Optimal risk mitigation strategies for cyber contagion in networks: A hybrid deep learning method

Yu Zhang[1], Zhuo Jin[2], Jiaqin Wei[3], and George Yin[4], *Life Fellow, IEEE*

***Abstract*—This paper presents a novel class of cyber security models based on SIR-type formulation. Our effort is on investigating optimal impulse controls arising from a cluster owner under exogenous cyber-attacks. We utilize the SIRS model from epidemiology to represent the spread of cyber-attacks within the cluster and evaluate the impact of protective measures. Within this framework, we determine the optimal defense strategy against effective hacking by formulating and solving a stochastic control problem with optimal switching. By employing dynamic programming principles, we derive a system of quasi-variational inequalities. Due to the inherent nonlinearity and complexity, a closed-form solution is not possible. We use a hybrid deep learning method to approximate the solution by simulating the optimal protection strategies. Finally, the effectiveness of the proposed hybrid deep learning method is validated by comparing it with the deep Galerkin method.**

***Index Terms*—SIR-type model, impulse control, numerical method, hybrid method, deep learning, stochastic approximation.**

I. Introduction

CYBER attacks have grown increasingly complex and widespread in recent years. As a result, we face unprecedented cybersecurity threats and challenges, including denial-of-service (DoS) attacks [20], malware [49], ransomware [20], blackmail [44], extortion [53], and more [11], [28]. Cybersecurity Ventures, a cybersecurity risk investment firm, estimated that the annual cost of cybercrime will rise to 10.5 trillion USD by 2025, compared to an estimate of 3 trillion USD by the World Economic Forum in 2015 [39].

The evolving nature of cyber risk, its potential to become systemic, and its behavioral characteristics make the quantification of cyber risk particularly challenging. Recently, we have seen significant progress in this area, particularly in insurance coverage. A foundational contribution was made by [21], who proposes a mathematical model to measure the loss reduction resulting from technical security investments and determine the optimal level of investment. Notable contributions related to insurance coverage include studies such as [4], [16], and [17]. Specifically, in [18], the authors investigate severe and extreme cyber claims using a combination of generalized Pareto modeling and regression tree methods. As emphasized by [54], the accumulation and propagation characteristics of cyber events can be formulated using network models from epidemiology. Subsequent studies adapted this idea to account for specific features of cyber risk, as detailed in [25], [26] and [40]. The article [37] examined optimal investment decisions in the context of mixed insurance and investment strategies for managing cyber risk. The response of defense systems to cyber-attacks was analyzed in [33], where it was modeled as a stochastic game involving a large number of interacting agents. A cluster-based method is developed to investigate the risk of cyber attacks in the continental United States in [38]. In [24], the authors reviewed cyber risk research across various disciplines, with a primary goal to aid researchers in the field of insurance and actuarial science to identify potential research gaps as well as to leverage useful models and techniques that have been considered in the literature.

Because of the page limitation, we will not be able to present all detailed mathematical developments such as convergence proofs and asymptotic studies. Rather, it seems to be most instructive for us to concentrate on the mathematical model descriptions and the algorithms we developed. The mathematical details will appear in a subsequent paper. In addition, we put emphasis on the examples. Our aim is to use the examples to illustrate the main ideas.

Our formulation in this paper stems from the well-known SIR models. Such epidemic models were first introduced by Kermack and McKendrick in [30], [31]. In recent years, the study on mathematical models has flourished. Much attention has been devoted to analyzing, predicting the spread, and designing controls of infectious diseases in host populations; see [3], [7], [8], [10], [19], [32], [34], [30], [31], [45], [50] and the references therein. The SIR (Susceptible-Infected-Removed) model is suitable for modeling some diseases with permanent immunity such as rubella, whooping cough, measles, smallpox, etc. For some of the most recent mathematical developments on SIR models, we refer the reader to some of our work [14], [15], [41] and references therein. The formulation begins with the so-called compartment models. In a SIR model, a homogeneous host population is subdivided into three epidemiologically distinct types of individuals, namely, susceptible class, the infective class, and the removed class. Then the spread of infection can be formulated by using a system of differential equations. Recognizing that random effect is not avoidable, it more realistic to assume that a population is subject to random disturbances. Thus renewed effort has been devoted to finding the corresponding

*The research of Jiaqin Wei was supported by the National Natural Science Foundation of China 12071146. The research of George Yin was supported in part by the National Science Foundation under grant DMS-2404508.

[1]Yu Zhang is with School of Mathematics and Statistics, Anhui Normal University, Wuhu, Anhui 241002, China, yzstmath@ahnu.edu.cn

[2]Zhuo Jin is with the Department of Actuarial Studies and Business Analytics, Macquarie University, 2109, NSW, Australia, zhuo.jin@mq.edu.au

[3]Jiaqin Wei is with the Key Laboratory of Advanced Theory and Application in Statistics and Data Science-MOE, School of Statistics, East China Normal University, Shanghai 200062, China, jqwei@stat.ecnu.edu.cn

[4]George Yin is with the Department of Mathematics, University of Connecticut, Storrs, CT 06269-1009, USA, gyin@uconn.edu

classification by means of stochastic models. Because of the importance, substantial effort has been devoted to the SIR models and their various variants.

