# A Fast Machine Learning for 5G Beam Selection for Unmanned Aerial Vehicle Applications

# A Fast Machine Learning for 5G Beam Selection for Unmanned Aerial Vehicle Applications

Wasswa Shafik
Computer Engineering Department, Yazd University, Yazd, Iran
wasswashafik@stu.yazd.ac.ir
S. Mojtaba Matinkhah*
Computer Engineering Department, Yazd University, Yazd, Iran
matinkhah@yazd.ac.ir
Mohammad Ghasemzadeh
Computer Engineering Department, Yazd University, Yazd, Iran
m.ghasemzadeh@yazd.ac.ir

Received: 04/Mar/2019 Revised: 24/Aug/2019 Accepted: 08/Nov/2019

**Abstract**
Unmanned Aerial vehicles (UAVs) emerged into a promising research trend applied in several disciplines based on the benefits, including efficient communication, on-time search, and rescue operations, appreciate customer deliveries among more. The current technologies are using fixed base stations (BS) to operate onsite and off-site in the fixed position with its associated problems like poor connectivity. These open gates for the UAVs technology to be used as a mobile alternative to increase accessibility in beam selection with a fifth-generation (5G) connectivity that focuses on increased availability and connectivity. This paper presents a first fast semi-online 3-Dimensional machine learning algorithm suitable for proper beam selection as is emitted from UAVs. Secondly, it presents a detailed step by step approach that is involved in the multi-armed bandit approach in solving UAV solving selection exploration to exploitation dilemmas. The obtained results depicted that a multi-armed bandit problem approach can be applied in optimizing the performance of any mobile networked devices issue based on bandit samples like Thompson sampling, Bayesian algorithm, and ε-Greedy Algorithm. The results further illustrated that the 3-Dimensional algorithm optimizes utilization of technological resources compared to the existing single and the 2-Dimensional algorithms thus close optimal performance on the average period through machine learning of realistic UAV communication situations.

**Keywords:** Unmanned Ariel Vehicle; Multi-Armed Bandit; Reinforcement Learning Algorithms; Beam selection.

## 1- Introduction

The dynamics and advances in technology in respect to scientific studies have emerged and influenced a number of fields for instance distributed compressed sensing [1], text mining applications [2], besides its increased manifestation in wireless, semi-online and online application, in particular, Unmanned Aerial Vehicles (UAV).

Unmanned Ariel vehicles (UAVs) are aircraft deprived of an anthropological pilot either onboard or off broad and a type of unmanned vehicle and are components of an unmanned aircraft system, to mention a UAV, a ground-based controller, and a system of communications. UAVs have got increased usage in a number of different industries like agri-businesses for precision agriculture operations, thermal imaging in road patrolling, path planning, rescue departments for post-natural disaster missions, targets monitoring, crack assessments among other notable benefits as well [3]- [7].

Machine learning entails quite numerous categories for example supervised learning, unsupervised learning, semi-Supervised learning and reinforcement learning. This reinforcement learning appears in two categories that is to say negative and positive with unties like Markov Decision Process-learning among others. Reinforcement learning is specified under simple reinforcement learning and the deep reinforcement learning having deterministic policy gradient algorithms simultaneously learn a Q-function and a policy through the use of Bellman equation given that Deep Q-Networks modernizes the Q-value function of a state for a specific action simply; this proposed model uses a simple reinforcement learning and multi armed bandit not deep reinforcement learning.

Fast machine learning techniques confirm an enhanced potential in learning patterns and extracting attributes from a complex dataset includes the use of other learnings like deep learning for pattern recognition and medication innovation, although others do surface explicit encounters

* Corresponding Author

that ought to be worked on embrace anthropologies in sensor synthesis, associate medicinal analyses, and experimental decision-makings [9].
In appreciation of the models availed in problem analysis, we are proposing an alternative way to problem attack to network issues analysis to the identified medical network scenarios basing on the existing studies, models, and architectures, enlighten on the medical application illustrated above. A simplified mathematical expression of the armed bandit scenarios has been used, validated, and, proofed calculations together with the illustration presented throughout this paper in comparison to the state-of-the-art models.

Some interesting questions are that machine learning tackles: Are there any correlations between crowdsourcing annotations with expert measurements to feed the fast machine learning training algorithms with quit satisfactory reliability? Can we use a hierarchical feature selection method for cancer detection? How we can make the multi-label biomedical compound more efficient? Is there any way to produce useful results by the use of computer games in embedding human intelligence-based tasks to train fast machine learning algorithms? Learning the lower-staged configuration of inventive data to achieve a more intangible portrayal is the central awareness of fast machine learning. How we design algorithms to implicitly capture the intricate associations and topographies of the larger-scale feedbacks?

The non-orthogonal schemes with multiple access capabilities for the 5G UAV social network that encompasses medical as well where they referred to the social network as communication. The author depicts that the systems of multiple access with no orthogonal ranges overtake extra numerous access of the same scheme capacities in the relation of summation capability, dynamism efficiency, and social phantom effectiveness. This accomplishes superior sum-rate on a subordinate loftiness that diminishes the inclusive dynamism outflow of the fifth generation UAV network.

The paper in the rest parts are structured as the following: Section 1.1 presents the motivation for this paper depicting early approaches and clearly showing the requisite for this study. In section 2, the study avails research literature focusing mainly on the UAVs' technological usage. In section 3, the detailed multi-armed bandit problem approach in comparison to other approaches in beam selection, in particular, we considered Thompson sampling, ε-greedy algorithm, the Bayesian Upper Confidence bounds bandits. Section 4 presents simulated results and discussion of the provided results provided. In section 5, we explicitly our simulation parameters and detailed discussion of beam selection and section 6 hold the conclusion of this article.

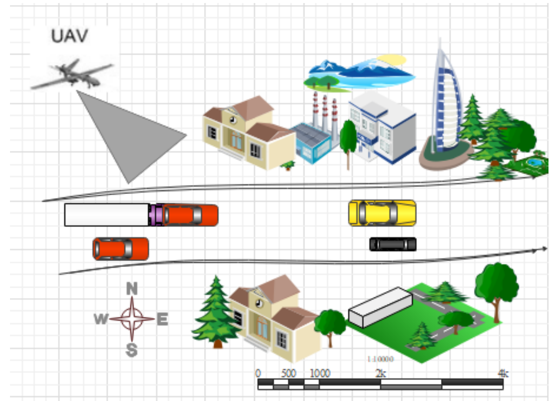


Fig. 1 The Simulation Atmosphere. The trolley and bus momentarily block the red car and black car correspondingly.

## 1-1- Motivation

UAVs preserves on advancing in capabilities and efficiency to increase its rapid evolution in different areas like medical, engineering, profitable, and entertaining applications, that is to say, aptitude to gather real-time data at cost-effectiveness, on-time delivery, on-time payloads deliveries. Regardless it's the sluggish extension in the multi- fields, some of the state-of-the-art studies do include a context-aware communication link [8], Improved communication security of UAVs [9], reduced communication delays through formation control [10], Multi-objective Placements [11], improved UAV wireless communication [12], mention but a few.
Based on the identified presented studies, it has been further identified that the beam that is emitted has not been given clear technological attention since current studies some of which focus on synthesis methods of the synchronized assemblage differential pulse-code modulation codec for UAV communication schemes [13], more, the application of UAV in communal protection communicating and optimizing of reportage [14]. Having discovered that beams emitted from the UAV stochastically and technologically behave like bandits, this paper, therefore, presents a multi-armed bandit' approach is selecting the best armed per the desire of the operator that has the approach to beam selection of UAVs.
In appreciation to the application of machine learning, the number of study analysis have been presented given a dynamic technological advancement, they led to the presence of the noise-resistant surface defect recognition tactics [15], color texture classification and identifications to solve difficulties involved in the texturing in computing to obtain better accuracy [16], according to the detailed study on learnings in remote sensing that categorizes the UAVs too included the concepts, apparatuses, and

encounters for the like a perceptive assignment investigation for emerging unmanned aerial vehicle boonies rifle sustenance [17]. Technologies involving expertise development monitoring, piloting tasks are seen in medical as well with optical brain imaging in conjunction with UAVs [18].

