Reinforcement Learning for Link Adaptation and Channel Selection in LEO Satellite Cognitive Communications

Muhammad Anjum Qureshi, Eva Lagunas, *Senior Member, IEEE* and Georges Kaddoum, *Senior Member, IEEE*

***Abstract*—In this letter, we solve the link adaptation and channel selection problem in next generation satellite cognitive networks under dynamically varying channel availability and time-varying channel statistics. Primary user (PU) activity in Low Earth Orbit (LEO) satellite cognitive communications forces the set of available transmission channels for a secondary user (SU) to vary dynamically over time. We consider the scenario where the channel state varies in a piecewise-stationary mode, referred to as quasi-static (block-fading) channels. We formalize the problem as a reinforcement learning problem, and propose Discounted Structured and Sleeping Thompson Sampling (dSTS), which maximizes the SU's throughput by selecting the optimum modulation and coding scheme (MCS) and the transmission channel under volatile and piecewise-stationary settings. When channel characteristics are unknown as well as piecewise-stationary, the proposed algorithm adapts the SU's link-rate by exploiting the structure of the transmission success probability in transmission rates over the selected available channel. Furthermore, channel state information (CSI) is absent and feedback is limited to 1-bit (success/failure).**

***Index Terms*—Satellite Communications, link adaptation, channel selection, discounted Thompson Sampling.**

I. Introduction

Low Earth Orbit (LEO) satellite has gained popularity due to its *quality of service* (QoS) in terms of high speed data rate and low transmission latency [1]. However, spectrum scarcity challenges LEO based satellite communications (SatCom) to serve pre-defined licensed users, referred to as *primary users* (PUs). Cognitive Radio Networks (CRNs) in LEO SatCom are able to serve dynamic users under certain conditions i.e., *secondary users* (SUs) can access unused PU channels, or coexist under some interference constraints [2]. Furthermore, Artificial Intelligence (AI)-enabled decision making for resource allocation (e.g., link adaptation referred to as *modulation and coding scheme* (MCS) and/or channel selection in wireless communications [3]–[7]) is a viable solution compared to traditional off-line solutions due to its autonomous cognitive decision making capabilities by interacting with the unknown environment in real-time [8].

For an SU in heterogeneous CRNs, spectrum sharing is often studied along with rate adaptation (also known as *link adaptation*). Various resource allocation strategies based on multi-armed bandits (MABs), a famous reinforcement learning (RL) technique for decision making in stochastic environments, are proposed to maximize the expected reward (also referred to as *throughput*), defined as the number of bits successfully transmitted over the unknown selected wireless channel [3]–[7], [9], [10]. These autonomous learning algorithms were proved to be more efficient than traditional resource allocation algorithms, which makes them a promising solution for future envisioned SatCom (e.g., 5G and 6G).

Link adaptation and channel selection in SatCom poses multiple challenges due to dynamic channel availability (i.e., PU activity) and time-varying channel characteristics (e.g., rain attenuation), and to device an efficient strategy, one needs a combination of variants of MAB and exploitation of inherent properties of the expected rewards. There exists a predefined correlation between the expected rewards and the transmission rates (e.g., the success probability decreases over increasing transmission rates, referred to as *monotonicity*), and exploiting this structural information converges faster to the optimal solution in RL algorithms [10]. Furthermore, *stochastic sleeping (volatile)* MABs [11], a variant of stochastic MAB where available arms vary dynamically over time, is applied in wireless scenarios when resources are not available all the time [7]. Last but not least, the transmission success probabilities vary slowly over time, referred to as *piecewise-stationary environment* in MAB literature [12], and learning algorithms need to be adapted to serve under such scenarios (e.g., [6], [13]).

In this work, we consider the SU's link adaptation and channel selection problem in an unknown and piecewise-stationary environment, where channel characteristics are time-varying and unknown to the SU, and the set of available channels varies over time. We cast the MCS-channel pair selection problem as an on-line learning problem, and propose a learning algorithm called Discounted Structured and Sleeping Thompson Sampling (dSTS), which handles the volatility of the wireless channel set and exploits the correlation between the transmission success probability and the transmission rates in piecewise-stationary setting over channels' gains.

