Predicting Malware Spread Using Stochastic Systems of Differential Equations

Jacob Atkinson, Zachary Babcock, Riley Morgan, Michael Orr

May 9, 2023

Abstract

Malware attacks continue to threaten the world, resulting in millions of dollars of damage every year. Models have been developed to portray the dynamics of malware spread, but haven't considered the stochasticity of malware attacks. This paper derives a new approach to modeling malware spread by incorporating stochasticity into a variety of epidemiology models. These stochastic systems of differential equations provide a time-dependent, dynamic, and stochastic representation of malware distribution in a network, allowing a simulation of the complex nature of malware spread with external factors. Through real-life case studies, the overall effectiveness of these models will be shown in predicting and mitigating malware spread.

1 Introduction

Malware refers to a type of software that is designed to impose harm on computer systems, networks, or devices. There exists a wide range of malicious programs, including, viruses, worms, Trojans, spyware, ransomware, and adware [2]. Malware attacks continue to threaten individuals, businesses, and governments worldwide, causing millions of dollars in damages each year, as well as the corruption of sensitive data [6].

In the past, there have been several large malicious attacks, highlighting the urgent need for effective and efficient models to analyze their behavior and how malware can spread over time. Recently, one of the most notable examples of malware was the WannaCry ransomware attack in 2017. The WannaCry attack affected over 200,000 computers in 150 countries, causing significant economic impact [1]. Other examples include the NotPetya attack and the SolarWinds attack which caused billions of dollars in damages and affected major government agencies and technology companies [4].

Malware is spread through a variety of channels, including email, file-sharing networks, social media, and infected websites. Once a computer or network is corrupted malware can attack in a few different ways. Malware can steal sensitive information, damage files, and disrupt system operations. To help mitigate the impact of malware, research has been conducted to better understand the dynamics of malware spread and devise effective countermeasures [9].

However, models that have been developed often oversimplify the complexity of malware spread, as they don't account for the inherent stochasticity of malware attacks [7]. The stochasticity in malware refers to the randomness and unpredictability of malware attacks. This unpredictability comes from a range of external factors, such as human behavior, software vulnerabilities, and network topology [5]. Therefore, there exists a need for a new approach to better capture the complexity of malware spread.

In this paper, we present a novel method for modeling the spread of malware. We achieve this by incorporating stochasticity into pre-established epidemiology malware models, including the SEIRS, MalSERIS, SIQVD, and SLBQR models [3, 6, 8, 10]. These models provide a more time-dependent, dynamic, and stochastic representation of malware distribution in a network, allowing for a more accurate simulation of malware spread with external factors.

This paper will highlight the urgent need for effective and efficient methods of predicting malware spread, which poses a significant threat to individuals, businesses, and governments worldwide [6]. A new approach to modeling malware spread is proposed, which accounts for the stochasticity of malware attacks and provides a more accurate simulation of malware spread with external factors. Stochasticity will be introduced into the following models: SEIRS, MalSEIRS, SIQVD, and SLBQR.

Through this approach, a better understanding of the dynamics of malware spread can be achieved.

2 Models

2.1 SEIRS

The SEIRS model is an elementary method for modeling the spread of malware within a computer network. In this section we show the differential equations of the SEIRS model. We also show the models parameters.

The SEIRS model consist of 4 differential equations with each equation modeling a phase of infection.

$$\frac{dS}{dt} = \mu N - \beta IS + \omega R - \mu S \tag{1}$$

$$\frac{dE}{dt} = \beta IS - \sigma E - \mu E \tag{2}$$

$$\frac{dI}{dt} = \sigma E - \gamma I - (\mu + \alpha)I \tag{3}$$

$$\frac{dR}{dt} = \gamma I - \omega R - \mu R \tag{4}$$

(1) represents the susceptible node. These are the computers on the network that have not been exposed to a malware carrying computer. In (2) we represent the exposed class of computer. A computer in this node has been exposed to malware but has not yet contracted the contagion. The (3) equation represents the infected group of computers. This group represents a computer that has become infected by the malware propagation. The final equation (4) in the SEIRS model represents the Recovered node of computers. A computer moves to the recovered class when the malware has been cleared from the system.

The table below includes the model parameters with a brief description of each.

SEIRS Parameters	
Symbol	Description
N	total population
μ	machine unavailability rate(not malware)
β	malware propagation rate
ω	loss of immunity rate
σ	malware execution rate
γ	malware removal rate
α	machine unavailability rate(malware)

We will simulate this model using Eulers method to approximate the differential equation compartments. We will then introduce stachasticity into the simulation with hopes of creating a model that more accurately describes the spread of malware.

2.2 MalSEIRS

This section will review the MalSEIRS [6] and will explain the reasons for the different compartments. We will also add a level of stochasticity to the model and show how the results vary.

The MalSEIRS is an updated and innovated model that is derived from the SEIR model. In comparison to the SEIR, the MalSEIRS will allow us to include the following:

1. infection, recovery, and loss of immunity rates to make them time dependent
2. inclusion of computer networks
3. and the concept of vaccine to shield the recovered nodes

Including these items into our equations will prove to be an innovative modification to the SEIR. Below are the differential equations that can be used to model the distribution of malware.

$$\frac{dS}{dt} = p \wedge -\beta(t)IS + \omega(t)R - (\mu - \phi)S \tag{5}$$

$$\frac{dE}{dt} = \beta(t)IS - \sigma E - \mu E \tag{6}$$

$$\frac{dE}{dt} = \sigma E - \gamma(t)I - (\mu + \alpha)I \tag{7}$$

$$\frac{dR}{dt} = (1 - p) \wedge + \gamma(t)I - \omega(t)R - \mu R + \phi S \tag{8}$$

In these equations, $\beta(t)$, $\omega(t)$, and $\gamma(t)$ are our infection rate, recovery removal rate, and loss of immunity rate.

Infection Rate
The infection rate can be denoted as:

$\beta(t) = \frac{\beta_0}{1+\xi I(t)}$

This equation is used to compute the infection rate over time, which takes into account an initial infection rate and a positive constant (ξ) that adjusts the speed of the decrease in the infection rate. This will allow us to adjust for the decay speed.

Loss of Immunity Rate
The loss of immunity rate can be denoted as:

$\omega(t) = |e^{-at} cos(\frac{2\pi t}{m})|$

The initial stages of a virus/malware attack can be unstable and inefficient in terms of the security measures adopted by susceptible and infected nodes. Overall, this can lead to the loss of immunity rate to oscillate. To avoid this, we can use the parameter (a) and parameter (m) to ensure the rate stays between 0 and 1.

Recovery Rate
The recovery rate can be denoted as:

$$\gamma(t) = tanh(\frac{(I(t)}{c})$$

This recovery rate also can vary with time based on the availability of treatment to be applied to the infected nodes. The recovery rate also depends on a positive constant (c) which is used to determine how fast the recovery rate will approach 1.

2.3 SIQVD

The SIQVD model, introduced by Yao et al., is a compartmental model that expands upon the basis of the SEIR model and considers five categories: susceptible (S), infected (I), quarantined (Q), vaccinated (V), and delay hosts (D)[15]. By tracking the number of devices in each status, the SIQVD model can simulate the spread of a malware attack over time and evaluate the effectiveness of different alleviation strategies.The model parameters can be seen in the table below with their description.

