On Softwarization of Intelligence in 6G Networks for Ultra-Fast Optimal Policy Selection: Challenges and Opportunities

Sherief Hashima, *Senior Member, IEEE,* Zubair Md Fadlullah, *Senior Member, IEEE*,
Mostafa M. Fouda, *Senior Member, IEEE*, Ehab Mahmoud Mohamed, *Member, IEEE,* Kohei Hatano,
Basem M. ElHalawany, *Senior Member, IEEE*, and Mohsen Guizani, *Fellow, IEEE.*

***Abstract*—The emerging Sixth Generation (6G) communication networks promising 100 to 1000 Gbps rates and ultra-low latency (1 millisecond) are anticipated to have native, embedded Artificial Intelligence (AI) capability to support a myriad of services, such as Holographic Type Communications (HTC), tactile Internet, remote surgery, etc. However, these services require ultra-reliability, which is highly impacted by the dynamically changing environment of 6G heterogeneous tiny cells, whereby static AI solutions fitting all scenarios and devices are impractical. Hence, this article introduces a novel concept called the softwarization of intelligence in 6G networks to select the most ideal, ultra-fast optimal policy based on the highly varying channel conditions, traffic demand, user mobility, and so forth. Our envisioned concept is exemplified in a Multi-Armed Bandit (MAB) framework and evaluated within a use case of two simultaneous scenarios (i.e., Neighbor Discovery and Selection (NDS) in a Device-to-Device (D2D) network and aerial gateway selection in an Unmanned Aerial Vehicle (UAV)-based under-served area network). Furthermore, our concept is evaluated through extensive computer-based simulations that indicate encouraging performance. Finally, related challenges and future directions are highlighted.**

***Index Terms*—6G, softwarization, optimization, Artificial Intelligence (AI), Multi-armed bandit (MAB).**

I. Introduction

With the race to provide Terabits per second (Tbps) data rates to mobile users to meet the dramatically increasing network traffic demands [1], the Sixth Generation networks (6G) are being conceptualized by researchers and practitioners of the ITU (International Telecommunication Union). The emerging 6G networks are anticipated to focus on mobile edge computing, whereby the core, edge networking, and computing functions will become seamlessly integrated. Furthermore, by leveraging the increasingly available computational opportunities and embedded intelligence [2], 6G networks are also expected to benefit from robust Artificial Intelligence (AI) capabilities. Although the contemporary data centers have recently enjoyed a significant paradigm shift during the Fifth Generation (5G) era toward virtualization and programmable, Software-Defined Networks (SDNs), the heterogeneous radio access networks of 5G and beyond are yet to fully utilize the tremendous power of programmable network functionality. For instance, many commodity routers or even Base Stations (BSs) may require hybrid Radio Access Networks (RANs) for on-demand network functionalities, e.g., Neighbor Discovery and Selection (NDS) in Device-to-Device (D2D) communication [3], optimal band allocation under mobility and blocking effect, relay probing [4], and so forth.

In 6G networks, even edge nodes are expected to engage in edge computing because of the increasing computing and energy resources. However, due to integrated terrestrial-aerial-satellite-underwater networks, the network dynamics of the heterogeneous access technologies in 6G networks will be much more volatile and unpredictable than even those in 5G networks. This is likely to severely impact the 6G services, such as tactile Internet, augmented reality, robotic surgery, and so forth, that demand ultra-reliability. Furthermore, the sheer number of ultra-dense tiny cells will make it more challenging to manage the network traffic in a scalable manner without the intervention of optimization and AI techniques available locally (i.e., at the BS/edge node level). However, accommodating all possible optimization and/or AI models in a single BS, let alone at an edge node, is a formidable research challenge and not deemed practical due to possibly high manufacturing, testing, deployment, and operational costs.

To mitigate some of the above challenges, this article presents the following contributions:

- We propose a programmable AI-based 6G system model deployment into Access Points (APs), BSs, and edge nodes in an on-demand manner to cope with the prevalent network conditions.
- We propose a class of AI/optimization schemes in the higher application plane of our system model, and then instantiate objects of the most relevant optimal policy type to any network topology (i.e., ranging from a macro

Sherief Hashima is with the RIKEN-Advanced Intelligence Project (AIP), Japan, and the Egyptian Atomic Energy Authority, Egypt (e-mail: sherief.hashima@riken.jp).

Zubair Md Fadlullah is with Western University, London, ON, Canada (email: Zubair.Fadlullah@lakeheadu.ca).

Mostafa M. Fouda is with Idaho State University, USA, and Center for Advanced Energy Studies (CAES), USA (e-mail: mfouda@ieee.org).

Ehab M. Mohamed is with Prince Sattam Bin Abdulaziz University, Saudi Arabia, and Aswan University, Egypt (email: ehab_mahmoud@aswu.edu.eg).

Kohei Hatano is with Kyushu University, Japan and with the RIKEN-AIP, Japan (email: hatano@inf.kyushu-u.ac.jp).

Basem M. ElHalawany is with Benha University, Egypt (basem.mamdoh@feng.bu.edu.eg).

Mohsen Guizani is with Mohamed Bin Zayed University of Artificial Intelligence, UAE (mguizani@ieee.org).

to a tiny cell), thereby transforming the "dumb" network equipment into AI-enabled, self-decision-making intelligent BSs or even smart edge nodes. The ideal policy type can be optimization modules (e.g., linear programming, convex/multi-objective optimization, and meta-heuristics) or any of the supervised, non-supervised, reinforcement learning or other sequential learning (e.g., Multi-Armed Bandit (MAB) [3], [5]) models.

- We provide a use case with two scenarios, involving NDS in a D2D-based network [3] and aerial gateway selection in an Unmanned Aerial Vehicle (UAV) or drone-based under-served area network [5], to extensively demonstrate the performance evaluation of our conceptualized softwarized intelligence-based optimal policy selection for ultra-fast decision-making in 6G networks.

