SCIENCE CHINA
Information Sciences

• LETTER •

November 2024, Vol. 67, Iss. 11, 219205:1–219205:2
https://doi.org/10.1007/s11432-024-4143-4

Optimal noise injection energy allocation and collection of energy harvesting attacker under quality-changeable environment

Lu DONG[1,2*], Tao HAN[1], Xin YUAN[3] & Chao DENG[4]

[1]*School of Cyber Science and Engineering, Southeast University, Nanjing* 211189*, China;*
[2]*Engineering Research Center of Blockchain Application, Supervision and Management, Southeast University, Ministry of Education, Nanjing* 211189*, China;*
[3]*School of Automation, Southeast University, Nanjing* 210096*, China;*
[4]*Institute of Advanced Technology, Nanjing University of Posts and Telecommunications, Nanjing* 210003*, China*

Received 8 May 2024/Revised 17 July 2024/Accepted 30 August 2024/Published online 18 October 2024

Advanced persistent threat (APT) is a powerful attack pattern in which attackers lurk in the system, waiting for opportune moments and launching attacks. Remote state estimation (RSE) is an essential scenario in the cyber-physical system (CPS). The attacker's objective is to take advantage of the vulnerability in the RSE system, expanding the estimation error of the remote estimator (RE). However, the lurking demand compels the attacker to consider harvesting energy from an external quality-changing environment, which also restricts the attacker. It is of great military interest to manage attack energy in an unreliable environment and with limited resources.

This problem usually faces the following challenges. First, due to the long-term lurking demand of attackers, wired power supply methods cannot be applied because they are easily detected. The attacker must obtain energy from the external environment. Second, the external environment is changeable. The attacker knows the real-time quality of the environment, i.e., the specific amount of energy that can be harvested. Third, the battery limitations. The attacker carries an energy-harvesting battery with limited capacity to store and release energy. Due to the different working frequencies of energy harvesting and attacking, the two actions cannot be performed simultaneously. Fourth, the error covariance of RE, the environmental quality, and the remaining energy in the battery are all continuous. It is difficult to determine the existence of an optimal solution, obtain the optimal solution, and explore the structural properties of the optimal solution.

Efforts have been made to address the challenges mentioned above. External energy such as radio frequency, solar, thermal, and flow can be harvested, which guarantees the wireless power supply [1]. The external environment transfer laws can be established as a Markov chain or Markov process by prior statistical information. Methods such as the Markov decision process (MDP) framework and deep learning algorithms can be used to solve the optimal strategy and its structure properties [2,3].

This study aims to find the optimal stationary energy allocation and collection strategy over an infinite time horizon so that the average estimation error of the RE is maximized. The key contributions of this paper are summarized as follows. First, an environment quality state conforming to a Markov process is introduced, and it is related to the efficacy of energy harvesting. The introduction of environmental quality makes the problem more realistic and enhances the granularity of strategies from specific environments. Second, the optimal energy allocation and collection problem is transformed into an MDP problem. The existence and the monotonic non-decreasing structural properties of the optimal stationary strategy are proven.

Problem formulation. Consider a discrete time-invariant linear physical process similar to [4]

$$x(k+1) = Ax(k) + w(k), \tag{1}$$

$$y(k) = Cx(k) + v(k), \tag{2}$$

where $k \in \mathbb{Z}_+$ refers to the time step, $x(k) \in \mathbb{R}^n, y(k) \in \mathbb{R}^m$ are the state of the physical process and the observation of the sensor. It is assumed that $w(k)$ and $v(k)$ are independent and identically distributed (i.i.d.) and zero-mean white Gaussian noise with covariance $Q \geqslant 0$ and $R > 0$, and pairs $(A, \sqrt{Q})$ and (A, C) are stabilizable and observable, respectively.

The communication channel is an unreliable additive white Gaussian noise (AWGN) channel with 4-quadrature amplitude modulation (4-QAM) and cyclic redundancy check (CRC). The symbol signal-to-interference-plus-noise ratio (SINR) is the ratio between transmission signal power and all interface signals at RE. The random variable $\gamma(k)$ equal to 1 or 0 indicates whether to receive a data packet without error at time k. The probability of correctly receiving an L bits sequence is given by

$$\Pr[\gamma(k) = 1] = (1 - 2\mathcal{Q}(\sqrt{\alpha \mathrm{SINR}}))^{\lceil \frac{L}{2} \rceil}, \tag{3}$$

where α is a constant, $\lceil \cdot \rceil$ is a mathematical function used for ceiling rounding and $\mathcal{Q}(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp\left(-\delta^2/2\right) \mathrm{d}\delta$. The

* Corresponding author (email: ldong90@seu.edu.cn)

specific transmission process of the RSE system and the attacker's interference details can be obtained in Appendix A. The updating rules of the remaining energy of battery $b(k)$ and environmental quality $q(k)$ are also included. Then, we can give the origin problem.

Problem 1 (Origin problem)**.** In the RSE system, the APT attacker collects or schedules its energy to maximize the average error of RE under a limited battery and quality-changeable environment.

$$\max_{\pi} J(\pi) = \lim_{K\to\infty} \inf \frac{1}{K}\mathbb{E}\left[\sum_{k=1}^{K} \mathrm{Tr}(P(k))\right], \tag{4}$$

$$\text{s.t. } 0 \leqslant \theta(k) \leqslant b(k), \tag{5}$$

where π is a stationary policy of the attacker, $\mathrm{Tr}(\cdot)$ is the trace of a matrix and $P(k)$ is the error covariance of RE at the time k.