Manuscript received Jan 05, 2023; accepted Jan 24, 2023. This work was supported in part by the Canada Research Chair Program Tier-II entitled "Towards a Novel and Intelligent Framework for the Next Generations of IoT Networks". The work of E. Lagunas has been partially supported by the Luxembourg National Research Fund (FNR) under the project SmartSpace (C21/IS/16193290). (*Corresponding author: Muhammad Anjum Qureshi.*)

Muhammad Anjum Qureshi is with the Department of Electrical and Electronics Engineering, Bilkent University, Ankara 06800 Turkey (e-mail: qureshi@ee.bilkent.edu.tr).

Eva Lagunas is with the Interdisciplinary Centre for Security Reliability and Trust, University of Luxembourg, Luxembourg (e-mail: eva.lagunas@uni.lu).

Georges Kaddoum is with the Department of Electrical Engineering, ETS, Montreal, Canada (e-mail: georges.kaddoum@etsmtl.ca).

This article has been accepted for publication in IEEE Communications Letters. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/LCOMM.2023.3240363

TABLE I: Comparison between the proposed dSTS and earlier works

	Learning Algorithms					
Properties / advantages	**RAGTS** [3]	**KL-UCB** [4]	**G-ORS** [5]	**CD-CoTS** [6]	**V-CoTS** [7]	**dSTS** (This work)
Structure exploitation	×	✓	✓	✓	✓	✓
Thompson Sampling	✓	×	×	✓	✓	✓
Volatile / Sleeping MABs	×	×	×	×	✓	✓
Time-varying channel statistics	✓	✓	✓	✓	×	✓

II. Related Work

Adaptive Modulation and Coding (AMC) is a well-known technique in CRNs and SatCom to adapt to the time-varying channel statistics [14], [15]. The receiver performs link quality estimation in the form of *signal-to-noise-ratio* (SNR) and feeds it back to the transmitter, which in turn selects the adequate MCS for the transmission. In CRNs, *channel state information* (CSI) is also used for adequate MCS selection. However, when SNR information is outdated or there is a limited feedback, it is hard to implement AMC by traditional methods (e.g., SNR based look-up tables) and necessitates AI-enabled learning algorithms to either predict CSI or directly adapt MCS without any CSI [3], [10], [16]. This motivates authors in [4], [5] to formulate the link adaptation and channel/mode selection problem as a MAB problem, and propose a frequentist algorithm based on Kullback-Leibler upper confidence bound (KL-UCB), which maximizes the average number of successfully transmitted packets over a given time horizon. In these works, authors exploited the unimodal structure of the expected rewards over rate-channel pairs. Authors in [9], presented the Bayesian counterpart for link rate selection, and proposed an algorithm based on Thompson Sampling (TS). As an extension to the work in [9], authors in [10] proposed a TS based algorithm which exploits the monotonicity of the success probability over the transmission rates.

In traditional MAB formulation [17], [18], an action is sequentially selected and a stochastic reward, which is drawn from an unknown but fixed stochastic distribution, is observed. When the action set is time-varying, an extension of MAB referred to as *volatile MAB*, is proposed. Both Bayesian and frequentist based algorithms are presented in [11] and [19], respectively. In [7], we proposed a Bayesian learning framework for rate and channel adaptation where the set of available actions varies dynamically over time. In piecewise-stationary or non-stationary MABs, the reward distributions of the actions vary over time, and an MAB based algorithm can readily be extended to handle such reward distributions by introducing a sliding window. Then, the reward estimation and action selection counters are estimated only for the current window, which ensures that the estimates are from most recent values and historical values are gradually eliminated, as presented in [4], [5]. Moreover, a sliding window based Bayesian counterpart for rate adaptation is presented in [3]. Since, sliding windows require additional memory to keep track of all the estimates in the window, a discounted TS is proposed which eliminates the memory requirement, it multiplies the estimates with the successive powers of a scalar discounted factor ($\gamma \in (0, 1]$), which in turn ensures that historical values die out [13]. A change detection based algorithm for piecewise-stationary MAB is presented in [12], the proposed *change-point detection based on Thompson sampling* (CD-TS) scheme is parametrized by w (size of the estimation window) as well as b (change detection parameter), which requires information about the expected rewards of the actions in hand and may be difficult to acquire in real world applications.

