Reinforcement Learning-assisted Threshold Optimization for Dynamic Honeypot Adaptation to Enhance IoBT Networks Security

Elnaz Limouchi
CEECS Department
Florida Atlantic University
Boca Raton, USA
elimouchi2012@fau.edu

Imad Mahgoub
CEECS Department
Florida Atlantic University
Boca Raton, USA
mahgoubi@fau.edu

2021 IEEE Symposium Series on Computational Intelligence (SSCI) | 978-1-7281-9048-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/SSCI50451.2021.9660066

***Abstract*—Internet of Battlefield Things (IoBT) is the application of Internet of Things (IoT) to a battlefield environment. IoBT networks operate in difficult conditions due to high mobility and unpredictable nature of battle fields and securing them is a challenge. There is increasing interest to use deception techniques to enhance the security of IoBT networks. A honeypot is a system installed on a network as a trap to attract the attention of an attacker and it does not store any valuable data. In this work, we introduce IoBT dual sensor gateways. We propose a Reinforcement Learning (RL)-assisted scheme, in which the IoBT dual sensor gateways intelligently switch between honeypot and real function based on a threshold. The optimal threshold is determined using reinforcement learning approach that adapts to nodes reputation. To focus on the impact of the mobile and uncertain behavior of IoBT networks on the proposed scheme, we consider the nodes as moving vehicles. We statistically analyze the results of our RL-based scheme obtained using ns-3 network simulation, and optimize value of the threshold.**

***Index Terms*—IoBT, IoT, security, deception technique, honeypot, reinforcement learning, threshold optimization**

I. Introduction

Over the last few years, Internet of Things (IoT) technology has experienced significant advancement in adoption and deployment. Advanced approaches in IoT technology are paving the way for Internet of Battle Field Things (IoBT), which is the utilization of IoT for military applications. IoBT networks are dynamic, unpredictable, with high mobility, and without infrastructure. The information is time sensitive, so decisions have to be made in a timely manner. They are bandwidth demanding, sensitive to outages, and jamming by adversaries. Such an environment causes many potential challenges in the context of security for researchers to design and develop effective solutions.

Security deception is a promising approach for cyber defense and has attracted researchers' interest. Defensive deception makes the defender able to anticipate the actions of attacker, conceal real resources, and also misdirect the attacker [1]. Honeypots have been widely used in different security applications such as malware analysis, detecting spams, and database security [2]. In a network, a honeypot is placed to engage the attention of attacker(s). Honeypots are supposed not to store any valuable data and not to be involved with everything happening in the network. Attackers waste their resources attempting to compromise the honeypot which does not hold any functional information. Meanwhile, in order to better secure the network, the actions of attackers can be extensively inspected.

An Intrusion Detection System (IDS) is a classical method to support the protection of network devices using low-cost solutions. IDS surveils the network based on different metrics, such as reputation management. A reputation is determined for each node according to the past interactions that it had with its neighbors.

Reinforcement learning algorithms try to comprehend a situation-to-action mapping where the main intention is to maximize a numerical reward. The learning agent needs to identify rewarding actions in a variety of situations by conducting trials.

In this paper, we introduce IoBT dual sensor gateways. We propose a Reinforcement Learning (RL)-assisted scheme, in which the IoBT dual sensor gateways intelligently switch between honeypot and real function based on a threshold. In order to make the decision, Intelligent Dual Function Sensor Gateway (IDFG) uses the information from reputation management module and compares a locally measured value to a threshold. If this value exceeds the threshold, IDFG will proceed as a honeypot. Observing the cost and rewards of this method, Bayesian optimization is used to optimize the value of the threshold. Bayesian optimization is most applicable for objective complex and noisy functions.

The remainder of this paper is structured as follows: Section II provides some related work on IoBT deception methods. In Section III we have an overview on Trust-based attacks against Reputation Management, the Drift-Diffussion

Office of the Secretary of Defense (OSD) Grant number: W911NF2010300

Model, and Bayesian optimization subjects. In Section IV our proposed reinforcement learning-based threshold optimization is presented. Section V discusses the results of our proposed scheme. Finally, Section VI concludes the paper.

II. Related Work

To develop defensive deception techniques employing honeypots, two main directions are significantly promising: game-theoretic and machine learning (ML)-based approaches.

