

# Vulnerabilities Caused by Metric-based Policies in Reinforcement Learning Based Covert Communication Under Steering Attack

Alyse M. Jones  
alysemjones@vt.edu  
Virginia Tech  
Blacksburg, Virginia, USA

Maice Costa  
mcosta@nexcepta.com  
Nexcepta, Inc.  
Gaithersburg, Maryland, USA

## Abstract

This paper explores the concept of timeliness in covert communications when faced with eavesdropping and jamming. We consider a transmitter-receiver pair communicating over a wireless channel where the choice of a resource block (frequency, time) to transmit is the result of a Reinforcement Learning policy. The eavesdropper aims to detect a transmission to perform a steering attack. Using two multiarmed bandit systems, we investigate the problem of minimizing the Age of Information (AoI) regret at the legit receiver, while maximizing the AoI regret at the adversary. We present an upper bound for regret and demonstrate through simulations the validity of the bound and the vulnerabilities introduced by the use of metric-guided policies such as age-aware policies.

## CCS Concepts

• Computing methodologies → Adversarial learning.

## Keywords

Age of Information, timeliness, status updates, multi-armed bandits, eavesdropping, covert communications

### ACM Reference Format:

Alyse M. Jones and Maice Costa. 2025. Vulnerabilities Caused by Metric-based Policies in Reinforcement Learning Based Covert Communication Under Steering Attack. In *Proceedings of the 2025 ACM Workshop on Wireless Security and Machine Learning (WiseML 2025)*, July 3, 2025, Arlington, VA, USA. ACM, New York, NY, USA, 6 pages. <https://doi.org/10.1145/3733965.3733973>

## 1 Introduction

This paper aims to **contribute to the characterization of the utility of information and the effect of timeliness in a decision process using Reinforcement Learning (RL) in an adversarial environment**. The importance of advancing the theoretic understanding in this topic is exacerbated by the prevalence of data-based processes and the integration of computation, communication, and control tasks in several areas, such as the Internet of Things (IoT), Cyber Physical Systems (CPS), and Intelligent Autonomous Systems (IAS). Having **relevant and timely information is paramount for effective decision making**, including but not limited to those applications bringing the latest technology to the battlefield, hence our focus on scenarios with the presence of an adversary.

The starting point is the concept of *Age of Information (AoI)*, defined as the time since the most recent data point was generated. AoI has received a great deal of attention in the past ten years (see [20] for an overview), and it remains **relevant to applications that require fresh information**. Although prior work has considered RL to optimize AoI (e.g., [2]), the effect of outdated information on the learning process of RL has received little attention, particularly when in the presence of an intelligent adversary.

We consider two inter-dependent intelligent systems modeled as multi-armed bandits, with one representing the primary user of the communication resources and the other representing the adversary. We define an age-based regret function for this coupled system and present an AoI-based algorithm where a transmitter aims to optimize its communication in the presence of an intelligent adversary. We provide an upper bound on the regret as a function of the number of suboptimal decisions during a sequence of time slots. Results show that non-adversaries using age-aware policies in the presence of an adversary can make it easier for the adversary to manipulate the non-adversaries decision-making, leading to less regret for the adversary and better steering attack results.

## 2 Related Work

The pioneer work in [13] introduced the concept of AoI. Initial work on AoI [13, 14, 19] focused on queuing models to establish new performance metrics for timeliness. Work by Ephremides et al. contributed to those initial steps, helping to sparkle interest from the wireless network research community [1, 10, 11]. The applications that require such timeliness metrics are numerous; refer to

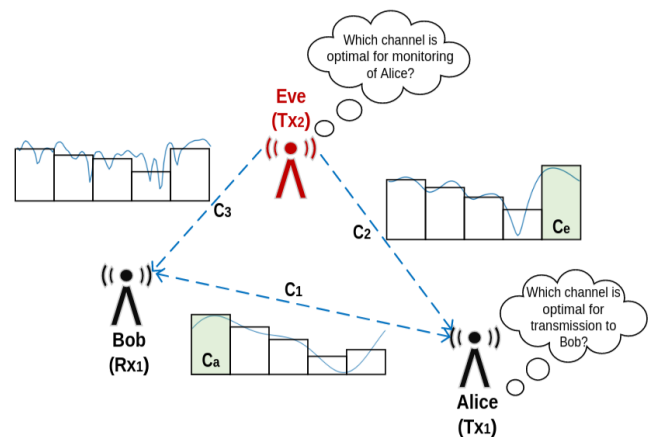


Figure 1: Covert communication model from [9] demonstrating Alice attempting to communicate to Bob on her best channel and the adversary Eve attempting to detect and steer Alice to her best channel for observing Alice.