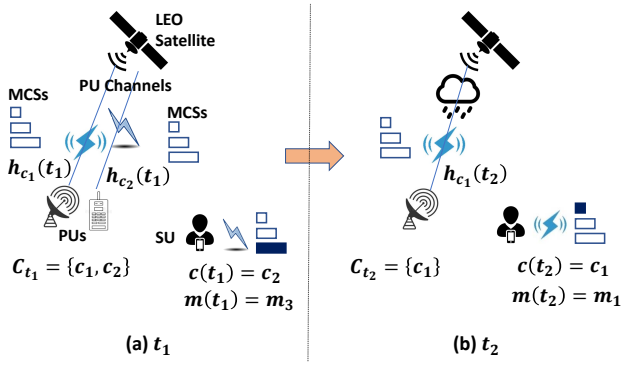

Fig. 1: System model and a toy example: (a) In round t_1, the channel set is $\mathcal{C}_{t_1} = \{c_1, c_2\}$ with channel states $h_{c_1}(t_1)$ and $h_{c_2}(t_1)$, respectively. SU selects MCS $m(t_1) = m_3$ to transmit over channel $c(t_1) = c_2$. (b) In round t_2, we have $\mathcal{C}_{t_2} = \{c_1\}$ with $h_{c_1}(t_2)$, SU selects $m(t_2) = m_1$ and $c(t_2) = c_1$.

Distinguished from prior works that either ignore the volatility in the channel set (e.g., see [4], [20]) or handle piecewise-stationary characteristics based on knowledge of the expected rewards and utilize sliding windows (e.g., see [4], [5], [12]), we propose an efficient Bayesian algorithm dSTS, which accounts for both volatility and piecewise-stationary settings (see Table I). To deal with evolving expected rewards (*piecewise-stationary settings*), dSTS is parametrized by a computationally inexpensive scalar quantity to discount historical values. To reduce the learning time, dSTS exploits the monotonicity of transmission success probabilities in the transmission rates over the chosen wireless channel (see Section IV). The performance of dSTS significantly exceeds that of well-known learning methods under such highly dynamic, volatile and time-varying communication scenarios (see Section V).

III. System Model and Problem Formulation

Let the set $\mathcal{M} = \{m_1, \ldots, m_M\}$ represent different MCSs with corresponding transmission rates $\mathcal{R}_M = \{r_1, \ldots, r_M\}$, while the set $\mathcal{C} = \{c_1, \ldots, c_C\}$ represents the collection of C channels. Without loss of generality, the transmission rates

satisfy $r_1 < \ldots < r_M$, and therefore, set $\mathcal{R}_M$ is ordered (and subsequently $\mathcal{M}$ is ordered). We consider the scenario where the channel statistics are unknown, and the channel is varying in a piecewise-stationary mode i.e., for a given channel $c \in \mathcal{C}$ with unknown statistics, there exists P_c ($P_c \geq 1$) channel state change points. The channel $h_c(t)$ remains the same during the block between two change points, and varies after a change point (see Fig. 2 in Section V).

A system model with a toy example of two wireless channels and three MCSs ($\mathcal{C} = \{c_1, c_2\}, \mathcal{M} = \{m_1, m_2, m_3\}$) is shown in Fig. 1. In Fig. 1(a), two PU channels are available in round t_1 (i.e., $\mathcal{C}_{t_1} = \{c_1, c_2\}$) with channel states $h_{c_1}(t_1)$ and $h_{c_2}(t_1)$, respectively. SU selects channel $c(t_1) = c_2$ and MCS $m(t_1) = m_3$ in round t_1. In Fig. 1(b), only one PU channel is available in round t_2 (i.e., $\mathcal{C}_{t_2} = \{c_1\}$) with channel state $h_{c_1}(t_2)$ (e.g., rain attenuation), SU selects MCS $m(t_2) = m_1$ to transmit over $c(t_2) = c_1$. Here, the SU opportunistically searches for free channels to utilize for its transmission, referred to as *interweave CRNs*. The set of these free channels varies over time due to PUs activity, hence adding volatility in the selection criteria. Furthermore, weather conditions, rain attenuation, etc., create dynamically changing channel conditions, which makes MCS-channel pair selection a challenging task in such dynamic scenarios.