Game theoretic methods are commonly applied to model strategies of an attacker and defender. Normally, the main objective of these models is to make attackers get confused or deceive them to select poor strategies. The method introduced in [3], is a two-player signaling game model. The proposed model a defender (sender) could have two roles either a normal node or a honeypot, while the attacker (receiver) has only one role. In [4], a honeypot-based deceptive attack and defence in an IoBT network is analyzed. The game theoretic model used to analyze the problem is a Bayesian signaling game of incomplete information. The Perfect Bayesian Equilibrium (PBE) solution is identified for both one-short and the repeated game scenarios.

Machine learning (ML)-based defensive deception techniques are mainly used to mislead or lure attackers by generating fake information or creating decoy objects which closely imitate real objects or information. A machine learning-based malware detection model for honeypots in an IoT network is proposed in [5]. The data originated by honeypots is then used as a dataset to dynamically train the machine learning model. Authors in [6] discuss the application of Artificial Intelligence(AI) for dynamic honeypots which are able to learn about the network environment and then perform the deployments accordingly. Since dynamic honeypots keep track of any changes in the network and update the current configurations based on the changes, authors conclude that AI-based techniques have significant potential to improve the reasoning and decision making approaches in the context of honeypot deployment. The proposed SSH/Telnet honeypot system for IoT networks in [7], uses reinforcement learning algorithms to capture more attacks and lure the attackers to download more malware.

In [8], authors formulize the IoBT network as a graph of graphs, from an adversary point of view, where adversaries can extract high level information of an IoBT network. Authors in [9], concentrate on the dynamic locations of honeypots to introduce a distributed honeypot scheme. This model recognizes the illegal attack flow by periodically changing the services which makes the attacker get confused to determine real services from honeypots.

In essence of addressing the trust and trustworthiness related issues in IoBT networks, a set of research guidelines are proposed in [10]. The proposed research directions mainly focus on two subjects, supporting trust assessment for known/unknown IoT assets and known IoBT assets and systems. In [11], honeypots are employed for reputation management in IoT networks considering the trade-off between reducing the detection error and battery lifetime of devices.

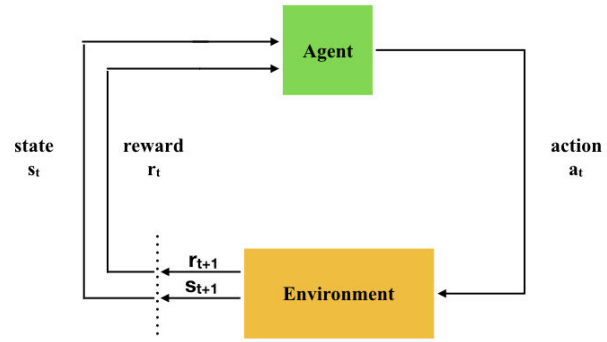

Fig. 1. Reinforcement Learning Block Diagram

In this paper, we propose a reinforcement learning-based scheme in which the Intelligent Dual Function Sensor Gateways (IDFGs) are able to choose between acting as a honeypot or a normal node based on a threshold. The threshold is determined by our proposed reputation management module, and is optimized by a Bayesian optimizer.

III. Methodology

A. Trust-based attacks against Reputation Management

Speaking of reputation management, an attacker can launch an attack in order to ruin the trust between nodes [11]. Mainly three types of attacks have been introduced that can be created against reputation management solutions [12]:

$Self-promoting\ attacks$: The attacker publishes better recommendation about itself to attract more data traffic. When the attacker receives traffic, it can then initiate selective-forwarding or sinkhole attacks.

$Bad-mouthing\ attacks$: The attacker plans to diminish the reputation of good nodes. For this purpose, the attacker may provide false trust values and destroy the reputation of the good node. This makes the other nodes not want to forward the traffic via that good node.

$Ballot-stuffing\ attacks$: The attacker supports other attackers by giving them false trust values to improve their reputation. In this way, these malicious nodes can commence a coordinated attack.

B. Drift-Diffusion Model

The drift-diffusion model (DDM) is considered to model the results of experiments with two-alternative forced choice (2AFC) in psychophysics [13], [14]. Considering a decision variable $z(t)$, which represents the sensory evidence accumulated to time t from a starting bias $z(0) = z_0$, the update equation is stated by (1).

$$z(t+1) = z(t) + \Delta z, \qquad \Delta z \sim N(\mu, \sigma^2) \quad (1)$$

where Δz is the increment of sensory evidence at time t. The decision scale is based on the added evidence, when it exceeds one of two decision thresholds, assumed at $-\theta_0 < 0 < \theta_1$.