## 2- Research Literature

This section presents a general analysis of the proposed UAV models and frameworks that have been claimed in data gathering tasks, secure remote connection of social media platforms among others.

### 2-1- UAV Technological Usage

The 5G technologies are the subsequent cohort of wireless communication, submission earlier swiftness, and more consistent acquaintances on smart medical devices and supplementary devices than interminably before. The convergence of multiple networking functions to achieve charge, power, and complexity reduction is one of the promising advantages of 5G. The 5G expected to assist power an enormous intensification in the IoT technologies, in case of the arrangement compulsory to transfer mammoth quantities of data to have a smart world. The 5G involves multiple entities with different ideas like the extraordinary record, truncated expectancy, effectiveness enlargement, and yield extension.

### 2-2- Search and Rescue Operations

Murphy et al. [19] proposed autonomous UAV can find missing people using the broadcasting motions fashioned by the comprehensive schemes for itinerant communications booth. The UAVs act as a universal mobile communications entity BS and persuade the disappeared individual's maneuver to effort to variety communication. They recycled a constraint-centered topographic-based path arrangement tactic to harvest a course for the unmanned aerial vehicles to navigate in the inflight transient finished predictable signals since a huge quantity of conceivable spring settings for abundant resolutions; this assisted the telephone enactment accordingly prominent the request of augmented complexity, and efficiency.

### 2-3- Smart City Scenarios

The need to automate this data gathering practice, a network of UAV is a suitable option for a vehicle are the inspiration for this paper through performance analysis of 5G network and beyond. Authors to mention, El-Sayed et al. [20] deployed UAVs performing like moveable edges following traffic proceedings and overcrowding situations prominent to the self-motivated attitudes, amplified mobility. The proposed traffic-awareness technique empowered the distribution of unmanned aerial vehicles in vehicular situations demonstrated the proposed technique to accomplish jam-packed social net analysis under diverse situations deprived of extra communication delay. This approach opened some technical questions that were not elaborated for instance what could the proposed model do in times of complex traffic from a big smart perspective [30].

### 2-4- A System with Pre-Computed Pre-coders

Meticulously within this subsection, M. Meng et al. [21] collective through the compensations of a fifth-generation mm-Wave detector, machine learning approaches with more methodological developments, for example, the multiple inputs multiple out to identify the "black flying" UAV. An operational resolution to disentangle the delinquent of distinguishing and categorizing a UAV by the fifth-generation mm-Wave detector of the Internet of Things, which has extraordinary applied solicitation assessment.

The existing technology improved the accessibility of social networks, rapid rate, nonetheless the next fifth-generation (5G) is anticipated to increase connectivity and increased data rate that multiples the existing one thousand times regardless of the location and data size. The advance in technology still has a deficiency of proper infrastructure as services that can accommodate all possible scenarios of mobility, this contributed mainly to its supposition in this study.

After the close discovery that numerous artificial 1ntelligence applications are emerging entailing UAVs with artificial capabilities including knowledge engineering, edge mobility performance assessments, a number of reinforcement learning issues agent Q-learning, Learning Automata and many more [22]–[24], numerous studies have focused on low complexities leaving a group of the multi-cells. Scrutinizing the intricate, vibrant relations of the number of dynamic drifts in the numerous booths are acquaint with the resilient control of interfering amongst immediate BS [25].

Furthermore, in case the non-orthogonal learning classifications are recycled of uplink learning, it is revealed that the precoding environment recycled by the BS in single-cell converting to besmirched through the station amongst that base station and the users in extra cells [26]. Nevertheless, in multiple cells do support the customer tools necessities to assess the conduit public data besides nourish them spinal to the BS schedulers for adaptabilities reserve supervision. This leads to a substantial rise of gesticulating overheads and feedbacks expectancy addicted to the complaisant networks of UAVs.

## 3- Multi-Armed Bandit and Beam Selection

This section presents the multi-armed bandit approach and most related arenas like Thompson sampling, ε-greedy algorithm, and Bayesian Upper confidence bounds (UCB) algorithm in solving medical devices. The MAB problematics are decisive problems in nature that well reveals the survey against the utilization quandary of medical computation. There exist numerous techniques to adopted using MAB containing no exploration connotation that the maximum unexperienced approach in 2-dimensional; vigorously this study follows but in 3-dimensional approach optimization per the model explained, according to 22 and 23 of the pseudo [28].

This concept proves that j-armed problems associated with the bandit might be elucidated by disentangling j-armed glitches. Contemplate a chronological verdict delinquent in that at correspondingly stage there are j potential actions, optimal action of j results in an estimation being occupied from the $j^{th}$ experimentation and the received the mathematical assessment of this observation as a reward. The annotations taken may provide useful information in forthcoming choices of actions. To exploit the contemporaneous assessment of the infinite tributary of rewards received, reduced in an approximately. The Bandit is attained from demonstrating difficulties, for example, an *i*-armed bandit, that is a slot engine with *i* arms, each subsequent in a mysterious, perchance divergent scattering of payoff s.

It is quite challenging to notice which arm (UAV beam) provides the utmost average return in technologies like determining the rates. However, by playing the numerous arms of the slot engine, the information on which arm is best may be obtained. Nonetheless, the observations taken to use to have information are similarly users' rewards. Striking equilibrium among gaining rewards with acquisition information a case, for instance, it is not garbed to uninterruptedly wrench the arm that has achieved best in the previous; subsequently, it might have to be situated that individual was impartial unfortunate with the preeminent arm.

Classically in this kind of problem as wireless communication to obtain, there is a time of gaining information, monitored by a period of narrowing down the arms, monitored by a period of playing the arm to be the best. More important modeling to illustrate these problems comes from operational trials in which there are p drones for a given mission.

It is assumed that response from the operation is instant so that the UAV efficiency based on the mission that the present UAV does an operation that is acknowledged once the other UAV requisite is preserved. It is in fact not acknowledged exactly which unique of the operational missions is best; nonetheless decide which operations to give each drone, remembering that the primary goal is to serve as many UAVs as possible. This may require a drone mission procedure that is not the one that seems utmost at the existing interval to attain info that might be of monotonous to forthcoming UAVs.

Assume that $y$ reward distributions be denoted by $D_1(v \mid \vartheta_1),....,D_y(v \mid \vartheta_y)$ where $\vartheta_1,...,....,\vartheta_y$ are parameters considering that the tenets are not identified exactly, nonetheless since combined erstwhile scattering is to be recognized and the study denotes it to be $P(\vartheta_1,...,...,\vartheta_y)$. Primarily, action $\theta_1$ will be chosen from a provided set $\{1,...,...,y\}$, observation $0_1$, the recompense for the initial phase, is taken from the scattering $D\theta_1$ might also be based on this information, action $\theta_2$ of the same action space, and an observation $0_2$, taken from $D\theta_2$. Let us also assume that given $\theta_m$, the parameters $\theta_1,...,...,\theta_y$, $0_m$ to be chosen from $D\theta_m$ autonomously of the previous. The choice instruction for this issue is a categorization $S=(\theta_1,\theta_2,\theta_3,....,...)$ of utilities adjusted to the interpretations; that is to say, it may be contingent on earlier engagements and interpretations putting in mind that the denotation of $0_m$ shows both past action and observation leading.

$$\theta_m(\theta_1,0_1,\theta_2,0_2,....,.....,....,\theta_{m-1},D_{m-1}) \quad (1)$$

There are possibilities discounted sequences, let us denote it to be K, $K=(\beta_1,\beta_2,....,...)$ such that the *i*th observation is discounted $\beta_i$ were $0 \le \beta_i \le 1$ for. The maximum expected rewards are presented as $E\sum_i^{\infty}\beta_i.0_i$ since the complete discounted return is $\sum_i^{\infty}\beta_i.0_i$. This delinquent is christened to be *i*-armed MAB problem continuously yields the identical recognized amount that is, the circulation D connected with one of the arms is debauched at an identified relentless.

