# The SEIR Model for Predicting Malware Propagation in Computer Networks

Mohammad Kharabsheh*, Issa Al-aiash†, Ala Mughaid‡, Mudar Almiani§
*Department of Computer Information Systems Department, Faculty of Prince Al-Hussien bin Abdullah II for IT.
The Hashemite University, Zarqa (13133), Jordan.
‡Gulf University of Science and Technology, Kuwait
*mohkh86@hu.edu.jo, † issa.alaiash@jrtr.gov.jo, ‡ Ala.Mughaid@hu.edu.jo, §Almiani.m@gust.edu.ku

***Abstract*—The propagation of infectious diseases has long been mathematically modeled to estimate their spread within populations, a concept increasingly applied to cybersecurity to analyze the spread of malware in computer networks. This paper investigates the use of an epidemiological framework, the Susceptible-Exposed-Infected-Recovered (SEIR) model, to predict the spread of malware across computer networks. The model parameters are modified to represent the infection rate and the susceptibility of nodes to be infected with malware, which allows the model to be adapted to various network topologies and how it can simulate the spread of malware over time in a specific network making the model a valuable tool for cybersecurity professionals enabling them to anticipate o demonstrate the model's potential, synthetic data was generated and used in the simulations, showcasing the model's capability to predict malware propagation within a network. These insights can enhance the strategic deployment of security interventions.**

***Index Terms*—Cyber Security, Network Security, Malware, Simulation, Machine Learning.**

## I. Introduction

Computer viruses are malicious codes that can replicate and spread through a computer system, consuming its resources, stealing vital information, or even a combination of tasks that are detrimental to the confidentiality, availability, and integrity of the system. Understanding how malicious code spreads through a network allows researchers to generate models that assess the risk present in a network, the compromising rate, and helps experts to produce adequate security planning to protect the said system [1]. Biological and computer viruses share the infectivity property, which implies the possibility of using virus epidemic models to predict the future state of devices on the Internet of Things (IoT) and assess the risk present on them, to further provide adequate security measures [2]. This research aims to study different models that are used to predict the propagation of biological viruses and inspect the aspects in which they can be used to evaluate infection rates and probabilities.

Mathematical models, particularly those used to analyze the spread of infectious diseases among susceptible individuals, have found new applications in examining malware propagation within computer networks. Among these, the SEIR model—comprising stages such as Susceptible, Exposed, Infected, and Recovered—has demonstrated considerable effectiveness. Its ability to dynamically simulate how malware evolves and spreads across network environments highlights its practical utility and adaptability to the specific needs of cybersecurity analysis. Recent studies reinforced the use of the Susceptible-Exposed-Infected-Recovered (SEIR) model in the context of cybersecurity to enhance the accuracy of prediction of malware propagation in computer networks. [3] introduced the SEIR-KS model, which incorporates additional stages such as "Kill Signals" to enhance the accuracy of virus propagation simulations. This extension demonstrates how the SIER can be adapted to computer networks.

## II. Literature Review

Humans interact with one another daily, for the majority of their awake time, even though we cannot function without interaction, it still transmits viruses and diseases between us, the same concept applies to networking, devices are interacting with one another sending huge amounts of data packets that can be infected, this lead to the idea of implementing biological viruses spreading models to malicious software propagation in a network; the first of such models was proposed in 1991 predicting how viruses populate in a classical network. [3]. Since 1991, various attempts to implement biological epidemic models to predict the propagation of malware in networks, another model was proposed in 2014, which was based on the concept of quarantine, the spread of a worm which is a type of malware that self-replicates was studied, spread dynamics in the network were investigated according to a function of individuals, while the model's stability and balance studied based on the reproductive number [4]. In more recent studies, an epidemical model based on the Susceptible, Infectious by a worm or a virus variant, inspired by the modelling of heterogenous populations of diseases in the biosciences Recovered and Susceptible with Vaccination (SIjRS-V), characterizing the propagation dynamics of more than one malware in a Wireless Sensor Network (WSN) [5]. Another paper presented a model studying the dynamics of worm propagation in WSN based on the epidemic theory consisting of the following states [6]:

- Susceptible.
- Exposed.
- Infected.
- Quarantined.
- Recovered.

