Factored MDP based Moving Target Defense with Dynamic Threat Modeling

Extended Abstract

Megha Bose
International Institute of Information Technology
Hyderabad, India
megha.bose@research.iiit.ac.in

Praveen Paruchuri
International Institute of Information Technology
Hyderabad, India
praveen.p@iiit.ac.in

Akshat Kumar
Singapore Management University
Bras Basah, Singapore
akshatkumar@smu.edu.sg

ABSTRACT

Moving Target Defense (MTD) has emerged as a proactive defense framework to counteract ever-changing cyber threats. Existing approaches often make assumptions about attacker-side knowledge and behavior, potentially resulting in suboptimal defense. This paper introduces a novel MTD approach, leveraging a Markov Decision Process (MDP) model that eliminates the need for prior knowledge about attacker intentions or payoffs. Our framework seamlessly integrates real-time attacker responses into the defender's MDP using a dynamic Bayesian network. We use a factored MDP model to enable a more comprehensive and realistic representation of the system having multiple switchable aspects and also accommodate incremental updates of an attack response predictor as new attack data emerges, ensuring adaptive defense. Empirical evaluations demonstrate the approach's effectiveness in uncertain scenarios with evolving as well as unknown attack landscapes.

KEYWORDS

Moving Target Defense; Markov Decision Process; Adaptive Strategy; Uncertainty

ACM Reference Format:
Megha Bose, Praveen Paruchuri, and Akshat Kumar. 2024. Factored MDP based Moving Target Defense with Dynamic Threat Modeling: Extended Abstract. In *Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), Auckland, New Zealand, May 6 – 10, 2024*, IFAAMAS, 3 pages.

1 INTRODUCTION

An inherent information asymmetry exists in cyber systems, favoring attackers due to their ability to conduct reconnaissance [13] on static systems. Attackers can probe systems, gaining insights into vulnerabilities and planning attacks over time. Moving Target Defense (MTD) [4] aims to counter this advantage by introducing uncertainty through dynamic alterations of system configurations, making it challenging for attackers to study the system and launch targeted attacks. However, MTD introduces overheads, including maintenance overhead and possible service disruptions [2], making it imperative to consider such switching costs involved.

Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), N. Alechina, V. Dignum, M. Dastani, J.S. Sichman (eds.), May 6 – 10, 2024, Auckland, New Zealand. © 2024 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

The MTD problem gets more complex as cyber systems evolve [5], as it becomes harder to know the attacker's intentions beforehand. Prior approaches, often based in game theory, assume an impractical amount of knowledge about the attacker side payoffs, leading to suboptimal defense strategies. This paper models MTD as a Factored Markov Decision Process (FMDP) [3], considering uncertainty over a set of attacker types and doesn't assume prior knowledge about attacker rewards and intentions.

2 PROPOSED METHOD

The FMDP's state space consists of all possible system configurations with adaptation aspects [2] as factors. Actions represent configurations switched to, and the defender obtains reward based on attack executed and the switching action taken. The *switching costs* (sc) depend solely on the current and next configurations and lie in $[0, 100]$. Transitions between states occur deterministically based on the switching action. We employ an attacker model of a binary *attacker response* variable φ that takes value 1 on defender observing a successful attack and 0 otherwise, captured into the FMDP using a dynamic Bayesian Network [9]. Similar to prior works [7, 10, 11], we consider that the system confronts various *attacker types*. Initially, the attacker-type distribution is unknown to the defender, but the defender maintains an estimation of it. We also account for an "unknown" attacker, encompassing all attacker types not known to the defender. An *attack success rate* denoted by $\mu(\tau, s) \in [0, 1]$ for each attacker type τ quantifies their proficiency in executing a successful attack in state s. An average *unit time system loss* $l(\tau, s) \in [0, 100]$ is experienced by the defender when dealing with a successful attack from attacker type τ in state s [7].