The remainder of this article is organized as follows. Section II presents our research motivation. Our considered problem of AI-softwarized 6G networks and envisioned softwarized system model including ultra-fast, on-demand policy selection methodology is described in Section III. Next, in Section IV, we provide a use case with two scenarios to demonstrate the effectiveness of employing various ultra-fast optimal policies frameworks as an efficient enabler of the future 6G applications under varying network conditions. The challenges and future directions are delineated in Section V. Finally, Section VI concludes the article.

II. Motivation: Ultra-fast Optimal Policy Selection in 6G Networks

In every successive generation of wireless communication networks, capacity and delay requirements are heavily stressed. Compared to their predecessors, the emerging 6G networks are anticipated to provide native, embedded intelligence to support ultra-fast handover under high mobility, determining actions of thousands of intelligent/re-configurable surfaces, smart energy consumption, and so forth. Given the massive connectivity support up to 10^7 devices/Km^2 with traffic capacity of up to 1 Gbps/m^2, embedded intelligence needs to be made scalable and available to both service providers and edge users to support new killer applications and human-centric services, space-air-ground-sea integrated networks, holographic communication, tactile Internet, remote surgery, augmented reality-based immersive computing, etc. These services will often require ultra-fast, distributed learning frameworks which may need to be rapidly changed depending on the varying scenarios. In contrast to the current paradigm where the pre-trained AI models are deployed to 6G-based user-devices, provisioning scenario-specific AI models to optimally cater to these varying network dynamics will pose a formidable research challenge. This is the primary motivation behind this work.

III. Proposed Softwarization of Intelligence in 6G Networks

We elucidate the critical need of softwarization of intelligence in the emerging 6G communication systems, and then describe our envisioned system model of heterogeneous softwarized networks as depicted in Fig. 1, where ultra-fast, online learning models may play a versatile role in the AI-enabled 6G network optimization compared to traditional learning and optimization models.

A. Why Softwarized 6G Networks?

Unlike 5G systems, the 6G networks are highly expected to seamlessly integrate terrestrial, aerial, satellite, and underwater networks with various radio access technologies and proliferate the intelligent computing and communication across these heterogeneous access networks, particularly at the edge of the network. However, embedded intelligence at the edge nodes cannot be static due to the highly variable behavior of heterogeneous bands/channels (ranging from legacy MHz/GHz range to THz carrier frequencies) under different-sized stationary/mobile blockers, high mobility of nodes, highly volatile traffic demands, zero-day cyber-attacks, etc. If those edge nodes are statically deployed, AI models may not be adequate to cope with these dynamics. For example, consider the D2D nodes in 6G networks that require relay node probing or neighbor discovery within near-real-time performance.

Several AI models have been used in the literature to improve the network performance, which are not limited to reinforcement learning models like MAB, Q-learning, or actor-critic but also belong to other smart learning techniques such as game theory, supervised, and unsupervised learning models. Theoretically, the softwarized architecture can select the proper model dynamically to achieve the best possible performance. Although various AI techniques can be implemented for the softwarized SDN network, our main focus in this article is to demonstrate a proof-of-concept with easy-to-understand use cases that need different models for fast decision making, no prior supervised training, adaptability to varying environments, lightweight (stateless), and fast convergence. Therefore, a list of various optimization, supervised and unsupervised, and reinforcement learning models including different MAB variants, are maintained by the softwarized network. For instance, for quasi-static nodes, stochastic MAB can be used as an optimal policy for rapid deployment onto the nodes and ultra-fast decision-making in contrast with traditional optimization and supervised learning models due to its maximal accumulative/long term reward strategy. Specifically, D2D devices need to be equipped with stochastic MAB models [3] to select the best neighbor. However, when the D2D nodes are mobile, an adversarial MAB model, which considers the adversary environment needs to be employed instead of the stochastic variant. Moreover, if we consider multi-band capable devices, contextual MAB comes to the scene, where the information obtained from one band can be adopted as contexts [3]. Upon various network conditions, the MAB policies (e.g., ϵ-greedy, Upper Confidence Bound (UCB), Thompson Sampling (TS), etc. [3], [5]) may be updated. By extending this example for many combinations of dynamic network scenarios, the D2D nodes need to have the capability of on-demand deployment of an intelligent model. Similarly, consider the wireless Access Point (AP) or home routers with dedicated operating software for routing, firewall, and security. Depending on the network dynamics, a capability for procuring the best possible module for managing network functions under the current network dynamism or dealing with the ongoing security threats should be provisioned. The 6G network equipment are considered to be programmable in our system model, which can download, in an on-demand manner, the necessary intelligent modules, which can range from optimization models (e.g., linear programming to convex optimization) or AI models (e.g., supervised/unsupervised/online learning) as illustrated in Fig. 1. Our envisioned system model goes beyond the existing SDN architectures that have yet to achieve the actual potential of programmable network functions.

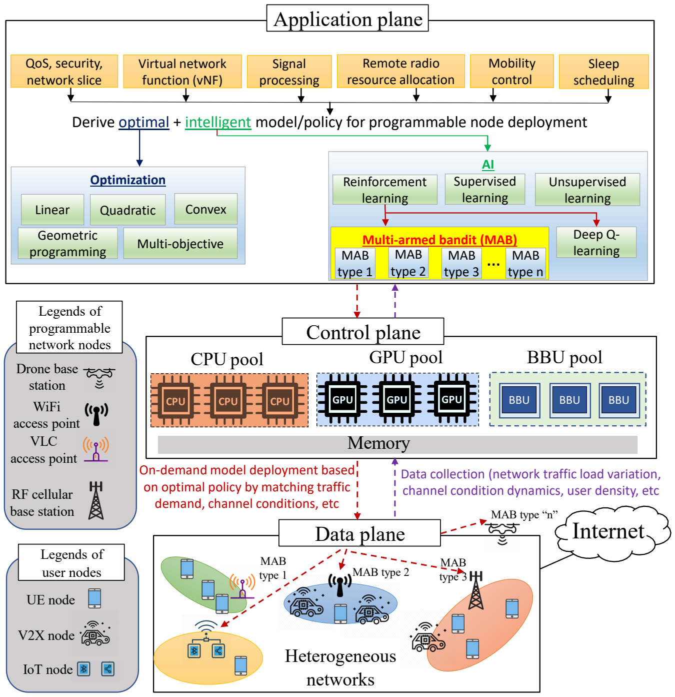


Fig. 1. Considered system model of heterogeneous softwarized networks whereby the Base Stations (BSs), Access Points (APs), and user devices can obtain on-demand, ultra-fast algorithmic policy selection to reflect the dynamically varying network conditions.