Let T represent the time horizon where sequential decisions are made in rounds $t \in \{1, \cdots, T\}$. At the beginning of each round t, channel sensing is performed (assuming the existence of a channel sensing mechanism in the system or the exploitation of a geo-location database for dynamic SUs to avoid sensing PUs) to determine the set of free channels $\mathcal{C}_t$. Then, the user chooses a channel $c(t) \in \mathcal{C}_t$ and MCS $m(t) \in \mathcal{M}$ for that channel, and transmits at a rate $r_{m(t)} \in \mathcal{R}_M$ over the selected channel with channel state $h_{c(t)}(t)$. After the transmission of $r_{m(t)}$ bits, it receives an indicator of ACK/NACK, ACK for successful transmission and NACK for failure [7]. We let $x_{m(t),c(t)}(t)$ represent this Bernoulli reward (1 for successful transmission, and 0 otherwise) with expected value $\varphi_{m(t),c(t)}(t)$. This $\varphi_{m,c}(t)$ represents the transmission success probability for the MCS-channel pair (m, c) with channel state $h_c(t)$ at round t. It is important to note that $[m(t), c(t)]$ represent the selected MCS-channel pair at round t, and therefore $\varphi_{m(t),c(t)}(t)$ is the success probability for that pair at round t, whereas $\varphi_{m,c}(t)$ represents the success probability for any pair $m \in \mathcal{M}, c \in \mathcal{C}$ at round t. Similarly, the throughput associated with the MCS-channel pair (m, c) is $\mu_{m,c}(t) = r_m \cdot \varphi_{m,c}(t)$. We call $\mu_{m,c}(t)/r_M$ the normalized throughput of MCS-channel pair (m, c).

Transmission success probabilities exhibit a monotonically decreasing structure over the set of transmission rates [10], where $\varphi_{1,c}(t) > \varphi_{2,c}(t) \ldots > \varphi_{M,c}(t)$ for a given channel c with channel state $h_c(t)$ for every t. The optimal MCS-channel pair at round t is denoted by $(m^*(t), c^*(t)) = \operatorname{argmax}_{m \in \mathcal{M}, c \in \mathcal{C}_t} \mu_{m,c}(t)$. Without loss of generality we assume that $(m^*(t), c^*(t))$ is unique for every t. Let $\mathcal{A}_t = \mathcal{M} \times \mathcal{C}_t$ represent the available action set in round t. For any T round available action sequence $\mathcal{A} = \{\mathcal{A}_1, \ldots, \mathcal{A}_T\}$ (since the available channel set is time-varying, each $t \in \{1, \cdots, T\}$ has different available channels to chose from), the expected regret is defined as

$$R_{\mathcal{A}}(T) = \mathbb{E}\left[\sum_{t=1}^{T} \Big(\mu_{m^*(t),c^*(t)}(t) - \mu_{m(t),c(t)}(t)\Big) \Big| \mathcal{A}\right]. \quad (1)$$

Algorithm 1 Discounted Structured and Sleeping Thompson Sampling (dSTS)

Input: M, C, T
Parameters: $\gamma \in (0, 1]$
Initialize: $S_{m,c} = 0, F_{m,c} = 0, \forall m \in \mathcal{M}, \forall c \in \mathcal{C}$
for $t = 1, 2, \cdots, T$ **do**
  Channels free from primary activity $\mathcal{C}_t$
  for $c \in \mathcal{C}_t$
    Draw $\boldsymbol{\phi}_c(t) \sim \mathbf{1}\big(\boldsymbol{\phi}_c(t) \in \Phi_c(t)\big) \prod_{m \in \mathcal{M}} \pi_{m,c}$
    $\theta_{m,c} = r_m \cdot \phi_{m,c}$, $\forall m \in \mathcal{M}$
  end for
  $[m(t), c(t)] = \operatorname{argmax}_{m \in \mathcal{M}, c \in \mathcal{C}_t} \ \theta_{m,c}$
  Transmission with rate $r_{m(t)}$ over channel $c(t)$
  observe ACK/NACK
  $x_{m(t),c(t)}(t) = 1$ if ACK
  $x_{m(t),c(t)}(t) = 0$ if NACK
  $S_{m,c} = \gamma \cdot S_{m,c}, F_{m,c} = \gamma \cdot F_{m,c} \quad \forall m \in \mathcal{M}, \forall c \in \mathcal{C}$
  $S_{m(t),c(t)} = S_{m(t),c(t)} + x_{m(t),c(t)}(t)$
  $F_{m(t),c(t)} = F_{m(t),c(t)} + \big(1 - x_{m(t),c(t)}(t)\big)$
  $t = t + 1$
end for

