virus, worm, trojan horse and ransomware attacks are compatible with the SEIQR model.

The initial assumptions for the number of forces and their kinetic attack rates are equal since I want to analyze the cyber effects during the battle.

$S_B = S_R = 150$
$E_B = E_R = 10$
$I_B = I_R = 20$
$Q_B = Q_R = 10$
$R_B = R_R = 10$

$A_B = A_R = 1$
$d_B = d_R = 0.1$
$\eta_B = \eta_R = 0.4$
$c_B = c_R = 0.4$

$\beta_U = \rho_U = 0.1$
$\beta_D = \rho_D = 0.01$

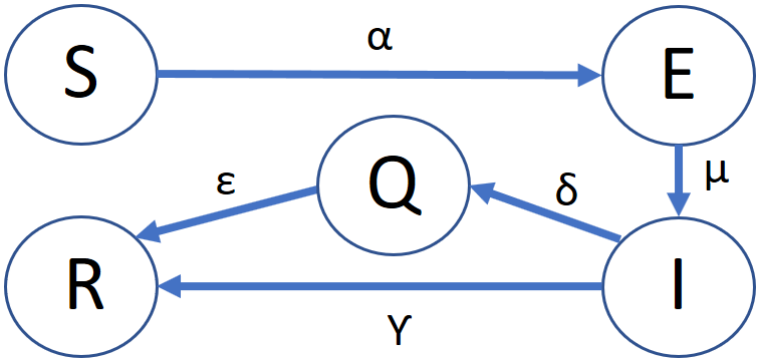


Figure 20: Cyber Effect Parameters in SEIQR Model

The cyber effect rates are alpha (spread rate from S to E), mu (spread rate from E to I), delta (spread rate from I to Q), gamma (spread rate from I to R) and epsilon (spread rate from Q to R) as shown in Figure 20.

Table 6: Sample Cyber Effect Values in SEIQR Model

	α	μ	δ	γ	ε
Blue	0.3	0.2	4	6	0.2
Red	0.2	0.3	6	4	0.3

The initial number of Blue and Red forces and both kinetic attack and cyber effect rates are given above. When I executed the differential equations 2.7, I obtain a graph as seen in Figure 21. The changes in number of units among compartments for Blue side are shown in Figure 22 and the changes in number of units among compartments for Red side are shown in Figure 23. We can conclude that under such circumstances, Blue side has superiority over Red side and in the end of the war, Blue wins.

According to the values above, the malware spreads among Red units faster than Blue units. Moreover, the rate of installing patches and recovering the damaged parts is faster among Blue units than Red units.

By doing minor changes in the values of cyber effect rates, I created a set of samples by executing my python code. My set consists of 207 sample values as shown in Appendix B. According to the graph I draw in line with the output after the execution of the equations, I can see which side has superiority over the other side.

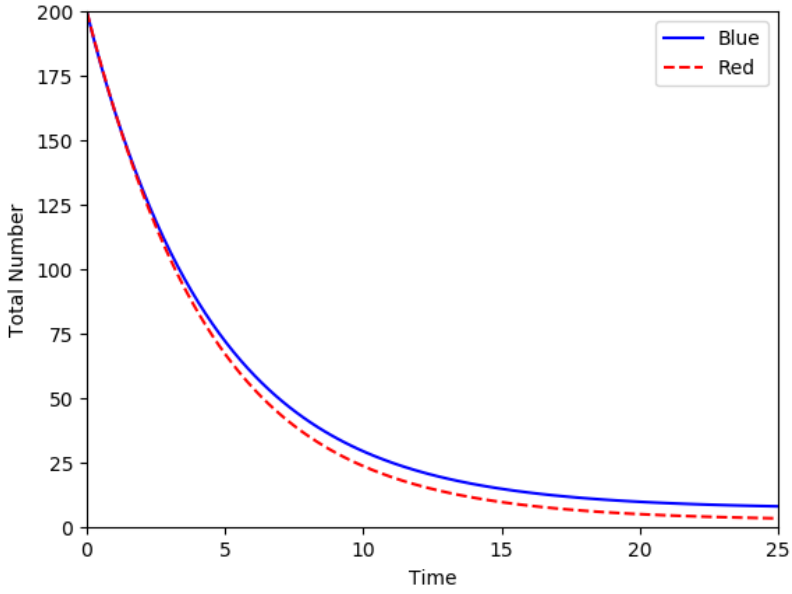


Figure 21: SEIQR Model Implementation

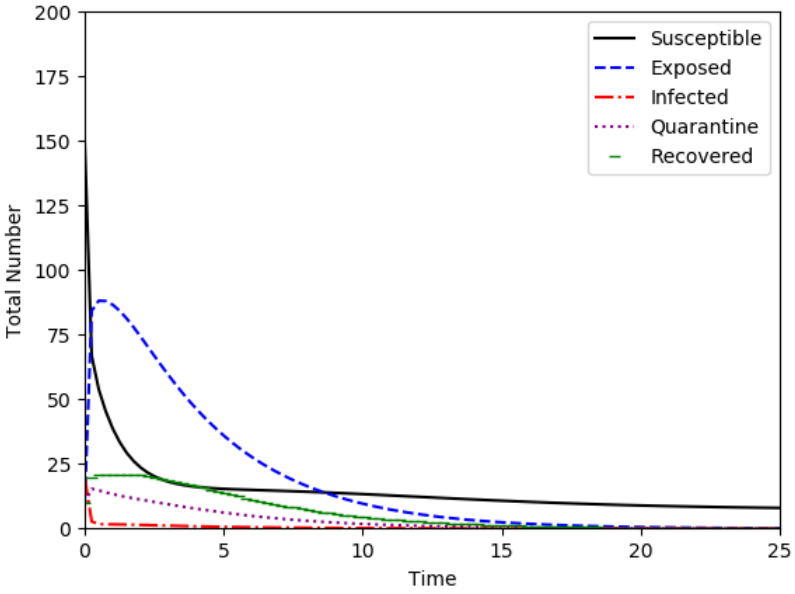


Figure 22: Compartmental Changes of Blue Units in SEIQR Model

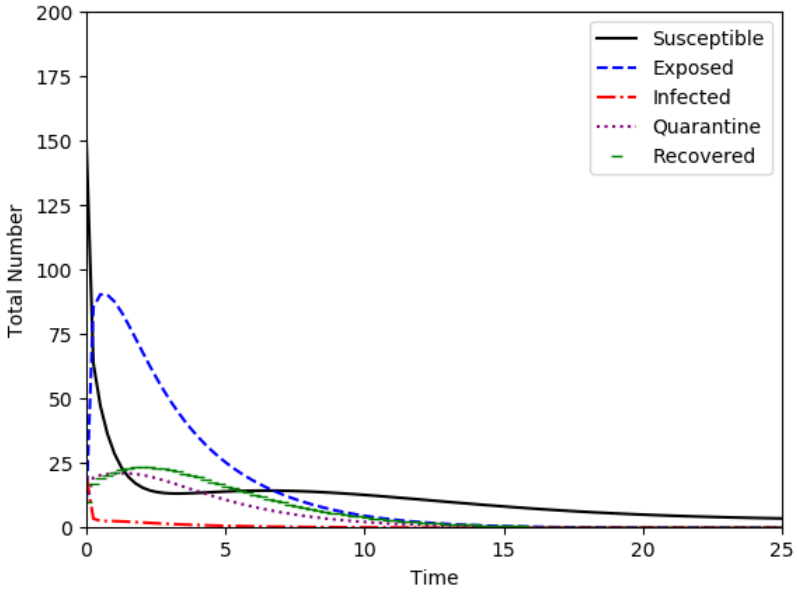


Figure 23: Compartmental Changes of Red Units in SEIQR Model

Conforming to the values in the dataset, I obtained a likelihood table of compartment transition rates as shown in Table 7. For example, in α_B table (upper left corner) 0.1 value is used for 32 times. Blue side won 18 times, while Red side won 14 time. In addition, in total number of 207 trials, Blue won 104 times, while Red won 103 times. Keeping these values in mind, the likelihood that the Blue won when $\alpha_B = 0.1$ is 18/104, whereas the likelihood that the Red won when $\alpha_B = 0.1$ is 14/103.

Although there are many values which can be assigned to the cyber effect parameters, in order to execute the model in Bayesian Network approach, I specified only a certain amount of values.

Table 7: Likelihood Table of Cyber Effect Parameters in SEIQR Model

Likelihood Table	Winner	
α_B	**Blue**	**Red**
0.1	18/104	14/103
0.2	26/104	43/103
0.3	37/104	2/103
0.4	22/104	13/103
0.5	1/104	31/103

Likelihood Table	Winner	
α_R	**Blue**	**Red**
0.1	20/104	42/103
0.2	2/104	1/103
0.3	23/104	20/103
0.4	2/104	1/103
0.5	57/104	39/103

Likelihood Table	Winner	
μ_B	**Blue**	**Red**
0.1	21/104	14/103
0.2	7/104	22/103
0.3	55/104	23/103
0.4	2/104	3/103
0.5	19/104	41/103

Likelihood Table	Winner	
μ_R	**Blue**	**Red**
0.1	21/104	54/103
0.2	3/104	16/103
0.3	3/104	2/103
0.4	30/104	8/103
0.5	47/104	23/103

Likelihood Table	Winner	
δ_B	**Blue**	**Red**
2	27/104	37/103
4	3/104	4/103
6	2/104	5/103
8	72/104	57/103

Likelihood Table	Winner	
δ_R	**Blue**	**Red**
2	26/104	66/103
4	4/104	3/103
6	70/104	31/103
8	4/104	3/103

Likelihood Table	Winner	
γ_B	**Blue**	**Red**
2	56/104	61/103
4	5/104	2/103
6	37/104	39/103
8	6/104	1/103

Likelihood Table	Winner	
γ_R	**Blue**	**Red**
2	99/104	51/103
4	2/104	5/103
6	1/104	6/103
8	2/104	41/103

Likelihood Table	Winner	
ε_B	**Blue**	**Red**
0.1	26/104	47/103
0.2	3/104	3/103
0.3	3/104	3/103
0.4	68/104	47/103
0.5	4/104	3/103

Likelihood Table	Winner	
ε_R	**Blue**	**Red**
0.1	73/104	49/103
0.2	24/104	11/103
0.3	2/104	4/103
0.4	2/104	4/103
0.5	3/104	35/103

By using the values in Table 7 and Bayesian Network formula 2.5, I obtained the probability of the winning side as described below.

Hypothesis = Blue wins (BW)
Data = (α_B=0.3, α_R=0.2, μ_B=0.2, μ_R=0.3, δ_B=4, δ_R=6, Υ_B=6, Υ_R=4, ε_B=0.2 and ε_R=0.3)

P(BW|Data)=
($P(\alpha_B=0.3 \mid BW)*P(\alpha_R=0.2 \mid BW)*P(\mu_B=0.2 \mid BW)*P(\mu_R=0.3 \mid BW)*P(\delta_B=4 \mid BW)*$
$P(\delta_R=6 \mid BW)*P(\Upsilon_B=6 \mid BW)*P(\Upsilon_R=4 \mid BW)*P(\varepsilon_B=0.2 \mid BW)*P(\varepsilon_R=0.3 \mid BW)*P(BW)$)
$/(P(\alpha_B=0.3)*P(\alpha_R=0.2)*P(\mu_B=0.2)*P(\mu_R=0.3)*P(\delta_B=4)*P(\delta_R=6)*P(\Upsilon_B=6)*P(\Upsilon_R=4)$
$*P(\varepsilon_B=0.2)*P(\varepsilon_R=0.3)$)

$$P(BW|Data) = \frac{\frac{37}{104} * \frac{2}{104} * \frac{7}{104} * \frac{3}{104} * \frac{3}{104} * \frac{70}{104} * \frac{37}{104} * \frac{2}{104} * \frac{3}{104} * \frac{2}{104} * \frac{104}{207}}{\frac{39}{207} * \frac{3}{207} * \frac{29}{207} * \frac{5}{207} * \frac{7}{207} * \frac{101}{207} * \frac{76}{207} * \frac{7}{207} * \frac{6}{207} * \frac{6}{207}} = 0.309$$

When I change the hypothesis in favor of Red's winning, similar calculations can be done as follows.