Therefore, to accomplish a finite rate for the predictable remuneration, Let us adopt

(a) each distribution, $D_i$ will be $i=1,...,...,y$, has finite first moment,

(b) also consider $\sum_i^{\infty}\beta_i$ p $\infty$, additional dualistic substantial circumstances of the discount sequence are

(c) the m-horizon unvarying reduction intended for which

$\beta_1 = .... = \beta_m = 1 \& \& \beta_{m+1} = \beta_{m+2} = .... = 0$ ,

(d) the geometric discount in which K, will be given by $K = (1, \beta, \beta^2, \beta^3, ....,...)$, logically meaning that $\beta_i = \beta^{i-1}$ for $i = 1, ..., ...,$ .

The payoff is basically $\sum_1^m 0_i$ the entirety of the very initial m clarifications. The problem converts a unique through a finite perspective which can in the method be disentangled by backward generation. Around is an interlude alteration in which the approaching m phases organized accomplishment at the twitch excluding for the alteration from the previous delivery to the advanced dissemination. The extravagance predominantly difficulties with symmetrical discount and autonomous arms, that is to say, prior scatterings P for that $\theta_1, ..., ..., \theta_y$ are autonomous accordingly that surveillance on the single-arm will not affect the acquaintance of the dissemination of slightly supplementary arm.

As an introductory to the explanation of the *i*-armed bandit with symmetrical reduction and autonomous arms, the study prerequisite to comprehend a modestly MAB as the nature of wireless complications is. The single or multi-armed bandit is really a bandit problem issue, but a single arm may be less useful during the dissemination of takings, besides so plays only an insignificant role. Firstly (1), the study further depicted how the MAB can be associated with a preventing rule problem.

Take responsibility that a prearranged arm has a concomitant categorization of unsystematic variables, $x_1, x_2, x_3, ....,....$ with recognized joint distribution satisfying $SUP_m X_m^+ \text{ p } \infty$. For the new arm, the returns are assumed firstly not considered from a known distribution with expectation $\lambda$ . The discount sequence in the case taken to be geometric P $P = (1, \beta, \beta^2, ....)$ $0 \text{ p } \beta \text{ p } 1$ was looking for a decision rule $\varphi M = (\theta_1, \theta_2, \theta_3)$ to maximize

$$\varphi(M) = E\left(\sum_1^\infty \beta^{i-1} 0_i \mid \varphi M\right) \quad (2)$$

The benefit of constructing this remark so that study might nowadays assume that the resolution imperative of $\varphi M$ ensures not be contingent over $0_i$ when $\theta_i = 2$ , since $0_i$ is acknowledged to stand $\lambda$ . Consequently, the study dons that the additional arm stretches a continuous reappearance of λ collectively interval it is dragged surveyed by (2).

First theorem: In case it is primarily optimum to habit the second arm in the sense that $SUP_{\varphi M} \varphi^* = \sup\{\varphi(\varphi M), \text{where } \theta_1 = 2\}$ , then it is optimal to use the second arm continuously and therefore the $\varphi M^* = \lambda / (1-\beta)$ . Second theorem: Consider that $\forall(\beta)$ designate the optimum degree of return for by means of the first arm at the concession $\beta$ . Let us prove all the assumptions in the fast lemma. Firstly, in the case $\in \text{f } 0$ to obtain the decision rule of M in that rule $\theta_1 = 2$, lastly $\varphi(M) \geq \varphi^* - \in$ this is given computed as

$$\varphi(\text{M}) = \lambda + \beta \,\text{E}\left(\sum_2^\infty \beta^{i-2} 0_i \mid M\right) = \lambda + \beta \,\text{E}\left(\sum_2^\infty \beta^{i-1} 0_{i+1} \mid M\right)$$

$$\lambda + \beta \,\text{E}\left(\sum_2^\infty \beta^{i-1} 0_i^{'} \mid M\right) = \lambda + \beta \,\text{E}\left(\sum_2^\infty \beta^{i-2} 0_i \mid M^1\right) \leq \lambda + \beta\varphi^*$$

Whereas the rule M shifted by 1, and $0_i^{'} = 0_{i+1}$ . Consequently, we have $\varphi^* - \in \leq \lambda + \beta\varphi^*$ subsequently $\in \text{p } 0$ subjective meaning, this implies

$$\varphi^* \leq \lambda / (1-\beta) \quad (3)$$

Nonetheless, this value obtained at (3) is achievable by using the second arm at the respective phase. This is also understood that theorem is likewise effective in the n-uniform reduction categorization. It is considered to be discount categorization P is supposed to be consistent in case it has a cumulative failure degree, i.e. in the case $\beta_m / \sum_m^\infty \beta_i$ is not decreasing on its definition domain. Notably, the above theorem is not factual for roughly reduction classifications is readily comprehended a case in the example in case $P = \{0.1, 1, 0, ..., ...,\}$ meaning that P is regular yet $x_1$ is exchangeable like at 10, other at 0 and $\lambda = 0$ formerly the solitary optimum stratagem is to track an initial wrench of that second arm with a wrench of the first arm. Accurately what assets of P is compulsory for the above theorem appears to be unidentified.

Considering lemma 2 nearby exist an optimum instruction for this problem, it is correspondingly the rule that customs second arm at completely stages, or the statute compatible to the terminating rule $L \geq 1$ that is optimum for the terminating rule delinquent with payoff and communicated as

$$H_m = \sum_1^m \beta^{i-1} x_1 + \lambda \sum_{m+1}^\infty \beta^{i-1} \quad (4)$$

$$\forall(\beta) = {}^{SUP}_{L\geq 1} \frac{E\left(\sum_1^L \beta^{i-1} x_i\right)}{E\left(\sum_1^L \beta^{i-1}\right)} \quad (5)$$

Therefore, the second arm (since arms have the same behavior, we consider arms to be the beams of the UAV) is optimal initially in case, $\lambda \geq \forall(\beta)$ let us prove the use first theorem and (4), it may perhaps contain the courtesy to pronouncement rules M specified by a discontinuing time L which characterizes the previous interval that first arm is recycled (5). The payoff s exhausting L tends to be

$$E\left(\sum_{1}^{L}\beta^{i-1}x_i+\lambda\sum_{L+1}^{\infty}\beta^{i+1}\right)$$

Where we considered $L=0$ producing $\lambda/1-\beta$. This implies that the second arm is optimal firstly in case all stopping rules are taken to be $L\geq 1$.

$$E\left(\sum_{1}^{L}\beta^{i-1}x_i+\lambda\sum_{L+1}^{\infty}\beta^{i-1}\right)\leq\lambda/(1-\beta)$$

$$E\left(\sum_{1}^{L}\beta^{i-1}x_i\leq\lambda E\sum_{L+1}^{\infty}\beta^{i-1}\right);\ E\left(\sum_{1}^{L}\beta^{i-1}x_i\,/\,E\sum_{;\,1}^{L}\beta^{i-1}\right)\leq\lambda$$