Wald's sequential probability ratio test (SPRT) checks if one of the hypotheses, H_0 or H_1, is supported by gathering samples x(t) up to the point a decision can be confidently made [15].

It is considered optimal since it minimizes the average sample size compared with all sequential experiments with the similar error probabilities. The DDM is a particular case of SPRT by setting the likelihoods as two equivariant Gaussians $N(\mu_1, \sigma)$, $N(\mu_0, \sigma)$ so that

$$log\frac{p(x|H_1)}{p(x|H_0)} = log\frac{e^{-(x-\mu_1)^2/2\sigma^2}}{e^{-(x-\mu_1)^2/2\sigma^2}} = \frac{\Delta\mu}{\sigma^2} + d \qquad (2)$$

where $\Delta\mu = \mu_1 - \mu_0$, and $d = \frac{{\mu_0}^2 - {\mu_1}^2}{2\sigma^2}$.

C. Reinforcement learning for optimal decision making

This cost function is linear in type I and II error probabilities $\alpha_1 = P(H_1|H_0) = E_1(e)$ and $\alpha_0 = P(H_0|H_1) = E_0(e)$, the decision error $e = \{0, 1\}$ for correct/incorrect trials. The cost function is also linear in the expected stopping times for each decision result.

$$C_{risk} := \frac{1}{2}(W_0\alpha_0 + cE_0[T]) + \frac{1}{2}(W_1\alpha_1 + cE_1[T]) \qquad (3)$$

with type I /II error costs $W0, W1 > 0$ and cost of time, c. The Bayes risk, C_{risk} has a unique minimum which follows from the error probabilities α_0, α_1, are monotonically decreasing and the expected stopping times $E_0[T]$, $E_1[T]$, are monotonically increasing with increasing threshold θ_0 or θ_1. For each pair $(W_0/c, W_1/c)$, there is thus a unique threshold pair (θ_0^*, θ_1^*) that minimizes C_{risk}. The reward can be formulated as (4)

$$R = \begin{cases} -W_0 - cT, & incorrect\ decision\ of\ hypothesis\ H_0 \\ -W_1 - cT, & incorrect\ decision\ of\ hypothesis\ H_1 \\ -cT, & correct\ decision\ of\ hypothesis\ H_0 or H_1. \end{cases} \qquad (4)$$

Over many decision trials, the average reward will become $< Ri >= -C_{risk}$, the negative of the Bayes risk. Reinforcement learning can then be utilized to find the optimal thresholds to maximize reward and thus optimize the Bayes risk. Exceeding many trials $n = 1, 2, ..., N$ with reward $R(n)$, the problem is to estimate these optimal thresholds (θ_0^*, θ_1^*) while maintaining minimal regret: the difference between the reward sum of the optimal decision policy and the the collected rewards sum.

$$f_{Regret}(N) = -NC_{risk}(\theta_0^*, \theta_1^*) - \sum_{n=1}^{N} R(n) \qquad (5)$$

In order to optimize the thresholds that maximize mean reward, a successful approach must combine reward averaging with learning.

D. Bayesian Optimization

Bayesian optimization is a powerful method to find the optimal thresholds by iteratively building a probabilistic model of the reward function that is used to lead future sampling [16]. Bayesian optimization typically uses a Gaussian process model, which provides a nonlinear regression model the mean reward and the reward variance with decision threshold. This model can then be used to determine the future threshold choice with maximizing an acquisition function of these quantities.

Selecting the decision thresholds is considered as a sampling problem, and expressed by maximizing an acquisition function of the decision thresholds taking into account the trade off between exploration and exploitation. Using the probability of improvement, the sampling is led towards regions of high uncertainty and reward by maximizing the chance of improving the present best estimate.