In this paper, our focus is on cyber security issues. We present a novel class of models based on the ideas from SIR formulations. We illustrate that SIR type of models can also be used in the study of cyber security related issues. We aim to address the challenge faced by cluster owners in balancing the costs of protecting their computer networks against cyber-attacks, with a particular focus on whether to regularly update or purchase security software. This involves a trade-off, where inadequate protection may lead to substantial financial losses due to cyber incidents, affecting both the cluster owner and its customers. Conversely, implementing active protection measures can be very costly. In addition, this paper models cyber risk using a stochastic epidemiological SIRS model, where the system switches between different dynamics based on two factors: the control exerted by the cluster owner (endogenous switching control) and hacking activities (exogenous and uncertain risks). In this framework, we consider an optimal impulse control problem for cyber risk management. Through dynamic programming principles, we derive a system of quasi-variational inequalities. Due to the inherent nonlinearity and complexity, no closed-form solution appears to be possible. As a viable alternative, we use a hybrid deep learning method to approximate the optimal strategy of a cluster owner. The effectiveness of the proposed method is demonstrated through a comparative study with the deep Galerkin method [46] in two specific scenarios: one with a constant attack and the other with Poisson attacks.

Recently, machine learning methods have been developed to handle cyber risk management; see, for example, [6], [22], [23], [42], [47], [48], and [51]. As seen in the recent years, deep learning and reinforcement learning are becoming increasingly popular in the field of risk management. In [52], the authors investigated network attacks related to intrusion detection, highlighting the limitations of existing datasets and suggesting future research directions for model development. Machine learning and deep learning methods for securing Internet of Things (IoT) technology were reviewed in [2]. In [5], the authors provided a comprehensive overview of various network attack types and offered an in-depth discussion on attack detection methods using deep learning techniques, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), and generative adversarial networks (GANs). Relevant literature on the application of reinforcement learning in cyber security includes [1], [12], and [43], among others. We remark that the deep-learning methods are deeply rooted to stochastic gradient methods and more generally to stochastic approximation; see Kushner and Yin [36] for a comprehensive treatment.

The main contributions of this paper are as follows. Unlike the controlled stochastic Kolmogorov systems considered in [55], we focus on an optimal impulse control problem encountered by a cluster owner under exogenous cyber-attacks. Although the epidemiological SIRS model used in our work represents the spread of cyber-attacks within the cluster bears that resemblance to the one in [27], our focus is on employing a hybrid deep learning Markov chain approximation method (see, for example, [13], [29], and [56]), whereas their approach utilizes the deep Galerkin method. The hybrid feature of the proposed method lies in an integration of Markov chain approximation and stochastic approximation algorithm. The Markov chain approximation method plays a key role in building iterative algorithms and finding initial values. Stochastic approximation is employed to search for the optimal neural network parameters within a bounded region defined by the Markov chain approximation method. Comparing with the existing numerical methods on stochastic control problems, our proposed deep learning algorithm has two main advantages. (1) The use of deep learning enables us to replace the optimization over the piecewise control grid for every state value by finding optimal parameters of neural networks for all state values. In this way, the number of computation nodes increases linearly with respect to the number of points in the state lattice. Consequently, the computation efficiency can be improved; (2) when the ranges of controls and states are not comparable, computational efficiency and accuracy are significantly impacted, as selecting an appropriate step size for the lattice becomes challenging. In contrast, neural networks enable the control strategy to take values within a continuous range, overcoming the difficulty of choosing an appropriate precision in control spaces with significant different scales. As a result, the accuracy of the numerical results can be improved.

The remainder of this paper is organized as follows. Section II presents the epidemiological SIRS dynamics used to model the contagion of cyber-attacks through the cluster. Section III briefly introduces a framework of the hybrid deep learning Markov chain approximation method. Section IV gives two numerical examples to illustrate the effectiveness of the proposed hybrid deep learning Markov chain approximation method. Finally, Section V concludes the paper with further remarks.

II. Model

Let $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathrm{P})$ be a complete filtered probability space, where $\{\mathcal{F}_t\}$ is a filtration satisfying the usual condition (i.e., right continuous, increasing, and $\mathcal{F}_0$ containing all the null sets). Denote by W a one-dimensional Brownian motion, which is viewed as an uncertainty to determine precisely the transmission rate of the virus inside the computer's cluster.

Following [27], we assume that the dynamics of the SIRS system evolves as

$$\begin{cases} dS_t = \left(\rho R_t - S_t\left(a_t\nu + I_t\beta + p_t\kappa\right)\right)dt - \sigma I_t S_t dW_t, \\ dI_t = a_t\nu S_t dt + \beta S_t I_t dt - I_t\gamma dt + \sigma I_t S_t dW_t, \\ dR_t = p_t\kappa S_t dt + \gamma I_t dt - \rho R_t dt, \end{cases} \tag{1}$$

where $(a_t)_{t\geq 0}$ is the hacker's strategy, which is a binary variable taking the value $a_t = 1$ if the hacker attacks the cluster or $a_t = 0$ if the hacking is inactive; $(p_t)_{t\geq 0}$ is the response of the cluster owner's to protect its network, which is also a binary control variable, taking the value $p_t = 1$ if

he/she develops a dedicated protection to this attack or $p_t = 0$ he/she remains with the benchmark level of protection. The $\nu > 0$ and $\kappa > 0$ are the intensity of attack and defense implementation, respectively.

The hacker's strategy $(a_t)_{t\geq 0}$ is defined by a binary variable $\tilde{\alpha} := \Big(a_0, (\tilde{\tau}_n)_{n\geq 0}\Big)$, where $a_0 \in \{0,1\}$ is the initial state and $(\tilde{\tau}_n)_{n\geq 0}$ are the switching times of the attack level, with $\tilde{\tau}_0 := 0$. Then the hacker's strategy $(a_t)_{t\geq 0}$ is defined as

$$a_t := \sum_{n\geq 0} \mathbf{1}_{\tilde{\tau}_{2n+1-a_0} \leq t < \tilde{\tau}_{2n+2-a_0}}. \tag{2}$$

Here, the random times $(\tilde{\tau}_n)_{n\geq 0}$ are assumed to be exogenous random variables, independent of the filtration $\{\mathcal{F}_t\}$.