Hypothesis = Red wins (RW)
Data = (α_B=0.3, α_R=0.2, μ_B=0.2, μ_R=0.3, δ_B=4, δ_R=6, Υ_B=6, Υ_R=4, ε_B=0.2 and ε_R=0.3)

P(RW|Data)=
(P(α_B=0.3 | RW)∗P(α_R=0.2 | RW)∗P(μ_B=0.2 | RW)∗P(μ_R=0.3 | RW)∗P(δ_B=4 | RW)∗
P(δ_R=6 | RW)∗P(Υ_B=6 | RW)∗P(Υ_R=4 | RW)∗P(ε_B=0.2 | RW)∗P(ε_R=0.3 | RW)∗P(RW))
/(P(α_B=0.3)∗P(α_R=0.2)∗P(μ_B=0.2)∗P(μ_R=0.3)*P(δ_B=4)∗P(δ_R=6)∗P(Υ_B=6)∗P(Υ_R=4)
∗P(ε_B=0.2)∗P(ε_R=0.3))

$$P(RW|Data) = \frac{\frac{2}{103} * \frac{1}{103} * \frac{22}{103} * \frac{2}{103} * \frac{4}{103} * \frac{31}{103} * \frac{39}{103} * \frac{5}{103} * \frac{3}{103} * \frac{4}{103} * \frac{103}{207}}{\frac{39}{207} * \frac{3}{207} * \frac{29}{207} * \frac{5}{207} * \frac{7}{207} * \frac{101}{207} * \frac{76}{207} * \frac{7}{207} * \frac{6}{207} * \frac{6}{207}} = 0.059$$

Probability of Blue wins = 0.309
Probability of Red wins = 0.059
Sum of Probabilities = 0.368

Likelihood of Blue wins = 0.309/0.368 = % 83.88
Likelihood of Red wins = 0.059/0.368 = % 16.12

Finally, we can conclude that given the values α_B=0.3, α_R=0.2, μ_B=0.2, μ_R=0.3, δ_B=4, δ_R=6, Υ_B=6, Υ_R=4, ε_B=0.2 and ε_R=0.3, Blue side wins the battle with %83.88 probability.

I used WEKA software in order to analyze my sample data set constituted by using SEIQR model. I obtained the confusion matrix as shown in Table 8 and results in Table 9.

Table 8: Confusion Matrix in SEIQR Model

Confusion Matrix		NaïveBayes		BayesNet	
		Real Values		Real Values	
		0	1	0	1
Predicted Values	0	82	22	89	15
	1	23	80	26	77
Total Number of Instances		207		207	
Correctly Classified Instances		162		166	
Accuracy		% 78.3		% 80.2	

According to the confusion matrix done with NaïveBayes, when the actual class is Blue (0), the predicted class is also Blue for 82 times (TP), but assessed wrongly for 23 times (FN). In addition, when the actual class is Red (1), the predicted class is also Red for 80 times (TN), but assessed wrongly for 22 times (FP). Furthermore, according to the confusion matrix done with BayesNet, when the actual class is Blue (0), the predicted class is also Blue for 89 times (TP), but assessed wrongly for 26 times (FN). In addition, when the actual class is Red (1), the predicted class is also Red for 77 times (TN), but assessed wrongly for 15 times (FP).

162 out of 207 instances are classified correctly in NaïveBayes and the accuracy is very high (% 78.3). On the other hand, 166 out of 207 instances are classified correctly in BayesNet and the accuracy is again very high (% 80.2).

Table 9: Results for the Bayesian Classifiers in SEIQR Model

Class	NaïveBayes		BayesNet	
	TP Rate	Precision	TP Rate	Precision
0	0.788	0.781	0.856	0.774
1	0.777	0.784	0.748	0.837
Weighted Average	0.783	0.783	0.802	0.805

In NaïveBayes classifier TP rate for class 0 and 1 is 0.788 and 0.777, while Precision for class 0 and 1 are 0.781 and 0.784 respectively. Besides, in BayesNet classifier TP rate for class 0 and 1 is 0.856 and 0.748, while Precision for class 0 and 1 are 0.774 and 0.837 respectively. These ratios are very high and this indicates that the model performs quite well.

ROC Curve for NaïveBayes is presented in Figure 24 and ROC Curve for BayesNet is presented in Figure 25. We can analyze that both NaïveBayes classifier and BayesNet classifier classify our dataset quite well. ROC curve is close to the upper left corner, so that we can figure out that our model performs really good.

PRC for NaïveBayes is presented in Figure 26 and PRC for BayesNet is presented in Figure 27. No-skill line is 0.5. We can analyze that both NaïveBayes and BayesNet classifier performs above the no-skill line. In these graphs the curve is close to the upper right corner. This also proves the good performance of the model.

By evaluating the results, we can make a conclusion that SEIQR model performs very well and can be used for further probabilistic predictions.

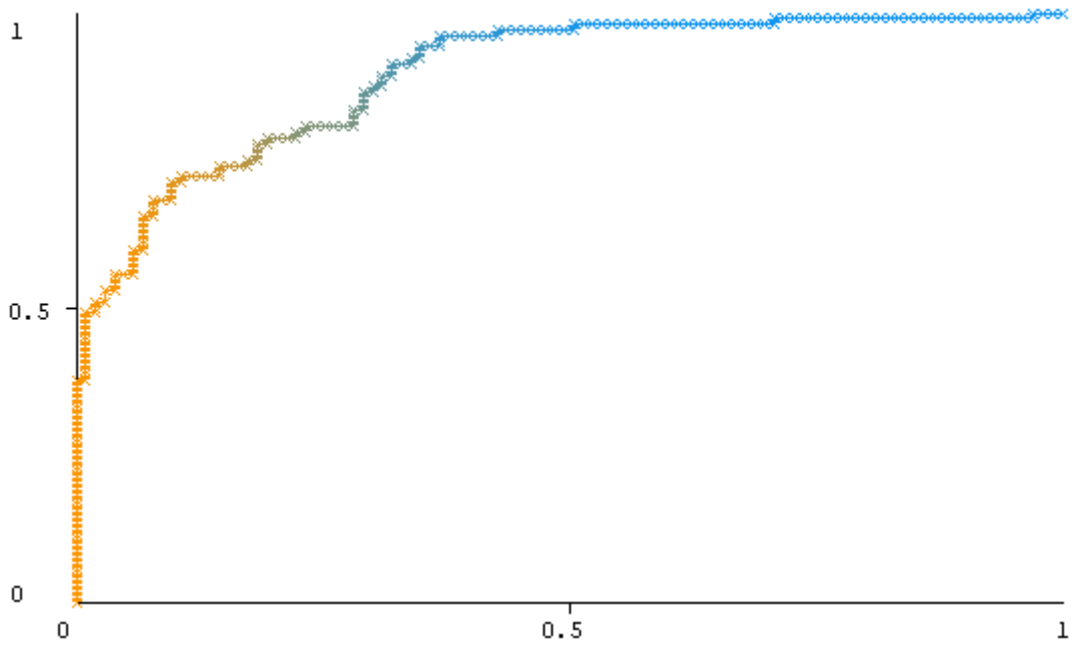


Figure 24: ROC Curve for NaïveBayes in SEIQR Model

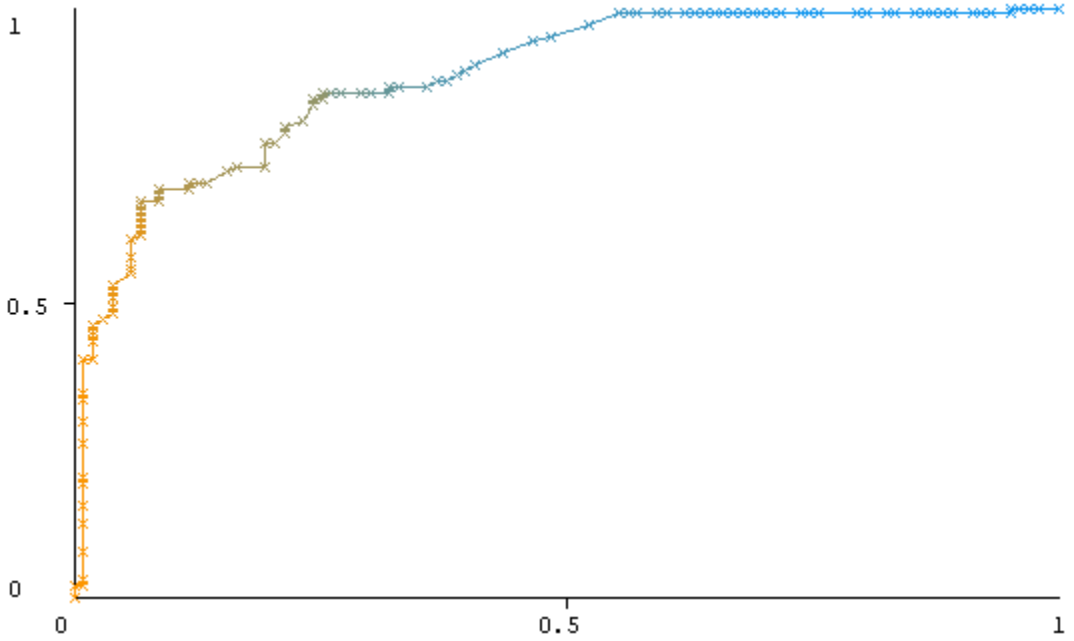


Figure 25: ROC Curve for BayesNet in SEIQR Model

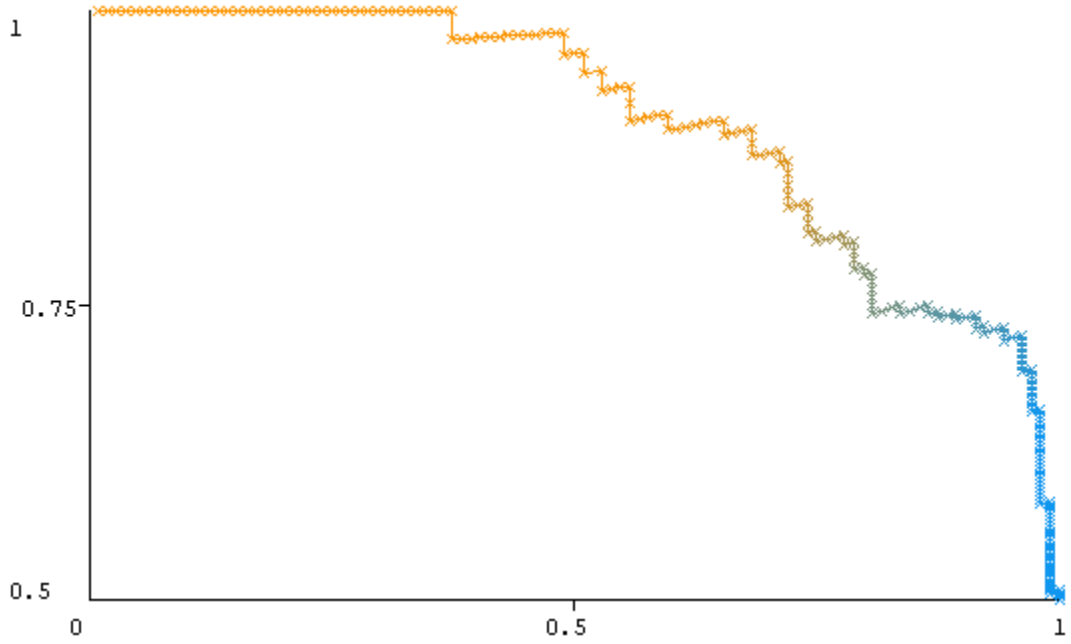


Figure 26: PRC for NaïveBayes in SEIQR Model

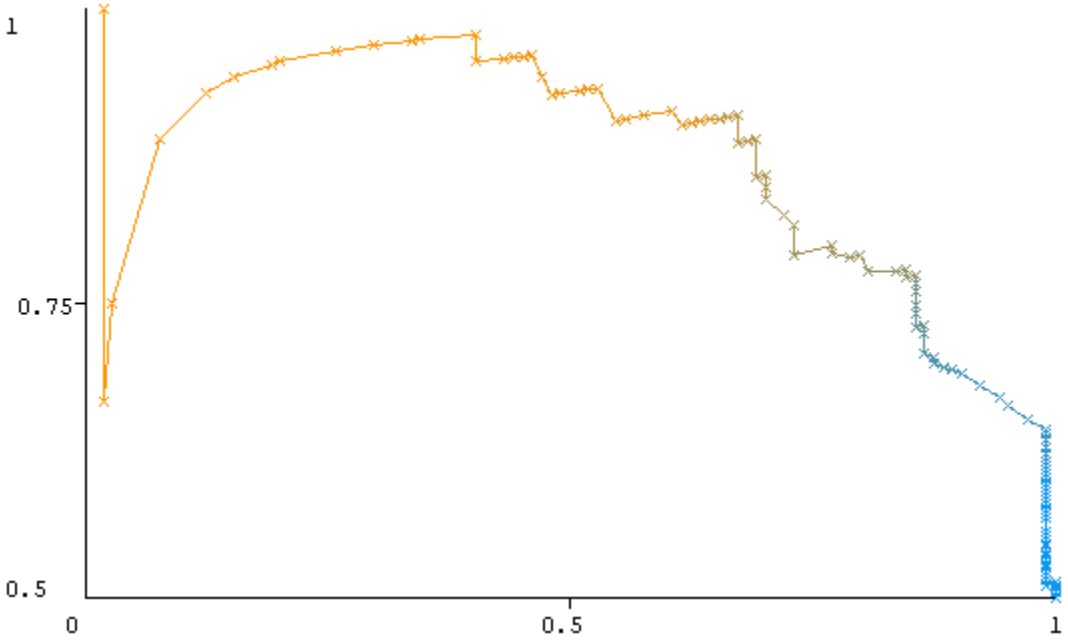


Figure 27: PRC for BayesNet in SEIQR Model

3.3. SEIR Model

Impersonation and social engineering attacks are compatible with the SEIR model.