IV. Proposed Method

A. System Model

Similar to other IoT environments, the topology of the IoBT can be configured in a variety of ways. In this work, we consider an IoBT network in which all the nodes are vehicles, therefore high speed mobility and constantly changing topology bring on even more challenges. In this system, we assume that all the nodes are equipped with a Global Positioning System (GPS) and have knowledge about their position and velocity. In order to exchange information with the other nodes in the transmission range (neighbors), each node is able to broadcast beacons which contain the position, velocity and ID information. In this way, each node can establish and update a neighboring table. In this work, we consider an adversary model that changes the mobility information exchanged via beaconing, with a focus on position and time data. First, the malicious node runs a random positioning function to falsify the information of its current position. To generate such a fake random position, the malicious node considers that the distance between the present and the random position should be less than or equal to the transmission range, R. Then, the malicious node attaches the generated fake position to a packet and broadcast it as a beacon. The Intelligent Attack Detection System (IADS) that we propose in this work, checks the information received from beacons and determines trust values of nodes in order to internally mange the reputations. Utilizing the observed reputations, nodes with Intelligent Dual Function Sensor Gateways (IDFGs) can intelligently decide whether to operate as a real function or a honeypot based on a decision threshold value. In this work, relying on the information from the proposed trust verification and reputation management module [17], we model the system as a Drift-Diffusion Reinforcement Learning environment [13], [14] and optimize the threshold value by Bayesian Optimization algorithm [18], [19]. The details about each module is presented in the following subsections.

B. Reputation Management

Our proposed Intelligent Attack Detection System (IADS) is required to determine the trustfulness of nodes in order to obtain the nodes reputation. We consider node h as an Intelligent Dual Function Sensor Gateway, which is able to switch between a regular node or a honeypot function. Node h examines the information received via beacons to verify the trustfulness of its neighbors. Then, it determines and assigns a trust value to each neighbor. Then, it determines a disrepute factor as the number of neighbors that have been recognized untrustable more than once within interval $t_{reputation}$. If the value of disrepute factor exceeds a threshold, θ, node h will act as a honeypot. As shown in Fig. 2, both parameters checking and reputation indicator modules contribute to reputation management.

1) Parameters Checking: When Node h receives a beacon message, it starts to analyze the contents to identify the mismatch in the information. The results of the examinations are described as binary values, 0 and 1. Basically, two parameters are considered to be checked in this method, position and time.

After receiving a packet from a neighbor, node h estimates the current position of this neighbor using previously received position information. Then, it compares the value of estimated position with the current position of this neighbor stated in the beacon. If these two values are unequal, then the information received from this neighbor is considered to be inaccurate. In this case, the position trust value for this neighbor will be set as 1 ($T_{position} \leftarrow 1$). Similarly, node h scans time information, if the timestamp of received packet is not current, the node that the packet is received from will be identified as untrusted ($T_{time} \leftarrow 1$).

2) Reputation Indicator: The main purpose of this unit is to derive the trustfulness of nodes which can be obtained by calculating the trust value as a discrete binary value (0 and 1). As mentioned earlier, node h is able to evaluate the trustfulness of its neighbor by inspecting the related received current and previous information. The trust value of position and time for each neighbor is calculated as 6:

$$T_{value} = T_{position} \otimes T_{time} \tag{6}$$

where T_{value} denotes trust value, and $T_{position}$ and T_{time} are position parameter and timestamp parameter, respectively.

Then the obtained trust values will be included into the neighboring table. Then, node h calculates the $Disrepute\ Factor$ which is defined as the number of neighbors that have had a trust value of 1 (inaccurate nodes) more than once within a time interval, $t_{reputation}$. This $Disrepute\ Factor$ is used by node h to decide whether to function normally or act as a honeypot.

C. Threshold Optimization

As mentioned earlier, we consider the optimization of decision threshold as reinforcement learning over single trial rewards extracted from averaged cost function. In this work, cost is bandwidth consumption that is defined in terms of redundant transmissions that is imposed by fake information sent from malicious node(s). Referring to Subsection III-B, we consider the case of optimization of one threshold (for equal thresholds θ_0=θ_1) where $c/W_1 = c/W_0$.