This is equivalent to $\forall(\beta)\leq\lambda$. The value $\forall(\beta)$ contingent only on $\beta$ and on the distribution of the returns from first arms $x_1,x_2,\ldots,\ldots$, everywhere, the situation signifies the indiff erence fact: the rate $\lambda$ for the second arm in the solitary of MAB that indiff erent amongst preparatory off s on the first arm and indicating the second arm wholly the segments. Let us yield to the *i*-armed outlaw by arithmetical reduction and self-determining arms obligating revenues symbolized by s

$$\text{First arm } x(1,1)\ \ldots\ldots x(1,2),\ldots,\ldots,\ldots \qquad (6)$$

$$\text{Second arm } x(2,1),\ldots\ldots,\mathrm{s}\,x(2,2),\ldots,\ldots \qquad (7)$$

$$\text{Arm } i,\ x(i,1),\ldots,,,x(i,2),\ldots(i,3),\ldots,\ldots$$

Suppose, the variables are reliant on flanked by commotions and that the first complete instants exist and are consistently constrained, $\sup_{i\geq 1,t\geq 1}E\,|\,x(i,t)|\leq\infty$ the deduction is $\beta$, where $0\leq\beta\leq 1$ so also considering (6) and (7), let us pursue a decision rule $\varphi=\theta_1,\theta_2,\theta_3,\ldots,\ldots,\ldots$. Therefore, to maximize the total discounted return it will be given by

$$\varphi(M)=E\left(\sum_{t=1}^{\infty}\beta^{t-1}0_t\,|\,M\right) \qquad (8)$$

For every arm (beam), computation of return ought to be articulated as

$$\forall_i={}^{SUP}_{L f 1}\ E\sum_{t=1}^{L}x(i,t)/E\sum_{t=1}^{L}\beta^{t-1}$$

$i=0,1,2\ldots\ldots,\ldots..$ Here, we suppress $\beta$ in the notation for $\forall\ \beta$ nice going to behold constant throughout.

Firstly (8), we proofed firstly the special case that around objectively two arms $(i=2)$ besides wherever altogether the arbitrary variables are decadent. We designate the outlays after the first arm to be $x(1),\mathrm{x}(2),\ldots,\ldots$, the second arm to be $z(1),\mathrm{z}(2),\ldots,\ldots$ among others. The stated two arms are bounded sequences of real numbers. For x, it will be expressed as

$$\forall_x={}^{SUP}_{i\geq 1}\ \sum_{1}^{i}\beta^{t-1}x(t)/\sum_{1}^{i}\beta^{t-1}\text{ and z}$$

$$\forall_z={}^{SUP}_{i\geq 1}\ \sum_{1}^{i}\beta^{t-1}z(t)/\sum_{1}^{i}\beta^{t-1}$$

Subsequently $x(t)$ assumed bounded, the series of $\sum_{i}^{m}\beta^{t-1}x(t)$ joins, consequently that there exists a value of $i$, possibly $\infty$ at which the supremum in the definition of $\forall_x$ is taken on as well us $\forall_z$. Suppose that $j$ is this value of $i$ so that $1\leq j\leq\infty$ considering this first lemma that the sequence of (6) is non-random and bounded. In case $\forall_x=\sum_{1}^{j}\beta^{t-1}x(t)/\sum_{1}^{j}\beta^{t-1}$, then for all $i\leq j$

$$\sum_{t=i}^{j}\beta^{t-1}x(\mathrm{t})\geq\forall_x\sum_{t-i}^{j}\beta^{t-1} \qquad (8)$$

And for j finite and $i$ f $j$, leading to

$$\sum_{t=j+1}^{j}\beta^{t-1}x(t)\leq\forall_x\sum_{t=j+1}^{j}\beta^{t-1} \qquad (9)$$

At this stage, we can now proof both (8) and (9) as

$$\sum_{t=i}^{j}\beta^{t-1}x(\mathrm{t})\geq\forall_x\sum_{t-i}^{j}\beta^{t-1},\text{and}\sum_{t=j+1}^{j-1}\beta^{t-1}x(t)\leq\forall_x\sum_{t=j+1}^{j-1}\beta^{t-1}$$

Subtracting the latter from the former contributes to (8) when i is less than or equal to j and gives (9) when this leads the equations to be simulated.

Let us use a Bernoulli MAB approach demonstrated as a point to $(\forall,\psi)$ given that a given medical device M obtains a reward probability $\{\theta_1,\ldots,\ldots\theta_{\mathrm{M}}\}$. At respectively interval phase $t$, we consider an accomplishment of a unique slit considered medical mechanism and obtain a remuneration of r. notably, $\forall$ are regularity of movements, independently mentioning to the interaction with one slot medical machine. The percentage of achievement is the predictable return, $Q(a)=E[r\,|\,a]=\theta$. In case the achievement *at* at the interval phase $t$ is on the $i^{th}$ medical machine, then $Q(a_t)=\theta_i$. the $\psi$ is a reward function. In the situation of Bernoulli MAB, during the study, distinguish a reward r in a *stochastic* approach was considered as well. At the stretch footstep $t$ $r_t=\psi(a_t)$, (at) may return reward 1 with a probability of 0 or other with it will be given by $Q(a_t)$. The aim here grows into to exploit the cumulative rewards computed as $\sum_{t=1}^{T}r_t$. Uncertainly the study demonstrated that it distinguishes the optimum accomplishment with the superlative reward, and then the aim is identical to minimalize the prospective regret without picking the optimal action.

This mainly means that the optimum reward probability $\theta^*$ of the optimal action $a^*$ will be depicted by

$$\theta^* = Q(a^*) = \underset{a \in \forall}{max}\ Q(a) = \underset{1 \le i \le M\theta_i}{max\,\theta_i}$$

The defeat utility $\xi$ is provided by the entire regrets that cannot be selected as the finest action up to the time step $T$

$$\xi = E\left[\sum_{t=1}^{T}(\theta * - Q(a_t))\right]$$

Another approach besides bandit, the ε-greedy algorithms exists to yield the superlative accomplishment utmost of the stretch, nonetheless guarantees unsystematic consideration infrequently. The accomplishment connotation is expectably bequeathing to the involvement by averaging the rewards interrelated to the objective achievement $a$ experimental to the current $t$ . Given that 1 is a binary pointer utility and $Nt(a)$ is the stretch of the stroke $a$ .

$$\hat{Q}_t(a) = \frac{1}{Zt(a)}\sum_{T=1}^{t} r_\tau 1\left[a_\tau = a\right] = Z_t(a) = \sum_{T=1}^{t} r_\tau 1\left[a_\tau = a\right]$$

The ε-greedy algorithm stress $\hat{a}_t^* = argmax\, a \in \forall \hat{Q}_t(a)$ es that means that a small probability $\epsilon$ considers a random action.

Table 1: Margin specifications

| Notations | Specification |
|---|---|
| Beta-Parameters $(\alpha = 1, \beta = 1)$ , $\alpha$ $\beta$ | Expected Reward Probability to be 50% successful beam Unsuccessful beam |
| Reward probabilities | {0.0, 0.1, 0.2, …, 0.9} |
| τ | Extensive negligible inputs |
| Device -system bandwidth | 1 GHz |
| M | 10 slot medical machines |
| $\alpha = 1000$ | Expected reward probability is 10% |
| $\beta$ =9000 | |
| solver Execution Times | 10000 steps |
| | ε-greedy algorithm |
| | UCB1 algorithm |
| | Thompson sampling |
| | Bayesian UCB Algorithm |
| 1st and 2nd solver Times | 500 steps |
| Simulator tool | MATLAB |

Bayesian UCB / UCB1 algorithm tends to have a different approach to the ε-greedy algorithm, assumes any prior on the reward circulation, and consequently depend on the variation of Hoeffding for an actual oversimplify guesstimate. Thompson sampling partakes an unpretentious awareness nevertheless then all excessive for responding to the MAB problem of devices where we select action an affording to the possibility that is optimal. Considering $\pi(a \mid hty)$ being the prospect of enchanting achievement a particular the antiquity $hty$ :

$$\pi(a \mid ht) = \text{¡}\ [Q(a) > Q(a'), \forall a' \ne a \mid hty]$$
$$= \text{£}\ R \mid hty[1(a = \underset{a \in \forall}{argmax}\ Q(a))]$$

Importantly, it is also regular to accept that $Q(a)$ follows the distribution of Beta in For the Bernoulli bandit, as $Q(a)$ is fundamentally the achievement opportunity $\theta$ in Bernoulli. It's important to note that all bandits have operation besides the assumption of the behaviors like Bernoulli, multi-armed among others.