Notation	Description
$S(t)$	Susceptible Population
$I(t)$	Infected Population
$Q(t)$	Quarantined Population
$V(t)$	Vaccinated Population
$D(t)$	Delay Hosts Population
$B(t)$	Infection Rate
ξ	Positive Constant of Infection Rate
γ	Recovery Rate for Infected
$\emptyset$	Vaccination Rate
ω	Loss of Immunity
θ	Recovery Rate for Quarantined
δ	Quarantined Rate
τ	Immunity Time

Table 1: SIQVD Model Parameters

These parameters are then used in our SIQVD model to formulate the following set of equations:

$$\begin{cases} \frac{dS}{dt} = -B(t)I(t)S(t) - \emptyset S(t) + \omega V(t-\tau) \\ \frac{dI}{dt} = B(t)I(t)S(t) - (\gamma+\delta)I(t) \\ \frac{dQ}{dt} = \gamma I(t) - \theta Q(t) \\ \frac{dV}{dt} = \emptyset S(t) + \gamma I(t) + \theta Q(t) - \omega V(t) \\ \frac{dD}{dt} = \omega V(t) - \omega V(t-\tau) \end{cases} \tag{9}$$

In this model, the infection rate parameter (β(t)) is time-dependent and uses the same equation mentioned in the previous model (Section 2.2). The rest of the values for the parameters used in this model can be found in Table 4. However, one of the limitations of the SIQVD model is that it does not consider a latency state, meaning the model assumes that the transition from susceptible to infected is immediate. This limitation suggests that the model is suitable for a specific type of malware, such as worms, and overlooks the option of warning victims not to execute the malicious file. The model also considers a constant loss of immunity rate, which means that a certain percentage of devices will never attain an appropriate protective state. While the model has it's limitations, it provides an effective way to analyze malware attacks.

To expand upon this model, we have introduced two stochastic variations to our base equations which will add a level of randomness to the model to simulate the events of the real-world. Our first variation implements the Geometric Brownian motion to our existing differential equations. The updated set of equations for the SIQVD model can be seen in equation 10.

$$\begin{cases} \frac{dS}{dt} = (-B(t)I(t)S(t) - \emptyset S(t)\omega V(t-\tau))dt + \\ \quad \sigma_1 S(t)dX_1 \\ \frac{dI}{dt} = (B(t)I(t)S(t) - (\gamma+\delta)I(t))dt + \sigma_2 I(t)dX_2 \\ \frac{dQ}{dt} = (\gamma I(t) - \theta Q(t))dt + \sigma_3 Q(t)dX_3 \\ \frac{dV}{dt} = (\emptyset S(t) + \gamma I(t) + \theta Q(t) - \omega V(t))dt + \\ \quad \sigma_4 V(t)dX_4 \\ \frac{dD}{dt} = (\omega V(t) - \omega V(t-\tau))dt + \sigma_5 D(t)dX_5 \end{cases} \tag{10}$$

The updated set of equations includes X_i (i = 1, 2, 3, 4, 5) and σ_i (i = 1, 2, 3, 4, 5) which represents the Brownian Motion and standard deviation, respectively. These additions will enhance our model's ability to evaluate the impact of random fluctuations on our simulation of malware attacks. Our objective is to investigate how these changes, which introduce stochasticity, influence the model's precision and accuracy in modeling real-world malware attacks.

The second variation of our base models integrates the Milstein method to our base differential equations. This method is commonly used to model systems that include random, unpredictable factors such as noise or environmental fluctuations. By incorporating the Milstein method into our models, we can better capture the uncertainty and variability of the real world, allowing us to more accurately predict how our system will behave under different conditions. Our new SIQVD model integrating the Milstein method can be seen in equation 11.

$$\begin{cases} \frac{dS}{dt} = (-B(t)I(t)S(t) - \emptyset S(t) + \omega V(t-\tau))dt + \\ \quad \sigma_1 S(t)dX_1 + (\sigma_1/2)S(t)(dX_1^2 - 1)dt \\ \frac{dI}{dt} = (B(t)I(t)S(t) - (\gamma+\delta)I(t))dt + \sigma_2 I(t)dX_2 + \\ \quad (\sigma_2/2)I(t)(dX_2^2 - 1)dt \\ \frac{dQ}{dt} = (\gamma I(t) - \theta Q(t))dt + \sigma_3 Q(t)dX_3 + \\ \quad (\sigma_3/2)Q(t)(dX_3^2 - 1)dt \\ \frac{dV}{dt} = (\emptyset S(t) + \gamma I(t) + \theta Q(t) - \omega V(t))dt + \\ \quad \sigma_4 V(t)dX_4 + (\sigma_4/2)V(t)(dX_4^2 - 1)dt \\ \frac{dD}{dt} = (\omega V(t) - \omega V(t-\tau))dt + \sigma_5 D(t)dX_5 + \\ \quad (\sigma_5/2)D(t)(dX_5^2 - 1)dt \end{cases} \tag{11}$$

The addition of the Milstein method adds to our existing Geometric Brownian motion modification by approximating the stochastic term using a Taylor series expansion and including a correction term that takes into account the second-order effects of the stochastic term. This additional correction term aids in improving the accuracy of the simulation. Overall, the inclusion of the Milstein method into our differential equations provides a powerful tool for simulating stochastic behavior.

2.4 SLBQR

The SLBQR model is another useful method for simulating malware attacks, particularly in mobile internet spaces. This model can accurately capture the complex patterns and characteristics of worm propagation. Qheng, Zhu, and Lai have developed a well-regarded SLBQR model that incorporates temporary immunity, vaccination, and quarantined strategies for worm propagation attacks in mobile spaces [10]. The model consists of five statuses: (S) denotes the susceptible status, where mobile devices are vulnerable to worm attacks; (L) indicates the latency status, where devices have been infected by worms but are not yet exhibiting symptoms; (B) signifies the breaking-out status, in which mobile devices are actively spreading the worm breakout; (Q) denotes the devices that are currently quarantined by a BS station; and (R) represents the recovered status, where devices have implemented some security countermeasure to temporarily avoid malicious worms. The defined parameters can be understood in the following table.

Notation	Description
$N(t)$	Total Population Size
$S(t)$	Susceptible Population
$L(t)$	Latent Population
$B(t)$	Breaking-Out Population
$Q(t)$	Quarantined Population
$R(t)$	Recovered Population
β	Status Transition Rate from S to L
α	Status Transition Rate from L to B
δ	Status Transition Rate from B to Q
λ	Status Transition Rate from Q to R
ϵ	Status Transition Rate from L to R
μ	The Natural Birth and Death Rate
τ	Immunity Time from R to S

Table 2: SLBQR Model Parameters

$$\begin{cases} \frac{dS(t)}{dt} = -\beta S(t)(L(t)+B(t)) - \mu S(t) + \mu N \\ \qquad + \epsilon L(t-\tau)e^{-\mu\tau} + \lambda Q(t-\tau)e^{-\mu\tau} \\ \frac{dL(t)}{dt} = \beta S(t)(L(t)+B(t)) - \mu L(t) - \alpha L(t) \\ \qquad - \epsilon L(t) \\ \frac{dB(t)}{dt} = \alpha L(t) - \delta B(t) - \mu B(t) \\ \frac{dQ(t)}{dt} = \delta B - \mu Q(t) - \lambda Q(t) \\ \frac{dR(t)}{dt} = \epsilon L(t) + \lambda Q(t) - \mu R(t) \\ \qquad - \epsilon L(t-\tau)e^{-\mu\tau} - \lambda Q(t-\tau)e^{-\mu\tau} \end{cases} \tag{12}$$

The SLBQR model parameters described in Table 5 allow for a deeper understanding of the complex interactions between the different groups in the model. These rates play a crucial role in determining the rate of propagation of the worm and the impact it has on the mobile space. For example, the transition rate from the susceptible status to the latent status (β) determines the rate at which devices become infected, while the transition rate from the breaking-out status to the quarantined status (δ) affects the number of devices that can be effectively isolated and prevented from spreading the worm further. The SLBQR model is unique in that it incorporates temporary immunity and vaccination, which can significantly impact the propagation of the worm. The interaction of these parameters with each other and with the different groups in the model can be further explored through the following equations.