Algorithm 1 solves for the optimal policy through the approximate linear programming formulation for the FMDP with constraints constructed by taking expectation over the possible values of the binary response variable φ. In each timestep, if any anomalous activity is detected, we run the re-estimation in Step 6 to update the MDP policy π based on the current estimate of attacker type probabilities P_{att}. The defender takes an action a according to the last calculated policy (Step 8), and receives response φ, the attacker type τ, and the next state (Step 9). Using the attack response φ, unit-time system loss $l(\tau, (\boldsymbol{s}, \boldsymbol{a}))$ (tuple $(\boldsymbol{s}, \boldsymbol{a})$ and next state $\boldsymbol{s'}$ have the same meaning as transitions are deterministic) due to the attacker type τ and the switching cost, we calculate the defender's reward. We maintain a value n for each attacker type τ, state s, and action tuple a, representing a weighted sum of the attack type's success in that state when that action was executed using a *weighing factor* $\beta > 1$. At timestep t, n is updated in Step 12 and the attacker-type

Algorithm 1 Adaptive Threat-Aware FMDP for MTD

1: **Initialize parameters** Total timesteps T, weighting factor β, start state s_0, constant $M = 200$, $n_{\tau,s,a} = 0\ \forall \tau, s, a$
2: **Load domain information** *(Dom)*: states, actions, attacker types, switching costs (sc), attack success rates (μ), unit time system losses (l)
3: **for** $t \leftarrow 1$ to T **do**
4: **if** re-estimation triggered **then**
5: $fmdp[\hat{P}_{att}, Dom] \leftarrow create\_fmdp(\hat{P}_{att}, Dom)$
6: $\pi \leftarrow lp\_solver(fmdp[\hat{P}_{att}, Dom], \hat{P}_{att}, Dom)$ ▷ Re-calculate FMDP policy π
7: **end if**
8: $\boldsymbol{a_t} \leftarrow \pi$ ▷ Choosing action $\boldsymbol{a_t}$ using last calculated policy
9: $\varphi_{t+1}, \tau_t, \boldsymbol{s_{t+1}} \leftarrow mtd\_sim(\boldsymbol{s_t}, \boldsymbol{a_t})$ ▷ Real-world Interaction
10: As $\boldsymbol{s_{t+1}}$ is fixed given $(\boldsymbol{s_t}, \boldsymbol{a_t})$, we use them interchangeably.
11: $r_t \leftarrow M - \mathbb{1}[\varphi_{t+1} = 1] l(\tau_t, (\boldsymbol{s_t}, \boldsymbol{a_t})) - sc(\boldsymbol{s_t}, \boldsymbol{a_t})$
12: $n_{\tau,\boldsymbol{s},\boldsymbol{a}} \leftarrow \frac{n_{\tau,\boldsymbol{s},\boldsymbol{a}}}{\beta} + \mathbb{1}[\varphi_{t+1} = 1, \tau = \tau_t, \boldsymbol{s} = \boldsymbol{s_t}, \boldsymbol{a} = \boldsymbol{a_t}]\ \forall \tau, \boldsymbol{s}, \boldsymbol{a}$
13: $\hat{P}_{att}(\tau|\boldsymbol{s}, \boldsymbol{a}) \leftarrow \frac{1}{\mathcal{N}} \frac{n_{\tau,\boldsymbol{s},\boldsymbol{a}}}{\mu(\tau,(\boldsymbol{s},\boldsymbol{a}))} \forall \tau, \boldsymbol{s}, \boldsymbol{a}$ ▷ Eq. 1
14: **end for**

probability estimates ($\hat{P}_{att}$) are updated in Step 13 as follows:

$$\hat{P}_{att}(\tau|\boldsymbol{s}, \boldsymbol{a}) = \frac{P(\tau, \varphi = 1, \boldsymbol{s}, \boldsymbol{a})}{P(\varphi = 1|\boldsymbol{s}, \boldsymbol{a}, \tau) P(\boldsymbol{s}, \boldsymbol{a})} = \frac{1}{\mathcal{N}} \frac{n_{\tau,\boldsymbol{s},\boldsymbol{a}}}{\mu(\tau, (\boldsymbol{s}, \boldsymbol{a}))} \tag{1}$$

where $n_{\tau,\boldsymbol{s},\boldsymbol{a}}$ is a temporally weighted estimate of the number of attack successes that happened in the state s, on taking action a under attacks by attacker type τ and $\mathcal{N}$ is the normalizing factor. We take μ of 1 for *unknown* attacker as the defender lacks information about its proficiency.