Optimal injection noise power schedule. The origin problem can be formulated as an MDP problem. First, the MDP framework is defined as a tuple $\{\mathbb{S}, \mathbb{A}, \mathbb{P}\{\cdot|\cdot,\cdot\}, r(\cdot|\cdot)\}$ and the details are provided in Appendix B. Then we obtain the following MDP problem.

Problem 2 (MDP problem)**.** Given initial state $s(1) \in \mathbb{S}$, we need to find an optimal policy $\pi^\star \in \Pi$ to maximize $J(s,\pi) = \lim\limits_{K\to\infty} \inf \frac{1}{K}\mathbb{E}[\sum\limits_{k=1}^{K} r(s(k),\theta(k))]$. That is

$$J(s,\pi^\star) = \max_{\pi\in\Pi} J(s,\pi). \tag{6}$$

Let $V_\pi : \mathbb{S} \to \mathbb{R}$ denote the state value function for a given policy $\pi \in \Pi$. The Bellman equation $J^\star(s) + V_\pi(s) = \max_{\theta\in\mathbb{A}(s)} \{r(s,\theta) + \sum_{s'\in\mathbb{S}} p(s'|s,\theta)V_\pi(s')\}$ and related iteration algorithms can be utilized to calculate the optimal value and action $J^\star(s), \theta^\star(s) \forall s \in \mathbb{S}$. According to [5], the following theorem can be derived in Appendixes C.1–C.3.

Theorem 1. For any fixed $b \in \mathbb{B}$ and fixed $q \in \mathbb{Q}$, there exists a limit inferior average optimal stationary policy $(\pi^\star)^\infty$ with the property that $(\pi^\star)^\infty$ is non-decreasing in τ and satisfies

$$J(s,\pi^\star) \geqslant J(s,\pi), \forall s \in \mathbb{S}, \pi \in \Pi. \tag{7}$$

Theorem 1 proves the existence and structural properties of the optimal energy harvesting and allocation stationary strategy for an attacker under a quality-changeable environment. Due to the continuity of the state space, acquiring the optimal policy requires the discretization of the state space according to the accuracy needs and the structural property. Moreover, the optimal structural property remains if a state dimension is unaffected by other state dimensions and actions (Appendix C.4).

Simulation example. A numerical example is developed to illustrate the structural property of the optimal noise injection energy allocation and collection strategy. The detailed parameter setting is provided in Appendix D, and the optimal policy is calculated by MDPtoolbox in MATLAB.

Figure 1 illustrates the optimal strategy with environmental quality transition matrix set to Q_5 in Appendix D. The horizontal axis τ represents the number of consecutive failed packet receptions up to the current time. We choose two scenarios with the remaining energy of the battery $b = 6$ and environmental quality states $q = 5$, respectively. Then, the relationship between optimal strategies and the rest states is given. It can be observed from the subfigures that given the environment quality q and the remaining energy of battery b, the optimal energy $\theta^\star(s)$ is non-decreasing with the increasing estimation error of the RE $P = h^\tau(\bar{P})$, thereby validating Theorem 1. Moreover, the optimal average reward $J(s,\pi^\star)$ is a critical metric for evaluating the attacker's strategy. We define 5 different transition matrices of environment quality in Appendix D and present their $J(s,\pi^\star)$. It is observed that the optimal average reward is significantly influenced when the environment transition matrices have a preference for a good-quality environment Q_3 or a bad-quality environment Q_4.

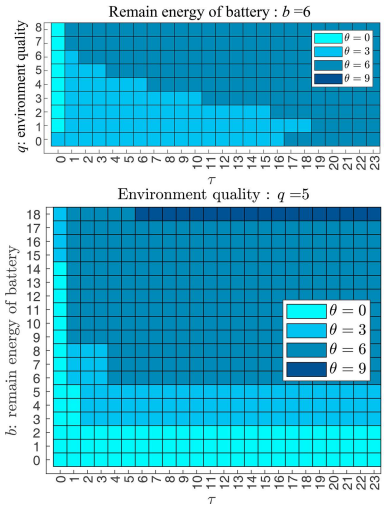

Figure 1 Optimal stationary policies given $b = 6$ and $q = 5$.

Conclusion. This study investigates the optimal energy allocation and collection strategy of an active attacker in an RSE system. The APT attacker is powered by an energy-harvesting battery in a quality-changeable environment, aiming to maximize the average estimation error of the RE. We transform the original problem into an MDP problem and prove that the optimal stationary policy exists and exhibits structural properties, extending it to more general situations, thereby reducing the complexity of this problem.

Acknowledgements This work was supported by the National Natural Science Foundation of China (Grant Nos. 62173251, 62203113).

Supporting information Appendixes A–D. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Gu X F, He L P, Wang S J, et al. Linear reciprocating motion energy harvester systems based on multiple operation modes: a review. J Electron Mater, 2023, 52: 2919–2931
2 You Q B, Ying C Y, Zhou X N, et al. Understanding adversarial attacks on observations in deep reinforcement learning. Sci China Inf Sci, 2024, 67: 152104
3 Peng L H, Cao X H, Sun C Y. Optimal transmit power allocation for an energy-harvesting sensor in wireless cyber-physical systems. IEEE Trans Cybern, 2021, 51: 779–788
4 Zhang S Z, Peng L H, Chang X Y, et al. Optimal energy allocation based on SINR under DoS attack. Neurocomput, 2024, 570: 127116
5 Puterman L M. Markov Decision Processes: Discrete Stochastic Dynamic Programming. New York: John Wiley & Sons, 2014