A naive method can be used is to continue playing with one alternative for numerous rounds to ultimately approximate the "accurate" recompense chance rendering to the common edict of huge records in computations. Nevertheless, this is moderately extravagant, and confidently does not assurance the superlative lasting recompense as anticipated.

At every time $t$ , we model an expected reward, $\tilde{Q}(a)$ since the prior distribution Beta $(\alpha i, \beta i)$ for each exploit. The superlative accomplishment is nominated among trials after the true recompense is experimental, informs the Beta dissemination consequently, which is fundamentally doing Bayesian implication to calculate the subsequent with the identified previous and the probability of attainment the experimented information as $aTSt = argmaxa \in \forall \tilde{Q}(a)$ and thus

$$\alpha_i \leftarrow \alpha_i + r_t 1[aTSt = a_i]$$
$$\beta_i \leftarrow \beta_i + (1 - r_t)1[aTSt = a_i]$$

To diminish the brink $p$ in time, consider extra assertive destined appraisal with additional rewards experiential. Customary of $p = t^{-4}$ . This assumption and innovation are done by the UCB1 given $u$ as the UCB, $u = U_t(a)$

$$U_t(a) = \sqrt{\frac{2logt}{Z_t(a)}}, a_t^{UCB1} = \underset{a \in \forall}{arg\ max}\ Q(a) + \sqrt{\frac{2\log t}{Zt(a)}}$$

This depicts that in Upper Confidence Bound algorithm; continuously inferior the avaricious achievement to exhaust the potentials the UCB given by

$$a_t^{UCB} = \underset{a \in A}{argmax}\ \hat{Q}_t(a) + \hat{U}_t(a)$$

## 4- Simulation Results and Discussion

Within this section, the demonstrations of results and conversation of the conveyed simulation results in figure 2a, figure 2b, figure 3a, figure 3b, figure 4 and, figure 5 with the exhaustive explanation of the algorithm presented.

### 4-1- Model Description

The mUBS considerably use a predictable stand $\beta$ of $\mathrm{B}$ is denoted as $|\mathrm{D}|$ distinctive, not non-orthogonal beams. The study adopted that the mUBS can solitary inferior a subsection of *bm* concurrently given that $bm \in \mathrm{N}\, bm \leq \mathrm{B}$, is considered to be immovable amounts. Some of the limitations identified on mmWave channel sparsity. The main reason for the mUBS is to choose a subsection of *bm* that will be maximizing the number of records efficaciously traditional by the imminent CR in the reportage capacity. The study further assumed that mUBS is not certain or no nothing about then environment.In situation, the simplicity of the system execution reduces as the operative necessities nothing to be configured like at each mUBS according to the environment. Therefore, the mUBS has to learn as the situation changes to select the subset of the beams. In this way, the UAV will be to account for every approaching CR in context to beams it emits. The model similarly takes in mind a discrete-interval situation, given that the mUBS modernizes it is beam-ray miscellany in the regular time setup/period in each a setup $t = 0,1\ldots\ldots, T,$ given that $T \in N$ considered to be a finite horizon, the following three activities are applied.

i. Given that set $W_t$ = {*CRu,i*}*i=1,…,….,* $\vartheta_t$ of *CRt* = |*Nt* | CR resembling buses, different IoTs mong others to the mUBS. The number of *CRt* of the CRt of the CRs fulfills the condition that CRt ≤ Vmax, considering the Vmax ϵ N will be the maximum number of the supported CRs contained by the analysis capacity. At the time of the cataloging, mUBS will be having the capacity to receive info about the context *it,i* of every imminent automobile *CRu,i* will be a three-dimensional flight path taken from the bounded context space or coverage area $\mathrm{K} = [0,1,2]^k$.

ii. The mUBS chooses a detachment of *bm* beams. The study similarly denotes that the regular of selected beams in the period in period t by St = *{st, j} j=1,…, bm* $\subseteq \beta$. Before the CR in the $W_t$ will be cognizant approximately the nominated beams complete CR interface.

iii. At the time, when the *CRu* will be within the range of mUBS coverage area according to google map and mUBS will be in a position to transmit data to any CR within the coverage area. Observation will be considered on the amount of data $A_{od,\,j}$(xt, i,t) CR *Cru* will be productively be received through selected beams $A_{od}$ *,j,j=1,.., bm*, till the expiration of *t.*

Considered the denotation of the random variable *rb(x)* the beam enactment of the b under the perspective of the *x*. It will be meaning that the extent of data *rb(x) a CR* with the perspective *x ϵ X* self-control be reception from the mUBS using the $b \in \beta$. We adopt that this indiscriminate adaptable is circumscribed based on [0,1, $M_{Aod}$], given $M_{Aod}$ will be the determined aggregate of the data that expected by CR. $M_{Aod}$ will be bounded by the determined frequency of the broadcast channel. We denoted the estimated value of the beam presentation of beam *b* in the context *x with* $\mu_b(x)$. The mUBS ambitions at choosing a subdivision of the *bm* which will exploit the anticipated obtain data at the CR. Therefore the optimal subdivision in interval

$t\forall_t^*(\mathrm{K}_t)$ $(K_t)=\{\ \forall\ {}^*_t j(X_t)\}j=1,..,bm \subseteq \beta$ .

Therefore the set $\forall_t^*(\mathrm{K}_t)$ will be depending on $X_t$*={xt,i}i=1,….,Vt* and *bm* satisfy.

$$\forall_t^*\left(K_t\right) \in \underset{b \in \beta \setminus \left(\mathrm{U}_{k=1}^{j=1}\left\{\forall_{t,\mathrm{k}}^*\left(K_t\right)\right\}\right)}{\arg\max} \sum_{i=1}^{\vartheta_t} \mu_b\left(kt,i\right)$$

Noting that j=1,…, bm. In case the mUBS will be knowing the expected beam performance $\mu_b(x)$. For every CR perspective *x* ϵ K and respectively *bm* b $\in \beta$, it will be selected optimum subcategory of the beam for every set of pending CR according to (1). To obtain an expected amount the data will be received over the sequence from 1 to the time T.

$$\sum_{t=1}^{T}\sum_{i=1}^{\vartheta_t}\sum_{j=1}^{bm} E\left[r\forall^*_{t,j(xt)}\left(xt,t\right)\right] = \sum_{t=1}^{T}\sum_{i=1}^{\vartheta_t}\sum_{j=1}^{bm} \mu\forall^*_{t,j(xt)}\left(xt,i\right) \quad (10)$$

The mUBS does not recognize the coverage area, it will be learning the expected performance $\mu_b(x)$. Over a given period (10).

To absorb these concerns, the mUBS has to attempt out diverse *bm* of miscellaneous CR context terminated interval also ensuring that the beams will be proved in being good. Lastly, the learning algorithm will be done some times for CR in the coverage area in the context of $X_t$, selecting the *St* of *bm*.