This work is licensed under a Creative Commons Attribution 4.0 International License. *WiseML 2025, Arlington, VA, USA*

© 2025 Copyright held by the owner/author(s).  
ACM ISBN 979-8-4007-1531-0/2025/06  
<https://doi.org/10.1145/3733965.3733973>

[20] for an introduction and survey. This shift to considering information timeliness also led to a new paradigm of communications that accounts for the *quality* of information, as opposed to *quantity* in the traditional approach [8, 16].

The impact of hostile interference on AoI was first addressed in [3], where the interaction is formulated as a non-zero-sum two player game to determine the transmission and interference power levels. The work in [18] proposed a dynamic game to study the selection of transmission times, focusing on medium access for the definition of utilities. The work in [7] extended [3] for the case with background noise, presenting the Nash equilibrium strategies as a function of the updating rate, and a Stackelberg equilibrium with the transmitter as a leader. Channel access and scheduling for AoI-focused transmissions in adversarial environments have also been considered in [4] and [17], which show that AoI can be minimized against adversaries.

Adversarial environments can contain a number of attacks, from jamming to eavesdropping. Of interest here is a ML-based attack called a *steering attack*, a form of policy manipulation where an intelligent adversary aims to direct a non-adversary towards a target policy through direct manipulation of the environment [9]. The algorithm in [9] showed that steering attacks in wireless communications are possible and should be defended against as it does not take many directed, intelligently crafted transmissions from an adversary to manipulate a non-adversaries learning policy. It is unclear how AoI affects the performance of a steering attack and should be studied for proper defense as more systems employ AoI.

## 3 System Model

### 3.1 Communication Model

We consider a transmitter (Alice) sending time-sensitive information to a receiver (Bob) in a hostile environment where an active adversary (Eve) can potentially eavesdrop and interfere with the attempted communication. Eve aims to steer Alice's policy to her target channel  $e^*$ , which is the channel through which Eve can best detect Alice. We assume that communication takes place using fixed and independent resource blocks, and Alice selects a resource block by selecting the transmit frequency at each time slot.

### 3.2 Adversary Model

We assume that the ill intended eavesdropper (Eve) is constantly listening to channels to detect ongoing communication in the link between Alice and Bob. While Alice selects a resource block to transmit by selecting the frequency, Eve selects a frequency to tune in and listen. We assume that channel propagation effects such as noise and fading may prevent Eve from detecting a transmission between Alice and Bob, an effect known as the *hidden node*. Eve's objective is to detect the transmission and steer Alice to transmit in the channel where its probability of detection is higher. The algorithm by which steering is performed without considering AoI, is presented in Algorithm 2 [9, Algorithm 1].

### 3.3 Status Updating Model

We consider a just-in-time model, where messages are generated immediately before transmission. In this case, there is no queuing of packets waiting for transmission. We assume Alice will not

---

#### Algorithm 1 Steering Attack Algorithm with Age-Aware Alice

---

**Result:**  $k_a^* \leftarrow$  optimal action from Alice's learned policy

```

1: for  $t = 1 : T$  do
2:    $limit(t) = \min_{k \in K} \frac{1}{\hat{\mu}_k(t-1)}$ 
3:   if  $aoi\_alice(t-1) > limit$  then
4:     Alice chooses her best channel  $a^*(t) = argmax(\hat{\mu})$ 
5:   else if  $aoi\_alice(t-1) \leq limit$  then
6:     Alice transmits using TS with Normal-Gamma prior
7:   end if
8:   Use Algorithm 2 for Eve and policy updates
9:   if Alice's transmission is successful then
10:     $aoi\_alice(t) = 1$ 
11:   else
12:     $aoi\_alice(t) = aoi\_alice(t-1) + 1$ 
13:   end if
14:   if Eve's attack is successful then
15:     $aoi\_eve(t) = 1$ 
16:   else
17:     $aoi\_eve(t) = aoi\_eve(t-1) + 1$ 
18:   end if
19: end for
20: Calculate  $R(T)$ 

```

---



---

#### Algorithm 2 Eve's steering attack against Alice from [9]

---

**Initialize:** Randomly generate  $C_1, C_2,$  and  $C_3$  with frequency response  $H_{a \rightarrow b}, H_{a \rightarrow e}, H_{e \rightarrow b}$

```

1: for  $t = 1 : T$  do
2:   Decide Alice's action through Thompson Sampling with Normal-Gamma prior
3:   Eve decides best channel for detection of Alice  $e^*(t)$  through MAB policy
4:   Eve estimates Alice's next move  $\hat{a}_{alice}(t)$  based on spectrum sensing detection
5:   Decide Eve's action  $a_e(t)$ 
6:   if  $\hat{a}_{alice}(t) = e^*(t)$  then
7:      $a_e(t) = N + 1 \leftarrow$  Eve listens
8:   else if  $\hat{a}_{alice}(t) \neq e^*(t)$  then
9:      $a_e(t) = \hat{a}_{alice}(t) \leftarrow$  Eve probes (i.e. attacks)
10:  end if
11:   $r_{alice}(t) = \frac{|H_{a \rightarrow b}|}{|H_{e \rightarrow b}|}$ , where  $|H_{e \rightarrow b}| = P_{eve} * |H_{e \rightarrow b}|$  if Eve is present
12:   $r_{eve}(t) = |H_{a \rightarrow e}|$ 
13:  Update MAB policies for Alice and Eve
14: end for