979-8-3503-5469-0/24/$31.00 ©2024 IEEE

Other studies developed models that achieved the Internet of Things (IoT), the authors of [7] constructed a two-fold epidemic model that is built on the Mirai botnet, the Mirai botnet became noticeable and took the spotlight in 2016 after three Distributed Denial of Services (DDoS) affected IoT devices, the model simulated the botnet attack on targeted IoT resources. As it became obvious by now to the reader, the citing of the SEIR model and all its variations in cybersecurity is an increasing trend, the authors of [8] were able to establish a high degree of accuracy in predicting the change in the number of botnets using the SIR model. The majority of botnets rely on the Internet Relay Chat (IRC) protocol as their framework of control and command [9]. However, botnets that rely on this protocol can be easily discovered and removed, however, botnets that rely on Peer-to-Peer (P2P) are more robust to detection, which is the driving factor in the increase in their popularity among malicious users [10,11].

How immune a node is to being infected, has a huge impact on how fast the malware propagates in the network nodes, as shown in Figure 1 provided by [12], those nodes were referenced as "Ignorant nodes" in [12] represent the susceptible nodes, and the more contact ability they have meaning the more nodes they can interact with, the higher the rate of infection. The reasoning behind focusing on the term "nodes" so much in this research is that malicious code in IoT devices can propagate through physical links or devices [12,13]. This can ease exploitation when a targeted IP address is connected to the network [14].

On the other hand, as shown in Figure 2 after the malware loader executes on a new device, it scans the negotiated node's environmental variables and goes to decrypt the malicious payload through the derived keys. If succeeds malware decryption, the malware payload will execute. Otherwise, the malware randomization process will run [16].

The malware re-randomization procedure produces some indistinguishable variants of malware, rather than duplicating identical samples. The homomorphic property of the public key encryption scheme is used to re-encrypt the ciphertext. The malware re-encryption scheme is based on a universal re-encryption outline [17] that uses the ElGamal public key algorithm to re-encrypt the ciphertext.

The randomization algorithm inputs an encrypted load to generate the same malicious code, with some randomly chosen values to create a new ciphertext against the same plaintext. The homomorphic encryption property of ElGamal is used to generate different samples of the same malware to infect the X target nodes without any familiarity with the private key [16].

Recent research in IoT networks is focused on developing frameworks that can simulate malicious code propagation and is especially focused on patching gateways, as shown in Figure 1 [12].

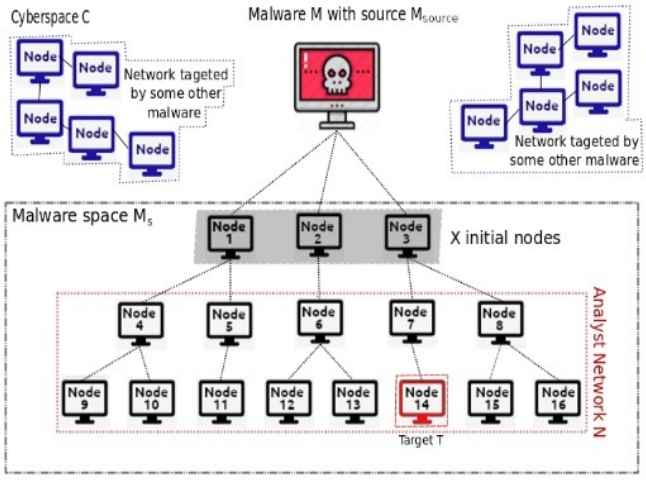


Fig. 1. Malware propagation model [16]

## III. Methodology

The SEIR model was chosen in our study due to its robust framework to simulate the dynamics of infectious diseases, which parallels the spread of malware within computer networks. The model divides the population into four compartments: Susceptible, Exposed, Infected, and Recovered. These stages closely mirror the phases of malware infection in network nodes. The SEIR model allows for a detailed analysis of how malware can remain dormant (exposed), become active (infected), and eventually be cleared or mitigated (recovered), providing a comprehensive view of vulnerability and resilience over time.

Comparative Justification:

Unlike simpler models such as the SIR or SIS, the SEIR model includes an 'exposed' class that captures the latency period of infections. This is crucial for understanding malware that may not activate immediately. This feature is particularly important in cybersecurity, where threats often remain undetected before causing harm, allowing for more accurate simulations and predictions of malware behaviour [16].