**The algorithm I Pseudocode of Proposed Reinforcement algorithm**

1. Input: $T, PT, K(\mathrm{t})$

2. Prime perspective divider: Create divider $P_T$ of perspective astronomical $[0,1,2]^k$ into $(\mathrm{P_T})^X$ hyper-cubes of duplicate amounts

3. Prime stands: Aimed by entirely $\mathrm{b} \in B$ and entirely the $h \in P_T$ stand $N_{b,h} = 0$

4. Prime guesses: Aimed at totally $\mathrm{b} \in B$ and wholly $h \in P_T$, set $\mu'_{b,h} = 0$

5. for each $t = 1, \ldots K$ do

6. Perceive car settings $K_t = \{k_t, i\}\ i = 1, \ldots, W_\mathrm{t}$

7. Discovery $H_t = \{h_t, i\}\ i = 1, \ldots, W_\mathrm{t}$ in that $k_{t,i} \in p_{t,i} \in p_T, i = 0,1,2\ldots, W_\mathrm{t}$

8. Calculate the usual of under-explored grins $\boldsymbol{B^{ue}_{Ht}(\mathrm{t})}$ trendy (5)

9. if $B^{ue}_{Ht}(\mathrm{t}) \neq 0$ -formerly Assessment

10. $u = acreage(B^{ue}_{Ht}(\mathrm{t}))$

11. if $u. \geq m$ then

12. choice $s_t, 1, \ldots, s_{t,\mathrm{m}}$ randomly from $B^{ue}_{Ht}(\mathrm{t})$

13. else

14. choice $s_t, 0,1,2\ldots, s_{t,\mathrm{u}}$ equally u beams from $B^{ue}_{Ht}(\mathrm{t})$

15. choice $s_{t,\mathrm{u}} + 1, \ldots, s_{t,u}$ equally the $(m, -u)$ *bm* from (6)
$b'_{1,}H_t(\mathrm{t}), \ldots, \mathrm{b_{m-u,}H_t(t)}$

16. else if

17. else

18. select $s_t, 1, \ldots, s_{t,\mathrm{m}}$ for instance the $\boldsymbol{m}$ *bm*
$b'_{1,}H_t(\mathrm{t}), \ldots, \mathrm{b_{m-u,}H_t(t)}$ as a (7)

19. end if

20. Detect the acknowledged date $r_{j,i}$ of every car $W_{t,\mathrm{i}} i = 1, \ldots, W_\mathrm{t}$ in every *bm* $s_{t,j}$, $\mathrm{j} = 0,1,2\ldots, m$

21. for $i = 1, \ldots, W_t$ do

22. for $\mathrm{j} = 1,2\ldots, m$ do

23. $\mu'_{s_{t,j},h_{t,j}} = \dfrac{\mu'_{x_{t,j},h_{t,j}} + r_{j,i}}{N_{x_{t,j},h_{t,i}} + 1}$ and $N_{s_{t,j},h_{t,\mathrm{i}}} = N_{s_{t,j},h_{t,j}} + 1$

24. end for

25. end for

26. end for

The variety of erudition motivation is depending on the history of the beams to be selected. The predictable magnitude of data expected by the vehicle will be certain as follows in case we consider the selection St,t=1,.., T of the algorithm.

$$\sum_{t=1}^{T}\sum_{i=1}^{\vartheta_t}\sum_{j=1}^{bm} E\left[r_{st,j}\left(xt,i\right)\right]$$

$$\sum_{t=1}^{T}\sum_{i=1}^{\vartheta_t}\sum_{j=1}^{bm} E\left[\mu_{st,j(kt)}\left(kt,i\right)-r_{st,j}\left(kt,i\right)\right] \qquad (11)$$

Therefore, the anticipated metamorphosis in the aggregate of acknowledged data accomplished and an algorithm will be the "regrets of learning" taken to be R considering both (10) and (11).

$$R\left(T\right)=E\left[\sum_{t=1}^{T}\sum_{i=1}^{\vartheta_t}\sum_{j=1}^{bm}\left(r\forall^{*}_{t,j(kt)}\left(kt,i\right)\right)-r_{st,j(kt)}\left(kt,i\right)\right]$$

$$R\left(T\right)=\left[\sum_{t=1}^{T}\sum_{i=1}^{\vartheta_t}\sum_{j=1}^{bm}\left(\mu\forall^{*}_{t,j(kt)}\left(kt,i\right)\right)-E\left[r_{st,j(kt)}\left(kt,i\right)\right]\right] (12)$$

## 4-2- Learning Technique

The prototypical *bm* assortment in an mUBS by way of a fast 3-dimension semi-online learning problem as depicted in the algorithm below since it allows identification of the best beams independently in a given interval although secretarial for energetic traffic and setting vicissitudes exhausting a condition in figure 2.

Hence, the mUBS necessities to recognize the superlative rays by prudently choosing subdivisions of *bm* over time. These tactic cataracts below the grouping of appropriate MAB problems. These difficulties moreover contain side evidence that disturbs the recompenses of the activities. A contextual MAB method escalates the mUBS doesn't objective absorb or choice which *bm* is unbeatable on steady, then over as an alternative it achievements extra info almost impending CR to ascertain that bms are the superlative under a prearranged traffics as the algorithm explains called the algorithm I Pseudocode of Proposed Reinforcement algorithm.

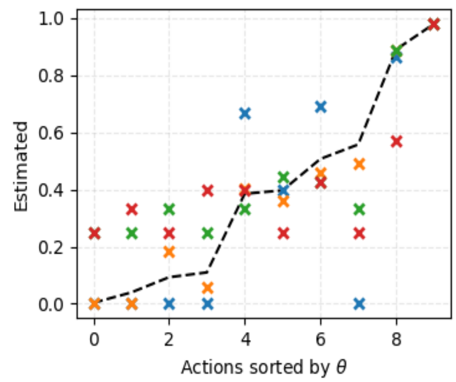


Fig. 2(a) Reward probability against the different estimated probability.

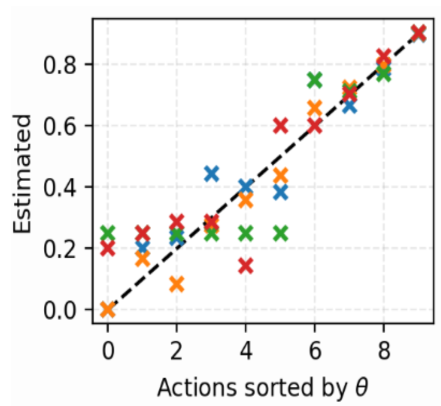


Fig. 2(b) Reward probability against the different estimated probability.

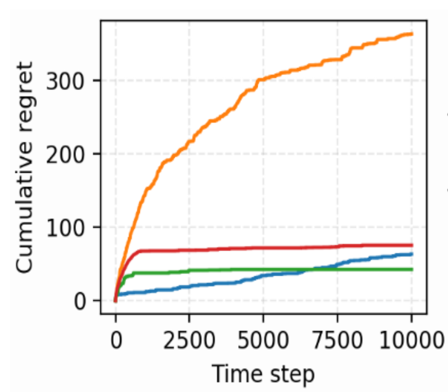


Fig. 3(a) Time steps against the dissimilar cumulative regrets.

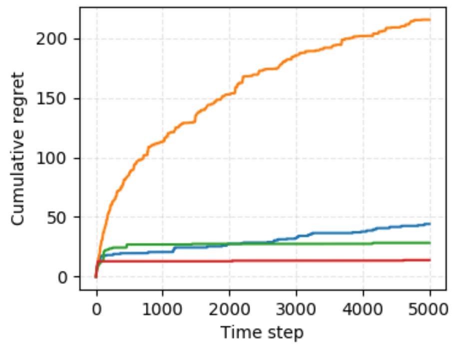


Fig.3(b) Time steps against the dissimilar cumulative regrets.

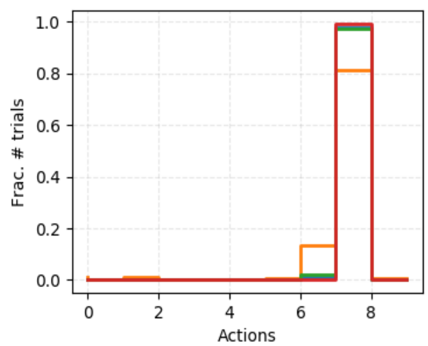


Fig. 4 A fraction Action is selected during the 5000 Execution.