The initial assumptions for the number of forces and their kinetic attack rates are equal since I want to analyze the cyber effects during the battle.

$S_B = S_R = 200$
$E_B = E_R = 20$
$I_B = I_R = 20$
$R_B = R_R = 5$

$\psi 1_B = \psi 1_R = 0.03$
$\psi 2_B = \psi 2_R = 0.3$
$\varphi_B = \varphi_R = 0.002$
$\theta_B = \theta_R = 0.002$
$\mu_B = \mu_R = 0.03$

$\beta_U = \rho_U = 0.1$
$\beta_D = \rho_D = 0.01$

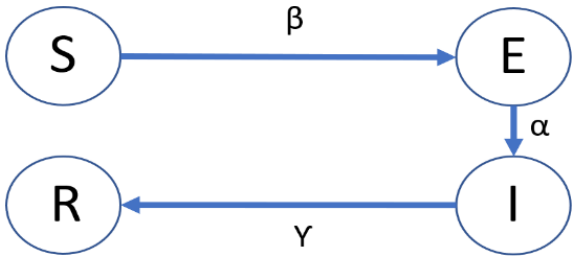


Figure 28: Cyber Effect Parameters in SEIR Model

The cyber effect rates are beta (spread rate from S to E), alpha (spread rate from E to I) and gamma (spread rate from I to R) as shown in Figure 28.

Table 10: Sample Cyber Effect Values in SEIR Model

	β	α	γ
Blue	0.3	0.4	0.3
Red	0.1	0.3	0.4

The initial number of Blue and Red forces and both kinetic attack and cyber effect rates are given above. When I executed the differential equations 2.8, I obtain a graph as seen in Figure 29. The changes in number of units among compartments for Blue side are shown in Figure 30 and the changes in number of units among compartments for Red side are shown in Figure 31. We can conclude that under such circumstances, Red side has superiority over Blue side and in the end of the war, Red wins.

According to the values obtained from the graphs, the malware spreads among Blue units faster than Red units. Moreover, the rate of installing patches and recovering the damaged parts is faster among Red units than Blue units.

By doing minor changes in the values of cyber effect rates, I created a set of samples by executing my python code. My set consists of 123 sample values as shown in

Appendix C. According to the graph I draw in line with the output after the execution of the equations, I can see which side has superiority over the other side.

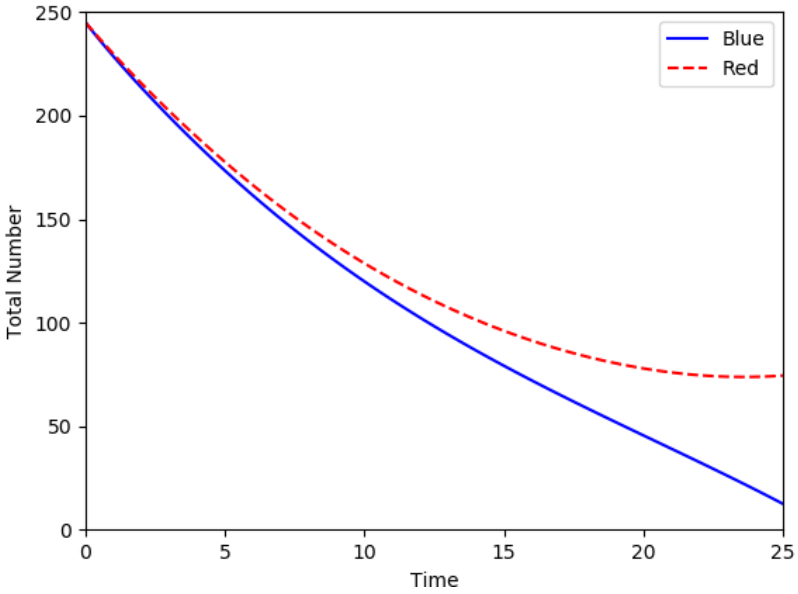


Figure 29: SEIR Model Implementation

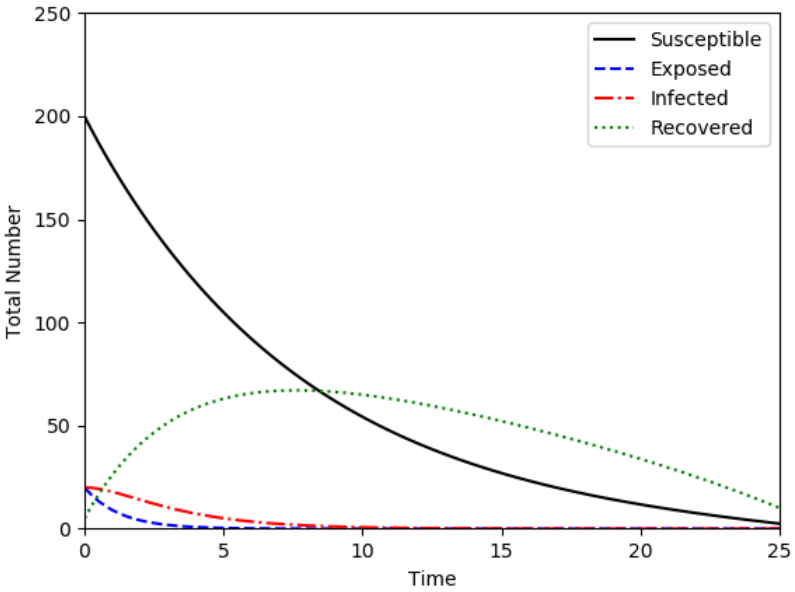


Figure 30: Compartmental Changes of Blue Units in SEIR Model

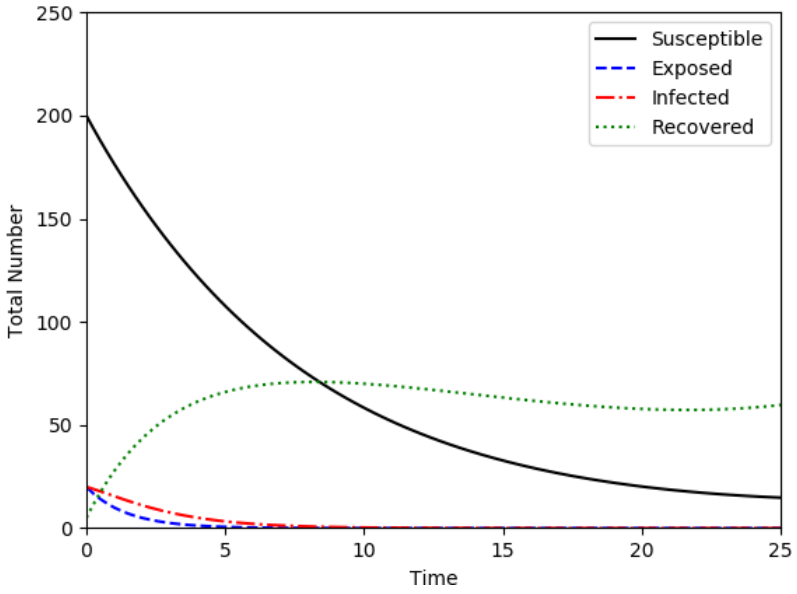


Figure 31: Compartmental Changes of Red Units in SEIR Model

Conforming to the values in the dataset, I obtained a likelihood table of compartment transition rates as shown in Table 11. For example, in β_B table (upper left corner) 0.1 value is used for 12 times. Blue side won 12 times, while Red side won 17 time. In addition, in total number of 123 trials, Blue won 64 times, while Red won 59 times. Keeping these values in mind, the likelihood that the Blue won when $\beta_B = 0.1$ is 12/64, whereas the likelihood that the Red won when $\beta_B = 0.1$ is 17/59.

Although there are many values which can be assigned to the cyber effect parameters, in order to execute the model in Bayesian Network approach, I specified only a certain amount of values.

Table 11: Likelihood Table of Cyber Effect Parameters in SEIR Model

Likelihood Table	Winner	
β_B	**Blue**	**Red**
0.1	12/64	17/59
0.2	8/64	12/59
0.3	17/64	16/59
0.4	9/64	4/59
0.5	18/64	10/59

Likelihood Table	Winner	
β_R	**Blue**	**Red**
0.1	13/64	20/59
0.2	16/64	13/59
0.3	3/64	7/59
0.4	10/64	4/59
0.5	22/64	15/59

Likelihood Table	Winner	
α_B	**Blue**	**Red**
0.1	23/64	22/59
0.2	7/64	4/59
0.3	15/64	13/59
0.4	16/64	16/59
0.5	3/64	4/59

Likelihood Table	Winner	
α_R	**Blue**	**Red**
0.1	17/64	36/59
0.2	5/64	10/59
0.3	17/64	8/59
0.4	16/64	4/59
0.5	9/64	1/59

Likelihood Table	Winner	
γ_B	**Blue**	**Red**
0.1	16/64	37/59
0.2	6/64	7/59
0.3	6/64	7/59
0.4	13/64	6/59
0.5	23/64	2/59

Likelihood Table	Winner	
γ_R	**Blue**	**Red**
0.1	36/64	15/59
0.2	14/64	7/59
0.3	10/64	12/59
0.4	2/64	15/59
0.5	2/64	10/59

By using the values in Table 11 and Bayesian Network formula 2.5, I obtained the probability of the winning side as described below.

Hypothesis = Blue wins (BW)
Data = (β_B=0.3, β_R=0.1, α_B=0.4, α_R=0.3, Υ_B=0.3, Υ_R=0.4)

P(BW|Data)=
(P(β_B=0.3 | BW)∗P(β_R=0.1 | BW)∗P(α_B=0.4 | BW)∗P(α_R=0.3 | BW)∗
P(Υ_B=0.3 | BW)∗P(Υ_R=0.4 | BW) ∗P(BW)) /
(P(β_B=0.3)∗P(β_R=0.1)∗P(α_B=0.4)∗P(α_R=0.3)*P(Υ_B=0.3)∗P(Υ_R=0.4))

$$\mathrm{P(BW|Data)} = \frac{\frac{17}{64} * \frac{13}{64} * \frac{16}{64} * \frac{17}{64} * \frac{6}{64} * \frac{2}{64} * \frac{64}{123}}{\frac{33}{123} * \frac{33}{123} * \frac{32}{123} * \frac{25}{123} * \frac{13}{123} * \frac{17}{123}} = 0.098$$

When I change the hypothesis in favor of Red's winning, similar calculations can be done as follows.