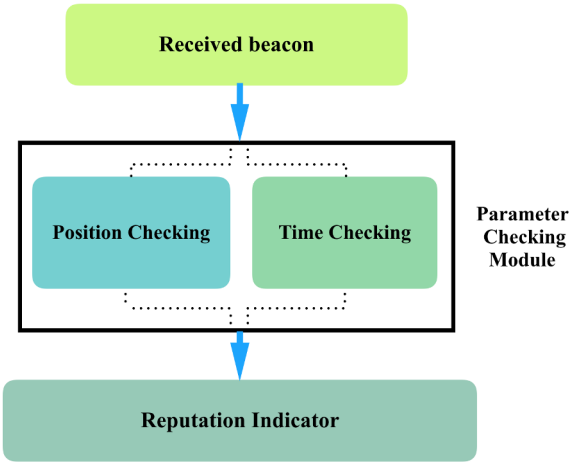

Fig. 2. Reputation Management Module Block Diagram

Within the time interval, $t_{reputation}$, and relying on the value of disrepute factor, if node h notices that the number of inaccurate nodes exceeds the threshold, θ, then it will switch to work as a honeypot. We use θ to represent the states which are disrepute factors. For this study, we consider $\theta \in$ {1, 2, 3, 4, 5}. The initial state is associated with threshold value of 5 where node h functions as a normal node. The action is either to act as a honeypot or a normal node. Considering the scenario of data dissemination, the rewards basically are dependent to overheard redundant transmissions. Based on that it can be concluded whether the previous action was correct or incorrect, and according to that the next state will be updated.

Bayesian optimization algorithm is deployed to observe the optimal threshold value by establishing a probabilistic model of the reward function to guide upcoming sampling process. The algorithm of Bayesian Optimization is described as follows:

Algorithm 1 Bayesian Optimization Algorithm

1: **for** N loops **do**:
2: Select the new threshold from optimizing acquisition function, θ_n
3: Make the decision with threshold θ_n to find reward $R(n)$
4: Augment data by including new samples S_n=(S_{n-1},θ_n, $R(n)$)
5: Update the statistical model (Gaussian process) of the rewards
6: **end for**

V. Results and Discussion

In this section, we describe the network simulation process and analyze the results. To simulate the network, ns-3 [20] is used. The simulation is run for an active period of 600

TABLE I
NETWORK SIMULATION PARAMETERS

Parameter	Value
Duration	600 seconds
Packet size	500 bytes
MAC/PHY protocol	IEEE 802.11p
Transmission range	250 meters
Layer 3 addressing	IPv4

seconds and the transmission range is considered to be 250 meters. Each packet has a data size of 500 bytes. To address the variation of signal strength due to multiple fading, "ns-3 Nakagami Propagation Loss Model" in combination with the "ns-3 Range Propagation Loss Model" is employed.

We consider the communication architecture based on the implemented Wireless Access in Vehicular Environments (WAVE) model in ns-3 [21]. This implemented model supports 802.11p MAC and PHY layers. It utilizes the 5.9 GHZ band with channel bandwidth of 10 MHZ and a data rate of 6 Mbps. The layer 3 addressing is based on IPv4 protocol. The mobility of nodes is generated applying ns-3.27 constant speed mobility model. We also use ns-3 random rectangle position model to locate nodes on a straight line.

We produce the adversary model based on changing the mobility information exchanged via beaconing. Certain nodes in the network have the ability to act either as a regular node or a honeypot. In this work we examine the network with one potential honeypot acting node.

We conduct the simulation to capture the results of 3 replications [22] running the position-based data dissemination scheme [23]. To allow adequate time for observation, the $t_{reputation}$ is considered to include 10 intervals of beaconing.

Three factors namely node density, disrepute factor, and beaconing interval are monitored in order to determine how these variables and their interaction affect the redundant transmission imposed by the adversary model. To test significance of the results from the simulation of the proposed model, analysis of variance is detailed in Table II. It is assumed that errors in the model are normally and independently distributed. On the other hand, as long as P-value for the model is less than 5 percent, the model would be significant. It means that at least one of the factor variations in the model is meaningfully affecting the response factor (output parameter) [22], which is redundant transmission. Table II indicates that the "disrepute factor" is the significant factor where its P-value is much less than 5 percent. Fig. 3 graphically visualizes the effectiveness of factors. It shows that the disrepute factor is the most effective, and the node density as the second most effective variable in determining the response variable of the model, redundant transmission. According the red Bonferroni Limit, which is 2.021, the impact of the other variables are not significant.