IV. The Proposed Algorithm

We propose dSTS, which exploits the monotone structure of the success probability in transmission rates and takes into account the time-varying availability (volatility) of PU channels as well as their change points by discounting the historical values from current estimations (pseudocode is given in Algorithm 1). Its main advantage lies in optimizing the channel selection from a dynamically varying channel set together with MCS selections under piecewise-stationary characteristics.

Let $\pi_{m,c} = Beta(S_{m,c}+1, F_{m,c}+1)$ represent the posterior distribution of $\varphi_{m,c}(t)$ for each MCS-channel pair (m, c), where $S_{m,c}(t)$ represents the number of successful transmissions and $F_{m,c}(t)$ represents the number of unsuccessful transmissions, when (m, c) was selected before round t. $\phi_{m,c}$ represents a random sample drawn from these distributions for each $(m, c) \in \mathcal{A}_t$. dSTS ensures that the samples satisfy the monotone structure of $\phi_{m,c}$ in transmission rates [7], [10]. For a channel c, let $\Phi_c(t) = \{(\phi_{1,c}, \ldots, \phi_{M,c}) \mid \phi_{m,c} \geq \phi_{m',c}, \forall r_m < r'_m,\ m, m' \in \mathcal{M}\}$ represent the selected sample set that satisfies monotonicity in round t. Primarily, for each $c \in \mathcal{C}_t$, dSTS takes samples $\boldsymbol{\phi}_c(t) = (\phi_{1,c}, \ldots, \phi_{M,c})$ such that $\boldsymbol{\phi}_c(t) \sim \mathbf{1}\big(\boldsymbol{\phi}_c(t) \in \Phi_c(t)\big) \prod_{m \in \mathcal{M}} \pi_{m,c}$, where $\mathbf{1}(\cdot)$ is the indicator function.

For each channel $c \in \mathcal{C}_t$, throughput samples $\theta_{m,c}$ are calculated by multiplying the obtained success probability samples $\phi_{m,c}$ with the corresponding transmission rates $r_m \in \mathcal{R}_M, \forall m \in \mathcal{M}$. Then, the MCS-channel pair with the maximum throughput sample is selected for transmission. After transmission over channel $c(t)$ with MCS $m(t)$ (corresponding rate $r_{m(t)}$), the 1-bit feedback $x_{m(t),c(t)}(t)$ is observed. For

This article has been accepted for publication in IEEE Communications Letters. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/LCOMM.2023.3240363

TABLE II: Synthetic Success Probabilities over MCS-channel pairs (inspired from [5], [7], [10])

	r_m in Mbits/s							
MCS	1	2	3	4	5	6	7	8
r_m	3	4.5	6	9	12	18	24	27
$\varphi_{m,1}$	0.95	0.85	0.75	0.65	0.55	0.45	0.35	0.25
$\varphi_{m,2}$	0.90	0.80	0.70	0.60	0.50	0.45	0.30	0.20
$\varphi_{m,3}$	0.85	0.75	0.65	0.55	0.45	0.35	0.25	0.15
$\varphi_{m,4}$	0.95	0.90	0.85	0.80	0.40	0.20	0.10	0.05
$\varphi_{m,5}$	0.75	0.65	0.55	0.45	0.35	0.25	0.15	0.05

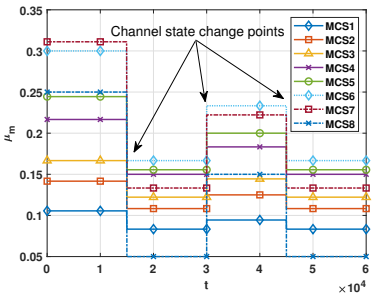

Fig. 2: The expected rewards for channel 1 as a function of t.