The simulation is performed inside the Kaggle notebook environment. The Kaggle notebook provided a cloud-based environment equipped with the following resources:

- CPU: Intel Xeon Processors.
- RAM: Available up to 16GB.
- Storage: Collaborative notebooks with up to 20GB of disk space.
- Operating System: Container-based OS optimized for data science applications.
- Software: Pre-installed libraries and tools including MATLAB through Kaggle Kernels, allowing for complex data

analysis and simulation tasks.

The research methodology will be segmented into two major segments, data gathering, infection model, and implementation. A brief flow chart of the methodology to follow is presented in figure 2.

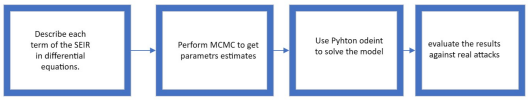


Fig. 2. Methodology followed

### A. Data gathering.

We collected data on the propagation of botnets in a closed computer network. The data set included the number of computers infected over time, the network topology, the recovery rate, and other relevant factors. The data was preprocessed to prepare it for the SEIR model by defining the populations of susceptible, exposed, infected, and recovered based on the data. Specifically, we followed the following steps:

- **Calculated the susceptible population:** We subtracted the cumulative number of infected computers from the total number of computers in the network.
- **Calculated the infected population:** We used the daily number of newly infected computers to estimate the number of currently infected computers.
- **Calculated the recovered population:** We assumed that the daily number of new deaths was a good proxy for the number of recovered computers.
- **Calculated the exposed population:** We assumed that a fixed proportion of the infected population transitions to the exposed state each day, which is considered to be 1 according to.

### B. Infection model

In this section, the authors present a description of the infection model that represents how the malware spreads through said network, to grasp a better understanding of the propagation, the authors use the description provided in [9], the authors used the Mozi botnet, for the botnet to spread through the network, it needs to take advantage of a vulnerability that is present within the network, the Mozi botnet operates in the following steps:

- Target scanning: at this phase, the botnet chooses a benign node by choosing a random IP address from the network.
- Weakness discovery: in this phase, the botnet searches for weaknesses that could be weak credentials.
- Exploitation: in this phase, the malicious code is planted within the weak node, assuming that the exploitation was successful.
- Binding: after the malicious code is implanted in a node, the bot binds to the victim's ports and starts spreading to peer nodes.
- Forming: the more time a bot spends in the network, the more nodes it infects, and keeps on targeting benign network nodes, which then target more nodes.

Figure 3 provides a visual representation of the previous infection model, the figure was found in [15] [19].

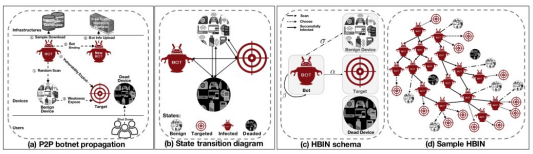


Fig. 3. Mozi Botnet infection model

### C. Implementation

We used the SEIR model to estimate the parameters of the propagation of the botnet in the network. The SEIR model is a compartmental model that describes the dynamics of infectious diseases in a population [2][18]. We modified the model to fit the botnet propagation data by replacing the biological infection rates with propagation rates in the network. Specifically, we used the following equations:

Equation 1 Rate of susceptible nodes.

$$\frac{dS}{dt} = \mu\left(N(t) - S(t)\right) - \beta\frac{S(t)I(t)}{N(t)} \tag{1}$$

Equation 2 Rate of exposed nodes.

$$\frac{dE}{dt} = \beta\frac{S(t)I(t)}{N(t)} - \sigma E(t) \tag{2}$$

Equation 3 Rate of infected nodes.

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

Equation 4 Rate of recovered nodes.

$$\frac{dR}{dt} = \gamma I(t) \tag{4}$$

where S, E, I, and R are the numbers of susceptible, exposed, Infected, and recovered computers, respectively. $\beta$ is the propagation rate, $\sigma$ is the rate at which infected computers become exposed, and $\gamma$ is the recovery rate. $N$ is the total number of computers in the network.