In this paper, we assume that the cluster owner can identify the current attack state a_t, but is unable to predict the hacker's strategy, i.e., the cluster owner is subjected to random switches in the environment. In each random time interval $[\tilde{\tau}_n, \tilde{\tau}_{n+1})$, characterized by a constant attack level $a_{\tilde{\tau}_n}$, the cluster owner's strategy p_t depends on this attack level $a_{\tilde{\tau}_n}$. From the perspective of switching times, the cluster owner's strategy consists of a sequence of increasing $\mathcal{F}_t$-stopping times $(\tau_n)_{n\geq 0}$, which depend on the random environment of the attack. Then the cluster owners strategy $(p_t)_{t\geq 0}$ is defined as

$$p_t := \sum_{n\geq 0} \mathbf{1}_{\tau_{2n+1-p_0} \leq t < \tau_{2n+2-p_0}}, \tag{3}$$

where $p_0 \in \{0,1\}$ is the initial state of protection and $(\tau_n)_{n\geq 0}$ are the switching-times of the protection level, with $\tau_0 := 0$.

Remark 2.1: The definition of the hackers strategy (2) and the cluster owners strategy (3) connects the attack (protection) switching pattern to the initial state, meaning that the subsequent attack (protection) periods and switching times are determined by whether the attack (protection) has been initiated. Additionally, it ensures that the attack (protection) states alternate throughout each time period. Specifically, we assume that the hacker constantly attacks the cluster, that is $\tilde{\tau}_n = \infty$ for any $n > 1$ and $a_0 = 1$. For more details of hackers strategy and cluster owners strategy, we refer the reader to [27].

To proceed, let $\mathcal{A}^p(\tilde{\alpha})$ be the set of admissible switching control of the cluster owner for a given strategy $\tilde{\alpha}$ of the hacker. For a protection strategy $\alpha \in \mathcal{A}^p(\tilde{\alpha})$, the dynamics of the system are given by

$$\begin{cases} S_t^{\alpha,\tilde{\alpha}} = s_0 + \int_0^t \rho R_s^{\alpha,\tilde{\alpha}} ds - \int_0^t S_s^{\alpha,\tilde{\alpha}} I_s^{\alpha,\tilde{\alpha}} \left(\beta ds + \sigma dW_s\right) \\ \quad - \sum_{\tau_n \leq t} \int_{\tau_n}^{\tau_{n+1}\wedge t} S_s^{\alpha,\tilde{\alpha}} \kappa p_s ds - \sum_{\tilde{\tau}_n \leq t} \int_{\tilde{\tau}_n}^{\tilde{\tau}_{n+1}\wedge t} S_t^{\alpha,\tilde{\alpha}} \nu a_s ds, \\ I_t^{\alpha,\tilde{\alpha}} = i_0 + \int_0^t I_s^{\alpha,\tilde{\alpha}} \Big((\beta S_s^{\alpha,\tilde{\alpha}} - \gamma) ds + \sigma S_s^{\alpha,\tilde{\alpha}} dW_s\Big) \\ \quad + \sum_{\tilde{\tau}_n \leq t} \int_{\tilde{\tau}_n}^{\tilde{\tau}_{n+1}\wedge t} S_t^{\alpha,\tilde{\alpha}} \nu a_s ds, \\ R_t^{\alpha,\tilde{\alpha}} = r_0 + \int_0^t \Big(I_s^{\alpha,\tilde{\alpha}} \gamma - \rho R_s^{\alpha,\tilde{\alpha}}\Big) ds \\ \quad + \sum_{\tau_n \leq t} \int_{\tau_n}^{\tau_{n+1}\wedge t} \kappa p_s S_s^{\alpha,\tilde{\alpha}} ds, \\ S_0 = s_0, I_0 = i_0, R_0 = r_0. \end{cases}$$

In this paper, an exogenous strategy $\tilde{\alpha}$ of the attacks is fixed, and we focus on deriving the optimal response strategy for the cluster owner, i.e., given initial state (s_0, i_0) and regime p_0, the cluster owner chooses an admissible switching control $\alpha = (\tau_n)_{n\geq 0} \in \mathcal{A}^p(\tilde{\alpha})$ that optimizes the following criteria for the cluster owner

$$\begin{aligned} & v^{\tilde{\alpha}}(s_0, i_0; p_0) \\ &= \inf_{\alpha \in \mathcal{A}^p(\tilde{\alpha})} \mathrm{E}\Big[\int_0^{+\infty} e^{-\delta t} \Big(c_I I_t^{\alpha,\tilde{\alpha}} + f(S_t^{\alpha,\tilde{\alpha}}, p_t)\Big) dt \\ &\quad + \sum_{n\geq 1} e^{-\delta \tau_n} g_{p_{\tau_{n-1}}, p_{\tau_n}}\Big], \end{aligned} \tag{4}$$

where $f(s,p) = c_V \kappa s p$ is the cost of the protection, c_V is the marginal cost of the protection, c_I is the marginal cost of the infected device, and $g_{0,1}, g_{1,0} > 0$ are fixed switching costs.