Hypothesis = Red wins (RW)
Data = (β_B=0.3, β_R=0.1, α_B=0.4, α_R=0.3, Υ_B=0.3, Υ_R=0.4)

P(RW|Data)=
(P(β_B=0.3 | RW)∗P(β_R=0.1 | RW)∗P(α_B=0.4 | RW)∗P(α_R=0.3 | RW)∗
P(Υ_B=0.3 | RW)∗P(Υ_R=0.4 | RW) ∗P(RW)) /
(P(β_B=0.3)∗P(β_R=0.1)∗P(α_B=0.4)∗P(α_R=0.3)*P(Υ_B=0.3)∗P(Υ_R=0.4))

$$P(RW|Data) = \frac{\frac{16}{59} * \frac{20}{59} * \frac{16}{59} * \frac{8}{59} * \frac{7}{59} * \frac{15}{59} * \frac{59}{123}}{\frac{33}{123} * \frac{33}{123} * \frac{32}{123} * \frac{25}{123} * \frac{13}{123} * \frac{17}{123}} = 0.880$$

Probability of Blue wins = 0.098
Probability of Red wins = 0.880
Sum of Probabilities = 0.978

Likelihood of Blue wins = 0.098/0.978 = % 10.02
Likelihood of Red wins = 0.880/0.978 = % 89.98

Finally, we can conclude that given the values β_B=0.3, β_R=0.1, α_B=0.4, α_R=0.3, Υ_B=0.3, Υ_R=0.4, Red side wins the battle with %89.98 probability.

I used WEKA software in order to analyze my sample data set constituted by using SEIR model. I obtained the confusion matrix as shown in Table 12 and results in Table 13.

Table 12: Confusion Matrix in SEIR Model

Confusion Matrix		NaïveBayes		BayesNet	
		Real Values		Real Values	
		0	1	0	1
Predicted Values	0	51	13	50	14
	1	13	46	12	47
Total Number of Instances		123		123	
Correctly Classified Instances		97		97	
Accuracy		% 78.86		% 78.86	

According to the confusion matrix done with NaïveBayes, when the actual class is Blue (0), the predicted class is also Blue for 51 times (TP), but assessed wrongly for 13 times (FN). In addition, when the actual class is Red (1), the predicted class is also Red for 46 times (TN), but assessed wrongly for 13 times (FP). Furthermore, according to the confusion matrix done with BayesNet, when the actual class is Blue (0), the predicted class is also Blue for 50 times (TP), but assessed wrongly for 12 times (FN). In addition, when the actual class is Red (1), the predicted class is also Red for 47 times (TN), but assessed wrongly for 14 times (FP).

97 out of 123 instances are classified correctly both in NaïveBayes and BayesNet and the accuracy is very high (% 78.86).

Table 13: Results for the Bayesian Classifiers in SEIR Model

Class	NaïveBayes		BayesNet	
	TP Rate	Precision	TP Rate	Precision
0	0.797	0.797	0.781	0.806
1	0.780	0.780	0.797	0.770
Weighted Average	0.789	0.789	0.789	0.789

In NaïveBayes classifier both TP rate and Precision for class 0 and 1 are 0.797 and 0.780 respectively. Besides, in BayesNet classifier TP rate for class 0 and 1 are 0.781 and 0.797 while Precision for class 0 and 1 are 0.806 and 0.770 respectively. These ratios are very high and this indicates that the model performs quite well.

ROC Curve for NaïveBayes is presented in Figure 32 and ROC Curve for BayesNet is presented in Figure 33. We can analyze that both NaïveBayes classifier and BayesNet classifier classify our dataset quite well. Furthermore, we can figure out that our model performs really good.

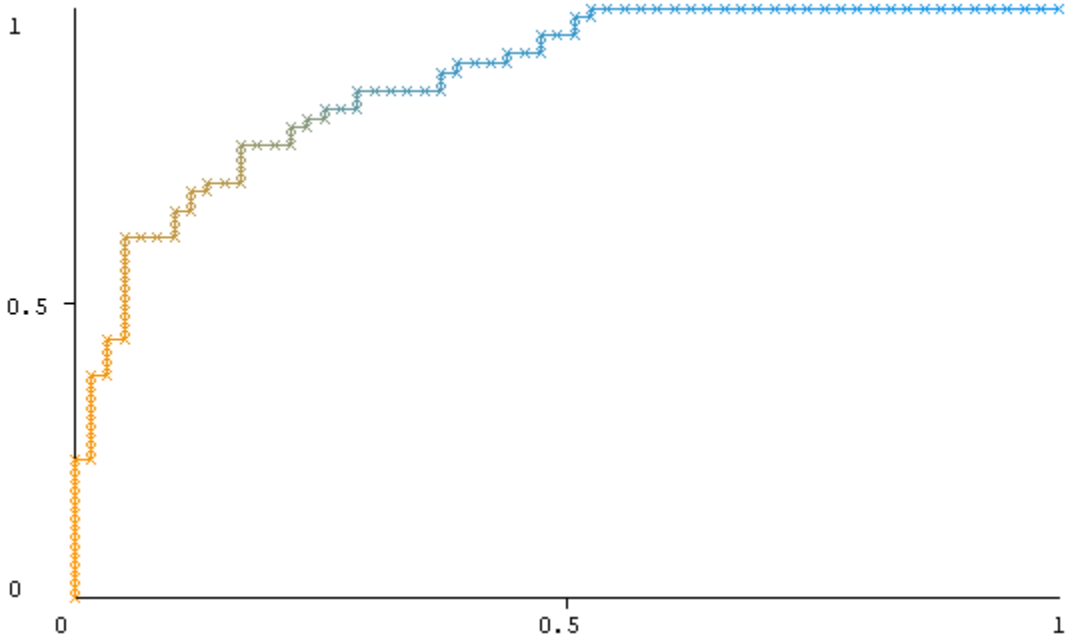


Figure 32: Compartmental Changes of Red Units in SEIR Model

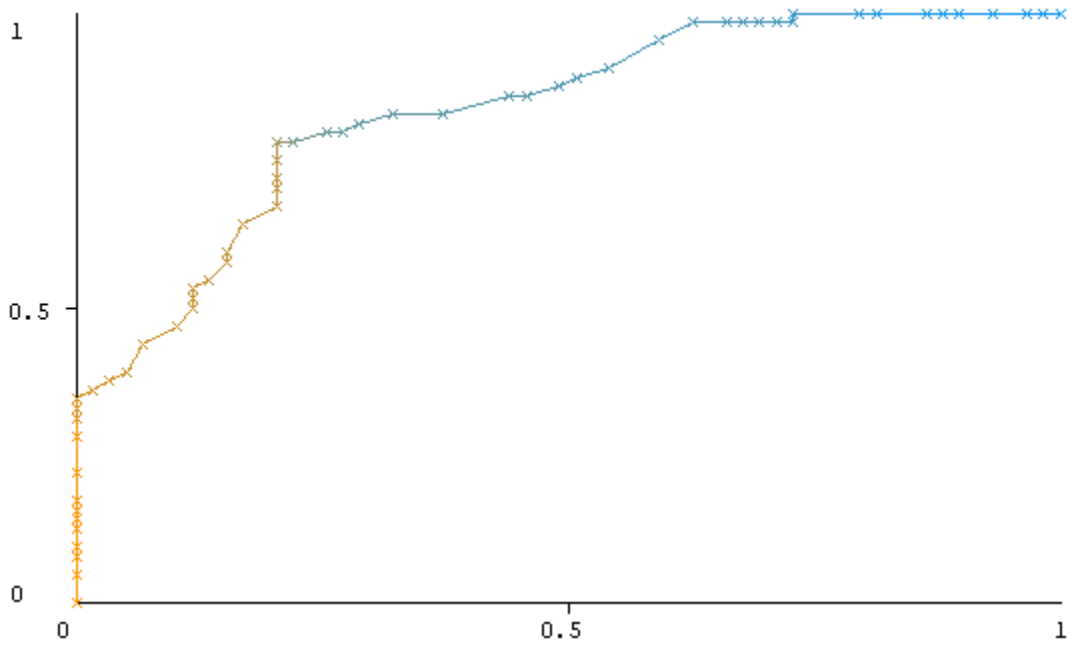


Figure 33: Compartmental Changes of Red Units in SEIR Model

PRC for NaïveBayes is presented in Figure 34 and PRC for BayesNet is presented in Figure 35. No-skill line is 0.52. We can analyze that both NaïveBayes and BayesNet classifier performs above the no-skill line.

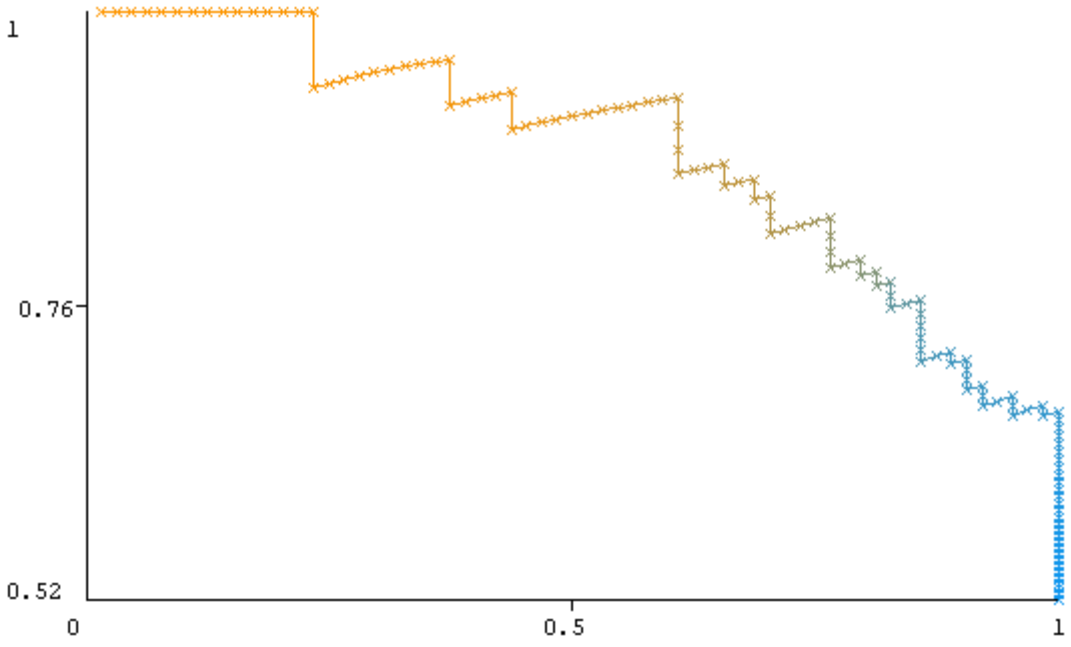


Figure 34: Compartmental Changes of Red Units in SEIR Model

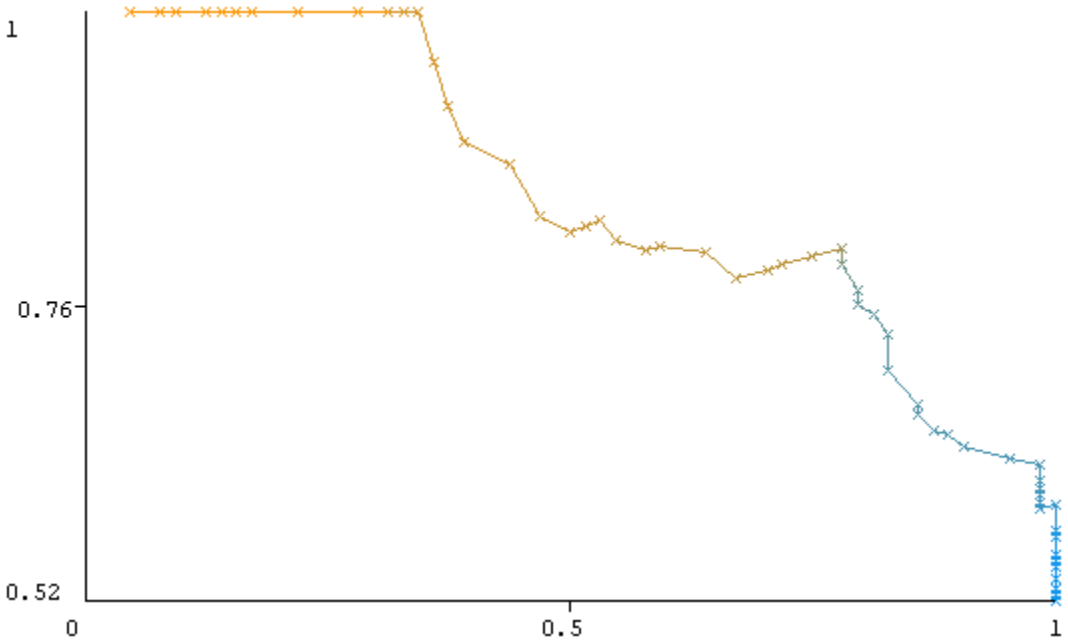


Figure 35: Compartmental Changes of Red Units in SEIR Model

ROC curves are well close to the upper left corner and PRC curves are close to the upper right corner both in NaïveBayes and BayesNet classifiers. By evaluating the results, we can make a conclusion that SEIR model performs very well and can be used for further probabilistic predictions.