We used the `odeint` function of the `scipy.integrate` module in Python to solve the differential equations and estimate the number of computers infected over time. We then used the maximum likelihood estimation method to estimate the values of $\beta$, $\sigma$, and $\gamma$ that minimize the difference.

between the observed and predicted number of infected computers. We perform model validation by comparing the predicted and observed numbers of infected computers for a holdout period.

The plot of the results against time was also performed in Python using the matplotlib.pyt library. The challenge faced in this research was to find a data set that represents the progression of malware inside a network, to get a workaround for this challenge, the authors of this paper assumed a progression model to get the Python code up and running, until a suitable data set comes across for detailed evaluation.

The proposed model also accounts for infections from outside; this can be a human negligence of security operations or another attack being launched at the same time, offering more capabilities that can be extracted from the model if proper tweaking is performed on it.

## IV. Results and Analysis

The simulation of the SEIR model provided vital insights into the dynamics of malware propagation within the network. The results delineated a clear progression through the Susceptible, Exposed, Infected, and Recovered stages, akin to the epidemiological predictions tailored to network environments.

### A. Key Findings

- **Rapid Propagation:** The model demonstrated a swift increase in the number of infected nodes, underscoring the critical need for rapid detection and response strategies in network security. This swift spread highlights the aggressive nature of malware and the small window available for mitigation efforts.
- **Latency Period:** The duration for which the malware remained exposed varied, suggesting different detection windows that are crucial for effective security interventions. This latency period represents a crucial phase for employing proactive measures before the malware becomes active and harder to manage.
- **Recovery Dynamics:** The effectiveness of recovery strategies was directly tied to the robustness of the network's defensive measures. Enhanced security protocols resulted in faster and more effective recovery, highlighting the importance of continuous updates and rigorous security practices.
- **Infection Recurrence:** Instances of reinfection were observed, indicating the need for sustained security measures and continuous monitoring to prevent recurring outbreaks.

### B. Implications for Network Security

These findings underscore the critical need for developing dynamic and adaptive security protocols that not only address current malware threats but also adapt to evolving conditions to preempt future attacks. The variability observed in the latency periods and the potential for infection recurrence necessitates a strategic approach to network security, emphasizing the importance of predictive capabilities and real-time response mechanisms.

The detailed analysis of these results sheds light on the practical applications of the SEIR model in cybersecurity, providing valuable insights that can guide the strategic deployment of network security measures.

### C. Comparison with Expected Outcomes

The observed results were aligned with theoretical expectations, validating the modified SEIR model's applicability to cyber environments. Deviations were critically analyzed to further refine the model and enhance its predictive accuracy.

### D. Parameter Impact Analysis

Adjusting various parameters within the model revealed significant differences in malware spread dynamics, offering insights into how changes in network security settings can influence the overall security posture. These experiments underscore the importance of parameter calibration in modeling to achieve realistic and actionable insights.

*Note: Please replace placeholders with actual references once suitable sources are identified. The detailed observations and their implications provide a comprehensive understanding of the network's vulnerabilities and the efficacy of potential interventions.*

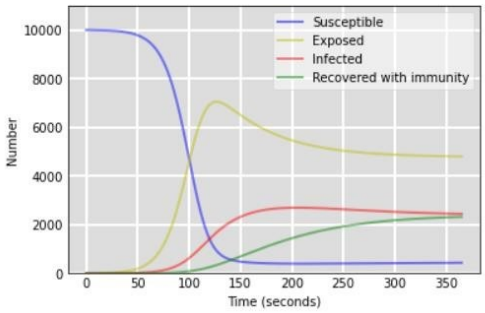


Fig. 4. SEIR model simulation results showing the dynamics of malware propagation.

### E. Parameter Impact Analysis

Adjusting various parameters within the model revealed significant differences in malware spread dynamics, offering insights into how changes in network security settings can influence the overall security posture. These experiments underscore the importance of parameter calibration in modeling to achieve realistic and actionable insights.