B. Envisioned Softwarized System Model

Now we provide our envisioned system model's groundwork. By referring to our considered system model in Fig. 1, we derive optimal and intelligent models and/or policies for deployment on to terrestrial/drone BSs, Radio Frequency (RF) BSs, Wireless Local Area Network (WLAN) APs, Visible Light Communication (VLC) BSs, User Equipment (UEs), Vehicle-to-other (V2X) nodes [6], and Internet of Things (IoT) devices. From the perspective of the application plane, the optimal or intelligent modules can facilitate one or more applications including Virtual Network Function (VNF) placement, Quality of Service (QoS) and security provisioning in network slice, signal and noise processing, dynamic allocation of remote radio resources, mobility prediction with network traffic control, sleep scheduling of BSs/user devices, etc. For each of these applications for the hybrid BSs and edge nodes in 6G during different times that experience different network dynamics, a unique optimization or intelligent model needs to be derived from providing the best solution with which BSs/edge nodes need to be programmed. The optimization and innovative models are executed through a pool of Central Processing Units (CPUs), supported by Graphics Processing Units (GPUs) to parallelize and scale up computing in the control plane. Also, baseband units are included in the control plane's hardware, which facilitates the radio resource virtualization and decouples the data plane. Nevertheless, in the data plane, the access points, based on the derived optimal and/or intelligent model/policy, observe the current traffic demand, channel conditions, and so forth and then decide which models are ideal for the current situation. Therefore, by matching the network dynamics, they proactively download the optimal/intelligent network policy stochastic MAB model for stationary D2D nodes while adversarial MAB model for the mobile UEs, and so forth [3]. Regarding the MAB models stated in our example use case, the portrayed system model benefits from the reusability of the MAB schemes for new programmable network nodes. Specifically, the application plane has to define a few primary instances of MAB model types. Moreover, if we consider multi-band capable devices, contextual MABs are more appropriate, where the information obtained from one band can be adopted as contexts of the MAB game played over the other band [3]. Upon current network topology and dynamics, network devices can simply be transformed to accommodate any AI model to combat the prevalent conditions by merely creating an instance or a program based on the base definition that it procures from the repository of AI modules in the application plane.

C. Envisioned Ultra-fast On-demand Policy Selection via Softwarized Intelligence

For ease of discussion, among various softwarized intelligence models maintained at the softwarized network controller, we highlight the MAB variants. While MAB is appealing for various network-centric decision-making in contrast with other optimization and supervised learning techniques, how the most relevant MAB model can be dynamically selected to cater to the highly varying network conditions needs to be decided. The advantage of MAB, as a sequential, online optimal policy selection algorithm, over the classical optimization techniques and existing AI methods, as depicted in Fig. 1, can be described in terms of its ultra-fast decision-making capability. The classical optimization techniques are not scalable with the highly varying network dynamics. As a result, it is often challenging to provide a closed-form expression on the existence and guarantee of an optimal solution for a well-defined, complex problem. Many of the constraints and conditions are often relaxed upon the utilized algorithm design to reach a sub-optimal solution. Furthermore, such optimization techniques are typically a one-shot process as they require centralized, oracle-like knowledge to ingest the entire dataset to give the optimal benchmark decision. On the other hand, a supervised learning model typically requires long training time as well as extensive and versatile

TABLE I
A POOL OF MAB TYPES AT THE DISPOSAL OF THE PROPOSED ON-DEMAND SELECTOR FOR RAPIDLY IDENTIFYING THE RELEVANT MAB VARIANT WITH THE OPTIMAL POLICY TO CATER TO EACH OF THE TWO SCENARIOS IN THE CONSIDERED USE CASE.

Network setting	Objective	Suitable MAB/policy type	Expected outcome
Wireless power transfer [7]	Efficiently charge all energy harvesters	Combinatorial MAB	Fair energy harvesting and QoS
Wireless Sensor Networks (WSNs) cooperative relay selection [8]	Enhance hierarchical WSNs energy efficiency	Multi-Player (MP-UCB)	Better relay selection strategy with higher energy efficiency
Online client scheduling [9]	Reduce the latency in federated learning	UCB policy & virtual queuing	High speed convergence with fairness constraints
Small cell caching [10]	Learning based caching for small cells	Multi-Player MAB (MP-MAB)	High computational complexity balance
Two hop relay selection [4]	Maximize the throughput of the network	Sleeping Contextual bandit (S-LinUCB)	Better relay selection strategy with higher energy efficiency
Millimeter Wave (mmWave) vehicular communications [6]	Fast beam tracking	Context and social aware machine learning	Near optimal performance
Machine-type communications [11]	Scheduling fast uplink transmissions	Sleeping MABs	A three fold latency reduction
Mode selection and resource allocation in D2D [4]	Adaptable reduced computational complexity	Combinatorial MABs	Efficient performance
Fast mmWave beam alignment [1]	Accurate and reliable mmWave beam alignment	Stochastic Correlated MAB	Optimal probable beam identification
Handover management [12]	Optimal BS selection during handover	Cascaded Bandits	Efficient dynamic and Received Signal Strength Indicator (RSSI) solutions
NDS in mmWave D2D [3]	mmWave D2D best NDS	Stochastic contextual MABs	Prolong network lifetime with good D2D link
UAV selection in disaster area [5]	Gateway UAV selection problem	Distributed, MP-MAB	Perfect gateway UAV selection
Server selection [13]	Optimal server selection in SDNs	UCB, ϵ greedy, softmax	Good average response time and reward score

training datasets. The lack of an adequate dataset, which is critical to train the existing machine/deep learning models, appears as a crucial barrier to maximize their predictive performance. Moreover, the performances of such supervised learning-based models are typically sub-optimal, and a lack of interpretation to why they provide such performances still raises a lot of concerns among researchers for their deployment on networking devices in contrast with the traditional straightforward, feedback-based decision making. Therefore, ultra-fast online learning techniques are essential to be deployed to the 6G users (e.g., BSs, home APs, mobile UEs, and so forth) for localized, distributed decision making. The type of MAB can also be changed in an on-demand manner to combat the sudden change in the network conditions experienced by the 6G users. Furthermore, the recent advances in regret analysis for the variants of MAB algorithms can be leveraged to demonstrate their tightly bounded performance guarantee. Thus, from hereon, we regard MAB as a more viable technique compared to the classical optimization and supervised learning counterparts for ultra-fast, on-demand, and optimal policy selection.