In the next chapter, I presented conclusions on mixed epidemic combat models (SIR, SEIQR and SEIR) and the probabilistic results of using these models on malware propagation during military operations with the help of Bayesian Network approach. Finally, I enumerated the probable future works.

3.4. Discussion

Three models were studied in this thesis, which are SIR, SEIQR and SEIR. The number of changes in the units of war (i.e., Suspected, Infected, Recovered) were computed by the proposed mathematical models. In each model, the effect of cyber effects was presented as a part of formula, which reflected its influence upon the overall changes in the number of units. There is also a specific part of the formula which demonstrates the kinetic effects of the forces.

In the brackets at beginning of each equation in the formula (2.6) of SIR Model, repeated below, the sign shows an increase or a decrease from one compartment to the other.

$$\frac{dS_B}{dt} = (-\varepsilon_B S_B I_B - \eta_B S_B R_B) - [\rho_U (S_Z + R_Z) + \rho_D I_Z] \frac{S_B}{S_B + I_B + R_B}$$

$$\frac{dI_B}{dt} = (\varepsilon_B S_B I_B - \eta_B I_B R_B) - [\rho_U(S_Z + R_Z) + \rho_D I_Z]\frac{I_B}{S_B + I_B + R_B}$$

$$\frac{dR_B}{dt} = (\eta_B S_B R_B + \eta_B I_B R_B) - [\rho_U(S_Z + R_Z) + \rho_D I_Z]\frac{R_B}{S_B + I_B + R_B}$$

$$\frac{dS_Z}{dt} = (-\varepsilon_Z S_Z I_Z - \eta_Z S_Z R_Z) - [\beta_U(S_B + R_B) + \beta_D I_B]\frac{S_Z}{S_Z + I_Z + R_Z} \quad (2.6)$$

$$\frac{dI_Z}{dt} = (\varepsilon_B S_Z I_Z - \eta_Z I_Z R_Z) - [\beta_U(S_B + R_B) + \beta_D I_B]\frac{I_Z}{S_Z + I_Z + R_Z}$$

$$\frac{dR_Z}{dt} = (\eta_Z S_Z R_Z + \eta_Z I_Z R_Z) - [\beta_U(S_B + R_B) + \beta_D I_B]\frac{R_Z}{S_Z + I_Z + R_Z}$$

To illustrate, as for the number of changes in the Infected group of Blue, $(\varepsilon_B S_B I_B - \eta_B I_B R_B)$ means that there is a transition from Susceptible to Infected, such that the number of units in the Infected group increases since the sign is positive. In addition, there is a transition from Infected to Recovered, which implies that the number of units in Infected group decreases since the sign is negative.

As for the kinetic effect section in the formula, for the number of changes in the Infected group of Blue units, $-[\rho_U(S_Z + R_Z) + \rho_D I_Z]$ indicates that the total number of units in Susceptible and Recovered groups of Red units attack with a high power but the number of units in the Infected group of Red units attack with a low power. The total effect of this kinetic value decreases the number of the Infected Blue units. A question that may arise at this point is why Red forces try not to increase the number of Infected Blue units. That is because the aim of the kinetic war is to destroy that unit. In particular, if we assume that the infected computer belongs to a cannon weapon, the virus in that computer affects the normal operation of that weapon and diminishes its power. It can be turned into a member of the Recovered group any time by installing updates/patches and fight again with high power. On the other hand, the aim here is to totally destroy that cannon and get away with that unit. Otherwise, the number of units in the Infected group increases as shown in Figure 36, which is an unwanted situation. Similarly, if this were the situation, total number of forces would change as shown in Figure 37.

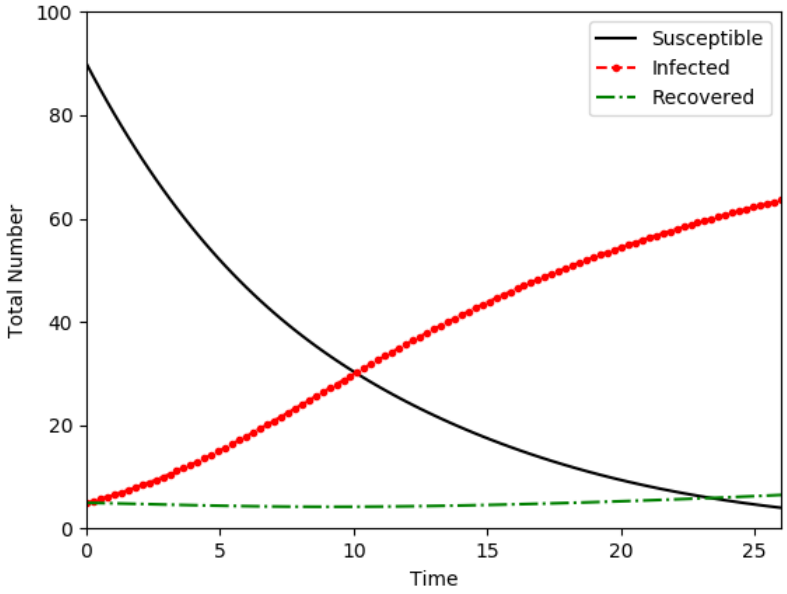


Figure 36: Compartmental Changes of Blue Units in Assumed SIR Model

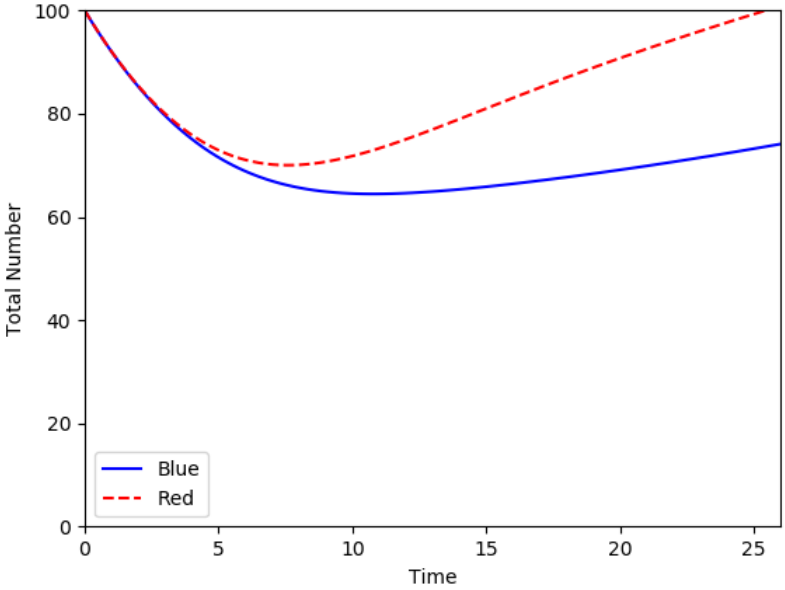


Figure 37: Assumed SIR Model Implementation

Accordingly, the implementation of the signs in the formula suggests alternative, context-dependent evaluations of the relationships among the forces. Future research should address those alternative evaluations in more detail.

CHAPTER 4

CONCLUSION AND FUTURE WORK

4.1. Conclusion

Warfare has been modelled since Lanchester, who introduced combat models in 1916. These models were used to figure out the initial numbers in units, attrition rates of warring forces and their remaining number of units in the end of the battle.

With the rise of malware and the damage they did in computer systems, the effects of malware have been a potential source of influence in conventional warfare. The adaptation of Kerman-McKendrick's SIR epidemic model for analysis of malware influence has been the most popular one. In addition, according to the features of cyber-attack, novel types of compartments were included or excluded to the epidemic model.

It has been always the most difficult part to gain data about the battle. For that reason, comparisons are mostly done with simulated data. In this thesis, the Lanchester aimed fire model is chosen because of its usability, flexibility and similarity to real-world situations. As for the epidemic model, there are three different models used in this thesis. Rootkit, DOS and replay attacks are compatible with the SIR model. Virus, worm, trojan horse and ransomware attacks are compatible with the SEIQR model. Impersonation and social engineering attacks are compatible with the SEIR model. These epidemic models are combined with the Lanchester combat model.

By using the formulas in these models, I assigned sample values to the parameters and executed to obtain the results. Then I composed a set of sample values with these results in order to make predictions about the outcome of the battle.

In the SIR Model, there are 75 sample values, which were used to make predictions. The cyber effect rates are infection spread rate and patch rate. By using the likelihood table, the probability of the hypothesis (in this situation, the winning chance of Blue or Red) with the light of given data (assigned values to cyber effect rates) was computed. I obtained confusion matrix and evaluated the usability of the model. In NaïveBayes classifier, the accuracy is % 81.3, while in BayesNet classifier the accuracy is % 66.6.

These results indicate that the performance of the model is good. In addition, weighted average of TP rate and Precision in NaïveBayes classifier are both 0.813, while weighted average of TP rate and Precision in BayesNet classifier are 0.667 and 0.668 respectively. These ratios are high enough and we can make a conclusion that this model separates the data into classes well. Finally, ROC curve and PRC were drawn to prove the high performance of the classifier.

In the SEIQR Model, there are 207 samples values, which were used to make predictions. The cyber effect rates are spread rate from S to E, from E to I, from I to Q, from I to R and from Q to R. By using the likelihood table, the probability of the hypothesis (in this situation, the winning chance of Blue or Red) with the light of given data (assigned values to cyber effect rates) was calculated. According to the confusion matrix, in NaïveBayes classifier, the accuracy is % 79.3, while in BayesNet classifier the accuracy is % 80.2. These results indicate that the performance of the model is very high. In addition, weighted average of TP rate and Precision in NaïveBayes classifier are both 0.783, while weighted average of TP rate and Precision in BayesNet classifier are 0.802 and 0.805 respectively. These ratios are high enough and we can make a conclusion that this model separates the data into classes well. Finally, ROC curve and PRC were drawn to prove the high performance of the classifier.

In the SEIR Model, there are 123 samples values, which were used to make predictions. The cyber effect rates are spread rate from S to E, from E to I and from I to R. By using the likelihood table, the probability of the hypothesis (in this situation, the winning chance of Blue or Red) with the light of given data (assigned values to cyber effect rates) was computed. According to the confusion matrix, the accuracy is % 78.86 in both NaïveBayes and BayesNet classifier. These results indicate that the performance of the model is very high. In addition, weighted average of TP rate and Precision in both NaïveBayes and BayesNet classifier are 0.789. This ratio is very high and we can make a conclusion that this model separates the data into classes well. Finally, ROC curve and PRC were drawn to prove the high performance of the classifier.

The cyber-attacks throughout the war aim to increase the number of Infected computers, while the aim of the kinetic attacks is to destroy the enemy forces, that is to decrease the number of any enemy compartment. For that reason, in the formulas of the models, the sign of the kinetic effect section which reflects the number of changes by time in the Infected group is taken negative.

By evaluating the results above, the judgement that we can make about SIR, SEIQR and SEIR model is that they perform well to help us compute the probability of the combat outcome. In addition, these models can be used for further probabilistic predictions to figure out the effects of malware propagation in a military operation.

4.2. Future Work

Lanchester models, which have been using to compute the combat outcome, have some assumptions at the beginning. For instance, both fighting states are constituted of homogenous units (i.e. tanks to tanks, artillery to artillery). This is actually far from simulating the real world problems. For that reason, a revised version of Lanchester model, which has the ability to taking into consideration the heterogeneous units, can be studied.

It is difficult to access and use the real world battle data. Therefore, the assumptions are done to execute the models. Privacy issues on these confidential data can be skipped to obtain a better performing combat and epidemic model. In this way, better predictions can be made to evaluate the ongoing battle.