Now, we define the operator $\mathcal{L}^{a,p}$ as

$$\begin{aligned} \mathcal{L}^{a,p} v(s,i;a,p) &:= (\rho(1-s-i) - s(p\kappa + a\nu + \beta i))\partial_s v \\ &+ (a\nu s - \gamma i + \beta s i)\partial_i v + \frac{\sigma^2}{2} s^2 i^2 \left(\partial_{ss} v + \partial_{ii} v - 2\partial_{is} v\right). \end{aligned}$$

For any $p, a \in \{0,1\}$, we derive the following system of variational inequalities

$$\begin{aligned} & \min\left[-\delta v(s,i;a,p) + \mathcal{L}^{a,p} v(s,i;a,p) + c_I i \right. \\ & \left. + f(s,p), v(s,i;a,\bar{p}) + g_{p,\bar{p}} - v(s,i;a,p)\right] = 0, \\ & \text{on the set } \mathcal{D} := \left\{(s,i) \in [0,1]^2, s+i \leq 1\right\}, \end{aligned} \tag{5}$$

where we define $\bar{p}$ by $\bar{p} = 0$ if $p = 1$, or $\bar{p} = 1$ if $p = 0$.

In what follows, our objective is devoted to employing a hybrid deep learning method to solve the equation (5).

III. Hybrid deep learning method

In this section, we outline briefly the Markov chain approximation method proposed in [35]. Based on this method, we will introduce the hybrid deep learning method developed by [13] and [29].

A. Markov chain approximation method

In this subsection, we construct transition probabilities of the Markov chain approximation method to establish an iterative computational scheme. Let $h > 0$ be a step size, and $\{\boldsymbol{\xi}_n^h, n \in \mathbb{Z}_+\}$ be a discrete-time controlled Markov chain with state space $\boldsymbol{S}_h$, where $\boldsymbol{S}_h$ is the h-grid of $\mathbb{R}^2$ defined by $\boldsymbol{S}_h := \left\{(k^1 h, k^2 h)^\top : k^i = 0, \pm 1, \ldots, i = 1, 2\right\}$. Let $\alpha^h = (\alpha_0^h, \alpha_1^h, \ldots)$ be the sequences of random variables that are the control actions at time $0, 1, \ldots$. Denote by $\mathrm{P}^h\left((\boldsymbol{y}, \boldsymbol{z}) \,|\alpha\right)$ the probability that $\boldsymbol{\xi}$ transits

from state $\boldsymbol{y}$ to state $\boldsymbol{z}$ with $\alpha \in \mathcal{A}^p(\tilde{\alpha})$. We say that α_n^h is admissible if it satisfies the following conditions:

(a) α^h is $\sigma\left\{\boldsymbol{\xi}_0^h, \ldots, \boldsymbol{\xi}_n^h, \alpha_0^h, \ldots, \alpha_{n-1}^h\right\}$-adapted.

(b) For any $\boldsymbol{x} \in \boldsymbol{S}_h$, we have

$$\begin{aligned} \mathrm{P}\left\{\boldsymbol{\xi}_{n+1}^h = \boldsymbol{x} \mid \mathcal{F}_n^h\right\} &= \mathrm{P}\left\{\boldsymbol{\xi}_{n+1}^h = \boldsymbol{x} \mid \boldsymbol{\xi}_n^h, \alpha_n^h\right\} \\ &= \mathrm{P}^h\left((\boldsymbol{\xi}_n^h, \boldsymbol{x}) \mid \alpha_n^h\right), \end{aligned}$$

where $\mathcal{F}_n := \sigma\left\{\boldsymbol{\xi}_0^h, \ldots, \boldsymbol{\xi}_n^h, \alpha_0^h, \ldots, \alpha_n^h\right\}$.

(c) For all $n \in \mathbb{Z}_+$, $\boldsymbol{\xi}_n^h \in \boldsymbol{S}_h$.

To simplify the notation, we denote

$$\begin{aligned} &\boldsymbol{x} := (x_1, x_2)^\top := (s, i)^\top, \\ &A := \rho(1-s-i) - s(p\kappa + a\nu + \beta i), \\ &B := a\nu s - \gamma i + \beta s i, \\ &C := \frac{1}{2}\sigma^2 s^2 i^2. \end{aligned}$$

Based on the notation mentioned above, we can discrete Equation (5) using finite difference method with the stepsize h. As a result, the transition probabilities of Markov chain approximation method are constructed as follows

$$\begin{aligned} &\mathrm{P}^h\left(\boldsymbol{x}, \boldsymbol{x} \pm h\mathbf{e}_1 \mid \alpha\right) = \frac{A^{\pm}}{h}, \\ &\mathrm{P}^h\left(\boldsymbol{x}, \boldsymbol{x} \pm h\mathbf{e}_2 \mid \alpha\right) = \frac{B^{\pm}}{h}, \\ &\mathrm{P}^h\left(\boldsymbol{x}, \boldsymbol{x} + h\mathbf{e}_1 + h\mathbf{e}_2 \mid \alpha\right) = \mathrm{P}^h\left(\boldsymbol{x}, \boldsymbol{x} - h\mathbf{e}_1 - h\mathbf{e}_2 \mid \alpha\right) = \frac{C}{h^2}, \\ &\mathrm{P}^h\left(\boldsymbol{x}, \boldsymbol{x} \mid \alpha\right) = 1 - \frac{|A|}{h} - \frac{|B|}{h} - \frac{2C}{h^2}. \end{aligned} \tag{6}$$