By incorporating parameters such as temporary immunity, vaccination, and quarantined strategies, the SLBQR model captures the complex interactions between different populations in the mobile space. However, the model is deterministic, assuming that the transitions between different statuses occur at fixed rates, which may not reflect the stochastic nature of real-world events. This limitation of the model can be addressed by introducing stochasticity, which would better reflect the unpredictable and random nature of real-world events. Therefore, a stochastic extension to the SLBQR model can be introduced, which will better simulate the uncertainty and variability of worm outbreaks in mobile spaces. This extension will involve incorporating randomness into the model, allowing for a more accurate representation of the effects of real-world events on worm propagation. The introduction of stochasticity into the model can be understood by the following updated equations.

$$
\begin{cases}
\frac{dS(t)}{dt} = (-\beta S(t)(L(t)+B(t)) - \mu S(t) + \mu N \\
\qquad + \epsilon L(t-\tau)e^{-\mu\tau} + \lambda Q(t-\tau)e^{-\mu\tau})dt \\
\qquad + \sigma_1 S(t) dX_1 \\
\frac{dL(t)}{dt} = (\beta S(t)(L(t)+B(t)) - \mu L(t) - \alpha L(t) \\
\qquad - \epsilon L(t))dt + \sigma_2 L(t) dX_2 \\
\frac{dB(t)}{dt} = (\alpha L(t) - \delta B(t) - \mu B(t))dt \\
\qquad + \sigma_3 B(t) dX_3 \\
\frac{dQ(t)}{dt} = (\delta B - \mu Q(t) - \lambda Q(t))dt \\
\qquad + \sigma_4 Q(t) dX_4 \\
\frac{dR(t)}{dt} = (\epsilon L(t) + \lambda Q(t) - \mu R(t) \\
\qquad - \epsilon L(t-\tau)e^{-\mu\tau} - \lambda Q(t-\tau)e^{-\mu\tau})dt \\
\qquad + \sigma_5 R(t) dX_5
\end{cases}
\tag{13}
$$

Where X_i, $(i = 1, 2, 3, 4, 5)$ represent standard Brownian motions and σ_i, $(i = 1, 2, 3, 4, 5)$ represents the amount of white noise in the model. Lower levels of white noise represent lower random fluctuations in the model, whereas higher levels of white noise represent higher random fluctuations.

The SLBQR model is a valuable method for simulating malware attacks in mobile internet spaces. With its ability to capture the complex patterns and characteristics of worm propagation, the SLBQR model is well-regarded for its incorporation of temporary immunity, vaccination, and quarantined strategies. The different statuses in the model, as well as the defined parameters, allow for a deeper understanding of the complex interactions between the different groups and the impact on the mobile space. By incorporating stochasticity into the model, its accuracy can be improved and the unpredictability of real-world events can be captured. The interaction of the various parameters and the groups in the model will be further explored through simulations later in the paper.

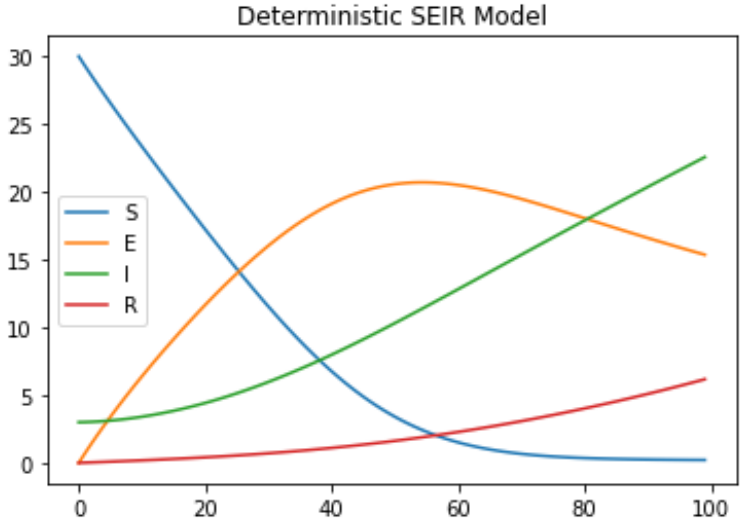
3 Results

3.1 SEIRS

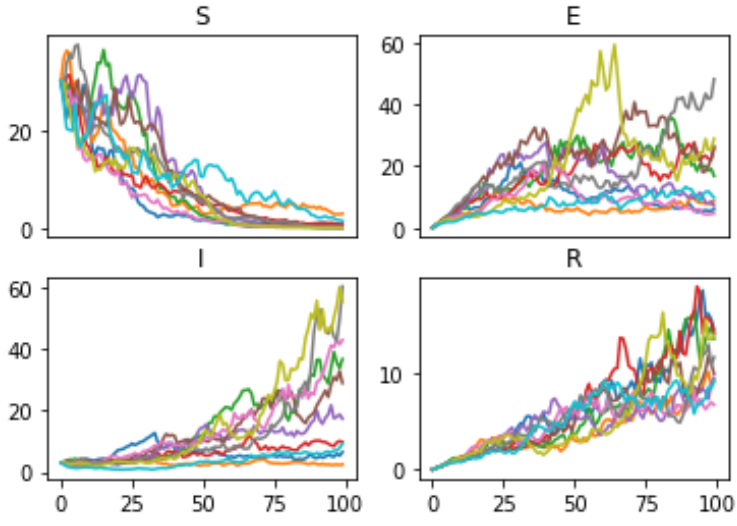
In our simulation of the SEIRS model we use a total population of 33 where 30 computers start in the susceptible node and 3 computers start as infected systems. We aim to simulate the spread of malware using the following parameters values.

SEIRS Parameters Values	
Symbol	Value
N	33
μ	0.1
β	0.8
ω	0.05
σ	0.9
γ	0.6
α	0.1

We first run our simulation as deterministic. This establishes a baseline and corroborates the results of [3]. In the plot below you can find each of the 4 model compartments simulations.



We can then add stochasticity into the model. Using a sigma value of 1 we aim to simulate multiple possible pseudo random outputs of the model.



With the approach of generating multiple possible outputs, we can take an average of all the outcomes. If our model has a high degree of accuracy this average should be the most likely possible output of each differential equation compartment.

3.2 MalSEIRS

We begin our analysis of the MalSEIR model by using these initial conditions.
S(0) = 30
E(0) = 2
I(0) = 1
R(0) = 0

Table 2 shows our initial parameters that were picked for our analysis. Similar to [6], we are assuming that the malware would be infectious at the first stage of the attack and that it would execute itself right after its installation, which is represented by a β_0 and σ close to the upper limit. As well as this, chosen values for a and m were given to indicate the nodes are more vulnerable to lose the immunity over time.

Symbol	Description	Value
β_0	Malware initial propagation rate	.8
ξ	Parameter for infection rate	1
μ	Machine unavailability rate provoked by other causes	.1
α	Machine unavailability rate provoked by malware	.2
c	Parameter for recovered rate	5
σ	Malware execution rate	.9
ϕ	Immunization rate	.46
p	Birth susceptibility rate	.1
$\wedge$	Number of new nodes	1
a	Parameter to control oscillations in loss of immunity rate	.1
m	Parameter to control oscillations in loss of immunity rate	1

Table 3: Initial Parameters for MalSEIR

In our MalSEIR model, we can add stochasticity to it to capture the randomness and unpredictability of the malware in the computer. Malware will behave in different ways depending on the user traffic, network traffic, configuration, and protection. This can help simulate these unpredictable behaviors and provide a more realistic representation of how malware may spread through the computer system. To incorporate stochasticity into our model, we will have to add additional data to our model. See below,

$$\frac{dS}{dt} = (p\wedge - \beta(t)IS + \omega(t)R - (\mu-\phi)S) * dt + k * \sqrt{dt} * \mathcal{N}(0,1) * S_{i-1} \tag{14}$$

$$\frac{dE}{dt} = (\beta(t)IS - \sigma E - \mu E) * dt + k * \sqrt{dt} * \mathcal{N}(0,1) * E_{i-1} \tag{15}$$

$$\frac{dI}{dt} = (\sigma E - \gamma(t)I - (\mu+\alpha)I) * dt + k * \sqrt{dt} * \mathcal{N}(0,1) * I_{i-1} \tag{16}$$

$$\frac{dR}{dt} = ((1-p)\wedge + \gamma(t)I - \omega(t)R - \mu R + \phi S) * dt + k * \sqrt{dt} * \mathcal{N}(0,1) * R_{i-1} \tag{17}$$

where k is our standard deviation.
Below we are going to look at our compartments while analyze three different values of β_0: .8, .5, and .2. Included in the results are also the added stochasticity. Starting at the top graphs, our $\beta_0 = .2$. Middle graph, $\beta_0 = .5$. Last, our Bottom graphs, $\beta_0 = .8$. Here our malware execution rate will remain at .9.