```

---

hold packets if they are ready to be transmitted, meaning that Alice is only silent when she has no packets to transmit. A packet transmission has fixed duration as determined by the system's resource block size. The decision to transmit a packet using a given channel depends on the policy that Alice is learning and optimizing. See Algorithm 1 for how Alice uses AoI in her policy to optimize her communication with Bob. When Alice sees that AoI is low based on her policy, she acts greedily and chooses to transmit on her current best channel to increase her chances of completing transmission to Bob. Otherwise, Alice explores through Thompson Sampling (TS). Meanwhile, Eve is also performing her steering attack.

## 4 Performance Analysis

### 4.1 Communication Timeliness

We denote with  $S_B(t)$  and  $S_E(t)$  the indicators of successful transmissions in the two links, Alice-Bob and Alice-Eve, respectively. The AoI increases by one unit when the transmission is not successful, and it is reset to one when the transmission is successful. We write the AoI at Bob and at Eve, respectively as

$$\begin{aligned} a_B^\pi(t) &= (1 - S_B(t))(a_B^\pi(t-1) + 1) + S_B(t) \\ a_E^\pi(t) &= (1 - S_E(t))(a_E^\pi(t-1) + 1) + S_E(t) \end{aligned} \quad (1)$$

This process can be described as a Discrete Time Markov Chain (DTMC) [6]. At state  $k$  the chain transition to state  $k+1$ , when no packet is received, or to state 1, when a packet is received and AoI is updated. The steady state distribution is  $p_k^{\Delta_i} = (1 - S_i)^{k-1} S_i$ , for all  $k$ , where  $i \in \{B, E\}$ . The average AoI is calculated as

$$\bar{\Delta}_i = \sum_{k=1}^{\infty} k(1 - S_i)^{k-1} S_i = \frac{S_i}{1 - S_i} \sum_{k=1}^{\infty} k(1 - S_i)^k. \quad (2)$$

As a result, we have  $\bar{\Delta}_i = 1/S_i$ . Since packets are not ‘aging’ in a queue, Alice would generate and transmit a packet in every resource block. However, introducing some uncertainty to confuse the adversary has advantages to Alice, and we are interested in better understanding how Alice can make decisions with these conflicting objectives, namely to deliver fresh information to Bob and to evade the attacks from Eve. We discuss this in Section 5.

### 4.2 Age Regret

Our system model is composed of two multi-armed bandit (MAB) systems coupled. One MAB represents the link between Alice and Bob, indicated with sub-index  $\{b, B\}$ , while the other MAB represents the link between Alice and Eve, indicated with  $\{e, E\}$ . We refer to the selection of a channel for transmission as the choice of a bandit arm. We assume that Alice should select a policy that will minimize the AoI at Bob, while maximizing the AoI at Eve.

We denote with  $t$  the current time slot, and  $T$  the total number of time slots in a sequence. The bandit arms are identified with index  $k \in \{1, \dots, K\}$ . Each arm is associated to a success probability  $\mu_k$ , and we define  $\mu_{min} = \min_k \mu_k$  and  $\mu^* = \max_k \mu_k$ . We identify the best arm with  $k^* = \arg \max_k \mu_k$ . Let  $k(t)$  be the arm selected at time slot  $t$  and  $\pi$  denote the policy that determines the selection of arms throughout the sequence of length  $T$ . We defined the regret as the difference between the age-regret for the communication link between Alice and Bob and the age-regret for the link between Alice and Eve, as follows.

*Definition 1. Age Regret for Covert Communication.* For a given policy  $\pi$ , the AoI at Bob at time  $t$  is denoted with  $a_B^\pi(t)$ , while the AoI at Eve at time  $t$  is denoted with  $a_E^\pi(t)$ . The AoI under the optimal (genie) policy is denoted with  $a_B^*(t)$  and  $a_E^*$  at Bob and Eve, respectively. We define Alice’s Age regret for a fixed interval of duration  $T$  time slots, as

$$R(T) = R_B(T) - R_E(T), \quad (3)$$

(I)	$k$ Identical	1	$k_2$ Worst	$k_1 - 1$ Best
(II)	$k$ Identical	$k_2$ Worst	$k_1$ Best	

**Figure 2:** [5, Lemma 3] proves that re-ordering optimal time slots to the end of a sequence can only increase average AoI.

where

$$R_B(T) = \sum_{t=1}^T \mathbb{E} [a_B(t) - a_B^*(t)], \quad (4)$$

$$R_E(T) = \sum_{t=1}^T \mathbb{E} [a_E(t) - a_E^*(t)]. \quad (5)$$

To define the bounds to the AoI regret, we first identify in the following lemmas the bounds in literature that have been applied to the case of a single multi-armed bandit system.