The adequacy of the regression model has been specified by residual analysis. In other words, checking the residuals is one of the main features of statistical modelling specifically in design of experiments. Since there is not any obvious

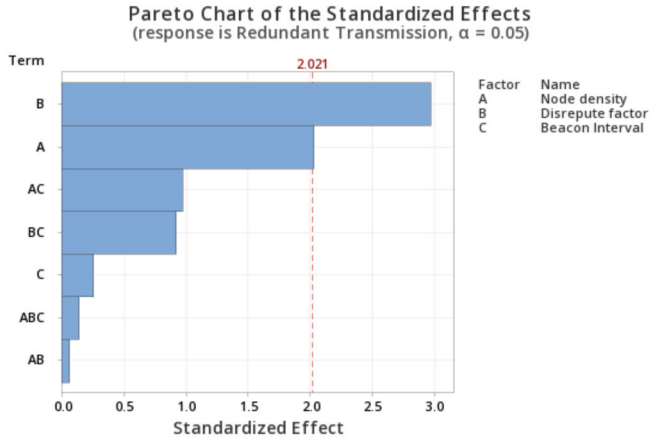

Fig. 3. Pareto Chart of the Standardized Effects

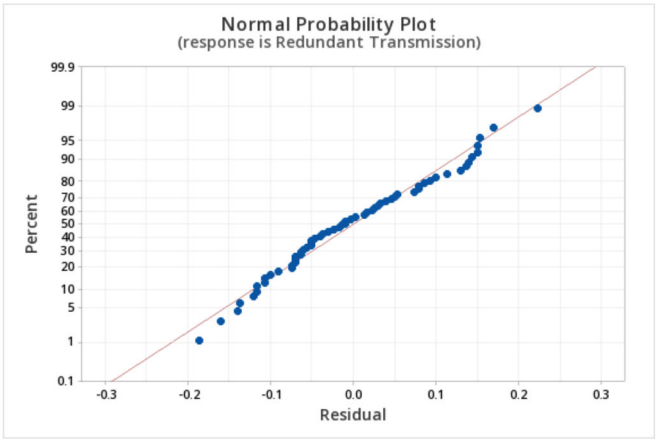

Fig. 4. Normal Probability Plot

pattern on residual plots, the adequacy of the regression model would be approved. Normal probability plot of residuals is shown in Fig. 4. Since there is no significant deviation from normality line, constant variance assumptions are satisfied and no action is required. It also indicates that errors are normally distributed, assumptions are reasonable and the regression model is well developed. Moreover, the normal probability plot exhibits that the experiment is not significantly affected by the noise.

Fig. 5 shows the histogram of Residual values. It shows that the residual values are almost normally distributed on both sides of the zero value. Hence, constant variance assumptions are met.

The Bayesian algorithm uses a Gaussian model which provides a nonlinear regression of the mean reward and the reward variance with decision threshold. Fig. 7 shows the plot of all samples (dots) and surrogate function across the domain (line) after Bayesian Optimization. The best result is associated with the threshold value of 2.

TABLE II
ANALYSIS OF VARIANCE

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	19	0.383952	0.020208	1.53	0.128
Linear	6	0.287593	0.047932	3.62	0.006
Node density	1	0.054602	0.054602	4.13	0.049
Disrepute factor	4	0.232110	0.058028	4.38	0.005
Beacon interval	1	0.000882	0.000882	0.07	0.798
2-way Interactions	9	0.081282	0.009031	0.68	0.720
Node density*Disrepute factor	4	0.009390	0.002347	0.18	0.949
Node density*Beacon Interval	1	0.012615	0.012615	0.95	0.335
Disrepute factor*Beacon Interval	4	0.059277	0.014819	1.12	0.361
3-Way Interactions	4	0.015077	0.003769	0.28	0.886
Node density*Disrepute factor*Beacon Interval	4	0.015077	0.003769	0.28	0.886