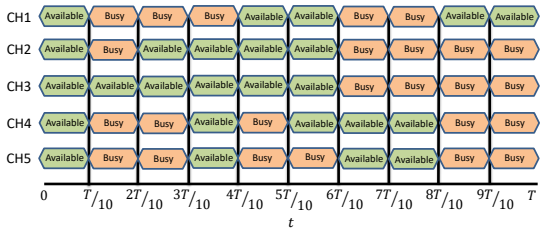

Fig. 3: Availability of wireless channels as a function of t.

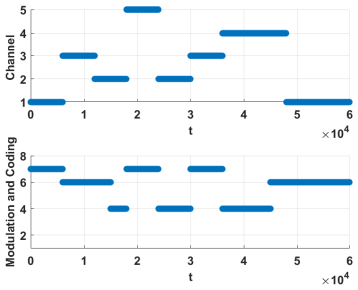

Fig. 4: The best (channel, MCS) pair as a function of t.

each MCS $m \in \mathcal{M}$ and channel $c \in \mathcal{C}$, values of $S_{m,c}$ and $F_{m,c}$ are discounted with factor γ to handle the piecewise-stationary distribution. The value of $\gamma \in (0, 1]$ ensures that historical values are gradually eliminated from estimations. Then, the estimates of the selected MCS-channel pair are updated i.e., $S_{m(t),c(t)}$ is incremented with 1 in case of successful transmission, whereas $F_{m(t),c(t)}$ is incremented with 1 in case of unsuccessful transmission. The proposed dSTS is able to minimize the regret in eq. 1 due to the fact that TS is asymptotically optimal for Bernoulli rewards (i.e., $x_{m(t),c(t)}(t)$), and discounted TS is an extension to TS for piecewise-stationary settings. Furthermore, by exploiting the monotone structure in success probabilities, dSTS converges faster to the optimal solution by eliminating the suboptimal selections that violate the monotone structure.

V. Numerical Evaluation

We set the number of MCSs (and corresponding transmission rates) to $M = 8$, the number of wireless channels $C = 5$, and the number of piecewise-stationary channel state change points is set to $P_c = 3$ (we chose this value for numerical illustration purposes for rain attenuations, cloud attenuations, etc., in LEO networks). Transmission success probabilities for MCS-channel pairs are shown in Table II. These channels are swapped randomly after passing each of the channel state change points, as shown in Fig. 2. We set the total number of rounds to $T = 60,000$, and the results are averaged over 10 runs. We set γ to 0.9999 for dSTS. We set the channel availability to change after every $\frac{T}{10}$ rounds, i.e, the number of available channels varies in rounds $\{6, 12, 18, 24, 30, 36, 42, 48, 54\} \times 10^3$, as shown in Fig. 3. We set the channel state change point to occur after every 15×10^3 rounds, and it is evident from Fig. 2 that the transmission success probabilities (and the expected rewards) vary in rounds $\{15, 30, 45\} \times 10^3$. Due to variations in the channel availability and the expected rewards, the optimal MCS-channel pair may vary after every 3×10^3 rounds, as shown in Fig. 4.

We compare the proposed algorithm with the following earlier works.

i) ORC (Oracle): Benchmark that knows the expected rewards beforehand and always selects the optimal MCS-channel pair in each round t.
ii) SW-KL-UCB-U [4]: Exploits the structural (unimodal) information of the expected rewards over the transmission rates and maintains a sliding window over rounds for reward estimates and selection counters. We set the sliding window size to $\tau = 3000$.
iii) V-CoTS [7]: V-CoTS selects a rate-channel pair over a time-varying channel set and exploits the monotonicity in the success probabilities over the transmission rates. However, it ignores the time-varying channel statistics.
iv) CD-CoTS [6]: We run a separate instance of CD-CoTS for each channel, CD-CoTS exploits the monotonicity in the success probabilities over the transmission rates and is based on change point detection in the expected rewards, due to which it can handle time-varying changes in the channel state. We set the specific selection, change-detection window size, and reward threshold to $F = 2000, w = 2000, b = 0.01$, respectively.