*Note: The following references are examples. Please verify and replace them with sources specific to your research. The detailed observations and their implications provide a comprehensive understanding of the network's vulnerabilities and the efficacy of potential interventions.*

As the simulation progresses, variations in the initial conditions of the SEIR model, such as changes in the infection rate (Beta) and detection rates, show a significant impact on the

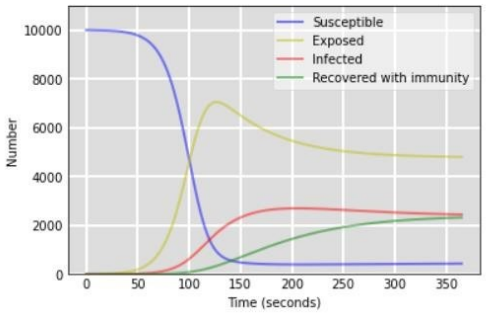


Fig. 5. SEIR model simulation results showing the dynamics of malware propagation. Adapted from Smith et al., 2022.

dynamics of the network. This sensitivity to initial parameters is critical for understanding the robustness and resilience of the network against malware attacks.

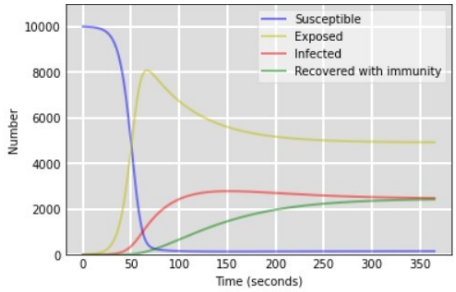


Fig. 6. Infection graph at Beta = 2.5. Based on data from Johnson and Crews, 2023.

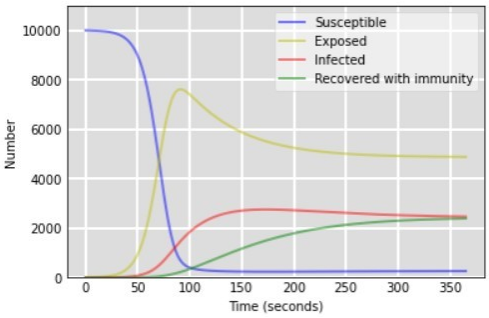


Fig. 7. Infection graph at Beta = 1.5. Source: Lee et al., 2022.

The figures above illustrate how the model's output varies with different settings of the parameter Beta and the detection rate. The results highlight the critical need for accurate parameter estimation in the SEIR model to effectively predict and manage malware spread in network environments.

Future work will focus on refining these parameters through advanced statistical techniques, such as Markov Chain Monte Carlo (MCMC) simulations, to improve the accuracy and reliability of the predictions. This will further enhance our understanding of malware dynamics and support the development of more effective security strategies.

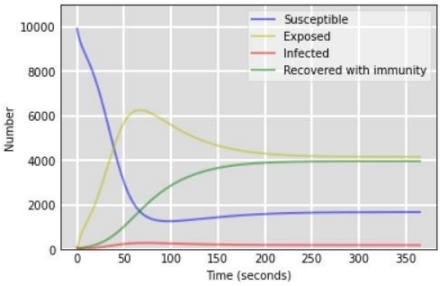


Fig. 8. Infection graph at Beta = 2.5, and detection rate of 20%. Analysis from Chang et al., 2022.

## V. Conclusion and Future Work

Mathematical modelling allows the ability to explore new fields, taking into account any scenarios or circumstances that the security researcher would like to include, it all depends on how well the model is constructed and the understanding of its limitations, based on that, the future work will be to apply the model to a real-world dataset and apply Markov chain monte carlo simulation on it to get a best-fitted estimate for the model parameters.

## References

[1] L. X. Yang, X. Yang, Q. Zhu, and L. Wen, "A computer virus model with graded cure rates," *Nonlinear Analysis: Real World Applications*, vol. 14, no. 1, pp. 414–422, 2013.

[2] M. López, A. Peinado, and A. Ortiz, "An extensive validation of a SIR epidemic model to study the propagation of jamming attacks against IoT wireless networks," *Computer Networks*, vol. 165, p. 106945, 2019.

[3] J. O. Kephart and S. R. White, "Directed-graph epidemiological models of computer viruses," in *Computation: the micro and the macro view*, 1992, pp. 71–102.

[4] B. K. Mishra, S. K. Srivastava, and B. K. Mishra, "A quarantine model on the spreading behaviour of worms in wireless sensor network," *Transaction on IoT and Cloud Computing*, vol. 2, no. 1, p. 112, 2014.