TABLE III
FITS AND DIAGNOSTICS FOR UNUSUAL OBSERVATIONS

Obs	Redundant Transmission	Fit	Resid	Std Resid	
46	0.57	0.3467	0.2233	2.38	R

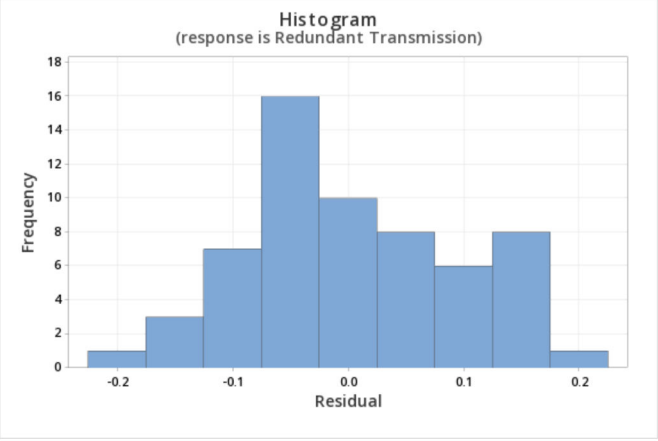

Fig. 5. Residual Histogram

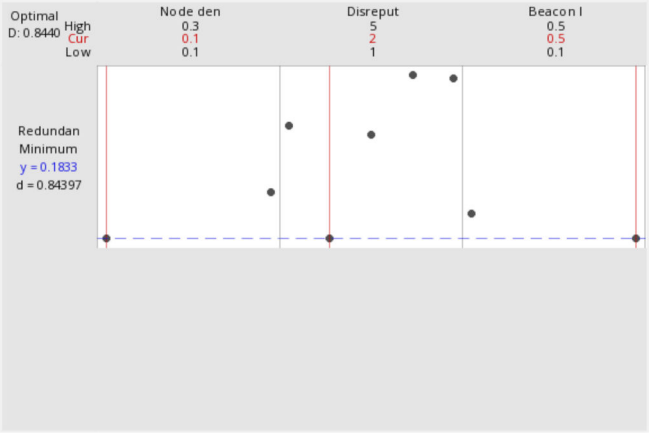

Fig. 6. Optimum Threshold Value

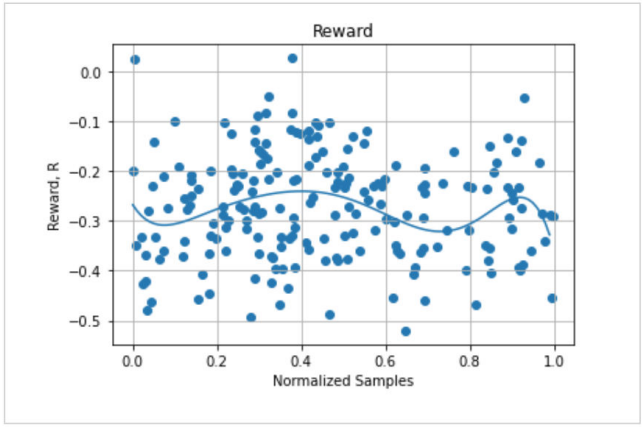

Fig. 7. Reward-cost

VI. CONCLUSION

The dynamics and extensive of the promising Internet of Battlefield Things (IoBT) bring up exclusive challenges to command and control, coincidental awareness, and the achievement of mission ambitions. In this paper we proposed a reinforcement learning-based scheme in which the IoBT Intelligent Dual Function Sensor Gateways (IDFGs) switch between honeypot and real function according to a threshold. The threshold optimization uses reinforcement learning approach. We introduced a reputation observing module whereby internal attacks were detected. The obtained disrepute factor characterized the value of the threshold. In order to address the highly mobile and uncertain nature of IoBT networks, we examined the scheme where all the nodes are vehicles. The communication of nodes is established based on IEEE 802.11P standard. The goal of this work is to optimize the decision threshold to minimize the redundant transmission due to the wrong information from malicious nodes that could negatively affect the process of data dissemination. We statistically analyzed the results from our proposed reinforcement learning-based model, and presented the value 2 as the Bayesian optimization resulted value of the threshold that IDFGs make decisions accordingly. As future work, we will extend and

evaluate the scheme for a heterogeneous IoBT network.

ACKNOWLEDGMENT

This work is done in Tecore Labs at Florida Atlantic University and funded by Office of the Secretary of Defense (OSD) Grant number W911NF2010300.

REFERENCES

[1] M. Zhu, A. H. Anwar, Z. Wan, J.-H. Cho, C. Kamhoua, and M. P. Singh, "Game-theoretic and machine learning-based approaches for defensive deception: A survey," *ArXiv*, vol. abs/2101.10121, 2021.

[2] S. Jajodia, V. S. Subrahmanian, V. Swarup, and C. Wang, *Cyber deception: building the scientific foundation*. Springer, 2016.