Fig. 5 shows that the number of successfully transmitted Mbits up to round T(i.e., $60,000$) by the Oracle is around 4.2×10^5 *Mbits*. The proposed dSTS algorithm achieves a performance close to that of Oracle by transmitting 3.9×10^5 *Mbits*. Whereas, the number of successfully transmitted Mbits by CD-CoTS, V-CoTS and SW-KL-UCB-U are around 3.6×10^5, 3.6×10^5 and 3.1×10^5 *Mbits*, respectively.

Fig. 6 compares the expected regrets (defined in eq. 1) of dSTS, CD-CoTS, V-CoTS and SW-KL-UCB-U. It is observed

This article has been accepted for publication in IEEE Communications Letters. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/LCOMM.2023.3240363

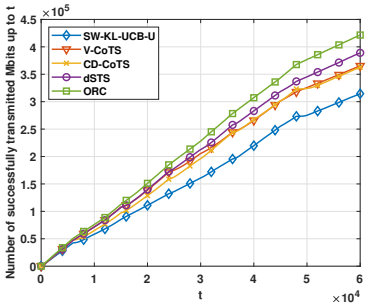

Fig. 5: Number of successfully transmitted Mbits up to t.

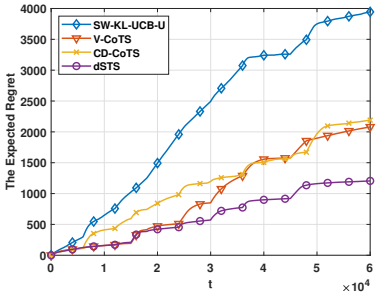

Fig. 6: The expected regret by round t.

that dSTS achieves the minimum expected regret. CD-CoTS also obtains a small regret by exploiting the correlation of the success probabilities and handling variations in the channel set as well as in the expected reward. Nonetheless, CD-CoTS requires knowledge of the minimum threshold of the expected reward changes, and the estimation window size, and requires memory to store values for all MCS-channel pairs for at-least the size of the estimation window. Meanwhile, dSTS only uses a scalar value γ to handle changes in the expected rewards. The improvement induced by dSTS compared to V-CoTS is due to the discounted TS, which performs better in piecewise-stationary environments, whereas V-CoTS fails to adapt to changes in the expected rewards due to channel state variations. SW-KL-UCB-U, a frequentist counterpart of dSTS, obtains a higher regret compared to other competitors. The expected regret obtained with dSTS is approximately 43% lower than that of V-CoTS, 45% lower than that of CD-CoTS, and 70% lower than that of SW-KL-UCB-U. It is also observed that, in contrast to stationary environments, the optimal MCS-channel pair varies with time in piecewise-stationary environments (see Fig. 4). Consequently, the optimal expect reward (throughput) varies with time, due to which the expected regret follows a staircase shape (see Fig. 6).

VI. Conclusion

In this letter, we proposed a discounted Bayesian learning algorithm for link adaptation and channel selection under unknown and piecewise-stationary channel statistics. Our proposed scheme, dSTS, handles the volatility in the channel set and exploits the structural information (monotonicity) of the transmission success probability in the transmission rates. Numerical evaluations demonstrate that dSTS handles the time-varying statistics and volatile settings, in terms of the expected regret and the expected cummulative throughput, better than state-of-the-art algorithms. The expected regret obtained with dSTS is approximately 43% lower than that of V-CoTS, 45% lower than that of CD-CoTS, and 70% lower than that of SW-KL-UCB-U.

References

[1] U. Khan, Z. Ali, E. Lagunas, A. Mahmood, M. Asif, A. Ihsan, S. Chatzinotas, B. Ottersten, O. A. Dobre, *et al.*, "Rate splitting multiple access for next generation cognitive radio enabled LEO satellite networks," *arXiv e-prints*, pp. arXiv–2208.03705, 2022.

[2] S. Chatzinotas, B. Evans, A. Guidotti, V. Icolari, E. Lagunas, S. Maleki, S. K. Sharma, D. Tarchi, P. Thompson, and A. Vanelli-Coralli, "Cognitive approaches to enhance spectrum availability for satellite systems," *Intl. J. Satellite Commun. Netw.*, vol. 35, no. 5, pp. 407–442, 2017.

[3] H. Qi, Z. Hu, X. Wen, and Z. Lu, "Rate adaptation with Thompson sampling in 802.11 ac WLAN," *IEEE Commun. Lett.*, vol. 23, no. 10, pp. 1888–1892, 2019.