**LEMMA 1.** *Number of draws of suboptimal arms [12, Theorem 2]: Let  $N_k(T)$  be the number of times the suboptimal arm  $k$  is drawn during a sequence of  $T$  time slots. Then for  $t > K$ ,*

$$\mathbb{E} [N_k(T)] \leq \mathcal{O}(K \log T). \quad (6)$$

**LEMMA 2.** *An upper bound on cumulative AoI [5, Lemma 3, Lemma 4]: Let  $N(T)$  be the total number of suboptimal draws during the sequence of length  $T$ , and  $\mathbb{E}[N(T)]$  its expectation. We define  $\mu^* = \max_k \mu_k$  and  $\mu_{min} = \min_k \mu_k$ . Reordering the sequence of actions so that optimal selections are last, as shown in 2, can only increase the expected accumulated AoI during the sequence. Furthermore, the cumulative expected AoI can be bounded as a function of  $N(T)$ ,*

$$\sum_{t=1}^T \mathbb{E} [a(t)] \leq f(T, \mu^*, \mu_{min}) + g(\mu^*, \mu_{min})N(T), \quad (7)$$

where the two functions acting as coefficients are

$$f(T, \mu^*, \mu_{min}) = \frac{T}{\mu^*} + \frac{1 - \mu^*}{\mu^* \mu_{min}} \quad (8)$$

$$g(\mu^*, \mu_{min}) = \frac{1}{\mu_{min}} + \frac{1}{\mu^*} \quad (9)$$

Given the definition of AoI regret in (3), and the bounds in Lemma 1 and Lemma 2, the upper bound for our system of two coupled MABs is stated in the following theorem.

**THEOREM 2.** *Let  $N_B(T) = \sum_{b \neq b^*} N_b(T)$  denote the total number of suboptimal selections in the link Alice-Bob during a sequence of  $T$  slots, and  $N_E(T)$  denote the analogous for the link Alice-Eve. Note that selection of one arm in the multi-armed bandit system representing the link to Bob automatically determined the arm at the system representing the link to Alice. We also define the gaps to the best bandit arm,  $\Delta_{max} \mu^* - \max_{k \neq k^*} \mu_k$ , and  $\Delta_{min} \mu^* - \min_{k \neq k^*} \mu_k$ . The regret  $R(T)$  defined as the difference in (3) is upper bounded as*

$$\begin{aligned} R(T) &\leq f(T, \mu_{B^*}^*, \mu_{min,B}) \mathbb{E} [N_B(T)] \\ &\quad + g(\mu_{B^*}^*, \mu_{min,B}) \Delta_{min,B} \mathbb{E} [N_B(T)]^2 \\ &\quad - \frac{T \Delta_{max,E}}{\mu_E^*} \mathbb{E} [N_E(T)]. \end{aligned} \quad (10)$$

PROOF. We use the equations describing the evolution of AoI in (1) to write the AoI regret as

$$R(T) = \sum_{t=1}^T \mathbb{E} [(1 - S_B(t))a_B(t-1) - (1 - S_B^*(t))a_B^*(t-1)] - \sum_{t=1}^T \mathbb{E} [(1 - S_E(t))a_E(t-1) - (1 - S_E^*(t))a_E^*(t-1)]. \quad (11)$$

In the first summation term in (11), we replace the AoI under the genie policy with the AoI under the policy  $\pi$ , and the result can only be larger, as the genie policy has the smallest possible expected value for AoI. For the second summation, we replace the AoI with the optimal value under the genie policy, and the resulting term can only be smaller. Combining the two terms, we have the inequality

$$R(T) \leq \sum_{t=1}^T \mathbb{E} [(1 - S_B(t))a_B(t-1) - (1 - S_B^*(t))a_B(t-1)] - \sum_{t=1}^T \mathbb{E} [(1 - S_E(t))a_E^*(t-1) - (1 - S_E^*(t))a_E^*(t-1)]. \quad (12)$$

We rearrange the terms to write

$$R(T) \leq \sum_{t=1}^T \mathbb{E} [(S_B^*(t) - S_B(t))a_B(t-1)] - \sum_{t=1}^T \mathbb{E} [(S_E^*(t) - S_E(t))a_E^*(t-1)]. \quad (13)$$