[3] J. Pawlick and Q. Zhu, "Deception by design: Evidence-based signaling games for network defense," *CoRR*, vol. abs/1503.05458, 2015. [Online]. Available: http://arxiv.org/abs/1503.05458

[4] Q. La, T. Q. S. Quek, J. Lee, S. Jin, and H. Zhu, "Deceptive attack and defense game in honeypot-enabled networks for the internet of things," *IEEE Internet of Things Journal*, vol. 3, pp. 1025–1035, 2016.

[5] R. Vishwakarma and A. Jain, "A honeypot with machine learning based detection framework for defending iot based botnet ddos attacks," *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, pp. 1019–1024, 2019.

[6] W. Z. A. Zakaria and M. M. Kiah, "A review on artificial intelligence techniques for developing intelligent honeypot," *2012 8th International Conference on Computing Technology and Information Management (NCM and ICNIT)*, vol. 2, pp. 696–701, 2012.

[7] B. I. P. F. e. a. Pauna, A., "On the rewards of self-adaptive iot honeypots," *Annals of Telecommunications*, vol. 74, 2019.

[8] J. Park, A. Mohaisen, C. Kamhoua, M. Weisman, N. O. Leslie, and L. Njilla, "Cyber deception in the internet of battlefield things: Techniques, instances, and assessments," in *WISA*, 2019.

[9] Y. Li, L. Shi, and H. Feng, "A game-theoretic analysis for distributed honeypots," *Future Internet*, vol. 11, no. 3, 2019. [Online]. Available: https://www.mdpi.com/1999-5903/11/3/65

[10] I. Agadakos, G. F. Ciocarlie, B. Copos, J. George, N. Leslie, and J. Michaelis, "Security for resilient iobt systems: Emerging research directions," in *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2019, pp. 1–6.

[11] Z. A. Khan and U. Abbasi, "Reputation management using honeypots for intrusion detection in the internet of things," *Electronics*, vol. 9, no. 3, 2020. [Online]. Available: https://www.mdpi.com/2079-9292/9/3/415

[12] Z. A. Khan, P. Herrmann, and J. M. Alcaraz-Calero, "Recent advancements in intrusion detection systems for the internet of things," *Sec. and Commun. Netw.*, vol. 2019, Jan. 2019. [Online]. Available: https://doi.org/10.1155/2019/4301409

[13] R. Ratcliff, "A theory of memory retrieval," *PSYCHOL. REV*, vol. 85, no. 2, pp. 59–108, 1978.

[14] N. F. Lepora, "Threshold learning for optimal decision making," in *Advances in Neural Information Processing Systems*, D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett, Eds., vol. 29. Curran Associates, Inc., 2016.

[15] A. Wald and J. Wolfowitz, "Optimum Character of the Sequential Probability Ratio Test," *The Annals of Mathematical Statistics*, vol. 19, no. 3, pp. 326 – 339, 1948. [Online]. Available: https://doi.org/10.1214/aoms/1177730197

[16] J. Snoek, H. Larochelle, and R. P. Adams, "Practical bayesian optimization of machine learning algorithms," in *Advances in Neural Information Processing Systems*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds., vol. 25. Curran Associates, Inc., 2012.

[17] L. Altoaimy and I. Mahgoub, "Mobility data verification for vehicle localization in vehicular ad hoc networks," in *2016 IEEE Wireless Communications and Networking Conference*. IEEE Press, 2016, p. 1–6. [Online]. Available: https://doi.org/10.1109/WCNC.2016.7564749

[18] R. J. Williams, "Simple statistical gradient-following algorithms for connectionist reinforcement learning," in *Machine Learning*, 1992, pp. 229–256.

[19] J. Peters and S. Schaal, "Reinforcement learning of motor skills with policy gradients." *Neural Networks*, vol. 21, no. 4, pp. 682–697, 2008. [Online]. Available: http://dblp.uni-trier.de/db/journals/nn/nn21.html/PetersS08

[20] "The network simulator ns-3," https://www.nsnam.org/.

[21] "Wave models," https://www.nsnam.org/docs/models/html/wave.html.

[22] "Minitab statistical software (2021). [computer software]," https://www.minitab.com/.

[23] M. Slavik, I. Mahgoub, and M. M. Alwakeel, "Analysis and evaluation of distance-to-mean broadcast method for {VANET}," *Journal of King Saud University - Computer and Information Sciences*, vol. 26, no. 1, pp. 153 – 160, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1319157813000293