## 4-3- Benchmark Algorithms and Metrics

In this sub section, it depicts the provision an exhaustive performance assessment by associating the proposed 3D-FML to numerous other algorithms. The succeeding particularizes on every yardstick (i.e., Optimal, 2-diamentional machine learning algorithm (2DFML), Dimensions typically measured are quality, time and cost. Where in this perspective, it shows the parameters that have been compared with) scheme:

- The Optimal

This category of the algorithm has a priori information about the predictable bm to be assessed µb(x) of every bm $b \in \beta$ in every circumstance $x \in X$ in addition henceforward contributes an upper bound to the additional procedures that satisfies all the properties in (1).

- The Upper Confidence Bound

This is an irregular of the conventional learning procedure [29], which it is applied and adjusted to the standard of the use-case. The algorithm learns from earlier perceived bm assessment, nevertheless deprived of enchanting keen on account context data. In every time, we chose to adopt UCB among other bandits chooses bm beams with the uppermost predictable UCB on their anticipated bm performance.

- The Random

Within this kind of the algorithm, it chooses random bm in separately period randonly. This algorithm may possibly explore every bm at least once or not. Formerly, the algorithm twigs over the bm with the utmost of the data received depending on the desire of the communication.

- The 2D-FML

This is another algorithm that has been currently availed to arena, where it uses reinforcement learning approach where the vehicles are known when they reach the coverage area of the base station [28].

## 4-4- Geometric Assessment

Firstly, an evaluation of a generic scenario is presented. Secondly, scrutinize the influence of numerous constraints, that is to say, the number of bms selected, the incidence of blockages, the arrival rate of the vehicles, and the underlying traffic patterns. Least otherwise stated, a consideration of the case where for instance as further channel is classified in table 2:

- The percentages at which the permanent blockages and temporary blockages block every one relates to at least the 20 % of all routes.
- The pattern of the traffics (for instance, the automobile (car, vehicle) on the arrival rate and the path chances) alternate per the distinctive traffic patterns demonstrated the by map presented.
- Subsequently Map simply delivers the distinctive everyday pattern of the traffic for 0.71 day (beginning at 06:00 to exactly 22:00); the algorithms were executed over a 0.71 day. Each execution was frequent over 22 times so that the presentation of results demonstrated 94% assurance intervals within the figures.

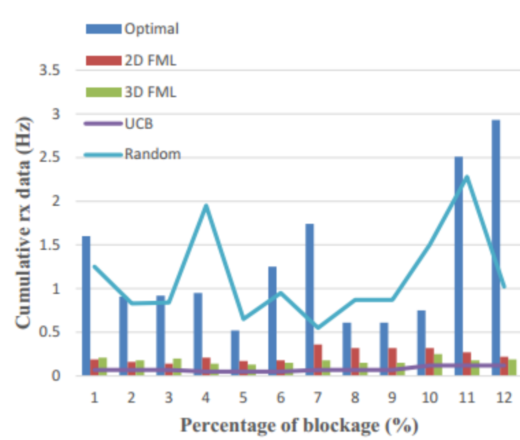

Fig. 5 The Impact of blockages on the accumulative data received for the arrival rate of specified λ and time (t).

## 4-5- The Impact of Blockages

Fig. 5 confirms that an increase in the received data given that predefined arrival percentage of λ in the instance of any given rate m nominated *bms* per interval like for instance 10%, 30%, 60%, 80%, 90% of the permanent obstructions in the structure. Evidently, as the percentages of the permanent obstructions in the system do increase, as the accumulative of data received decrease. On behalf of slightly percentages of the permanent blockage, the proposed 3D-FML overtakes wholly non-optimal algorithms. The accumulative of the data received attained by 3D-FML deceits amongst 16.75%, and 18.32% complex than that succeeded by the next-best algorithm upper confidence bound. Furthermore, the 3D-FML's algorithm accomplished results diverge from that of an optimal simply by at utmost 3.76%.

## 4-6- The Live daily traffic patterns

Traffics has different patterns per different users, The daily traffics is occasionally a inaccuracy, as the peak retro repeatedly continues additional sixty minutes and the "rush hour" denotes to the measurements of traffics, not the promptness of its stream. A rush-hour maybe 6:00pm to10:00 am and 4:00pm to 8:00 pm depending on the set up of the city. Traffic periods could fluctuate from city to city, from area to area, district to district, and seasonally. Application developers availed applications to alert the user on the proceeding pattern depending on the place, Waze.

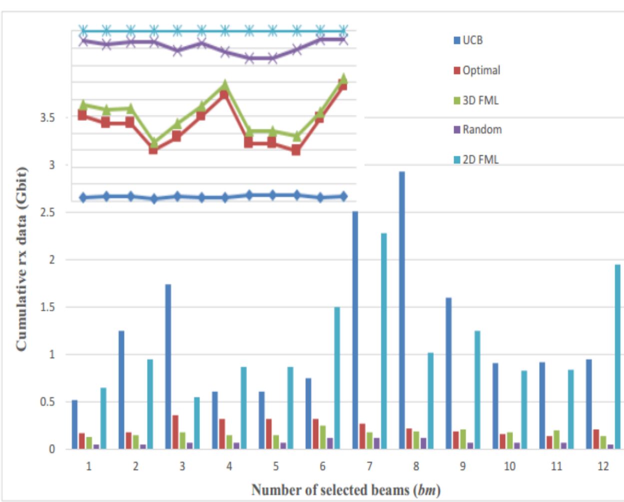


Fig. 6 Impact on the received bm data for arrival rate on cumulative received specified data.

The most well-known traffic app is free and perhaps its biggest draw is the real-time traffic information provided by users includes Google Maps, INRIX, MapQuest, Apple Maps, Traffic Spotter, USA Traffic Cameras, TomTom GPS Navigation Traffic among others. The national highway traffic safety administration recorded that utmost misfortunes transpire throughout "rush hour," altered period of time. And according to the national highway traffic safety administration, the Saturday is the supreme treacherous day of the week to drive, predominantly since there are more compartments and more drunk drivers on the highway than any additional day.

## 4-7- The Impact of the numeral Selected Beams

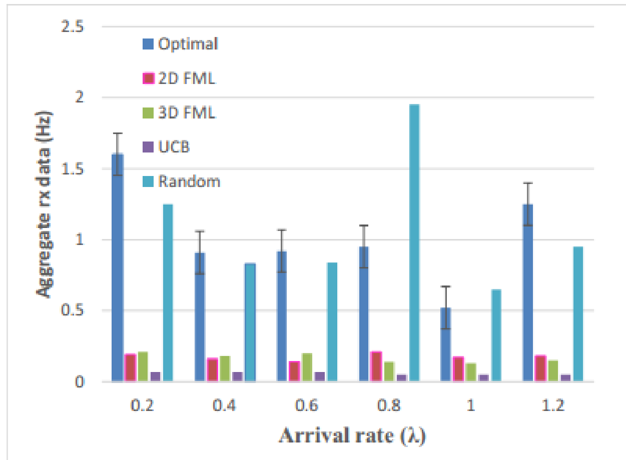


Fig.7 Impact on the arrival rate λ of data based on the cumulative received data.

Analysis of the influence of the amount of beams selection bm per retro on the accumulative of the data received. Fig. 3 displays the accumulative of the data received accomplished with an arrival rate $\lambda$ for diverse bm $\in \{0,1,2,4,8,9,\cdots..\}$. As the quantity of concurrently beam increases selected, the accumulative of the data received increases. Nonetheless, the greater the quantity of bms, the greater is the hardware hurdle and energy consumptions at the mUBS. For diverse value of bm, the accumulative of the data received achieved by 3D-fast machine leaning is amid 11.75% and 20.88% upper than that attained by the next-best algorithms like upper confidence bound and merely up to 5.76% is the lower than that attained by an Optimal as figure 7 depicts.