(a) Switching Costs

	C_1	C_2	C_3	C_4
C_1	0	20	60	100
C_2	20	0	90	50
C_3	60	90	0	20
C_4	100	50	20	0

(b) Attacker Type Capabilities, Attack Success Rates, and Unit Time System Losses (true values for the *unknown* type)

	C_1	C_1	C_1	C_1
v_{MH}	{PHP, MySQL}	{Python, MySQL}	{PHP}	{Python}
μ_{MH}	0.32	0.32	0.36	0.36
l_{MH}	61	43	66	29
v_{DH}	{MySQL}	{MySQL}	{Postgres}	{Postgres}
μ_{DH}	0.7	0.7	0.65	0.65
l_{DH}	43	43	50	50
$v_{unknown}$	{PHP, MySQL}	{MySQL}	{PHP}	{}
$\mu_{unknown}$	0.78	0.7	0.87	0.0
$l_{unknown}$	100	100	100	0

Figure 1: Configurations $C_1 = (PHP, MySQL)$, $C_2 = (Python, MySQL)$, $C_3 = (PHP, Postgres)$, $C_4 = (Python, Postgres)$

3 EXPERIMENTS

Inspired by previous works [6, 7, 11], we employ the National Vulnerability Database (NVD) data [1] from the years 2020 to 2022 and Common Vulnerability Scoring System (CVSS) [8] scores to establish the experimental framework for a web app with states characterized by language and database: {(*PHP*, *MySQL*), (*Python*, *MySQL*), (*PHP*, *PostgreSQL*), (*Python*, *PostgreSQL*)}. Unit time losses and attack success rates are computed based on the CVSS scores [7]. v represents technologies that can be attacked by an attacker type in given state. Attacker types of Mainstream Attacker (*MH*), Database Hacker (*DH*) and *unknown* attacker are considered.

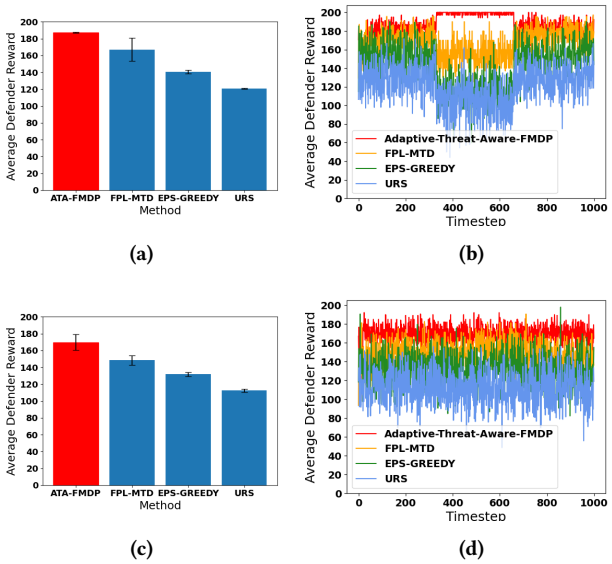

Figure 2: Evolving Attack Landscape where (1) (a, b) **Between timesteps** 330 **and** 660**, the** *unknown* **attacker prevails (2)** (c, d) **Strategic attackers resort to the most adverse strategy based on defender policy estimated till current timestep.**

Defense Methods Compared: Our approach, which we call 'Adaptive Threat-Aware Factored MDP', is compared to methods that use a similar amount of prior information: (1) A bandit-based approach ($FPL-MTD$) [12]: This has proved as a strong baseline amongst approaches that do not consider prior knowledge regarding attackers, (2) $EPS-GREEDY$ approach: Epsilon-greedy based exploration-exploitation and (3) Uniform Random Strategy (URS).