We use the equivalent system as described in [15]. Define  $\{U(t)\}_{t \geq 0}$  as independent and identically distributed random variables, uniformly distributed in  $(0, 1)$ . With this definition,  $\mathbb{E} [(S^*(t) - S(t))] = \sum_{k \neq k^*} \mathbb{1}\{k(t) = k\} \mathbb{P}(\mu_k \leq U(t) \leq \mu^*)$ , where  $\mathbb{1}\{\cdot\}$  is an indicator function [5]. Using the independence between  $a(t-1)$  and the result of the selected arm at time  $t$ , we rewrite the upper bound as

$$R(T) \leq \sum_{t=1}^T \mathbb{E} [(a_B(t-1)) \times \sum_{b \neq b^*} (\mu_b^* - \mu_b) \mathbb{P}(\mathbb{1}\{b(t) = b\} = 1)] - \sum_{t=1}^T \mathbb{E} [(a_E^*(t-1)) \times \sum_{e \neq e^*} (\mu_e^* - \mu_e) \mathbb{P}(\mathbb{1}\{e(t) = e\} = 1)]. \quad (14)$$

Note that the AoI under the genie policy is a geometric random variable. Replacing  $\mu_b$  with the minimum over all suboptimal arms  $b \neq b^*$ , and replacing  $\mu_e$  with the maximum over all suboptimal arms  $e \neq e^*$ , and using previously defined  $\Delta_{min}$ ,  $\Delta_{max}$ ,  $N_B(T)$ , and  $N_E(T)$  we have an upper bound as

$$R(T) \leq \left( \sum_{t=1}^T \mathbb{E} [(a_B(t-1))] \right) \Delta_{min,B} \mathbb{E}[N_B(T)] - \frac{T}{\mu_E^*} \Delta_{max,E} \mathbb{E}[N_E(T)]. \quad (15)$$

Using the result in (7), we obtain the desired result in (10).  $\square$

## 5 Numerical Results

All simulations were performed with  $T = 1000$  steps and  $n = 10000$  Monte Carlo trials. Each link was modeled as a Rayleigh frequency fading channel, as set up in [9], discretized into  $N = 10$  channels,

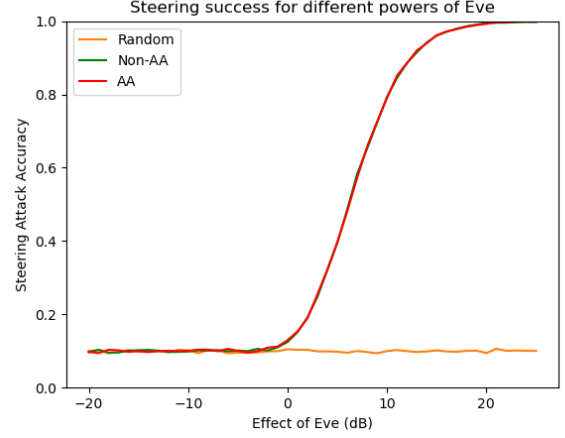


Figure 3: Success probability of Eve’s steering attack against Alice as Eve increases her transmission power.

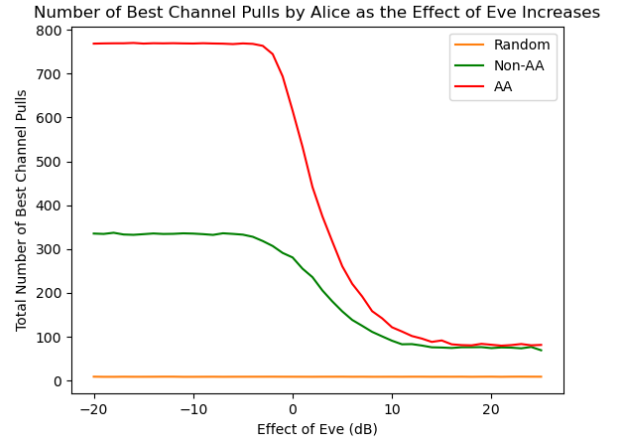


Figure 4: Number of times Alice pulled her best channel  $a^*$  as Eve’s transmission power increases.

and then normalized to be between 0.05 and 0.9. Eve was modeled as an additive barrage jammer that, when transmitting on the same channel as Alice, raised the noise floor at the receiver Bob (in dB). Algorithm 1 was tested for increasing powers of Eve (from -20 dB to 25 dB) and compared against a non-age-aware (non-AA) policy for Alice to demonstrate how an age-aware (AA) Alice differs in performance. A random policy for Alice was also tested as a baseline performance for both the non-AA and AA policies.