Next, we present a high-level description of how the softwarized on-demand selection of MAB is possible by referring to Table I. Note that each problem is matched with an appropriate MAB technique, e.g., single player, Multi-Player (MP) [5], [8], combinatorial [7], sleeping [4], contextual [3], correlated bandits [1], cascaded MABs [12], and so on. For example, single player, stochastic MABs and also contextual MABs have been identified to solve mmWave D2D NDS problem in our earlier work [3]. In small caching scenarios, MP-MABs emerge as the most appropriate type [10]. MP-MABs have also been leveraged for the Gateway UAV selection problem in disaster scenarios [5]. On the other hand, sleeping contextual MABs have exhibited encouraging performance for two-hop relay probing in mmWave networks [4]. For handover problem handling, cascaded bandits are the most suitable promising solution [12] due to different rewards obtained by the learner given the location. Moreover, in machine-centric communications, where some devices are inactive while others continue to operate, the sleeping bandits can effectively describe the scenario [11]. Also, a MAB-based server selection method is discussed in [13] for SDNs, which provided better response time and payoff scores. Now, we can consider that our envisioned on-demand selector, as shown in Fig. 1, can choose from a pool of these MAB frameworks and their various policy implementations, along with other optimization and supervised/unsupervised/reinforcement learning techniques. When the network dynamics changes, the on-demand selector will proactively and rapidly select another AI model which is more ideal to adapt with the new network topology/conditions, and accordingly reprogram the network nodes to begin using the new AI model.

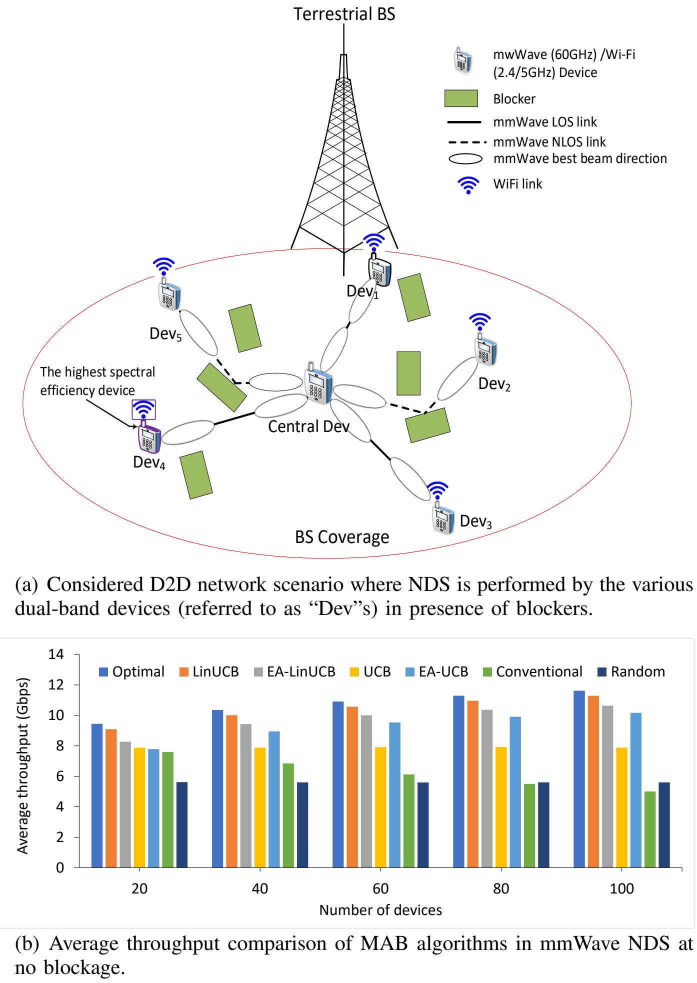
IV. AN ILLUSTRATIVE USE CASE AND PERFORMANCE EVALUATION

To confirm the effectiveness of using MAB variants with different policies for ultra-fast inference in 6G networks, in this section, we present an illustrative use case consisting of of two distinct scenarios that occur simultaneously, managed by the softwarized network controller. The topology of the first scenario is based on mmWave-enabled D2D topology to extend the coverage area. When such D2D relay-based coverage expansion is not adequeate in under-served areas, we consider a UAV-based communication network topology to further extend the coverage area in the second scenario. The first scenario handles the NDS task of D2D nodes [3] while the second scenario focuses on the aerial gateway selection in UAV-based under-served (e.g., disaster-affected) area communication [5]. The on-demand selector, in the control plane of the softwarized network, acquires the current network and demand information (e.g., mobility of users, traffic rate variation, traffic demands of users, indoor or outdoor scenario, QoS expectation, security expectation, and so forth), and then can compare which MAB type from Table I best describes the current network dynamics. Based on this, for either scenarios, the on-demand selector chooses the most ideal MAB algorithmic implementation with the optimal algorithmic policies that include UCB (Upper Confidence Bound), TS (Thompson Sampling), meta-TS, and so forth.

A. Scenario 1: Optimal Policy Selection in mmWave-enabled D2D Neighbor Discovery Service

The first scenario of our considered use case is depicted in Fig. 2(a). When the proposed on-demand selector of our softwarized network controller identifies that the mmWave-enabled D2D nodes are performing NDS tasks, it chooses contextual and non-contextual budget constrained MABs [3].