In subfigures (a, c), we compare the reward among the methods, and subfigures (b, d) present the evolution of defender rewards over 1000 timesteps. In an evolving attack landscape where between timesteps 330 and 660, the *unknown* attacker prevails (Fig. 2a, 2b), our approach performs 13%, 34%, and 56% better than $FPL-MTD$, $EPS-GREEDY$, and URS, respectively. Our approach quickly adapts to the new scenario and is able to switch to a favorable configuration, maintaining a high defender reward. In an evolving attack landscape with strategic attackers (Fig 2c, 2d), our approach performs 15%, 30%, and 52% better than $FPL-MTD$, $EPS-GREEDY$, and URS, respectively. It consistently outperforms other methods across the 1000 timesteps.

REFERENCES

[1] Harold Booth, Doug Rike, and Gregory Witte. 2013. The national vulnerability database (nvd): Overview. (2013).

[2] Jin-Hee Cho, Dilli P Sharma, Hooman Alavizadeh, Seunghyun Yoon, Noam Ben-Asher, Terrence J Moore, Dong Seong Kim, Hyuk Lim, and Frederica F Nelson. 2020. Toward proactive, adaptive defense: A survey on moving target defense. *IEEE Communications Surveys & Tutorials* 22, 1 (2020), 709–745.

[3] Carlos Guestrin, Daphne Koller, Ronald Parr, and Shobha Venkataraman. 2003. Efficient solution algorithms for factored MDPs. *Journal of Artificial Intelligence Research* 19 (2003), 399–468.

[4] Sushil Jajodia, Anup K Ghosh, Vipin Swarup, Cliff Wang, and X Sean Wang. 2011. *Moving target defense: creating asymmetric uncertainty for cyber threats*. Vol. 54. Springer Science & Business Media.

[5] Charles A Kamhoua, Christopher D Kiekintveld, Fei Fang, and Quanyan Zhu. 2021. *Game theory and machine learning for cyber security*. John Wiley & Sons.

[6] Henger Li, Wen Shen, and Zizhan Zheng. 2020. Spatial-Temporal Moving Target Defense: A Markov Stackelberg Game Model. In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems* (Auckland, New Zealand) *(AAMAS '20)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 717–725.

[7] Henger Li and Zizhan Zheng. 2023. Robust Moving Target Defense Against Unknown Attacks: A Meta-Reinforcement Learning Approach. In *Decision and Game Theory for Security: 13th International Conference, GameSec 2022, Pittsburgh, PA, USA, October 26–28, 2022, Proceedings* (PIttsburgh, PA, USA). Springer-Verlag, Berlin, Heidelberg, 107–126. https://doi.org/10.1007/978-3-031-26369-9_6

[8] Peter Mell, Karen Scarfone, and Sasha Romanosky. 2006. Common vulnerability scoring system. *IEEE Security & Privacy* 4, 6 (2006), 85–89.

[9] Kevin Patrick Murphy. 2002. *Dynamic bayesian networks: representation, inference and learning*. University of California, Berkeley.

[10] Sailik Sengupta and S. Kambhampati. 2020. Multi-agent Reinforcement Learning in Bayesian Stackelberg Markov Games for Adaptive Moving Target Defense. *ArXiv* abs/2007.10457 (2020), 9. https://api.semanticscholar.org/CorpusID:220665563

[11] Sailik Sengupta, Satya Gautam Vadlamudi, Subbarao Kambhampati, Adam Doupé, Ziming Zhao, Marthony Taguinod, and Gail-Joon Ahn. 2017. A Game Theoretic Approach to Strategy Generation for Moving Target Defense in Web Applications.. In *AAMAS*, Vol. 1. International Foundation for Autonomous Agents and Multiagent Systems, 178–186.

[12] Vignesh Viswanathan, Megha Bose, and Praveen Paruchuri. 2022. Moving Target Defense under Uncertainty for Web Applications. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 1750–1752.

[13] Tarun Yadav and Arvind Mallari Rao. 2015. *Technical Aspects of Cyber Kill Chain*. Springer International Publishing, 438–452. https://doi.org/10.1007/978-3-319-22915-7_40