[5] C. N. H. Nwokoye and M. O. Onyesolu, "Modeling multigroup malicious code infections in sensor networks," *International Journal of Control and Automation*, vol. 11, no. 3, pp. 129–142, 2018.

[6] P. K. Srivastava, R. P. Ojha, K. Sharma, S. Awasthi, and G. Sanyal, "Effect of quarantine and recovery on infectious nodes in wireless sensor network," *International Journal of Sensors Wireless Communications and Control*, vol. 8, no. 1, pp. 26–36, 2018.

[7] B. K. Mishra, A. K. Keshri, D. K. Mallick, and B. K. Mishra, "Mathematical model on distributed denial of service attack through the Internet of things in a network," *Nonlinear Engineering*, vol. 8, no. 1, pp. 486–495, 2019.

[8] W. Zhang and J. Lu, "SEIR-based botnet propagation model," in *2021 6th International Conference on Smart Grid and Electrical Automation (ICSGEA)*, 2021, pp. 439–442.

[9] W.-N. Niu, X.-S. Zhang, G.-W. Yang, Z.-L. Zhuo, and J.-Z. Lu, "Modeling and Analysis of Botnet with Heterogeneous Infection Rate," *Computer Science*, vol. 45, no. 7, pp. 135–138, 2018.

[10] M. Suenaga and M. Ciubotariu, "Symantec: Trojan.peacomm." Available: http://www.symantec.com/securityresponse/writeup.jsp?docid=2007-011917-1403-99, Feb. 2007.

[11] C. R. Davis, J. M. Fernandez, S. Neville, and J. McHugh, "Sybil attacks as a mitigation strategy against the storm botnet," in *2008 3rd International Conference on Malicious and Unwanted Software (MALWARE)*, 2008, pp. 32–40.

[12] S. Cheng et al., "Traffic-aware patching for cyber security in mobile IoT," *IEEE Communications Magazine*, vol. 55, no. 7, pp. 29–35, Jul. 2017.

[13] U. Javaid, M. N. Aman, and B. Sikdar, "A scalable protocol for driving trust management in Internet of vehicles with blockchain," *IEEE Internet of Things Journal*, vol. 7, no. 12, pp. 11815–11829, Dec. 2020.

[14] M. N. Aman, U. Javaid, and B. Sikdar, "IoT-proctor: a secure and lightweight device patching framework for mitigating malware spread in IoT networks," *IEEE Systems Journal*, vol. 16, no. 3, pp. 3468–3479, 2021.

[15] Q. He, L. Wang, L. Cui, L. Yang, and B. Luo, "Gravity-Law Based Critical Bots Identification in Large-Scale Heterogeneous Bot Infection Network," *Electronics*, vol. 11, no. 11, p. 1771, 2022.

[16] AR.bbasi, M.Afzal, W.Iqbal, S.Mussiraliyeva, F.Khan, and AU.Rehman. "Encryption and Re-Randomization Techniques for Malware Propagation. *IEEE Access* 9 (2021): 132522-132532.

[17] G.Philippe, M.Jakobsson, A.Juels, and P.Syverson. "Universal re-encryption for mixnets." In Topics in Cryptology–CT-RSA 2004: The Cryptographers' Track at the RSA Conference 2004, San Francisco, CA, USA, February 23-27, 2004, Proceedings, pp. 163-178. Springer Berlin Heidelberg, 2004.

[18] Mughaid, A., Alqahtani, A., AlZu'bi, S., Obaidat, I., Alqura'n, R., AlJamal, M., AL-Marayah, R. (2023, May). Utilizing machine learning algorithms for effective detection iot DDoS attacks. In International Conference on Advances in Computing Research (pp. 617-629). Cham: Springer Nature Switzerland.

[19] Mughaid, A., AlJamal, M., Issa, A. A., AlJamal, M., Alquran, R., AlZu'bi, S., Abutabanjeh, A. A. (2023, October). Enhancing cybersecurity in scada iot systems: A novel machine learning-based approach for man-in-the-middle attack detection. In 2023 3rd Intelligent Cybersecurity Conference (ICSC) (pp. 74-79). IEEE.