In the scenario of Fig. 2(a), a trade-off exists between investigating more nearby devices for increasing the spectral efficiency of the D2D link and decreasing its achievable throughput due to extensive beamforming training overhead. Conventionally, direct NDS is employed by examining all available nearby devices by the central one, as shown in the figure, to select the highest spectral efficiency device. However, this scheme suffers from a low achievable throughput. Also, it does not consider the limited battery capacity of the nearby devices when the residual energy of the selected nearby device should be reserved only for its essential activities. The network controller, then, needs to select and deploy an optimal algorithm depending on the specific needs of this scenario. Based on the comparative scenarios and their corresponding policies learned by the controller as listed in Table I, the proposed on-demand policy selector identifies the budget-constrained single player MAB to optimally address this problem. The central device acts as the bandit player that targets maximizing its achievable spectral efficiency, which behaves as the reward of the bandit, by utilizing the nearby devices as the arms of the bandit. By using multi-band standardized WiGig devices, i.e., containing both 2.4/5 GHz WiFi and 60 GHz mmWave bands as shown in Fig. 2(a), WiFi contexts can be used to further enhance the mmWave D2D NDS process. This is empowered by the direct relationship between WiFi and mmWave link statistics. Thus, based on Table I, our proposed on-demand policy selector adopts the linear Contextual MAB algorithm, called LinUCB. To optimize this algorithmic choice, it also adopts two variants of a budget-constrained, non-contextual MAB implementation, referred to as Energy-Aware UCB (EA-UCB) and Energy-Aware LinUCB (EA-LinUCB) policies. The WiFi contexts include the instantaneous value of the WiFi Received Signal Strength (RSS) in addition to its mean and variance up to time t when the central device performs NDS.

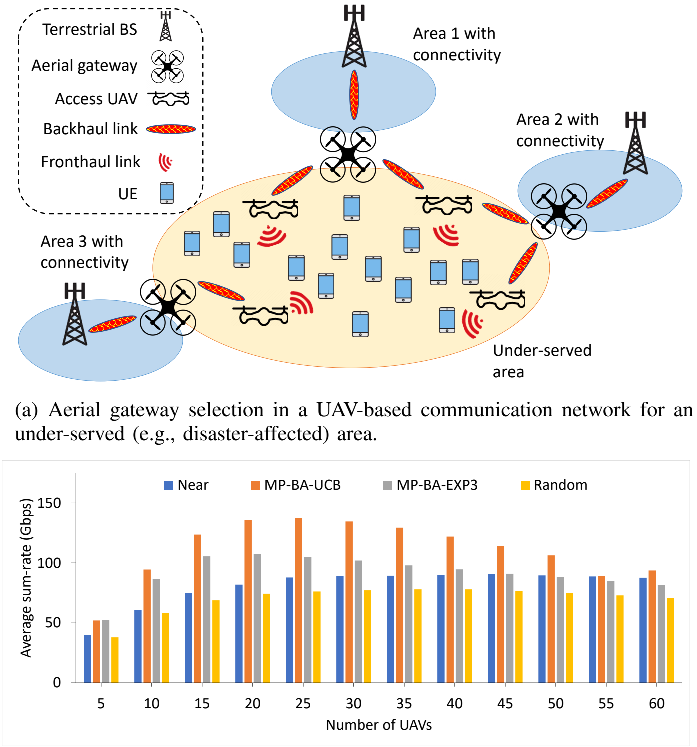


(a) Considered D2D network scenario where NDS is performed by the various dual-band devices (referred to as "Dev"s) in presence of blockers.

(b) Average throughput comparison of MAB algorithms in mmWave NDS at no blockage.

Fig. 2. Scenario 1: D2D-based neighbor discovery service. The results indicate that the proposed on-demand selector chooses LinUCB, which exhibits near-optimal performance for growing numbers of devices, and outperforms conventional and random methods besides other MAB variants.

Next, we describe our simulation results with the aforementioned MAB adoption in this scenario. In the conducted simulations, 20 to 100 dual-band devices are uniformly distributed around the central device in an area of $125 \times 125m^2$. Also, a perfect mmWave beam alignment is assumed. The TX (transmission) power for mmWave and WiFi modules are considered to be 10 and 20 dBm, and their operating frequencies are set to 60 GHz and 5GHz with 2.16 GHz and 40 MHz bandwidths, respectively. The beamforming training time is set to 0.28 s while a time horizon of 1000 is assumed. The path losses (standard deviation) for WiFi and mmWave LoS (Line of Sight) and NLoS (Non Line of Sight) paths are set to 2.32 (6), 2.22 (10.3), and 3.88 (14.6) , respectively. The initial energy levels of the devices are randomly selected in the range of [0.1, 1] Joule, and the noise power is set to -70 dBm. Fig. 2(b) demonstrates the average throughput of the compared



(a) Aerial gateway selection in a UAV-based communication network for an under-served (e.g., disaster-affected) area.

(b) Average sum-rate for different numbers of access UAVs using 20 aerial gateways and 60° beamwidth.

Fig. 3. Scenario 2: aerial gateway selection in UAV-based communication for an under-served area. The results illustrates why proposed on-demand selector chooses MP-BA-UCB, which outperforms conventional (near) and random methods as well as another MAB variant (MP-BA-EXP3) for various numbers of UAVs.

MAB schemes in addition to the conventional direct NDS as well as the random NDS policy. The results in Fig. 2(b) indicate that the average throughput of the conventional NDS decreases with the growing number of nearby devices. This happens due to the extensive beamforming training overhead. The compared MAB algorithms are not only known to have a fast convergence compared to the optimal and conventional methods, but also experience the lowest beamforming training overhead. This is because they need to investigate only a single nearby device at a time that significantly improves the achievable average throughput. Thus, the results in Fig. 2(b) elucidate that the on-demand policy selector chooses LinUCB since it knows from the pool of MAB variants in Table I that it demonstrates near-optimal performance for a growing number of devices, and also outperforms conventional and random methods along with the other energy-aware MAB variants.