Figs. 3-6 show the performance of Alice and Eve when Alice behaves with a random policy, non-AA policy, and AA policy and is under a steering attack from Eve. When Alice adopts the random policy, Eve’s steering attack fails, as shown in Fig. 3. Conversely, when Alice employs a MAB, Eve is able to achieve her steering attack more efficiently, and the attack is successful in both the Non-AA and AA policy cases. However, based on Figs. 4, 5, and 6, the behavior of Alice and Eve differs for the non-AA and AA policies. As shown in Fig. 4, when Alice is AA, AoI is helpful in Alice’s learning process when Eve’s effect (power) is not strong and not effective to Alice’s policy (below 0 dB), as demonstrated by

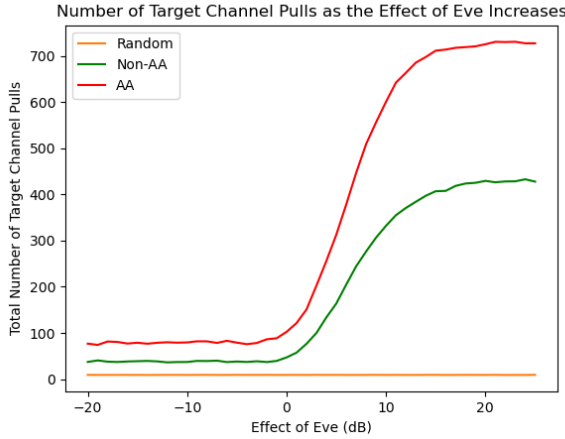


Figure 5: Number of times Alice pulls Eve’s target channel  $e^*$  as Eve’s transmission power increases.

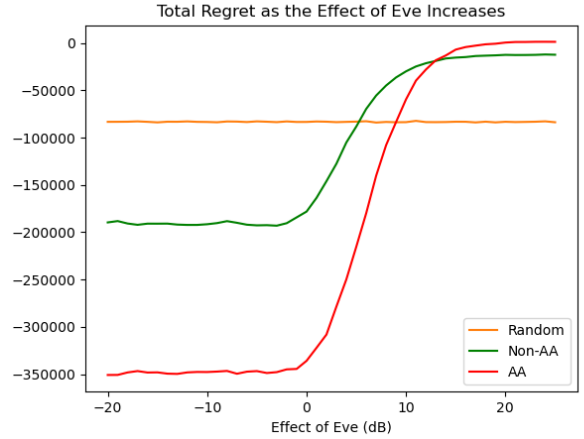


Figure 7: Total regret as Eve increases her transmission power for different policy settings.

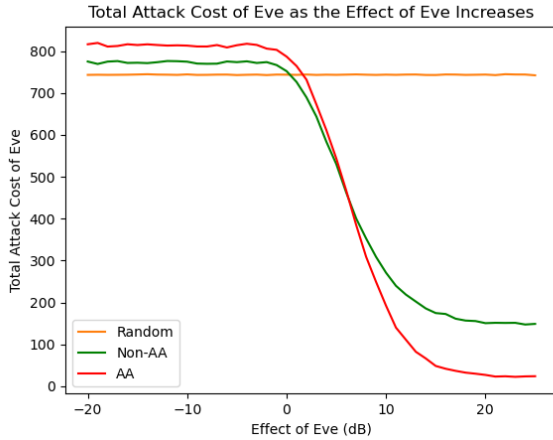


Figure 6: Cost of Eve’s attack as her transmission power increases. Cost is the number of times Eve transmitted to achieve her steering goal.

the 700 plus pulls by Alice of her best channel  $a^*$  as compared to approximately 350 pulls in the Non-AA results. When Eve’s effect is above 0 dB, Figs. 4 and 5 show that Eve can manipulate Alice into favoring the target channel  $e^*$  over her optimal channel  $a^*$ . At high power, Eve forces Alice to pull  $e^*$  nearly twice as often under AA compared to non-AA. Furthermore, Fig. 6 shows that Eve is able to achieve this feat at a lower attack cost, i.e. less number of transmissions, when Alice is AA. Hence, these results show that though Alice improves performance like in [5] under little to no attack, but feature-based policies may be exploited negatively when an attacker is strong enough. The main intuition behind this phenomenon and the reason AA becomes vulnerable when an attacker like Eve is present is because when Alice prioritizes AoI, she is only exploiting and is no longer exploring other possibilities and changes in the environment. When Alice prioritizes AoI in her policy and her AoI is low, she is greedily choosing her best channel according to her policy. With no attacker in the environment or when Eve is not powerful enough, Alice greedily chooses her best

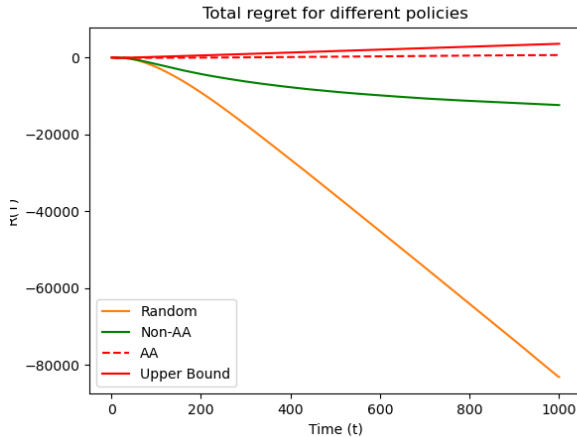
channel  $a^*$ . As Eve becomes more powerful in her attacks, Alice’s policy is manipulated more towards  $e^*$ , fooling Alice in thinking  $e^*$  is  $a^*$ . Thus, Alice will also think choosing  $e^*$  means lower AoI and Alice will select Eve’s target channel more often, missing the opportunity to explore again. This makes it easier for the attacker to perform a steering attack and results in better regret for Eve and worse regret for Alice, as shown in Figs. 7-9.