In the Figure 2 group, we notice the oscillating that is happening between the susceptible and recovered groups. We also notice how the infected, and exposed groups increase as β_0 grows in value.

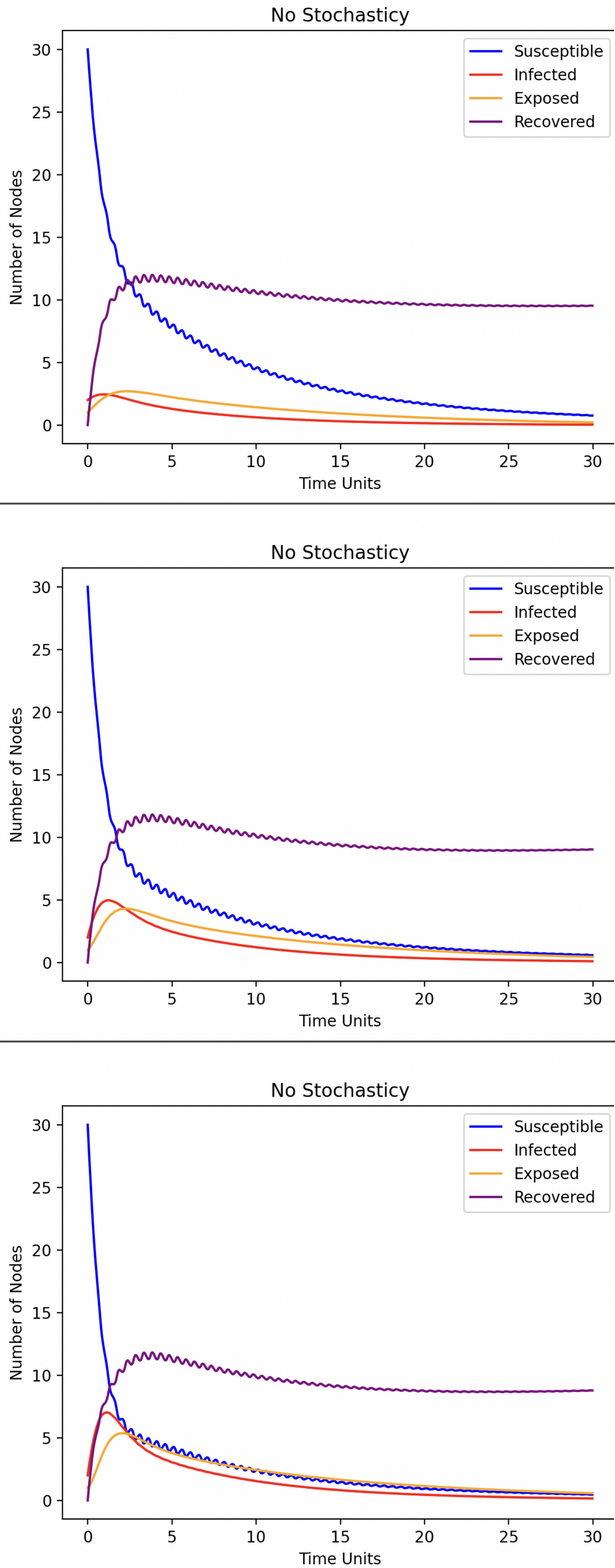


Figure 1: No stochasticity $\beta_0 = .2, .5, .8$

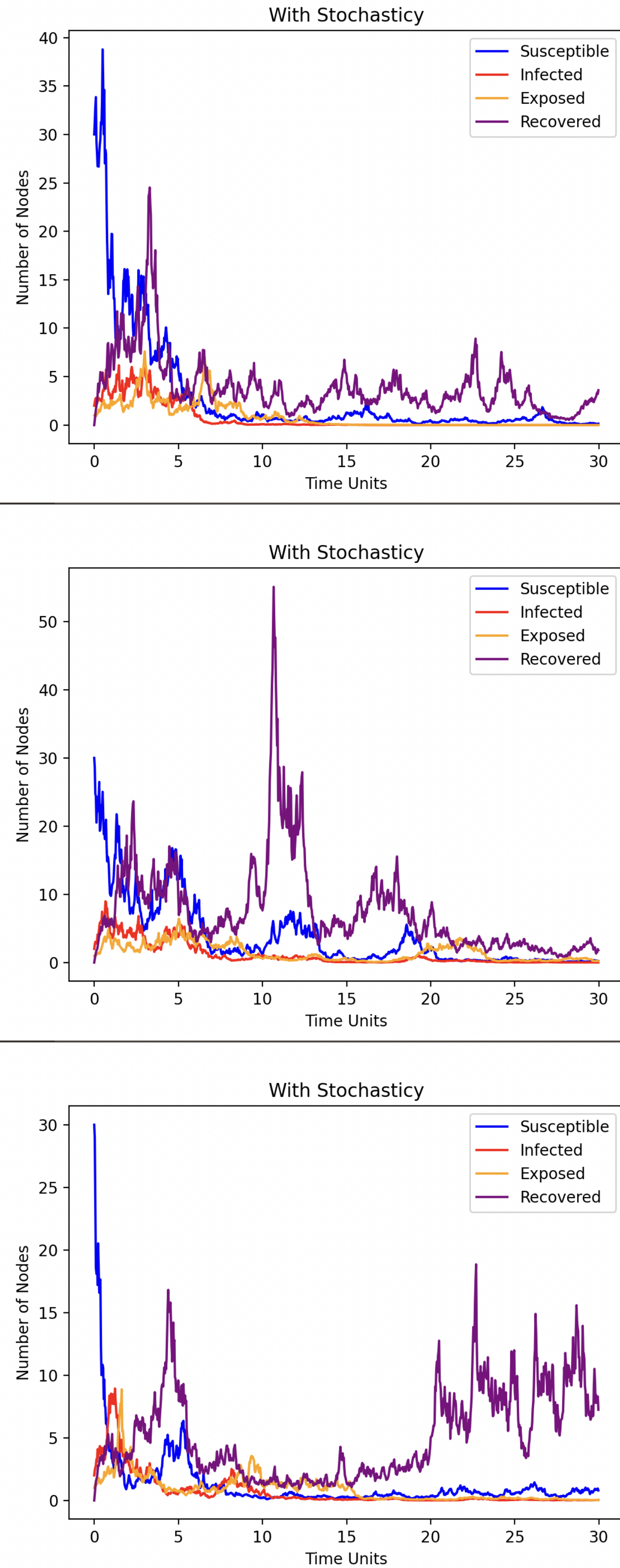


Figure 2: With stochasticity $\beta_0 = .2, .5, .8$

Right off the bat we notice that increasing our β_0 will result in higher exposed and infected compartments. As stated earlier, β_0 is our infection rate so it supports our claim that the more β_0 is raised, the more exposed and infected our nodes become.

In addition to adding stochasticity to our model, enhancing the accuracy and stability of our solution can be made by implementing the Milstein Method. We can improve our approximations by applying this other numerical method. How might the model look with this stability included?