B. Scenario 2: Aerial gateway selection for UAV-based communication network in an under-served area

To further extend the communication range, in the second scenario of our use case depicted in Fig. 3(a), the formation of a UAV-based wireless communication network is assumed for an under-served area. The distributed UAVs are split into access nodes and aerial gateways. The access nodes are used to perform data collection from the under-served users. Furthermore, the aerial gateways relay the collected data from access UAVs to the nearest terrestrial BS. Thus, each access UAV should select and fly towards an aerial gateway to maximize its achievable throughput while minimizing its flight energy. The proposed on-demand selector identifies the multi-player (MP) budget-constrained MAB to optimally address this issue. In this scenario, the access UAVs act as the players of the MAB game to maximize their achievable throughputs acting as the rewards of bandit game, while aerial gateways are considered as the bandit-arms. Moreover, each access UAV selfishly plays the game as information is neither available beforehand nor exchanged among the distributed UAVs in such a fully decentralized setting. To implement the specific policies for the multi-player budget-constrained MAB, two Battery-Aware (BA) policy implementations using UCB (stochastic type) and Exponential-weight algorithm (EXP3) (adversarial type) are adopted [5], referred to as MP-BA-UCB and MP-BA-EXP3, respectively. Based on the past history, the proposed on-demand policy selector identifies the MP-BA-UCB as the viable policy to be rapidly deployed to the UAV nodes for an effective aerial gateway selection. By only observing their achievable throughputs, access UAVs can learn the interference/collision patterns and enhance their gateway UAV selections while the game continues to execute. For the simulation setup, a post-disaster area of dimension $750 \times 750m^2$ is assumed where access UAVs are uniformly distributed inside this area for rescue services. The gateway UAVs are uniformly distributed around this area in a circle of 1250 m diameter. The mmWave TX power is set to 10 dBm with 60 GHz central frequency and 2.16 GHz bandwidth. The noise power is set to -120 dBm. The hovering and flying engine powers of the UAV are set to 4 and 2 watts, respectively. Moreover, the hovering time is set to 120 s, while the flying speed is adjusted to 40 Km/h. The total battery capacity of the access UAV is set to 400,000 Joules. Fig. 3(b) demonstrates the average system rate comparison among these two MAB variants and two benchmark algorithms, i.e., near and random aerial gateway selection schemes. In the near selection scheme, the closest aerial gateway is always chosen by the access UAV. In the random scheme, the aerial gateway selection is performed arbitrarily. Without any loss of generality, the mmWave communication links are assumed among the UAVs with a beamwidth of 60°, and 20 Gateway aerials are considered in this scenario.

In Fig. 3(b), when using a low number of access UAVs, the average system rate of the MAB algorithms increases. Then after reaching a certain point, it slightly drops with the growing number of access UAVs. This comes from the low interference experienced by the small number of access UAVs. However, as the number of access UAVs is increased beyond the number of ariel UAVs, i.e., 20 UAVs, a high interference is experienced by access UAVs. Note that MP-BA-UCB achieves much better performance than MP-BA-EXP3. The poor aerial gateway selection policy of MP-BA-EXP3 can be explained by its nearly equal weights assignment to the UAVs during each trial. To analyze the performance of the adopted MAB algorithms, the analysis of regret (the cumulative rewards of the best arm in hindsight) is useful; however this is beyond the scope of this article. For simplicity, consider that the algorithm to be successful if its regret is $\mathcal{O}(T)$ after T trials (meaning

TABLE II
EXECUTION TIMES OF COMPARED MAB ALGORITHMS IN THE TWO SCENARIOS.

Scenario 1 (mmWave-enabled D2D NDS)					**Scenario 2 (UAV-based network gateway selection)**		
Number of devices	**UCB**	**EA-UCB**	**Lin-UCB**	**EA-LinUCB**	**Number of UAVs**	**MP-BA-UCB**	**MP-BA-EXP3**
20	0.1 ms	0.1ms	0.3ms	0.3ms	10	1.5ms	1.7ms
40	0.1 ms	0.1ms	0.4ms	0.4ms	20	1.6ms	1.8ms
60	0.1ms	0.1ms	0.6ms	0.6ms	30	1.7ms	1.9ms
80	0.2 ms	0.2 ms	0.8ms	0.8ms	40	1.9ms	2ms
100	0.2 ms	0.2ms	0.9ms	0.9ms	60	2.0ms	2.1ms

that the average regret per trial converges to zero). The adopted MAB models with UCB or TS have $\mathcal{O}(logT)$ regret bounds.

The execution times of the MAB-based approaches utilized in the two scenarios are summarized in Table II. We recorded the MATLAB R2020 b execution time of these MAB techniques against different numbers of devices and UAVs, respectively. The utilized machine specifications consist of an Intel core i7-8565U CPU (Central Processing Unit) and 8 GB (Giga Bytes) RAM (Random Access Memory). From the table, execution times of the proposed algorithms are within milliseconds range suited to 6G millisecond latency. On the other hand, when traditional Integer Linear Programming (ILP)-based optimization solver is used for the mmWave topology, the execution time is in the order of seconds for 10 nodes and is not attempted for higher number of nodes in [14]. Similarly, for scenario 2, the execution time of a traditional application of an optimization algorithm also results in execution times of order of seconds and exponentially increases for an increase of coverage areas [15]. Thus, the high execution times of the traditional optimization algorithms reflect a high time complexity in contrast with those of the considered MAB techniques. This corroborates with our proposed optimal policy selector's choice of deploying the aforementioned MAB models for fast, sequential decision making in the considered scenarios.

V. CHALLENGES AND FUTURE DIRECTIONS

In this section, we describe the challenging research topics for further investigation in softwarized intelligence, particularly using ultra-fast policy selection and deployment in 6G networks, by exploiting MAB frameworks and similar sequential/online algorithms.