Using the above constructed Markov chain (6), for a given control strategy α, we can define the change of objective value as

$$S(\boldsymbol{x}, v, \alpha) \approx \sum_{\boldsymbol{y} \in \boldsymbol{S}_h} v(\boldsymbol{y}) \mathrm{P}^h(\boldsymbol{x}, \boldsymbol{y} \mid \alpha) + c_I i + f(s, p).$$

The optimal control strategy and value function are given by

$$v(\boldsymbol{x}) = \inf_{\alpha \in \mathcal{A}^p(\tilde{\alpha})} S(\boldsymbol{x}, v, \alpha), \quad \alpha^* = \arg\inf_{\alpha \in \mathcal{A}^p(\tilde{\alpha})} S(\boldsymbol{x}, v, \alpha).$$

B. Deep learning Markov chain approximation method

Here, we present the hybrid deep learning Markov chain approximation method. Within the framework of deep learning method, the control variable is approximated using neural networks and evaluated in a lattice structure. Let Θ be the set of all weights and biases in neural networks, and $N(\boldsymbol{x}|\Theta)$ be the neural network control. For simplicity, we assume that the state variable's range is discretized into n points for the neural network setup, such that the state variable is approximated by $\{\boldsymbol{x}_i\}_{i=1}^n$. Given the state lattice $\{\boldsymbol{x}_i\}_{i=1}^n$, we define the global improvement function G as $G := G\left(v(\boldsymbol{x}_1), v(\boldsymbol{x}_2), \ldots, v(\boldsymbol{x}_n)\right)$. For more details of the global improvement function G, we refer the readers to [13] and [29].

To proceed, let θ_k^h and $N\left(\boldsymbol{x}, \theta_k^h\right)$ be the k-th iterative optimal parameters and control strategy, respectively. The system of dynamic programming equations in the k-th iteration follows

$$v^k\left(\boldsymbol{x}_i\right) = S\left(\boldsymbol{x}_i, v^{k-1}, \hat{\alpha}^k\left(\boldsymbol{x}_i\right)\right), 1 \le i \le n,$$

where v^{k-1} is an iterative value function obtained from the previous iteration, and $\hat{\alpha}^k(\boldsymbol{x}) = N\left(\boldsymbol{x}|\Theta^k\right)$. Here

$$\begin{aligned} \Theta^k &= \arg\min_{\Theta} G\left(S^k\left(\boldsymbol{x}_1\right), S^k\left(\boldsymbol{x}_2\right), \ldots, S^k\left(\boldsymbol{x}_n\right)\right), \\ S^k(\boldsymbol{x}) &= S\left(\boldsymbol{x}, v^{k-1}, N(\boldsymbol{x}|\Theta)\right). \end{aligned}$$

C. Algorithm summary

With the implementation details explained above, the pseudo-code of the proposed hybrid deep learning method is summarized in Algorithm 1.

Algorithm 1 Framework of the hybrid deep learning method

Input:

The state lattices for deep learning algorithm, $\{\boldsymbol{x}_i\}_{i=1}^n$;

The state lattices for Markov chain approximation method, $\{\boldsymbol{y}_j\}_{j=1}^{n'}$;

The initial values of value functions, $U^0(\boldsymbol{y}_j)$ and $v^0(\boldsymbol{x}_i)$;

The set of computation precision, ϵ;

The maximal number of learning times, $\tilde{N}$;

Output:

The approximation of optimal control $N(\boldsymbol{x}_i, \theta_k)$;

1: Obtain the optimal control $\hat{\alpha}^k(\boldsymbol{y}_j)$ by Markov chain approximation method;

2: Obtain the initial value of the parameter $\theta_{k,0}$ by

$$\theta_{k,0} = \mathrm{argmin}_\theta \sum_{j=1}^{n'} \left(\|\hat{\alpha}^k(\boldsymbol{y}_j) - N(\boldsymbol{y}_j, \theta)\|\right)^2;$$

3: Obtain the iterative control strategy by minimizing G by SGD method;

4: Iterate the value function $U^k(\boldsymbol{y}_j)$ by $U^k(\boldsymbol{y}_j) = S\left(\boldsymbol{y}_j, U^{k-1}, N(\boldsymbol{y}_j, \theta_k)\right)$;

5: Iterate the value function $v^k(\boldsymbol{x}_i)$ by $v^k(\boldsymbol{x}_i) = S\left(\boldsymbol{x}_i, v^{k-1}, N(\boldsymbol{x}_i, \theta_k)\right)$;

6: **while** $k < \tilde{N}$ **do**

7: **if** $\sum_{i=1}^n \left(v^k(\boldsymbol{x}_i) - v^{k-1}(\boldsymbol{x}_i)\right)^2 < \epsilon$ **then**

8: Stop;

9: **else**

10: Go to Step 1;

11: **end if**

12: **end while**

13: **return** $N(\boldsymbol{x}_i, \theta_k)$;

IV. Two numerical examples

In this section, we demonstrate the proposed algorithm under two attack scenarios: (1) A constant attack of the hacker; (2) Exogenous attacks switches given by a Poisson process.

A. Case 1: A constant hacker attack

In this section, we apply the deep learning Markov chain approximation method to solve a specific scenario involving a constant hacker attack with $a = 1$. Following [27], we assume that the time period is $T = 30$, the time step is $h = 0.125$, the contagion rate is $\beta = 0.04$, the recovery rate is $\gamma = 0.02$, the replacement rate is $\rho = 0.002$, the intensity of the attack is $\nu = 0.05$, the volatility of the SIRS system is $\sigma = 0.2$, the actualisation parameter is $\delta = 0.2$. We begin with only susceptible devices and no corrupted devices, with initial conditions $S_0 = 1, I_0 = 0$. In addition, we assume the efficiency of the protection is $\kappa = 0.03$, the marginal cost of protection is $c_V = 0.05$, the marginal cost of infected device is $c_I = 0.01$. The switching costs are given by $g_{01} = 0.001v(s, i, 1, 0)$ and $g_{10} = 0.001v(s, i, 1, 1)$. As for the part of deep learning method,

the number of hidden layers is 2, the number of vertices in each layer is 20, and the learning rate is 10^{-4}.