The fighting states have the same type of topology in their network system. A computer is connected to a certain amount of other computers. Bearing this in mind, an epidemic can spread only a fixed number of computers. However, in real world scenario, the number which a computer is connected changes according to its being a computer or server or being in a small scale network or in a big scale network. A novel model can be constituted for taking into account these factors.

Different types of malware other than I studied in this thesis can be examined according the diagram in Figure 7. Dynamic strategic learning model (Duffey, 2017) can be studied to display the effects of changing situations during the battle. The effects of the intelligence and its relation with the cyber-attack on the combat outcome can be studied as well. Furthermore, the situation that the cyber-attack starts earlier or later than the kinetic attack can also be evaluated to better reflect the real-time battle scenarios.

Defense-in-depth is a cyber-security strategy. It provides a layered mechanism to protect important data and systems. The approach of this strategy can be investigated in further studies on this subject.

REFERENCES

Batistela, C. M., & Piqueira, J. R. C. (2018). SIRA Computer Viruses Propagation Model: Mortality and Robustness. International Journal of Applied and Computational Mathematics, 4(5), 128.

Bendovschi, A. (2015). Cyber-attacks–trends, patterns and security countermeasures. Procedia Economics and Finance, 28, 24-31.

Bracken, J. (1995). Lanchester models of the Ardennes campaign. Naval Research Logistics (NRL), 42(4), 559-577.

Clausen, S. (2003). Warfare Can Be Calculated. In Mathematics and War (pp. 216-238). Birkhäuser, Basel.

Clausewitz, C., & Maude, F. N. (1982). On war. Penguin UK.

Coulson, S. G. (2018). Lanchester modelling of intelligence in combat. IMA Journal of Management Mathematics, 30(2), 149-164.

Das, S. K. (2019). Fitting Heterogeneous Lanchester Models on the Kursk Campaign. arXiv preprint arXiv:1903.06666.

Deitchman, S. J. (1962). A Lanchester model of guerrilla warfare. Operations Research, 10(6), 818-827.

del Rey, A. M. (2015). Mathematical modeling of the propagation of malware: a review. Security and Communication Networks, 8(15), 2561-2579.

DoD, US Department of Defense Cyber Strategy, 2018.

Draeger, J., & Öttl, S. (2018). Malware Epidemics Effects in a Lanchester Conflict Model. arXiv preprint arXiv:1811.01892.

Duffey, R. B. (2017). Dynamic theory of losses in wars and conflicts. European Journal of Operational Research, 261(3), 1013-1027.

Engel, J. H. (1954). A verification of Lanchester's law. Journal of the Operations Research Society of America, 2(2), 163-171.

Fenton, N., & Neil, M. (2012). Risk assessment and decision analysis with Bayesian Networks. Crc Press.

Fricker Jr, R. D. (1998). Attrition models of the Ardennes campaign. Naval Research Logistics (NRL), 45(1), 1-22.

Johnson, I. R., & MacKay, N. J. (2011). Lanchester models and the Battle of Britain. Naval Research Logistics (NRL), 58(3), 210-222.).

Kabakchieva, D. (2013). Predicting student performance by using data mining methods for classification. Cybernetics and information technologies, 13(1), 61-72.

Kermack, W. O., & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character, 115(772), 700-721.

Lanchester, F. W. (1916). Aircraft in warfare: The dawn of the fourth arm. Constable limited.

Lindsay, J. R. (2013). Stuxnet and the limits of cyber warfare. Security Studies, 22(3), 365-404.

Lucas, T. W., & Dinges, J. A. (2004). The effect of battle circumstances on fitting Lanchester equations to the Battle of Kursk. Military Operations Research, 9(2), 17-30.

MacKay, N. J. (2006). Lanchester combat models. arXiv preprint math/0606300.

Mirmoeini, F., & Krishnamurthy, V. (2005, March). Reconfigurable Bayesian Networks for adaptive situation assessment in battlespace. In Proceedings. 2005 IEEE Networking, Sensing and Control, 2005. (pp. 810-815). IEEE.

Mishra, B. K., & Ansari, G. M. (2012). Differential Epidemic Model of Virus and Worms in Computer Network. IJ Network security, 14(3), 149-155.

Mishra, B. K., & Prajapati, A. (2013). Modelling and simulation: cyber war. Procedia Technology, 10, 987-997.

Murray, J. D. (2002). Mathematical biology: Interdisciplinary applied mathematics.

Nielsen, T. D., & Jensen, F. V. (2009). Bayesian Networks and decision graphs. Springer Science & Business Media.

Pettit, L. I., Wiper, M. P., & Young, K. D. S. (2003). Bayesian inference for some Lanchester combat laws. European Journal of Operational Research, 148(1), 152-165.

Schramm, H. C., & Gaver, D. P. (2013). Lanchester for cyber: The mixed epidemic-combat model. Naval Research Logistics (NRL), 60(7), 599-605.

Simmons, C., Ellis, C., Shiva, S., Dasgupta, D., & Wu, Q. (2014, June). AVOIDIT: A cyber attack taxonomy. In 9th Annual Symposium on Information Assurance (ASIA'14) (pp. 2-12).

Sfikas, I. D. (2017). Model of warfare. Journal of Computations & Modelling, 7(1), 99-114.

Wiper, M. P., Pettit, L. I., & Young, K. D. (2000). Bayesian inference for a Lanchester type combat model. Naval Research Logistics (NRL), 47(7), 541-558.

APPENDICES

APPENDIX A

SIR MODEL SAMPLE SET

@relation sir_malware_propogation

@attribute epsilon_Blue real
@attribute epsilon_Red real
@attribute eta_Blue real
@attribute eta_Red real
@attribute class {0, 1}

@data
0.004,0.005,0.0005,0.0005,0
0.003,0.005,0.0005,0.0005,0
0.002,0.005,0.0005,0.0005,0
0.006,0.005,0.0005,0.0005,1
0.007,0.005,0.0005,0.0005,1
0.008,0.005,0.0005,0.0005,1
0.003,0.004,0.0005,0.0005,0
0.003,0.002,0.0005,0.0005,1
0.003,0.001,0.0005,0.0005,1
0.005,0.005,0.0004,0.0005,1
0.005,0.005,0.0003,0.0005,1
0.005,0.005,0.0002,0.0005,1
0.005,0.005,0.0006,0.0005,0
0.005,0.005,0.0007,0.0005,0
0.005,0.005,0.0008,0.0005,0
0.005,0.005,0.0005,0.0004,0
0.005,0.005,0.0005,0.0003,0
0.005,0.005,0.0005,0.0002,0
0.005,0.005,0.0005,0.0006,1
0.005,0.005,0.0005,0.0007,1
0.005,0.005,0.0005,0.0008,1
0.004,0.007,0.0008,0.0001,0
0.003,0.006,0.0001,0.0005,0
0.002,0.008,0.0005,0.0002,0
0.006,0.007,0.0001,0.0001,0
0.006,0.003,0.0009,0.0002,1
0.007,0.004,0.0003,0.0009,1
0.008,0.002,0.0008,0.0002,1
0.009,0.003,0.0001,0.0003,1
0.008,0.004,0.0006,0.0005,1
0.003,0.005,0.0001,0.0009,0
0.002,0.006,0.0001,0.0001,0
0.001,0.005,0.0005,0.0005,0
0.009,0.005,0.0005,0.0005,1
0.001,0.002,0.0005,0.0005,0
0.001,0.003,0.0005,0.0005,0
0.001,0.004,0.0005,0.0005,0
0.001,0.006,0.0005,0.0005,0
0.001,0.007,0.0005,0.0005,0
0.001,0.008,0.0005,0.0005,0
0.001,0.009,0.0005,0.0005,0

0.002,0.001,0.0005,0.0005,1
0.003,0.001,0.0005,0.0005,1
0.004,0.001,0.0005,0.0005,1
0.005,0.001,0.0005,0.0005,1
0.006,0.001,0.0005,0.0005,1
0.007,0.001,0.0005,0.0005,1
0.008,0.001,0.0005,0.0005,1
0.009,0.001,0.0005,0.0005,1
0.005,0.001,0.0005,0.0005,1
0.005,0.002,0.0005,0.0005,1
0.005,0.003,0.0005,0.0005,1
0.005,0.004,0.0005,0.0005,1
0.005,0.006,0.0005,0.0005,0
0.005,0.007,0.0005,0.0005,0
0.005,0.008,0.0005,0.0005,0
0.005,0.009,0.0005,0.0005,0
0.001,0.001,0.0004,0.0001,1
0.007,0.009,0.0008,0.0001,0
0.008,0.009,0.0009,0.0001,0
0.009,0.009,0.0008,0.0001,0
0.007,0.006,0.0009,0.0001,1
0.008,0.007,0.0009,0.0001,1
0.009,0.008,0.0006,0.0002,1
0.009,0.009,0.0002,0.0007,1
0.004,0.008,0.0002,0.0007,0
0.004,0.007,0.0003,0.0007,0
0.004,0.007,0.0004,0.0007,0
0.008,0.004,0.0007,0.0006,1
0.002,0.004,0.0009,0.0006,0
0.008,0.004,0.0009,0.0001,1
0.007,0.003,0.0006,0.0004,1
0.003,0.007,0.0004,0.0006,0
0.003,0.006,0.0003,0.0007,0
0.003,0.008,0.0002,0.0008,0

APPENDIX B

SEIQR MODEL SAMPLE SET

@relation seiqr_malware_propogation

@attribute a_Blue real
@attribute a_Red real
@attribute mu_Blue real
@attribute mu_Red real
@attribute delta_Blue real
@attribute delta_Red real
@attribute gama_Blue real
@attribute gama_Red real
@attribute epsilon_Blue real
@attribute epsilon_Red real
@attribute class {0, 1}

@data
0.2,0.1,0.1,0.1,2,2,2,2,0.1,0.1,1
0.3,0.1,0.1,0.1,2,2,2,2,0.1,0.1,1
0.4,0.1,0.1,0.1,2,2,2,2,0.1,0.1,1
0.5,0.1,0.1,0.1,2,2,2,2,0.1,0.1,1
0.1,0.2,0.1,0.1,2,2,2,2,0.1,0.1,0
0.1,0.3,0.1,0.1,2,2,2,2,0.1,0.1,0
0.1,0.4,0.1,0.1,2,2,2,2,0.1,0.1,0
0.1,0.5,0.1,0.1,2,2,2,2,0.1,0.1,0
0.1,0.1,0.2,0.1,2,2,2,2,0.1,0.1,1
0.1,0.1,0.3,0.1,2,2,2,2,0.1,0.1,1
0.1,0.1,0.4,0.1,2,2,2,2,0.1,0.1,1
0.1,0.1,0.5,0.1,2,2,2,2,0.1,0.1,1
0.1,0.1,0.1,0.2,2,2,2,2,0.1,0.1,0
0.1,0.1,0.1,0.3,2,2,2,2,0.1,0.1,0
0.1,0.1,0.1,0.4,2,2,2,2,0.1,0.1,0
0.1,0.1,0.1,0.5,2,2,2,2,0.1,0.1,0
0.1,0.1,0.1,0.1,4,2,2,2,0.1,0.1,1
0.1,0.1,0.1,0.1,6,2,2,2,0.1,0.1,1
0.1,0.1,0.1,0.1,8,2,2,2,0.1,0.1,1
0.1,0.1,0.1,0.1,2,4,2,2,0.1,0.1,0
0.1,0.1,0.1,0.1,2,6,2,2,0.1,0.1,0
0.1,0.1,0.1,0.1,2,8,2,2,0.1,0.1,0
0.1,0.1,0.1,0.1,2,2,4,2,0.1,0.1,0
0.1,0.1,0.1,0.1,2,2,6,2,0.1,0.1,0
0.1,0.1,0.1,0.1,2,2,8,2,0.1,0.1,0
0.1,0.1,0.1,0.1,2,2,2,4,0.1,0.1,1
0.1,0.1,0.1,0.1,2,2,2,6,0.1,0.1,1
0.1,0.1,0.1,0.1,2,2,2,8,0.1,0.1,1
0.1,0.1,0.1,0.1,2,2,2,2,0.2,0.1,0
0.1,0.1,0.1,0.1,2,2,2,2,0.3,0.1,0
0.1,0.1,0.1,0.1,2,2,2,2,0.4,0.1,0
0.1,0.1,0.1,0.1,2,2,2,2,0.5,0.1,0
0.1,0.1,0.1,0.1,2,2,2,2,0.1,0.2,1
0.1,0.1,0.1,0.1,2,2,2,2,0.1,0.3,1
0.1,0.1,0.1,0.1,2,2,2,2,0.1,0.4,1
0.1,0.1,0.1,0.1,2,2,2,2,0.1,0.5,1
0.2,0.1,0.2,0.1,2,2,2,2,0.1,0.1,1
0.2,0.1,0.3,0.1,2,2,2,2,0.1,0.1,1
0.2,0.1,0.4,0.1,2,2,2,2,0.1,0.1,1
0.2,0.1,0.5,0.1,2,2,2,2,0.1,0.1,1
0.2,0.1,0.2,0.2,2,2,2,2,0.1,0.1,1
0.2,0.1,0.2,0.3,2,2,2,2,0.1,0.1,1
0.2,0.1,0.2,0.4,2,2,2,2,0.1,0.1,0