- Resource allocation in Low Power Wide Area Network (LoRAWAN) networks, Non-Orthogonal Multiple Access (NOMA), underwater relay selection and routing, and underlay D2D communication can greatly benefit from using relevant MAB variants to suit a diverse range of scenarios. Also, the on-demand controller's decisions in the proposed softwarized network can be greatly improved for routing information by utilizing MABs with ideal policies.
- Reflecting Intelligent Surfaces (RIS) and meta-learning are two attractive, relatively new areas where sequential learning algorithms may be highly effective. RIS is a two-dimensional surface composed of several passive reconfigurable meta-material elements, which reflect the incident signal by introducing a controllable phase shift. Recently, RIS has been used in communication networks for several purposes, including coverage enhancement, relaying, and physical-layer security. However, there are several challenges for achieving the potential of such structure, where AI-based techniques, particularly on-demand deployment of MAB models, can help learn the proper phase shift for each element under the discrete-phase shift assumption.
- Meta-learning may be a helpful technique in inductive bias's selection automation. It leverages valuable information or active observations from tasks that are expected to be related to the future tasks of interest. Such learning can facilitate the AI model training with a significantly lower amount of data and time. Hence, the meta-learning policies for MAB may be effectively leveraged for satisfying the ultra-low latency requirement of 6G network nodes.

VI. CONCLUSION

Intelligent decision-making is anticipated to be a key embedded feature in the upcoming 6G networks that will realize innovative future applications. Since these services have ultra-reliable requirements easily impacted by varying network dynamics, on-demand ultra-fast learning techniques emerge as a formidable research challenge. In this article, we addressed this challenge, and proposed a softwarized network consisting of an on-demand policy selector that considers the ongoing network dynamics and accordingly chooses the best intelligence module for that particular network setting. Unlike the classical optimization and supervised learning methods, online/sequential learning techniques such as MAB algorithms with different policies, were illustrated to be viable sequential learning techniques by the proposed on-demand selector for 6G node deployment. A use case with two scenarios was presented comprising NDS in a D2D network and aerial gateway selection in a UAV network, respectively. Extensive computer-based simulation results demonstrated that the selected MAB variant for both scenarios significantly outperforms both the conventional techniques and other MAB variants. Thus, the reported results clearly indicate the optimal policy selection capability of our proposed on-demand selector. As a caveat, it is worth noting that for deploying the models in an on-demand manner, there could be some connectivity issues causing the AI models not to be timely updated that may cause the target routers/network nodes to be rendered dysfunctional.

To combat such a corner case, we may assume a default, basic functionality of programmable routers to cope with such scenarios. How to optimally generalize such a default functionality is left open as a future work for 6G softwarized networks and programmable routers.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Numbers JP21K14162 and JP22H03649 and also supported by the National Science Foundation of the USA under Award No. 2210252 and Prince Sattam bin Abdulaziz University, KSA, project No. (PSAU/2023/R/1444).

References

[1] W. Wu, N. Cheng, N. Zhang, P. Yang, W. Zhuang, and X. Shen, "Fast mmwave beam alignment via correlated bandit learning," *IEEE Transactions on Wireless Communications*, vol. 18, no. 12, pp. 5894–5908, 2019.

[2] X. Pan, X. Wang, B. Tian, C. Wang, H. Zhang, and M. Guizani, "Machine-learning-aided optical fiber communication system," *IEEE Network*, vol. 35, no. 4, pp. 136–142, 2021.

[3] S. Hashima, K. Hatano, H. Kasban, and E. Mahmoud Mohamed, "Wi-Fi assisted contextual multi-armed bandit for neighbor discovery and selection in millimeter wave device to device communications," *Sensors*, vol. 21, no. 8, p. 2835, 2021.

[4] E. M. Mohamed, S. Hashima, K. Hatano, S. A. Aldossari, M. Zareei, and M. Rihan, "Two-hop relay probing in WiGig device-to-device networks using sleeping contextual bandits," *IEEE Wireless Communications Letters*, vol. 10, no. 7, pp. 1581–1585, 2021.

[5] E. M. Mohamed, S. Hashima, A. Aldosary, K. Hatano, and M. A. Abdelghany, "Gateway selection in millimeter wave UAV wireless networks using multi-player multi-armed bandit," *Sensors*, vol. 20, no. 14, p. 3947, 2020.

[6] D. Li, S. Wang, H. Zhao, and X. Wang, "Context-and-social-aware online beam selection for mmwave vehicular communications," *IEEE Internet of Things Journal*, vol. 8, no. 10, pp. 8603–8615, 2021.

[7] Y. Xing, Y. Qian, and L. Dong, "A multi-armed bandit approach to wireless information and power transfer," *IEEE Communications Letters*, vol. 24, no. 4, pp. 886–889, 2020.

[8] J. Zhang, J. Tang, and F. Wang, "Cooperative relay selection for load balancing with mobility in hierarchical WSNs: A multi-armed bandit approach," *IEEE Access*, vol. 8, pp. 18 110–18 122, 2020.

[9] W. Xia, T. Q. S. Quek, K. Guo, W. Wen, H. H. Yang, and H. Zhu, "Multi-armed bandit-based client scheduling for federated learning," *IEEE Transactions on Wireless Communications*, vol. 19, no. 11, pp. 7108–7123, 2020.

[10] X. Xu, M. Tao, and C. Shen, "Collaborative multi-agent multi-armed bandit learning for small-cell caching," *IEEE Transactions on Wireless Communications*, vol. 19, no. 4, pp. 2570–2585, 2020.

[11] S. Ali, A. Ferdowsi, W. Saad, N. Rajatheva, and J. Haapola, "Sleeping multi-armed bandit learning for fast uplink grant allocation in machine type communications," *IEEE Transactions on Communications*, vol. 68, no. 8, pp. 5072–5086, 2020.

[12] C. Wang, J. Yang, H. He, R. Zhou, S. Chen, and X. Jiang, "Neighbor cell list optimization in handover management using cascading bandits algorithm," *IEEE Access*, vol. 8, pp. 134 137–134 150, 2020.

[13] H.-A. Tran, S. Souihi, D. Tran, and A. Mellouk, "MABRESE: A new server selection method for smart SDN-based CDN architecture," *IEEE Communications Letters*, vol. 23, no. 6, pp. 1012–1015, 2019.