	Triggering Error	Max # of steps
Control Fit	10^{-3}	10000
Gradient Descent	10^{-5}	5000
Global Iteration	10^{-6}	50000

We apply the deep learning Markov chain approximation method and get one path of S, I without protection and with optimal protection in Figure 1. As illustrated in Figure 1, the cluster owner initially refrains from strengthening the system's protection until $t = 9.29$ (the first green vertical line). Given that the cost of infection is higher than the cost of switching the protection system, the cluster owner chooses to implement protective measures at this point to mitigate the cost of further infections. Subsequently, at $t = 22.37$ (the second green vertical line), the cost of maintaining protection becomes prohibitively high, prompting the cluster owner to reduce the protection level from $p = 1$ to $p = 0$ in order to lower protection costs. By alternating between protection states, the cluster owner effectively keeps the number of damaged devices low at $t = 30$, in contrast to the outcomes observed under the no-protection strategy. Throughout this process, the cluster owner manages the number of infected devices (represented by the yellow curve) more efficiently than in the scenario with no protection (represented by the blue curve). Consequently, the number of susceptible devices that have not been compromised by the attack (depicted by the red curve) decreases at a slower rate compared to the no-protection strategy (represented by the pink curve). This observation is consistent with the findings presented in [27].

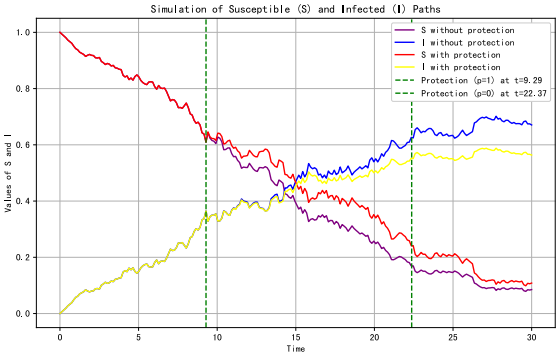

Fig. 1. Optimal trajectory of S and I with protection and switching v.s. no protection strategy. Case 1: A constant attack.

B. Case 2: Exogenous Poisson attacks

We now consider a scenario where the initiation and termination of attacks follow a Poisson process with intensity $\lambda = 0.1$. Following [27], the system is assumed to start at time 0 under an active attack ($a = 1$) and without any protective measures ($p = 0$). The protection efficiency is set to $\kappa = 0.02$, with the marginal cost of protection given by $c_V = 0.04$ and the marginal cost associated with infected devices being $c_I = 0.01$. The switching cost for activating protection (from $p = 0$ to $p = 1$) is defined as $g_{01} = 0.01v(s, i, a, 0)$, while the cost for deactivating protection (from $p = 1$ to $p = 0$) is $g_{10} = 0.001v(s, i, a, 0)$, for all $a \in \{0, 1\}$.

We employ the hybrid deep learning algorithm to simulate the trajectories of S and I under both protection and non-protection strategies, with the results illustrated in Figure 2. It is observed that the cluster owner initially permits the attack to propagate and postpones the activation of the protection mechanism until time $\hat{\tau}_1 = 7.15$ (first pink dotted vertical line). This delayed intervention is primarily due to the relatively high switching cost of initiating protection compared to the marginal cost of infection in the early stage. At time $\hat{\tau}_1 = 7.15$, the protection mechanism is activated. The attacker subsequently disengages at $\tilde{\tau}_1 = 13.20$ (first blue dotted vertical line). Despite the cessation of the attack, the cluster owner maintains the protection regime to further mitigate the residual infection risk within the network. The protective measures are eventually withdrawn at $\hat{\tau}_2 = 19.10$, shortly before the onset of the next stochastic attack at $\tilde{\tau}_2 = 22.35$ (last blue dotted vertical line). Notably, the protection mechanism is not re-engaged after the second attack occurs, suggesting that the adopted switching strategy effectively balances containment of the infection and control cost, thereby avoiding unnecessary protective actions while maintaining system resilience.

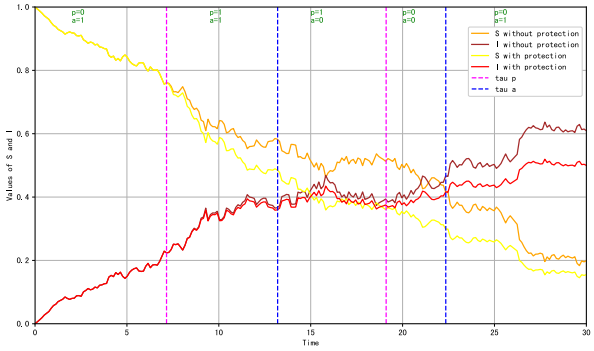

Fig. 2. Optimal trajectory of S and I with protection and switching v.s. no protection strategy. Case 2: Poisson attacks.