[4] R. Combes and A. Proutiere, "Dynamic rate and channel selection in cognitive radio systems," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 5, pp. 910–921, May 2015.

[5] R. Combes, J. Ok, A. Proutiere, D. Yun, and Y. Yi, "Optimal rate sampling in 802.11 systems: Theory, design, and implementation," *IEEE Trans. Mobile Comput.*, vol. 18, no. 5, pp. 1145–1158, 2018.

[6] H. Tang, X. Hou, J. Wang, and J. Song, "Joint link rate selection and channel state change detection in block-fading channels," in *2021 IEEE Global Commun. Conf. (GLOBECOM)*, pp. 1–6, IEEE, 2021.

[7] M. A. Qureshi and C. Tekin, "Rate and channel adaptation in cognitive radio networks under time-varying constraints," *IEEE Commun. Lett.*, vol. 24, no. 12, pp. 2979–2983, 2020.

[8] Y. Gao, E. Hossain, G. Y. Li, K. Sowerby, C. Regazzoni, and L. Zhang, "IEEE TCCN special section editorial: Evolution of cognitive radio to AI-Enabled radio and networks," *IEEE Trans. Cogn. Commun. Netw.*, vol. 6, no. 1, pp. 1–5, 2020.

[9] H. Gupta, A. Eryilmaz, and R. Srikant, "Low-complexity, low-regret link rate selection in rapidly-varying wireless channels," in *Proc. IEEE Conf. Comput. Commun.*, pp. 540–548, IEEE, 2018.

[10] H. Gupta, A. Eryilmaz, and R. Srikant, "Link rate selection using constrained Thompson sampling," in *Proc. IEEE Conf. Comput. Commun.*, pp. 739–747, 2019.

[11] A. Chatterjee, G. Ghalme, S. Jain, R. Vaish, and Y. Narahari, "Analysis of Thompson sampling for stochastic sleeping bandits.," in *Proc. Conf. Uncertainty in Artif. Intell.*, 2017.

[12] Y. Cao, Z. Wen, B. Kveton, and Y. Xie, "Nearly optimal adaptive procedure with change detection for piecewise-stationary bandit," in *The 22nd Intl. Conf. Artif. Intell. and Statistics*, pp. 418–427, PMLR, 2019.

[13] M. Zhou, T. Wang, and S. Wang, "Spectrum sensing across multiple service providers: A discounted Thompson sampling method," *IEEE Commun. Lett.*, vol. 23, no. 12, pp. 2402–2406, 2019.

[14] Q. Liu, S. Zhou, and G. B. Giannakis, "Cross-layer combining of adaptive modulation and coding with truncated ARQ over wireless links," *IEEE Tran. Wireless Commun.*, vol. 3, no. 5, pp. 1746–1755, 2004.

[15] D. Tarchi, G. E. Corazza, and A. Vanelli-Coralli, "Adaptive coding and modulation techniques for mobile satellite communications: A state estimation approach," in *6th Adv. Satellite Multimedia Systems Conf. (ASMS) and 12th Signal Process. for Space Commun. Workshop (SPSC)*, pp. 36–43, IEEE, 2012.

[16] X. Wang, H. Li, and Q. Wu, "Optimizing adaptive coding and modulation for satellite network with ML-based CSI prediction," in *2019 IEEE Wireless Commun. and Netw. Conf. (WCNC)*, pp. 1–6, IEEE, 2019.

[17] T. L. Lai and H. Robbins, "Asymptotically efficient adaptive allocation rules," *Adv. Appl. Math.*, vol. 6, no. 1, pp. 4–22, 1985.

[18] W. R. Thompson, "On the likelihood that one unknown probability exceeds another in view of the evidence of two samples," *Biometrika*, vol. 25, pp. 285–294, 1933.

[19] R. Kleinberg, A. Niculescu-Mizil, and Y. Sharma, "Regret bounds for sleeping experts and bandits," *Mach. Learn.*, vol. 80, no. 2-3, pp. 245–272, 2010.

[20] M. A. Qureshi and C. Tekin, "Online Bayesian learning for rate selection in millimeter wave cognitive radio networks," in *Proc. IEEE Int. Conf. Comput. Commun.*, 2020.