(A,B,C, and D represent our original differential equations at the beginning of the section)

$$\frac{dS}{dt} = A+k*\sqrt{dt}*\mathcal{N}(0,1)*S_{i-1}+(k/2)*S_{i-1}*((\mathcal{N}(0,1)^2)-1)*dt \tag{18}$$

$$\frac{dE}{dt} = B+k*\sqrt{dt}*\mathcal{N}(0,1)*E_{i-1}+(k/2)*E_{i-1}*((\mathcal{N}(0,1)^2)-1)*dt \tag{19}$$

$$\frac{dI}{dt} = C+k*\sqrt{dt}*\mathcal{N}(0,1)*I_{i-1}+(k/2)*I_{i-1}*((\mathcal{N}(0,1)^2)-1)*dt \tag{20}$$

$$\frac{dR}{dt} = D+k*\sqrt{dt}*\mathcal{N}(0,1)*R_{i-1}+(k/2)*R_{i-1}*((\mathcal{N}(0,1)^2)-1)*dt \tag{21}$$

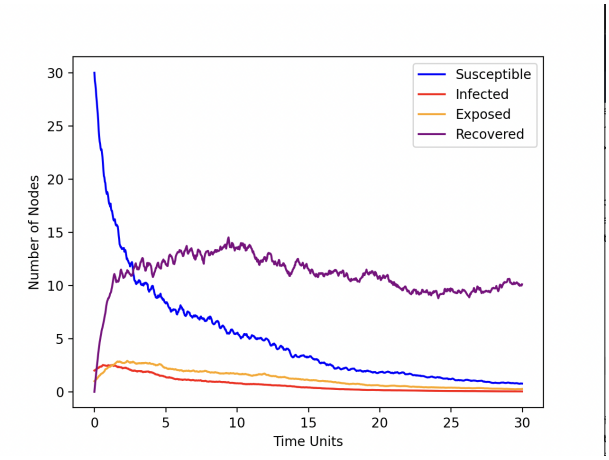


Figure 3: Milstein .1 SD

We notice that when applying this method, the progression of the virus in the computer system appears similar to the progression of the disease without stochasticity. This could suggest that the deterministic components of the model are the dominant factors in driving the dynamics of the system.

Notice when the standard deviation (k) raises to .5 there is little difference.

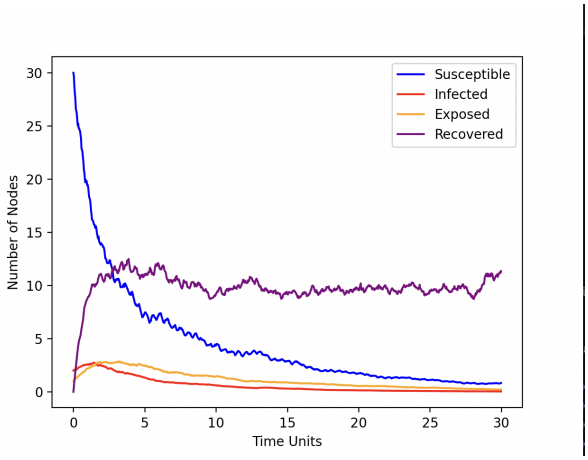


Figure 4: Milstein .5 SD

3.3 SIQVD

We can simulate the spread of malware attacks over time by utilizing both the deterministic and stochastic equations from the SIQVD model, as shown in equations 9, 10, and 11. To generate our results, we will refer to the values provided in Table 4.

Parameter	Value
$S(0)$	30
$I(0)$	3
$Q(0)$	0
$V(0)$	0
$D(0)$	0
β	0.8
ξ	1
γ	0.5
$\varnothing$	0.46
ω	0.05
θ	0.6
δ	.2
τ	5
σ_i	[0, 0.1, 0.2, 0.3]

Table 4: SIQVD Model Parameters

The deterministic model results, as shown in figure 5,

demonstrate a synchronized increase in vaccinated nodes and decrease in susceptible nodes. Eventually, each compartment reaches an equilibrium and remain at a constant state. Furthermore, we observed that the model's behavior was smooth and did not respond to any events over time as it did not include randomness.

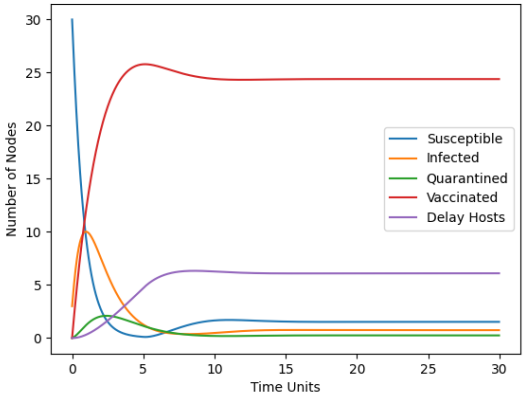


Figure 5: Deterministic SIQVD

On the other hand, when we introduce stochasticity to our SIQVD equations, we immediately observe major changes in the simulations.

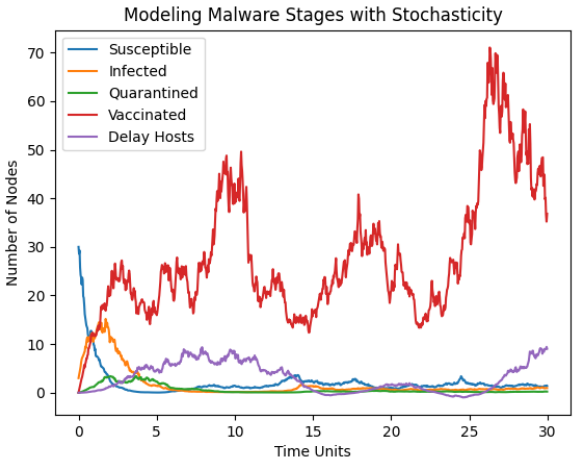
Figure 6a depicts the simulation using the same parameters with the introduction of the Geometric Brownian motion seen in equation 10. In these results, we notice that each compartment continuously changes as it responds to the white noise in the model. While there are differences compared to the deterministic SIQVD model, we notice similar trends in the number of nodes in each compartment over time.

Figures 6b, 6c, and 6d illustrate the impact of different values of σ on the fluctuating values in each compartment. By increasing the σ value, we observe greater randomness in our results and larger variations in the outcomes.

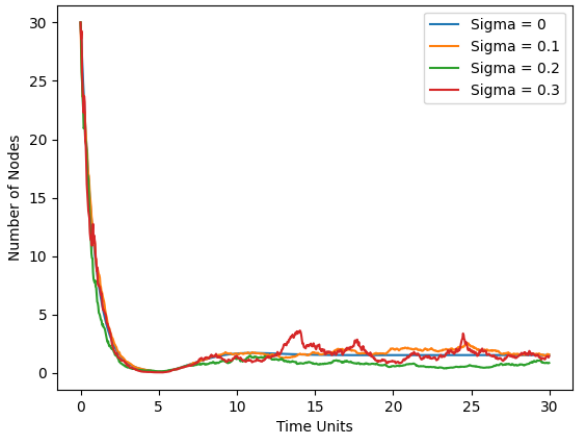
Our last simulation using the SIQVD compartmental model used the Milstein method to add stochasticity as seen in equation 11. The results from the simulations using these equations are found in figures 7a, 7b, 7c, and 7d. Figure 7a displays movement in the values of each compartment over time within the SIQVD model. We notice that the Milstein method equations lead to a similar overall trend as compared to the previous two models. However, this model has more volatility throughout the time-span of the simulation. We notice the vaccinated and delay hosts compartments oscillate in terms of the number of nodes which differs from the previous models.

Additionally, when changing the σ values of $S(t), I(t)$, and $V(t)$ we observe slight changes in our results in comparison to the Geometric Brownian motion model. While $S(t)$, and $I(t)$ follow similar trends, the $V(t)$ component displays a smaller number of nodes with increased σ values, as opposed to the Geometric Brownian motion model where an increase in σ values results in a higher number of nodes.