V. CONCLUSIONS

This paper addresses the optimal impulse control problem encountered by a cluster owner in the context of exogenous cyber-attacks, utilizing the epidemiological SIRS model to represent the propagation of attacks within the cluster and evaluate the impact of defensive measures. We formulate the problem as a stochastic control problem with optimal switching, resulting in a system of quasi-variational inequalities. Due to the inherent nonlinearity and complexity of these inequalities, we employ the hybrid deep learning method to obtain their solutions. Two specific scenarios are presented to demonstrate the effectiveness of the proposed method, which is further validated through a comparison with the deep Galerkin method.

REFERENCES

[1] A. M. K. Adawadkar and N. Kulkarni. Cyber-security and reinforcement learninga brief survey. Engineering Applications of Artificial Intelligence, 114: 105116, 2022.

[2] M. A. Al-Garadi, A. Mohamed, A. K. Al-Ali, et al. A survey of machine and deep learning methods for internet of things (IoT) security. IEEE Communications Surveys & Tutorials, 22(3): 1646–1685, 2020.

[3] M. E. Alexander, C. Bowman, S. M. Moghadas, R. Summers, A. B.Gumel, B. M. Sahai, A vaccination model for transmission dynamics of influenza, *SIAM J. Appl. Dyn. Syst.*, 3 (2004), no. 4, 503-524.

[4] K. Awiszus, T. Knispel, I. Penner, G. Svindland, A. Voss, and S. Weber. Modeling and pricing cyber insurance: Idiosyncratic, systematic, and systemic risks. European Actuarial Journal, 13(1):1–53, 2023.

[5] D. S. Berman, A. L. Buczak, J. S. Chavis, et al. A survey of deep learning methods for cyber security. Information, 10(4):122, 2019.

[6] L. Buczak and E. Guven. A Survey of data mining and machine learning methods for cyber security. IEEE Communications Surveys & Tutorials, 18(2):1153–1176, 2016.

[7] F. Ball, D. Sirl, An SIR epidemic model on a population with random network and household structure, and several types of individuals, *Adv. in Appl. Probab.*, 44 (2012), no. 1, 63-86.

[8] F. Brauer, C. C. Chavez, *Mathematical models in population biology and epidemiology*, Springer-Verlag New York, 2012.

[9] Y. Cai, Y. Kang, M. Banerjee, W. Wang, A stochastic SIRS epidemic model with infectious force under intervention strategies. *J. Differential Equations* 259 (2015), no. 12, 7463-7502.

[10] V. Capasso, *Mathematical Structures of Epidemic Systems*, Springer-Verlag, Berlin, 1993.

[11] D. Craigen, N. Diakun-Thibault, and R. Purse. Defining cybersecurity. Technology Innovation Management Review, 4(10):13–21, 2014.

[12] E. Cengiz and M. Gök. Reinforcement learning applications in cyber security: A review. Sakarya University Journal of Science, 27(2):481–503, 2023.

[13] X. Cheng, Z. Jin, and H. Yang. Optimal insurance strategies: A hybrid deep learning Markov chain approximation approach. ASTIN Bulletin: The Journal of the IAA, 50(2):449–477, 2020.

[14] N.T. Dieu, D.H. Nguyen, N.H. Du, and G. Yin, Classification of asymptotic behavior in a stochastic SIR model, *SIAM Journal on Applied Dynamic Systems*, **15** (2016), 1062–1084.

[15] N. Du, A. Hening, N. Nguyen, and G. Yin, Hybrid stochastic epidemic SIR models with hidden states, *Nonlinear Analysis Hybrid Systems*, 49 (2023), Paper No. 101368, 21 pp.

[16] M. Eling and N. Loperfido. Data breaches: Goodness of fit, pricing, and risk measurement. Insurance: Mathematics and Economics, 75:126–136, 2017.

[17] M. Eling and W. Schnell. Capital requirements for cyber risk and cyber risk insurance: An analysis of Solvency II, the US risk-based capital standards, and the Swiss Solvency Test. North American Actuarial Journal, 24(3):370–392, 2020.

[18] S. Farkas, O. Lopez, and M. Thomas. Cyber claim analysis using Generalized Pareto regression trees with applications to insurance. Insurance: Mathematics and Economics, 98:92–105, 2021.

[19] M. Gathy, C. Lefevre, Claude From damage models to SIR epidemics and cascading failures,*Adv. in Appl. Probab.,* 41 (2009), no. 1, 247-269.

[20] B. B. Gupta and O. P. Badve. Taxonomy of DoS and DDoS attacks and desirable defense mechanism in a cloud computing environment. Neural Computing and Applications, 28(12):3655–3682, 2017.

[21] L. A. Gordon and M. P. Loeb. The economics of information security investment. ACM Transactions on Information and System Security (TISSEC), 5(4): 438–457, 2002.

[22] C. Gomes, Z. Jin, and H. Yang. Insurance fraud detection with unsupervised deep learning. Journal of Risk and Insurance, 88(3):591–624, 2021.

[23] P. Garcia-Teodoro, J. Diaz-Verdejo, G. Maciá-Fernández, and E. Vázquez. Anomaly-based network intrusion detection: Techniques, systems and challenges. Computers & Security, 28(1-2):18–28, 2009.

[24] R. He, Z. Jin, and J. S. H. Li. Modeling and management of cyber risk: A cross-disciplinary review. Annals of Actuarial Science, 1–40, 2024.

[25] C. Hillairet and O. Lopez. Propagation of cyber incidents in an insurance portfolio: counting processes combined with compartmental epidemiological models. Scandinavian Actuarial Journal, 1–24, 2021.

[26] C. Hillairet, O. Lopez. L. dOultremont, and B. Spoorenberg. Cybercontagion model with network structure applied to insurance. Insurance: Mathematics and Economics, 107:88–101, 2022.