Fig. 7 shows the total regret as Eve increases transmission power in her steering attack. When Alice is AA, Eve accumulates large regret when her transmission power is under 0 dB because it is not enough to manipulate Alice’s policy. Conversely, Alice achieves lower regret because the extra exploitation caused by the prioritization of AoI causes her to choose her target channel  $a^*$  more often. However, when Eve’s power is high, Eve achieves less regret because Alice is being steered to  $e^*$  and Eve is able to do so with less transmissions (i.e. cost). Alice, consequently, accumulates more regret because she is not achieving her goal in choosing  $a^*$  and is steered to a suboptimal result. Figs. 8 and 9 show a closeup of the regret for Alice and Eve, respectively, over the horizon  $T$  when Eve’s transmission power is 20dB. Alice achieves larger regret when adopting AA versus non-AA policy, while Eve achieves lower regret with AA versus non-AA due to Alice continuously exploiting  $e^*$ .

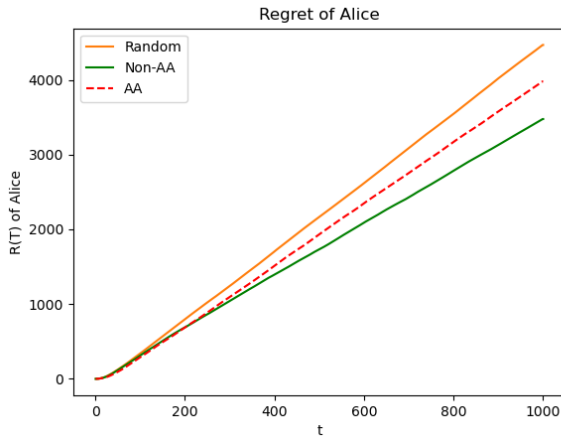
Lastly, Fig. 10 plots the upper bound found in Equation 15 against the Non-AA and AA policies found empirically through Algorithm 1 to verify its validity. As shown, the bound holds since the total regret never exceeds this boundary for the entire horizon  $T$ .

## 6 Conclusion

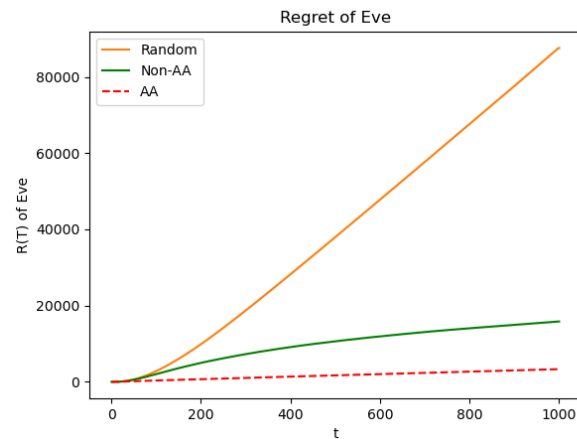
This paper discussed the trade-offs involving communication reliability, timeliness, and stealthiness, characterizing the AoI in a hostile RF environment with an active adversary performing a steering attack. We have defined AoI-based regret functions, and modeled the interactions between the channel user Alice and the adversary Eve as two interdependent multi-armed bandit systems. Alice aims to learn a transmission policy that selects time and frequency blocks to minimize the AoI regret at its receiver Bob. Eve aims to disrupt the communication and steer Alice to transmit in its selected frequency.



**Figure 10: Total regret at a transmission power of 20 dB for Eve for different policy settings. Upper bound is provided to validate Equation 15.**



**Figure 8: Total regret for Alice when Eve has her transmission power at 20dB for different policy settings.**



**Figure 9: Total regret for Eve when Eve has her transmission power at 20dB for different policy settings.**

Through simulation, we compared the Age-Aware (AA) and non-AA policies of Alice under different effects of the adversary Eve. We show that a metric-based policy as AA policy renders the user more susceptible to manipulation by the adversary due to increased predictability of the selected actions. Though metric-based policies can improve performance, it should not come at the cost of security and further research should be performed to investigate how to design robust policies that mitigate this vulnerability. Future work includes investigating how to design a policy for Alice that can be age-aware while simultaneously evading Eve's steering attack.