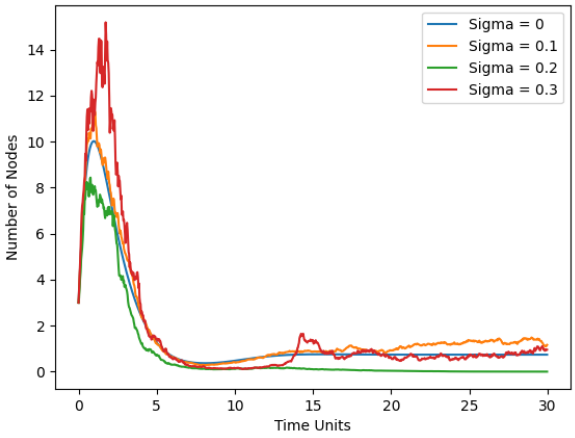
While both deterministic and stochastic models can provide valuable insights into malware attacks, the stochastic models offers a more realistic approach to modeling real-world events. Our simulations using the stochastic model continually change, better reflecting variations over time, making it a powerful tool for accurately modeling malware attacks.



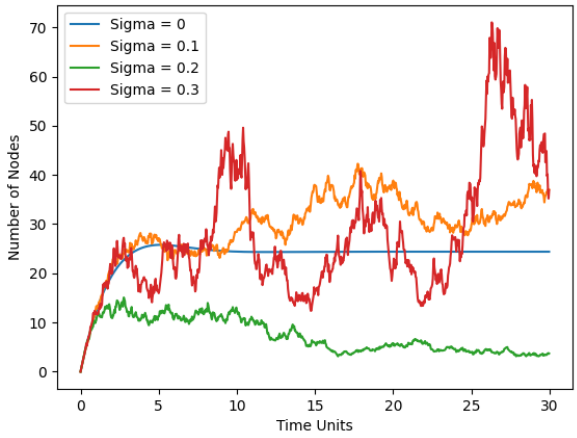
(a) Stochastic SIQVD (Geometric Brownian Motion)



(b) Different Values of σ for $S(t)$ Geometric Brownian Motion

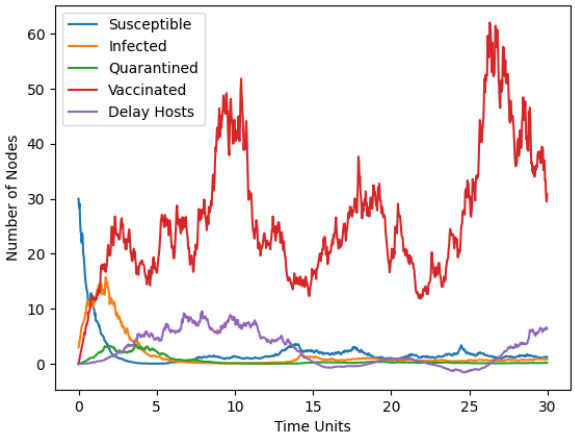


(c) Different Values of σ for $I(t)$ Geometric Brownian Motion

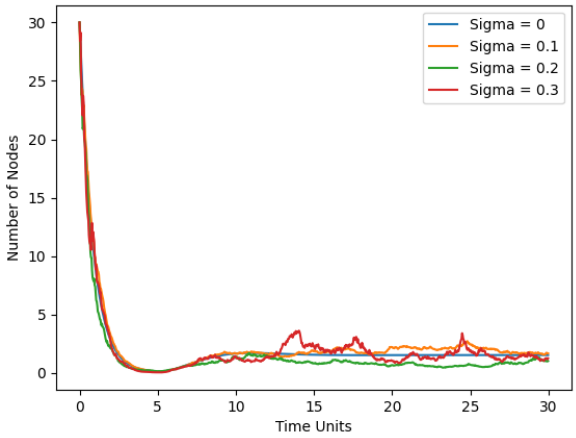


(d) Different Values of σ for $V(t)$ Geometric Brownian Motion

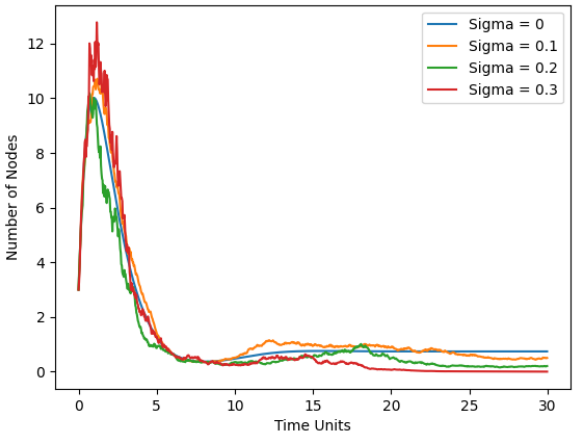
Figure 6: SIQVD Geometric Brownian Simulations



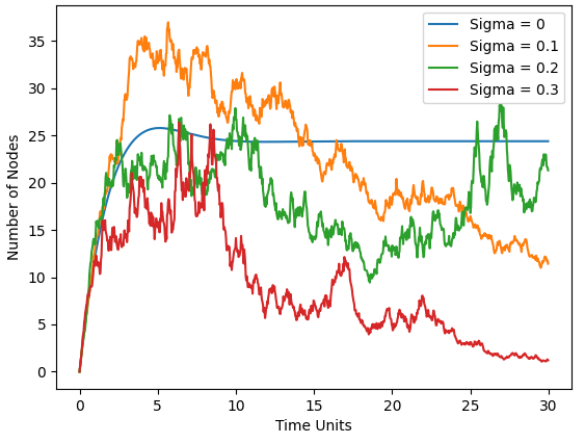
(a) Stochastic SIQVD (Milstein Method)



(b) Different Values of σ for $S(t)$ Milstein Method



(c) Different Values of σ for $I(t)$ Milstein Method



(d) Different Values of σ for $V(t)$ Milstein Method

Figure 7: SIQVD Milstein Method Simulations

3.4 SLBQR

Using the equations from 12 and 13, an analysis of the SLBQR model will be performed to display how adding stochasticity to a pre-existing model can effectively simulate malware spread. The following parameters and initial conditions will be used for consistency with the other models introduced in this paper [6].

Parameter	Value
$N(0)$	33
$S(0)$	30
$L(0)$	2
$B(0)$	1
$Q(0)$	0
$R(0)$	0
β	0.8
α	0.9
δ	0.6
λ	0.6
ϵ	0.46
μ	0.3
τ	5
σ_i	varies

Table 5: SLBQR Model Parameters

By incorporating σ_i for $(i = 1, 2, 3, 4, 5)$ as the white noise, we are able to determine how different levels of white noise will affect the simulation. This will account for the level of randomness and unpredictability in a given scenario, as malware attacks are often distinct. The simulation of the deterministic equations presented in 12 in comparison to the stochastic equations in 13 is shown in figures 8a and 8b.

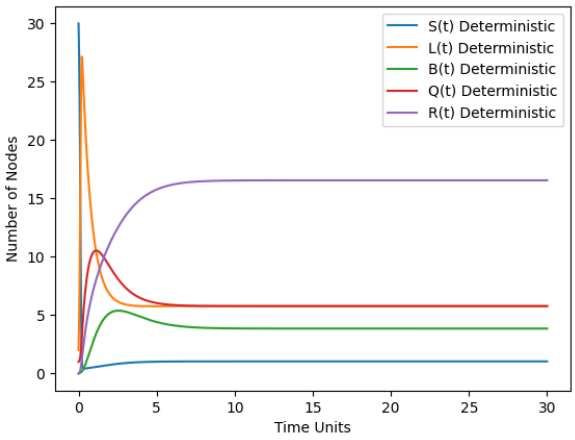
In figures 8a and 8b there are two graphs representing the SLBQR model with identical parameters. In the graph with no stochasticity, the simulation is able to effectively show how the number of nodes throughout a thirty time unit period is affected by a malicious attack. However, there is little to no variation in the model, and it quickly stabilizes after five time units. This isn't a practical application of malicious attacks as they are often long and continuously changing. Introducing a σ value of 0.2, the stochastic figure shown on the right represents the SLBQR model with a small to moderate amount of white noise. The stochastic simulation runs similarly to the deterministic model, however, the variation that is experienced in malicious attacks is better represented. The stochastic model also appears to converge to the same values as the deterministic model, however, it better represents the continuous behavior of malware attacks. Depending on how much variation a malware attack experiences, the stochastic model's σ in 13 can be altered to account for different levels of white noise. Alternate values of σ are shown in figures 8c, 8d, 8e, and 8f for the different population groups.