[27] C. Hillairet, T. Mastrolia, and W. Sabbagh. Optimal impulse control for cyber risk management. arXiv preprint arXiv:2410.17706, 2024.

[28] M. Husák, J. Komárková, E. Bou-Harb, and P. Celeda. Survey of attack projection, prediction, and forecasting in cyber security. IEEE Communications Surveys & Tutorials, 21(1):640–660, 2019.

[29] Z. Jin, H. Yang, and G. Yin. A hybrid deep learning method for optimal insurance strategies: Algorithms and convergence analysis. Insurance: Mathematics and Economics, 96:262–275, 2021.

[30] W. Kermack, A. McKendrick, Contributions to the mathematical theory of epidemics (part I), *Proc. Royal Soc. Ser. A,* 115 (1927), 700-721.

[31] W. Kermack, A. McKendrick, Contributions to the mathematical theory of epidemics (part II), *Proc. Royal Soc. Ser. A*, 138 (1932), 55-83.

[32] D.H. Knipl, G. Rost, J. Wu, Epidemic spread and variation of peak times in connected regions due to travel-related infections-dynamics of an antigravity-type delay differential model. *SIAM J. Appl. Dyn. Syst.*, 12 (4) (2013), 1722-1762.

[33] V. N. Kolokoltsov and A. Bensoussan. Mean-field-game model for botnet defense in cyber-security. Applied Mathematics & Optimization, 74:669–692, 2016.

[34] I. Kortchemski, A predator-prey SIR type dynamics on large complete graphs with three phase transitions, *Stochastic Process. Appl.,* 125 (2015), no. 3, 886-917.

[35] H. Kushner and P. G. Dupuis. Numerical Methods for Stochastic Control Problems in Continuous Time, Springer Science, New York, 2013.

[36] H.J. Kushner and G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*, 2nd Ed., Springer-Verlag, New York, 2003.

[37] A. Mazzoccoli and M. Naldi. Robustness of optimal investment decisions in mixed insurance/investment cyber risk management. Risk Analysis, 40(3):550–564, 2020.

[38] B. Ma, T. Chu, and Z. Jin. Frequency and severity estimation of cyber attacks using spatial clustering analysis. Insurance: Mathematics and Economics, 106:33–45, 2022.

[39] S. Morgan. Cybercrime to cost the world 10.5 trillion annually by 2025. Cybercrime Magazine, 13(11), 2020.

[40] B. Nguyen. Modelling cyber vulnerability using epidemic models. In Simultech, 232–239, 2017.

[41] D. Nguyen, G. Yin, and C. Zhu, Long-term analysis of a stochastic SIRS model with general incidence rates, *SIAM Journal on Applied Mathematics*, **80** (2020), 814–838.

[42] T. T. Nguyen and G. Armitage. A survey of techniques for internet traffic classification using machine learning. IEEE Communications Surveys & Tutorials, 10(4):56–76, 2008.

[43] T. T. Nguyen and V. J. Reddi. Deep reinforcement learning for cyber security. IEEE Transactions on Neural Networks and Learning Systems, 34(8): 3779–3795, 2021.

[44] T. Rid and P. McBurney. Cyber-weapons. The RUSI Journal, 157(1):6–13, 2012.

[45] F. Selley, A. Besenyei, I.Z. Kiss, P.L. Simon, Dynamic control of modern, network-based epidemic models. *SIAM J. Appl. Dyn. Syst.,* 14 (2015), no. 1, 168-187.

[46] J. Sirignano and K. Spiliopoulos. DGM: A deep learning algorithm for solving partial differential equations. Journal of Computational Physics, 375:1339–1364, 2018.

[47] A. Sperotto, G. Schaffrath, R. Sadre, C. Morariu, A. Pras, and B. Stiller. An overview of IP flow-based intrusion detection. IEEE Communications Surveys & Tutorials, 12(3):343–356, 2010.

[48] J. M. Torres, C. I. Comesaña, and P. J. García-Nieto. Machine learning techniques applied to cybersecurity. International Journal of Machine Learning and Cybernetics, 10(10):2823–2836, 2019.

[49] J. P. Tailor and A. D. Patel. A comprehensive survey: Ransomware attacks prevention, monitoring and damage control. International Journal of Scientific Research, 4:2321–2705, 2017.

[50] W. Wang, X. Q. Zhao, Basic reproduction numbers for reaction-diffusion epidemic models. *SIAM J. Appl. Dyn. Syst.,* 11 (2012), no. 4, 1652-1673.

[51] S. X. Wu and W. Banzhaf. The use of computational intelligence in intrusion detection systems: A review. Applied Soft Computing, 10(1):1–35, 2010.

[52] Y. Xin, L. Kong, Z. Liu, et al. Machine learning and deep learning methods for cybersecurity. IEEE Access, 6: 35365–35381, 2018.

[53] A. Young and M. Yung. Cryptovirology: extortion-based security threats and countermeasures. In Proceedings 1996 IEEE Symposium on Security and Privacy, 129–140, 1996.

[54] G. Zeller and M. A. Scherer. Is accumulation risk in cyber systematically underestimated? European Actuarial Journal, 14(17), 2024.

[55] Y. Zhang, Z. Jin, and J. Wei. A hybrid deep learning method for controlled stochastic Kolmogorov systems with regime-switching//2024 10th International Conference on Control, Decision and Information Technologies (CoDIT), 970–975, 2024.

[56] Y. Zhang, Z. Jin, J. Wei, and G. Yin. A hybrid deep learning method for finite-horizon mean-field game problems. Automatica, to appear, 2025.