## References

- [1] Maice Costa, Marian Codreanu, and Anthony Ephremides. 2014. Age of information with packet management. In *2014 IEEE Int. Symp. on Info. Theory*. 1583–1587.
- [2] Dhillon et al. 2020. A Reinforcement Learning Framework for Optimizing Age of Information in RF-Powered Communication Systems. *Trans. on Commun.* 68, 8 (2020), 4747–4760.
- [3] Gam D. Nguyen et al. 2017. Impact of hostile interference on information freshness: A game approach. In *2017 15th Int. Symp. on Model. and Opt. in Mobile, Ad Hoc, and Wireless Netw. (WiOpt)*. 1–7. doi:10.23919/WIOPT.2017.7959909
- [4] Y. Yang et al. 2022. Game-Based Channel Access for AoI-Oriented Data Transmission Under Dynamic Attack. *IEEE Internet of Things Journal* 9, 11 (2022), 8820–8837.
- [5] Santosh Fatale, Kavya Bhandari, Urvidh Narula, Sharayu Moharir, and Manjesh K. Hanawal. 2022. Regret of Age-of-Information Bandits. *IEEE Transactions on Communications* 70, 1 (2022), 87–100.
- [6] Emmanouil Fountoulakis, Themistoklis Charalambous, Nikolaos Nomikos, Anthony Ephremides, and Nikolaos Pappas. 2022. Information freshness and packet drop rate interplay in a two-user multi-access channel. *Journal of Commun. and Netw.* 24, 3 (2022), 357–364.
- [7] Andrey Garnaev, Wuyang Zhang, Jing Zhong, and Roy D. Yates. 2019. Maintaining Information Freshness under Jamming. In *IEEE INFOCOM 2019 - IEEE Conf. on Computer Commun. Workshops (INFOCOM WKSHPs)*. 90–95.
- [8] Tilahun M. Getu, Georges Kaddoum, and Mehdi Bennis. 2023. Making Sense of Meaning: A Survey on Metrics for Semantic and Goal-Oriented Communication. *IEEE Access* 11 (2023), 45456–45492.
- [9] Alyse M. Jones, Harpreet S. Dhillon, and William C. Headley. 2024. A Framing of Eavesdropper Policy Manipulation Attacks on RL-enabled Wireless Systems. In *Proc. IEEE GLOBECOM 2024*.
- [10] Clement Kam, Sastry Kompella, and Anthony Ephremides. 2013. Age of information under random updates. In *2013 IEEE Int. Symp. on Info. Theory*. 66–70.
- [11] Clement Kam, Sastry Kompella, and Anthony Ephremides. 2014. Effect of message transmission diversity on status age. In *2014 IEEE Int. Symp. on Info. Theory*. 2411–2415.
- [12] Emilie Kaufmann, Nathaniel Korda, and Rémi Munos. 2012. Thompson Sampling: An Asymptotically Optimal Finite-Time Analysis. In *Algorithmic Learning Theory*, Nader H. Bshouty, Gilles Stoltz, Nicolas Vayatis, and Thomas Zeugmann (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 199–213.
- [13] Sanjit Kaul, Roy Yates, and Marco Gruteser. 2012. Real-time status: How often should one update?. In *Proceedings IEEE INFOCOM*. 2731–2735.
- [14] Sanjit K. Kaul, Roy D. Yates, and Marco Gruteser. 2012. Status updates through queues. In *2012 46th Annual Conf. on Info. Sciences and Systems (CISS)*. 1–6.
- [15] Subhashini Krishnasamy, Rajat Sen, Ramesh Johari, and Sanjay Shakkottai. 2016. Regret of Queueing Bandits. In *Advances in Neural Info. Proc. Sys.*, D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett (Eds.), Vol. 29. Curran Associates, Inc. [https://proceedings.neurips.cc/paper\\_files/paper/2016/file/430c3626b879b4005d41b8a46172e0c0-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2016/file/430c3626b879b4005d41b8a46172e0c0-Paper.pdf)
- [16] Nived Rajaraman, Rahul Vaze, and Goonwanth Reddy. 2021. Not Just Age but Age and Quality of Information. *IEEE Journal on Sel. Areas in Commun.* 39, 5 (2021), 1325–1338.
- [17] Abhishek Sinha and Rajarshi Bhattacharjee. 2022. Optimizing Age-of-Information in Adversarial and Stochastic Environments. *IEEE Trans. on Info. Theory* 68, 10 (2022), 6860–6880.
- [18] Yuanzhang Xiao and Yin Sun. 2018. A dynamic jamming game for real-time status updates. In *IEEE INFOCOM 2018 - IEEE Conf. on Comp. Commun. Workshops (INFOCOM WKSHPs)*. 354–360.
- [19] Roy D. Yates and Sanjit Kaul. 2012. Real-time status updating: Multiple sources. In *2012 IEEE Int. Symp. on Info. Theory Proc.* 2666–2670.
- [20] Roy D. Yates, Yin Sun, D. Richard Brown, Sanjit K. Kaul, Eytan Modiano, and Sennur Ulukus. 2021. Age of Information: An Introduction and Survey. *IEEE Journal on Sel. Areas in Commun.* 39, 5 (2021), 1183–1210.