0.2,0.1,0.2,0.5,2,2,2,2,0.1,0.1,0
0.2,0.1,0.2,0.1,4,2,2,2,0.1,0.1,1
0.2,0.1,0.2,0.1,6,2,2,2,0.1,0.1,1
0.2,0.1,0.2,0.1,8,2,2,2,0.1,0.1,1
0.2,0.1,0.2,0.1,2,4,2,2,0.1,0.1,1
0.2,0.1,0.2,0.1,2,6,2,2,0.1,0.1,1
0.2,0.1,0.2,0.1,2,8,2,2,0.1,0.1,1
0.2,0.1,0.2,0.1,2,6,4,2,0.1,0.1,1
0.2,0.1,0.2,0.1,2,6,6,2,0.1,0.1,0
0.2,0.1,0.2,0.1,2,6,8,2,0.1,0.1,0
0.2,0.1,0.2,0.1,2,6,2,4,0.1,0.1,1
0.2,0.1,0.2,0.1,2,6,2,6,0.1,0.1,1
0.2,0.1,0.2,0.1,2,6,2,8,0.1,0.1,1
0.2,0.1,0.2,0.1,2,6,2,2,0.2,0.1,1
0.2,0.1,0.2,0.1,2,6,2,2,0.3,0.1,1
0.2,0.1,0.2,0.1,2,6,2,2,0.4,0.1,1
0.2,0.1,0.2,0.1,2,6,2,2,0.5,0.1,0
0.2,0.1,0.2,0.1,2,6,2,2,0.4,0.2,1
0.2,0.1,0.2,0.1,2,6,2,2,0.4,0.3,1
0.2,0.1,0.2,0.1,2,6,2,2,0.4,0.4,1
0.2,0.1,0.2,0.1,2,6,2,2,0.4,0.5,1
0.2,0.3,0.3,0.4,2,6,2,2,0.4,0.1,0
0.2,0.3,0.3,0.5,2,6,2,2,0.4,0.1,0
0.2,0.3,0.3,0.1,4,6,2,2,0.4,0.1,0
0.2,0.3,0.3,0.1,6,6,2,2,0.4,0.1,1
0.2,0.3,0.3,0.1,8,6,2,2,0.4,0.1,1
0.2,0.3,0.3,0.1,8,2,2,2,0.4,0.1,1
0.2,0.3,0.3,0.1,8,4,2,2,0.4,0.1,1
0.2,0.3,0.3,0.1,8,8,2,2,0.4,0.1,1
0.2,0.3,0.3,0.1,8,6,4,2,0.4,0.1,0
0.2,0.3,0.3,0.1,8,6,6,2,0.4,0.1,0
0.2,0.3,0.3,0.1,8,6,8,2,0.4,0.1,0
0.2,0.3,0.3,0.1,8,6,2,4,0.4,0.1,1
0.2,0.3,0.3,0.1,8,6,2,6,0.4,0.1,1
0.2,0.3,0.3,0.1,8,6,2,8,0.4,0.1,1
0.2,0.3,0.3,0.1,8,6,2,2,0.1,0.2,1
0.2,0.3,0.3,0.1,8,6,2,2,0.1,0.3,1
0.2,0.3,0.3,0.1,8,6,2,2,0.1,0.4,1
0.2,0.3,0.3,0.1,8,6,2,2,0.1,0.5,1
0.2,0.3,0.3,0.4,2,6,2,2,0.1,0.5,1
0.2,0.3,0.3,0.4,4,6,2,2,0.1,0.1,1
0.2,0.3,0.3,0.4,6,6,2,2,0.1,0.1,1
0.2,0.3,0.3,0.4,8,6,2,2,0.1,0.1,1
0.2,0.3,0.3,0.4,8,2,2,2,0.1,0.1,0
0.2,0.3,0.3,0.4,8,4,2,2,0.1,0.1,0
0.2,0.3,0.3,0.4,8,6,2,2,0.1,0.1,0
0.2,0.3,0.3,0.4,8,8,2,2,0.1,0.1,0
0.2,0.3,0.3,0.4,8,6,4,2,0.1,0.1,0
0.2,0.3,0.3,0.4,8,6,6,2,0.1,0.1,0
0.2,0.3,0.3,0.4,8,6,8,2,0.1,0.1,0
0.2,0.3,0.3,0.4,8,6,2,4,0.1,0.1,1
0.2,0.3,0.3,0.4,8,6,2,6,0.1,0.1,1
0.2,0.3,0.3,0.4,8,6,2,8,0.1,0.1,1
0.2,0.3,0.3,0.4,8,6,2,2,0.2,0.1,0
0.2,0.3,0.3,0.4,8,6,2,2,0.3,0.1,0
0.2,0.3,0.3,0.4,8,6,2,2,0.4,0.1,0
0.2,0.3,0.3,0.4,8,6,2,2,0.5,0.1,0
0.2,0.3,0.3,0.4,8,6,2,2,0.4,0.2,0
0.2,0.3,0.3,0.4,8,6,2,2,0.4,0.3,0
0.2,0.3,0.3,0.4,8,6,2,2,0.4,0.4,0
0.2,0.3,0.3,0.4,8,6,2,2,0.4,0.5,0
0.3,0.1,0.3,0.4,8,6,2,2,0.4,0.1,0
0.3,0.2,0.3,0.4,8,6,2,2,0.4,0.1,0
0.3,0.3,0.3,0.4,8,6,2,2,0.4,0.1,0
0.3,0.4,0.3,0.4,8,6,2,2,0.4,0.1,0
0.3,0.5,0.3,0.4,8,6,2,2,0.4,0.1,0
0.3,0.5,0.1,0.4,8,6,2,2,0.4,0.1,0
0.3,0.5,0.2,0.4,8,6,2,2,0.4,0.1,0
0.3,0.5,0.3,0.4,8,6,2,2,0.4,0.1,0
0.3,0.5,0.4,0.4,8,6,2,2,0.4,0.1,0
0.3,0.5,0.5,0.4,8,6,2,2,0.4,0.1,0
0.3,0.5,0.3,0.1,8,6,2,2,0.4,0.1,1
0.3,0.5,0.3,0.2,8,6,2,2,0.4,0.1,0
0.3,0.5,0.3,0.3,8,6,2,2,0.4,0.1,0
0.3,0.5,0.3,0.4,8,6,2,2,0.4,0.1,0
0.3,0.5,0.3,0.5,8,6,2,2,0.4,0.1,0
0.3,0.5,0.3,0.5,2,6,2,2,0.4,0.1,0
0.3,0.5,0.3,0.5,4,6,2,2,0.4,0.1,0
0.3,0.5,0.3,0.5,6,6,2,2,0.4,0.1,0
0.3,0.5,0.3,0.5,8,6,2,2,0.4,0.1,0
0.3,0.5,0.3,0.5,8,2,2,2,0.4,0.1,0
0.3,0.5,0.3,0.5,8,4,2,2,0.4,0.1,0
0.3,0.5,0.3,0.5,8,6,2,2,0.4,0.1,0
0.3,0.5,0.3,0.5,8,8,2,2,0.4,0.1,0
0.3,0.5,0.3,0.5,8,6,4,2,0.4,0.1,0
0.3,0.5,0.3,0.5,8,6,6,2,0.4,0.1,0
0.3,0.5,0.3,0.5,8,6,8,2,0.4,0.1,0
0.3,0.5,0.3,0.5,8,6,6,4,0.4,0.1,0
0.3,0.5,0.3,0.5,8,6,6,6,0.4,0.1,0
0.3,0.5,0.3,0.5,8,6,6,8,0.4,0.1,0

0.3,0.5,0.3,0.5,8,6,6,2,0.1,0.1,0
0.3,0.5,0.3,0.5,8,6,6,2,0.2,0.1,0
0.3,0.5,0.3,0.5,8,6,6,2,0.3,0.1,0
0.3,0.5,0.3,0.5,8,6,6,2,0.4,0.1,0
0.3,0.5,0.3,0.5,8,6,6,2,0.5,0.1,0
0.3,0.5,0.3,0.5,8,6,6,2,0.4,0.2,0
0.3,0.5,0.3,0.5,8,6,6,2,0.4,0.3,0
0.3,0.5,0.3,0.5,8,6,6,2,0.4,0.4,0
0.3,0.5,0.3,0.5,8,6,6,2,0.4,0.5,0
0.4,0.5,0.1,0.5,8,6,6,2,0.4,0.2,0
0.4,0.5,0.2,0.5,8,6,6,2,0.4,0.2,0
0.4,0.5,0.3,0.5,8,6,6,2,0.4,0.2,0
0.4,0.5,0.4,0.5,8,6,6,2,0.4,0.2,0
0.4,0.5,0.5,0.5,8,6,6,2,0.4,0.2,0
0.4,0.5,0.5,0.1,8,6,6,2,0.4,0.2,1
0.4,0.5,0.5,0.2,8,6,6,2,0.4,0.2,0
0.4,0.5,0.5,0.3,8,6,6,2,0.4,0.2,0
0.4,0.5,0.5,0.4,8,6,6,2,0.4,0.2,0
0.4,0.5,0.5,0.5,8,6,6,2,0.4,0.2,0
0.4,0.5,0.5,0.5,2,6,6,2,0.4,0.2,0
0.4,0.5,0.5,0.5,4,6,6,2,0.4,0.2,0
0.4,0.5,0.5,0.5,6,6,6,2,0.4,0.2,0
0.4,0.5,0.5,0.5,8,6,6,2,0.4,0.2,0
0.4,0.5,0.5,0.5,8,2,6,2,0.4,0.2,0
0.4,0.5,0.5,0.5,8,4,6,2,0.4,0.2,0
0.4,0.5,0.5,0.5,8,6,6,2,0.4,0.2,0
0.4,0.5,0.5,0.5,8,8,6,2,0.4,0.2,0
0.4,0.5,0.5,0.5,8,2,2,2,0.4,0.2,0
0.4,0.5,0.5,0.5,8,2,4,2,0.4,0.2,0
0.4,0.5,0.5,0.5,8,2,6,2,0.4,0.2,0
0.4,0.5,0.5,0.5,8,2,8,2,0.4,0.2,0
0.4,0.5,0.5,0.5,8,2,6,4,0.4,0.2,0
0.4,0.5,0.5,0.5,8,2,6,6,0.4,0.2,1
0.4,0.5,0.5,0.5,8,2,6,8,0.4,0.2,1
0.4,0.5,0.5,0.5,8,2,6,8,0.1,0.2,1
0.4,0.5,0.5,0.5,8,2,6,8,0.2,0.2,1
0.4,0.5,0.5,0.5,8,2,6,8,0.3,0.2,1
0.4,0.5,0.5,0.5,8,2,6,8,0.4,0.2,1
0.4,0.5,0.5,0.5,8,2,6,8,0.5,0.2,1
0.4,0.5,0.5,0.5,8,2,6,8,0.4,0.1,1
0.4,0.5,0.5,0.5,8,2,6,8,0.4,0.3,1
0.4,0.5,0.5,0.5,8,2,6,8,0.4,0.4,1
0.4,0.5,0.5,0.5,8,2,6,8,0.4,0.5,1
0.5,0.5,0.5,0.5,8,2,6,8,0.4,0.5,1
0.5,0.1,0.5,0.5,8,2,6,8,0.4,0.5,1
0.5,0.2,0.5,0.5,8,2,6,8,0.4,0.5,1
0.5,0.3,0.5,0.5,8,2,6,8,0.4,0.5,1
0.5,0.4,0.5,0.5,8,2,6,8,0.4,0.5,1
0.5,0.5,0.1,0.5,8,2,6,8,0.4,0.5,0
0.5,0.5,0.2,0.5,8,2,6,8,0.4,0.5,1
0.5,0.5,0.3,0.5,8,2,6,8,0.4,0.5,1
0.5,0.5,0.4,0.5,8,2,6,8,0.4,0.5,1
0.5,0.5,0.5,0.1,8,2,6,8,0.4,0.5,1
0.5,0.5,0.5,0.2,8,2,6,8,0.4,0.5,1
0.5,0.5,0.5,0.3,8,2,6,8,0.4,0.5,1
0.5,0.5,0.5,0.4,8,2,6,8,0.4,0.5,1
0.5,0.5,0.5,0.5,2,2,6,8,0.4,0.5,1
0.5,0.5,0.5,0.5,4,2,6,8,0.4,0.5,1
0.5,0.5,0.5,0.5,6,2,6,8,0.4,0.5,1
0.5,0.5,0.5,0.2,8,4,6,8,0.4,0.5,1
0.5,0.5,0.5,0.2,8,6,6,8,0.4,0.5,1
0.5,0.5,0.5,0.2,8,8,2,8,0.4,0.5,1
0.5,0.5,0.5,0.2,8,2,4,8,0.4,0.5,1
0.5,0.5,0.5,0.2,8,2,6,8,0.4,0.5,1
0.5,0.5,0.5,0.2,8,2,8,8,0.4,0.5,1
0.5,0.5,0.5,0.2,8,2,6,2,0.4,0.5,1
0.5,0.5,0.5,0.2,8,2,6,4,0.4,0.5,1
0.5,0.5,0.5,0.2,8,2,6,6,0.4,0.5,1
0.5,0.5,0.5,0.2,8,2,6,8,0.1,0.5,1
0.5,0.5,0.5,0.2,8,2,6,8,0.2,0.5,1
0.5,0.5,0.5,0.2,8,2,6,8,0.3,0.5,1
0.5,0.5,0.5,0.2,8,2,6,8,0.4,0.5,1
0.5,0.5,0.5,0.2,8,2,6,8,0.5,0.5,1
0.5,0.5,0.5,0.5,8,2,6,8,0.5,0.5,1