Now, seeking to improve the overall predictions, the Milstein method will be incorporated into our SLBQR stochastic model as seen in equations 13. Adjusting these equations to incorporate the Milstein approximation by adding in the Taylor Series expansion, the following equations are obtained:

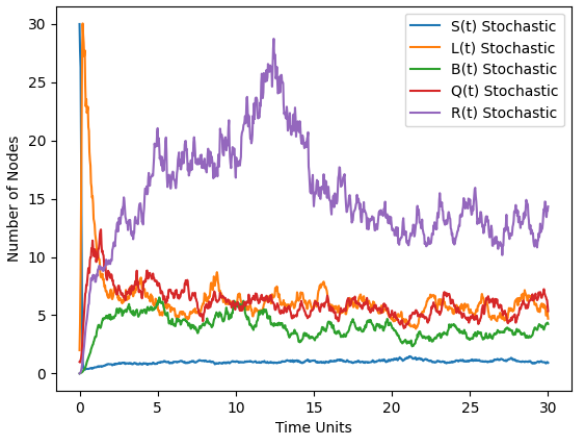
$$\begin{cases} \frac{dS(t)}{dt} = (-\beta S(t)(L(t)+B(t)) - \mu S(t) + \mu N \\ \qquad + \epsilon L(t-\tau)e^{-\mu\tau} + \lambda Q(t-\tau)e^{-\mu\tau})dt \\ \qquad + \sigma_1 S(t)dX_1 + \frac{\sigma_1}{2}S(t)(dX_1^2 - 1)dt \\ \frac{dL(t)}{dt} = (\beta S(t)(L(t)+B(t)) - \mu L(t) - \alpha L(t) \\ \qquad - \epsilon L(t))dt + \sigma_2 L(t)dX_2 \\ \qquad + \frac{\sigma_2}{2}L(t)(dX_2^2 - 1)dt \\ \frac{dB(t)}{dt} = (\alpha L(t) - \delta B(t) - \mu B(t))dt \\ \qquad + \sigma_3 B(t)dX_3 + \frac{\sigma_3}{2}B(t)(dX_3^2 - 1)dt \\ \frac{dQ(t)}{dt} = (\delta B - \mu Q(t) - \lambda Q(t))dt \\ \qquad + \sigma_4 Q(t)dX_4 + \frac{\sigma_4}{2}Q(t)(dX_4^2 - 1)dt \\ \frac{dR(t)}{dt} = (\epsilon L(t) + \lambda Q(t) - \mu R(t) \\ \qquad - \epsilon L(t-\tau)e^{-\mu\tau} - \lambda Q(t-\tau)e^{-\mu\tau})dt \\ \qquad + \sigma_5 R(t)dX_5 + \frac{\sigma_5}{2}R(t)(dX_5^2 - 1)dt \end{cases} \tag{22}$$

As seen in figure 9 the Milstein simulations are very similar to those that are in figure 8. However, the white noise doesn't quite effect the Milstien simulations like it does in the classic stochastic simulations. This is due to the Taylor series approximation method that helps the values converge better to a point. This can allow the introduction of more white noise, and can help improve the reliability of the overall model.

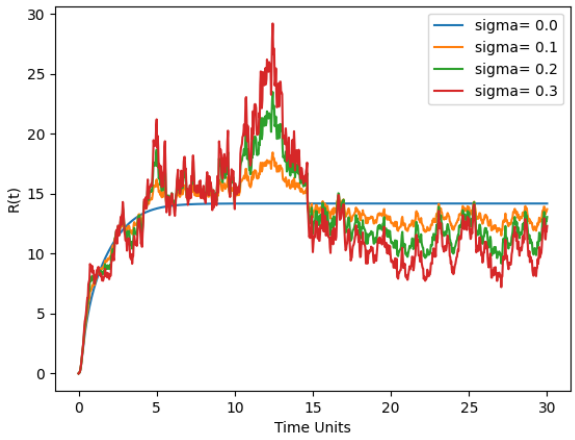
Dependent on the value for σ, it is shown how much the stochastic SLBQR model varies from the original deterministic SLBQR model (shown by the value $\sigma = 0.0$). The SLBQR stochastic model is a powerful tool for sim-