[14] G. H. Sim, M. Mousavi, L. Wang, A. Klein, and M. Hollick, "Joint relaying and spatial sharing multicast scheduling for mmWave networks," in *2020 IEEE 21st International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM)*, 2020, pp. 127–136.

[15] J. Sabzehali, V. K. Shah, Q. Fan, B. Choudhury, L. Liu, and J. H. Reed, "Optimizing number, placement, and backhaul connectivity of multi-UAV networks," *arXiv preprint arXiv:2111.05457*, 2021.

Biographies

Sherief Hashima is currently a postdoctoral researcher with the computational learning theory team, RIKEN-AIP, Japan. He also holds the position of Associate Professor with the Department of Engineering and Scientific Equipment, Nuclear Research Center (NRC), Egyptian Atomic Energy Authority (EAEA), Egypt. He was a visiting researcher at EJUST Center, Kyushu University, Japan. His research interests include wireless communications, machine learning, online learning, Massive MIMO, B5G, and 6G systems, image processing, millimeter waves, and Internet of things. He is a technical committee member in many international conferences and a reviewer in many international conferences, journals and transactions. He is an IEEE senior member and AAAI member.

Zubair Md Fadlullah [M'11, SM'13] is currently an Associate Professor with the Computer Science Department, Lakehead University, and a Research Chair of the Thunder Bay Regional Health Research Institute (TBRHRI), Thunder Bay, Ontario, Canada. He was an Associate Professor at the Graduate School of Information Sciences (GSIS), Tohoku University, Japan, from 2017 to 2019. He received his Ph.D. degree in Information Sciences from Tohoku University, Japan in 2011. His main research interests are in the areas of emerging communication systems, UAV based systems, smart health technology, cyber security, game theory, and smart grid.

Mostafa M. Fouda [M'11, SM'14] is currently an Assistant Professor with the Department of Electrical and Computer Engineering at Idaho State University, ID, USA. He also holds the position of Associate Professor at Benha University, Egypt. He received his Ph.D. degree in Information Sciences from Tohoku University, Japan in 2011. His research interests include cyber security, machine learning, IoT, and 6G networks. He has served on the technical committees of several IEEE conferences. He is also a Reviewer in several IEEE Transactions and Magazines. He is an Editor of IEEE Transactions on Vehicular Technology (TVT) and an Associate Editor of IEEE Access.

Kohei Hatano received Ph.D. from Tokyo Institute of Technology in 2005. Currently, he is an associate professor at Faculty of Arts and Science in Kyushu University. He is also the leader of the Computational Learning Theory team at RIKEN AIP. His research interests include machine learning, computational learning theory, online learning and their applications

Ehab Mahmoud Mohamed (Member,IEEE) received the B.E. and M.E. degrees in electrical engineering from South Valley University,Egypt, in 2001 and 2006, respectively, and the Ph.D. degree in information science and electrical engineering from Kyushu University, Japan, in 2012. From 2013 to 2016, he has joined Osaka University, Japan, as a Specially Appointed Researcher. Since 2017, he has been an Associate Professor with Aswan University, Egypt. He has also been an Associate Professor with Prince Sattam Bin Abdulaziz University, Saudi Arabia, since 2019. His current research interests include 5G, B5G and 6G networks, cognitive radio networks, millimeter wave transmissions,Li-Fi technology, MIMO systems, and underwater communication. He is a technical committee member of many international conferences and a reviewer of many international conferences, journals, and transactions. He is the General Chair of the IEEE ITEMS'16 and IEEE ISWC'18.

Basem M Elhalwany (Senior Member, IEEE) received the master's degree from Benha University, Banha, Egypt and the Ph.D. degree from Egypt-Japan University of Science and Technology, New Borg El Arab, Egypt, in 2011 and in 2014, respectively, both degrees in electronic and communication engineering. He is an Associate Professor with the Faculty of Engineering, Shoubra,Benha University. He has authored or coauthored more than 40 high-quality research papers in international leading journals and primer conferences. He was a Research Fellow with Smart Sensing and Mobile Computing Laboratory, Shenzhen University, Shenzhen, China, and EJUST Center, Kyushu University, Fukuoka, Japan. His research interests include performance analysis, resource management, and optimization in wireless networks, NOMA, and machine learning applications in communication. Dr. ElHalawany is a Technical Committee Member in many international conferences and a Reviewer in many international conferences, journals, and transactions.

Mohsen Guizani [S’85-M’89-SM’99-F’09] received the B.S. (with distinction) and M.S. degrees in electrical engineering, the M.S. and Ph.D. degrees in computer engineering from Syracuse University, Syracuse, NY,USA, in 1984, 1986, 1987, and 1990, respectively.He is currently a Professor at the Computer Science and Engineering Department in Qatar University,Qatar. Previously, he served in different academic and administrative positions at the University of Idaho, Western Michigan University, University of West Florida, University of Missouri-Kansas City, University of Colorado-Boulder, and Syracuse University. His research interests include wireless communications and mobile computing, computer networks, mobile cloud computing, security, and smart grid. He is currently the Editor in-Chief of the IEEE Network Magazine, serves on the editorial boards of several international technical journals and the Founder and Editor-in-Chief of Wireless Communications and Mobile Computing journal (Wiley). He is the author of nine books and more than 600 publications in refereed journals and conferences. He guest edited a number of special issues in IEEE journals and magazines. He also served as a member, Chair, and General Chair of a number of international conferences. Throughout his career, he received three teaching awards and four research awards. He is the recipient of the 2017 IEEE Communications Society Wireless Technical Committee (WTC) Recognition Award, the 2018 AdHoc Technical Committee Recognition Award for his contribution to outstanding research in wireless communications and Ad-Hoc Sensor networks and the 2019 IEEE Communications and Information Security Technical Recognition (CISTC) Award for outstanding contributions to the technological advancement of security. Hewas the Chair of the IEEE Communications Society Wireless Technical Committee and the Chair of the TAOS Technical Committee. He served as the IEEE Computer Society Distinguished Speaker and is currently the IEEE ComSoc Distinguished Lecturer. He is a Fellow of IEEE and a Senior Member of ACM.418