APPENDIX C

SEIR MODEL SAMPLE SET

```
@relation seir_malware_propogation

@attribute beta_Blue real
@attribute beta_Red real
@attribute alfa_Blue real
@attribute alfa_Red real
@attribute gama_Blue real
@attribute gama_Red real
@attribute class {0, 1}

@data
0.2,0.1,0.1,0.1,0.1,0.1,1
0.3,0.1,0.1,0.1,0.1,0.1,1
0.4,0.1,0.1,0.1,0.1,0.1,1
0.5,0.1,0.1,0.1,0.1,0.1,1
0.1,0.2,0.1,0.1,0.1,0.1,0
0.1,0.3,0.1,0.1,0.1,0.1,0
0.1,0.4,0.1,0.1,0.1,0.1,0
0.1,0.5,0.1,0.1,0.1,0.1,0
0.1,0.1,0.2,0.1,0.1,0.1,1
0.1,0.1,0.3,0.1,0.1,0.1,1
0.1,0.1,0.4,0.1,0.1,0.1,1
0.1,0.1,0.5,0.1,0.1,0.1,1
0.1,0.1,0.1,0.2,0.1,0.1,0
0.1,0.1,0.1,0.3,0.1,0.1,0
0.1,0.1,0.1,0.4,0.1,0.1,0
0.1,0.1,0.1,0.5,0.1,0.1,0
0.1,0.1,0.1,0.1,0.2,0.1,0
0.1,0.1,0.1,0.1,0.3,0.1,0
0.1,0.1,0.1,0.1,0.4,0.1,0
0.1,0.1,0.1,0.1,0.5,0.1,0
0.1,0.1,0.1,0.1,0.1,0.2,1
0.1,0.1,0.1,0.1,0.1,0.3,1
0.1,0.1,0.1,0.1,0.1,0.4,1
0.1,0.1,0.1,0.1,0.1,0.5,1
0.2,0.3,0.1,0.1,0.1,0.1,0
0.2,0.4,0.1,0.1,0.1,0.1,0
0.2,0.5,0.1,0.1,0.1,0.1,0
0.2,0.1,0.2,0.1,0.1,0.1,1
0.2,0.1,0.3,0.1,0.1,0.1,1
0.2,0.1,0.4,0.1,0.1,0.1,1
0.2,0.3,0.5,0.1,0.1,0.1,1
0.2,0.4,0.1,0.2,0.1,0.1,0
0.2,0.5,0.1,0.3,0.1,0.1,0
0.2,0.1,0.2,0.4,0.1,0.1,0
0.2,0.1,0.3,0.5,0.1,0.1,0
0.2,0.1,0.4,0.1,0.1,0.1,1
0.2,0.1,0.5,0.1,0.1,0.2,1
0.2,0.1,0.1,0.1,0.2,0.3,1
0.2,0.1,0.1,0.1,0.3,0.4,1
0.2,0.1,0.1,0.1,0.4,0.5,1
0.3,0.2,0.2,0.1,0.4,0.2,0
0.3,0.2,0.3,0.1,0.4,0.3,1
0.3,0.2,0.4,0.1,0.4,0.4,1
0.3,0.2,0.5,0.1,0.4,0.5,1
0.3,0.2,0.1,0.2,0.4,0.1,0
0.3,0.2,0.1,0.3,0.4,0.2,0
0.3,0.2,0.1,0.4,0.1,0.3,1
```

0.3,0.2,0.1,0.5,0.2,0.4,0
0.3,0.2,0.1,0.1,0.3,0.5,1
0.3,0.2,0.1,0.1,0.4,0.1,0
0.3,0.2,0.2,0.5,0.5,0.1,0
0.3,0.2,0.3,0.4,0.1,0.1,0
0.3,0.2,0.4,0.3,0.2,0.1,0
0.3,0.2,0.5,0.2,0.3,0.1,0
0.3,0.2,0.2,0.1,0.4,0.1,0
0.3,0.2,0.3,0.2,0.5,0.1,0
0.3,0.2,0.4,0.3,0.1,0.1,1
0.3,0.2,0.5,0.4,0.2,0.1,0
0.3,0.2,0.2,0.5,0.3,0.1,0
0.3,0.2,0.3,0.1,0.4,0.1,0
0.3,0.2,0.3,0.1,0.4,0.2,1
0.3,0.3,0.3,0.1,0.3,0.3,1
0.3,0.4,0.3,0.1,0.2,0.4,1
0.3,0.5,0.3,0.1,0.1,0.5,1
0.3,0.1,0.3,0.1,0.5,0.1,0
0.3,0.2,0.3,0.1,0.4,0.2,1
0.3,0.3,0.3,0.1,0.3,0.3,1
0.3,0.4,0.3,0.1,0.2,0.4,1
0.3,0.5,0.3,0.1,0.1,0.5,1
0.3,0.1,0.3,0.1,0.4,0.1,0
0.4,0.5,0.4,0.3,0.1,0.1,1
0.4,0.5,0.4,0.3,0.2,0.1,0
0.4,0.5,0.4,0.3,0.3,0.1,0
0.4,0.5,0.4,0.3,0.4,0.1,0
0.4,0.5,0.4,0.3,0.5,0.1,0
0.4,0.5,0.4,0.3,0.1,0.2,1
0.4,0.5,0.4,0.3,0.2,0.2,1
0.4,0.5,0.4,0.3,0.3,0.2,0
0.4,0.5,0.4,0.3,0.4,0.2,0
0.4,0.5,0.4,0.3,0.5,0.2,0
0.5,0.5,0.1,0.3,0.5,0.3,0
0.5,0.5,0.2,0.3,0.5,0.3,0
0.5,0.5,0.3,0.3,0.5,0.3,0
0.5,0.5,0.4,0.3,0.5,0.3,0
0.5,0.5,0.5,0.3,0.5,0.3,0
0.5,0.5,0.4,0.1,0.5,0.3,1
0.5,0.5,0.4,0.2,0.5,0.3,1
0.5,0.5,0.4,0.3,0.5,0.3,0
0.5,0.5,0.4,0.4,0.5,0.3,0
0.5,0.5,0.4,0.5,0.5,0.3,0
0.5,0.5,0.4,0.1,0.1,0.4,1
0.5,0.5,0.4,0.2,0.2,0.4,1
0.5,0.5,0.4,0.3,0.3,0.4,1
0.5,0.5,0.4,0.5,0.5,0.4,0
0.5,0.5,0.4,0.1,0.1,0.3,1
0.5,0.5,0.4,0.2,0.2,0.3,1
0.5,0.5,0.4,0.3,0.3,0.3,1
0.5,0.5,0.4,0.4,0.4,0.3,0
0.5,0.5,0.4,0.5,0.5,0.3,0
0.5,0.4,0.3,0.4,0.5,0.2,0
0.4,0.3,0.3,0.4,0.5,0.2,0
0.3,0.2,0.3,0.4,0.5,0.2,0
0.2,0.1,0.3,0.4,0.5,0.2,0
0.5,0.4,0.2,0.4,0.5,0.2,0
0.5,0.4,0.1,0.4,0.5,0.2,0
0.5,0.4,0.3,0.4,0.4,0.2,0
0.5,0.4,0.3,0.4,0.3,0.2,0
0.5,0.4,0.3,0.4,0.2,0.2,0
0.5,0.4,0.3,0.4,0.1,0.2,1
0.1,0.5,0.1,0.2,0.1,0.5,1
0.1,0.2,0.1,0.2,0.1,0.4,1
0.1,0.3,0.1,0.2,0.1,0.4,1
0.1,0.4,0.1,0.2,0.1,0.4,1
0.1,0.5,0.1,0.2,0.1,0.4,1
0.1,0.2,0.1,0.3,0.1,0.4,1
0.1,0.2,0.1,0.4,0.1,0.4,1
0.1,0.2,0.1,0.5,0.1,0.4,1
0.1,0.2,0.1,0.2,0.1,0.3,1
0.2,0.3,0.2,0.2,0.1,0.5,1
0.2,0.3,0.2,0.3,0.2,0.5,1
0.3,0.3,0.3,0.4,0.3,0.5,1
0.4,0.4,0.3,0.4,0.4,0.5,0
0.5,0.5,0.4,0.5,0.5,0.5,0

TEZ İZİN FORMU / THESIS PERMISSION FORM

ENSTİTÜ / INSTITUTE

Fen Bilimleri Enstitüsü / Graduate School of Natural and Applied Sciences [ ]

Sosyal Bilimler Enstitüsü / Graduate School of Social Sciences [ ]

Uygulamalı Matematik Enstitüsü / Graduate School of Applied Mathematics [ ]

Enformatik Enstitüsü / Graduate School of Informatics [X]

Deniz Bilimleri Enstitüsü / Graduate School of Marine Sciences [ ]

YAZARIN / AUTHOR

Soyadı / Surname : ŞENGÜL
Adı / Name : Zafer
Bölümü / Department : Siber Güvenlik

TEZİN ADI / TITLE OF THE THESIS (**İngilizce** / English) : Modelling the Effects of Malware Propagation on Military Operations by Using Bayesian Network Framework

TEZİN TÜRÜ / DEGREE**:** **Yüksek Lisans** / Master [X] **Doktora** / PhD [ ]

1. **Tezin tamamı dünya çapında erişime açılacaktır. /** Release the entire work immediately for access worldwide. [ ]
2. **Tez iki yıl süreyle erişime kapalı olacaktır.** / Secure the entire work for patent and/or proprietary purposes for a period of **two year. *** [X]
3. **Tez altı ay süreyle erişime kapalı olacaktır.** / Secure the entire work for period of **six months**. ***** [ ]

***** *Enstitü Yönetim Kurulu Kararının basılı kopyası tezle birlikte kütüphaneye teslim edilecektir.*
A copy of the Decision of the Institute Administrative Committee will be delivered to the library together with the printed thesis.

Yazarın imzası / Signature **Tarih** / Date