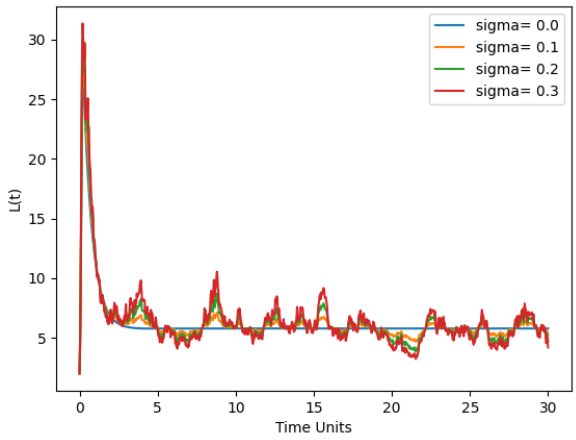
(a) Deterministic SLBQR



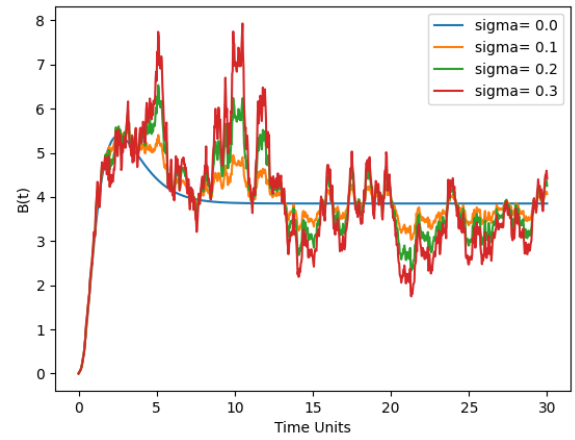
(b) Stochastic SLBQR



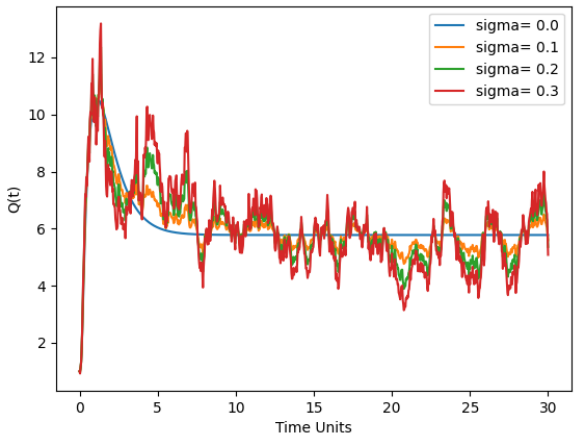
(c) Different Values of σ for $R(t)$



(d) Different Values of σ for $L(t)$

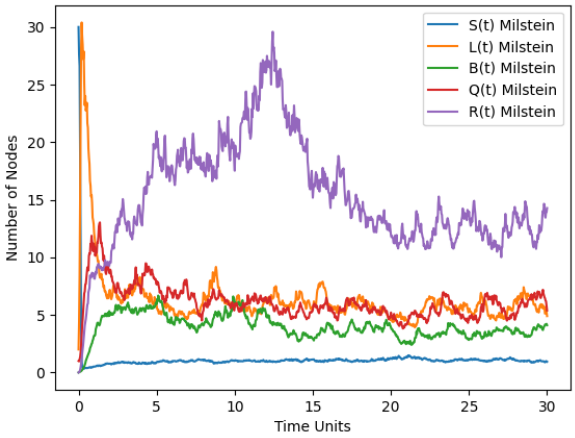


(e) Different Values of σ for $B(t)$

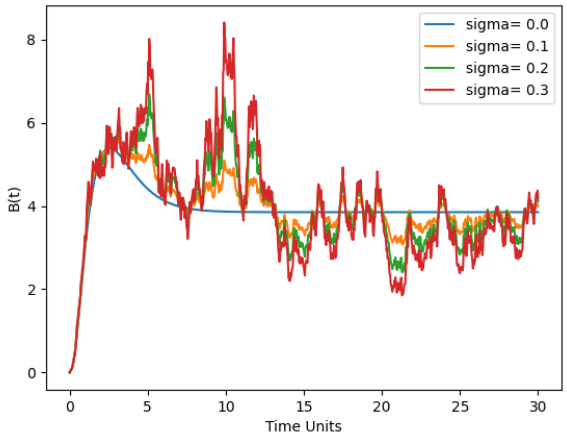


(f) Different Values of σ for $Q(t)$

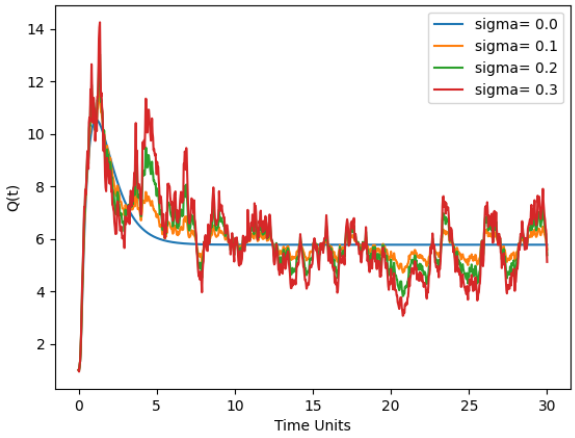
Figure 8: SLBQR Deterministic and Stochastic Simulations



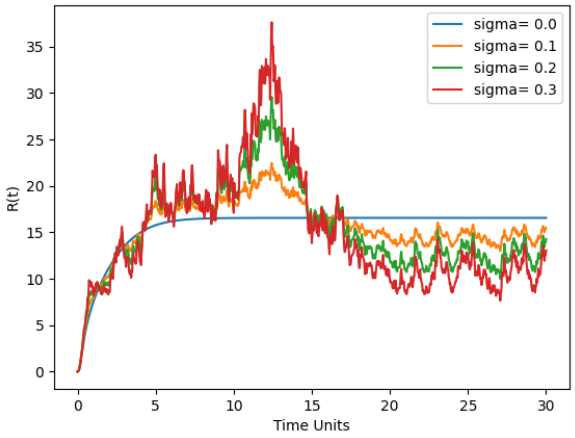
(a) Stochastic SLBQR (Milstein Method)



(b) Different Values of σ for $B(t)$ Milstein Method



(c) Different Values of σ for $Q(t)$ Milstein Method



(d) Different Values of σ for $R(t)$ Milstein Method

Figure 9: SLBQR Milstein Method Simulations

ulating the spread of malware in a population. By incorporating white noise into the model, we can better represent the unpredictable and continuously changing nature of malicious attacks. The comparison between the deterministic and stochastic simulations shows that the stochastic model better represents the real-world behavior of malware attacks. Additionally, the flexibility of the model allows for the exploration of different levels of white noise to simulate different scenarios. Overall, the SLBQR stochastic model provides a valuable contribution to the field of malware analysis and can be used to inform decision-making in the development of effective cybersecurity measures.

4 Conclusion

The objective of this paper was to introduce a new approach to modeling malware spread by incorporating the stochasticity of malware attacks into pre-established epidemiology malware models. We have shown that current models do not capture the inherent complexity of malware spread due to the randomness and unpredictability of malware attacks. To address this issue, we have proposed an approach that provides a more time-dependent, dynamic, and stochastic representation of malware distribution in a network, which allows for a more accurate simulation of malware spread with external factors.

Through this approach, our results show that our models, which include the SEIRS, MalSEIRS, SIQVD, and SLBQR models, are more effective in predicting and mitigating malware spread than previous models that do not incorporate stochasticity. The stochasticity of malware attacks, such as human behavior, software vulnerabilities, and network topology, are taken into account, which provides a more accurate representation of malware spread in a network.

Overall, our research highlights the urgent need for effective and efficient methods of predicting and mitigating malware attacks, which pose a significant threat to individuals, businesses, and governments worldwide. Our proposed approach provides a promising avenue for further research, and it has the potential to improve current malware mitigation strategies.

References

[1] Maxat Akbanov, Vassilios G. Vassilakis, and Michael D. Logothetis. "WannaCry Ransomware: Analysis of Infection, Persistence, Recovery Prevention and Propagation Mechanisms". In: *Journal of Theoretical and Applied Information Technology* 97 (5 2019), pp. 1–3. URL: https://doi.org/10.26636/jtit.2019.130218.

[2] Chandini S B, Rajendra A B, and Nitin Srivatsa G. "A Research on Different Types of Malware and Detection Techniques". In: *International Journal of Recent Technology and Engineering* 8 (2S8 2019), pp. 1–2. URL: https://www.ijrte.org/wp-content/uploads/papers/v8i2S8/B11550882S819.pdf.

[3] Ottar N. Bjørnstad et al. "The SEIRS model for infectious disease dynamics". In: *Nature Methods* 17 (2020), pp. 557–558. DOI: 10.1038/s41592-020-0856-2. URL: https://doi.org/10.1038/s41592-020-0856-2.

[4] Sharifah Fayi. "What Petya/NotPetya Ransomware Is and What Its Remediations Are". In: *Advances in Cybersecurity: Proceedings of the 2017 Central European Cybersecurity Conference (CECC 2017)*. Springer, 2018. DOI: 10.1007/978-3-319-77028-4_15. URL: https://doi.org/10.1007/978-3-319-77028-4%5C_15.

[5] R. Mahalakshmi, S. Jeevanandham, and S. V. Gokulnath. "Dynamic Network Topology Through Malware Attack Detection on Software Defined Network". In: *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. IEEE, 2018, pp. 1871–1875. DOI: 10.1109/ICACCI.2018.8554646. URL: https://doi.org/10.1109/ICACCI.2018.8554646.

[6] Isabella Martı´nez Martı´nez et al. "MalSEIRS: Forecasting Malware Spread Based on Compartmental Models in Epidemiology". In: *Journal of Cybersecurity and Privacy* 2021 (2021), p. 5415724. URL: https://doi.org/10.1155/2021/5415724.

[7] Curtis Storlie et al. "Stochastic Identification of Malware with Dynamic Traces". In: *Annals of Applied Statistics* 8.4 (2014), pp. 2174–2199. DOI: 10.1214/13-AOAS703. URL: https://doi.org/10.1214/13-AOAS703.

[8] Yu Yao et al. "An Epidemic Model of Computer Worms with Time Delay and Variable Infection Rate". In: *Mathematical Problems in Engineering* 2018 (2018), pp. 1–11. DOI: 10.1155/2018/9756982. URL: https://doi.org/10.1155/2018/9756982.

[9] Xulong Zhang and Yong Li. "Modelling and Analysis of Propagation Behavior of Computer Viruses with Nonlinear Countermeasure Probability and Infected Removable Storage Media". In: *Mathematical Problems in Engineering* 2020 (2020), pp. 1–12. DOI: 10.1155/2020/8814319. URL: https://doi.org/10.1155/2020/8814319.

[10] Yonghua Zheng, Jianhua Zhu, and Chaoan Lai. "A SEIQR Model considering the Effects of Different Quarantined Rates on Worm Propagation in Mobile Internet". In: *Mathematical Problems in Engineering* 2020 (2020), pp. 1–13. DOI: 10.1155/2020/8161595. URL: https://doi.org/10.1155/2020/8161595.