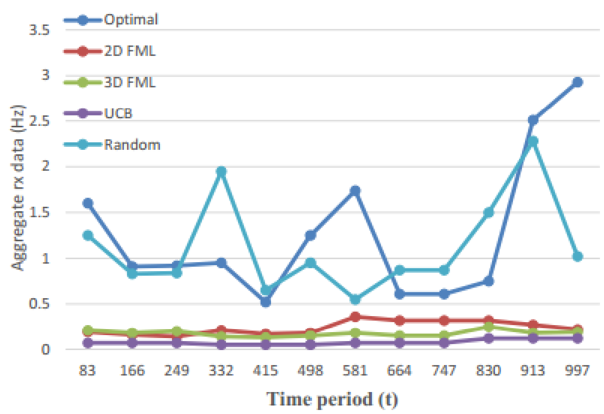


Fig. 8 The aggregates data received for arrival ratio with λ, and the β.

## 4-8- The Impact of arrival rate

The impact of arrival rate depends on some considerable factors, labor-intensive formation and struggle- dynamic assessments to necessitate abundant more than 1.16667 hours. Additionally, combat-driving examinations might only internment possessions of permanent blockages, nonetheless not from the temporary obstacles. The subplot in Fig. 4 depicts the ordinary of the expected data (rx) through arrival rate λ. Regular data (rx) tends to the data terminated altogether the automobiles in the systems up to this historical. These figures demonstrate fast machine learning's rapid learning and adaptation capabilities. Precisely, the three dimensional fast machine learning completes over 89% of performance of optimum in thirty nine times. This illustrates how quick 3D- fast machine learning congregates to near-optimal of the beam selections.

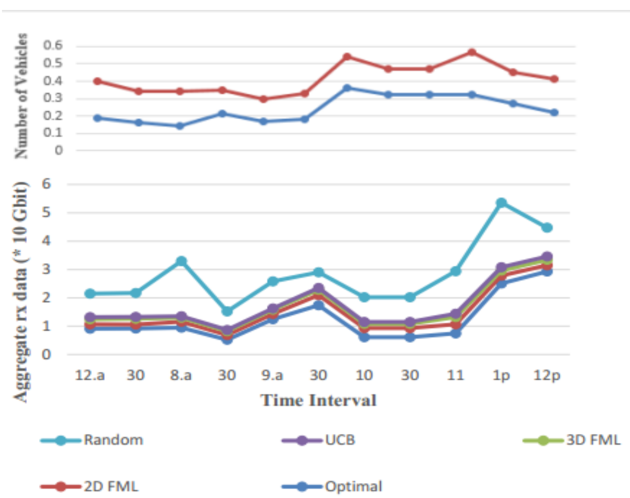


Fig. 9 Data for 172800 minutes of a-live day-to-day Traffic pattern.

## 4-9- The Average Received data

The variations in the figure results from the numeral of vehicles in the systems. Explicitly, the aggregated of the data received rises with the quantity of automobiles, the impacts of automobile arrival rates and traffics. As predictable, an optimal elasticities of the additional upper bounds because of a priori consociate of the anticipated *bm* performance. The proposed 3D-FML algorithm outclassed the additional algorithms as identified for example upper confidence bound, and the random. Observation achieved comes out to be that Fast Machine Learning's assessments rapidly tactics that of an optimum contained by the first hundred periods although the additional algorithms accomplish at least 22% eviler than 3D-fast machine learning. Throughout the comparisons, mean absolute deviation for all the compared algorithms and the proposed is in precise demonstrated in the table 3.

Table 2: Margin specifications

| Parameters | Values and Notations |
|---|---|
| Carrier Wave | 28 Gigahertz |
| Scheme Bandwidth | 1 Gigahertz |
| Transmit Power | 30 dBm |
| Path loss Model (dB) | $32.4+17.3\log_{10} d(m)$ $+20\log_{10}(f_c(GHz))+\xi$ |
| Noise figure | 4 dB at mmBS; 7 dB vehicle |
| UAV's beam width | $36^0$ |
| Simulator tool | MATLAB |
| Thermo Noise | $-174\,dBm/Hz$ |

The result illustrated in this paper was contained using the parameters put in table 2 compared to models theorized [28]; the exploration is rummage-sale to acclimatize to structural subtleties, for instance, the arrival of obstructions and vicissitudes in traffic configurations. The algorithm recognizes obstructions by estimating the cumulative acknowledged data of each automobile for the individually designated *bm* where the current models have not that trait. Furthermore, the algorithm acclimates the traffic's arrays through learning affiliation amongst the path of advent and, the acknowledged dAsadiata. Results, it chooses the beams, which exploit the inclusive system aptitude. Consequently, it affords additional extra to the infrastructures with sophisticated traffic and henceforward, attends an outstanding numeral of UAV.

Table 3: Mean Absolute Deviation results

| Parameters | Mean Absolute Deviation | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| Optimal | 1.60 | 0.91 | 0.92 | 0.95 | 0.52 | 1.25 |
| 3D FML | 0.19 | 0.16 | 0.14 | 0.21 | 0.17 | 0.18 |
| 2D FML | 0.21 | 0.18 | 0.20 | 0.14 | 0.13 | 0.15 |
| UCB | 0.07 | 0.07 | 0.07 | 0.05 | 0.05 | 0.05 |
| Random | 1.25 | 0.83 | 0.84 | 1.95 | 0.65 | 0.95 |
| Parameters Extension | 7 | 8 | 9 | 10 | 11 | 12 |
| Optimal | 1.74 | 0.61 | 0.61 | 0.75 | 2.51 | 2.93 |
| 3D FML | 0.36 | 0.32 | 0.32 | 0.32 | 0.27 | 0.22 |
| 2D FML | 0.18 | 0.15 | 0.15 | 0.25 | 0.18 | 0.19 |
| UCB | 0.07 | 0.07 | 0.07 | 0.12 | 0.12 | 0.12 |
| Random | 0.55 | 0.87 | 0.87 | 1.50 | 2.28 | 1.02 |

The focal awareness linked within figure 3 shows that it's conceivable to putrefy a multifarious supervisory of MAB problem into a categorization of uncomplicated pronouncements, where each verdict of the sequence is solved utilizing the MAB. The mean of the first arm can be obtained in various conducts including ranked bandit approach retains the agreeable traits of preparatory the attention by a sample of the intergalactic and then concentrating gradually on the most propitious capacity, at the diverse weighing machine, permitting to the assessments eventually undertaking an indigenous pursuit about the global optimal functions.

The Thompson sampling riggings the impression of probability matching this is due to the reward approximations $\hat{Q}(a)$ are appraised from subsequent deliveries, any of these prospects is corresponding to the possibility that the correspondent action is optimum, comfortable on pragmatic $hty$ . It's observed that exploration is needed since information is valuable, this means that no exploration is much effective using greedy algorithms, random exploration on the ε-greedy algorithm, and current exploration is seen in Upper confidence bounds and Thompson illustrated in figure 2(a) and 2(b).

A multi-armed bandit approach to wireless communication problems can be a sign of an efficient optimization method compared to traditional statistics as a state vogueish figure 3(a), besides vogueish figure 3(b).

## 4-10- Heuristic approaches in instigating MAB.

Other concepts like heuristic approach in instigating MAB tryouts are approachable to countenance stretchy enticement circulations that may knob the varieties of concerns that ascend intangible claims, the principles of the preconception's apparels are known nevertheless the identities of special cases of reward distributions.

The contributions of the utility optimization are a period of categorized enthusiastic processes anticipated for broad-spectrum exploration creations with diverse algorithmic instantiations contingent on whether the estimates are noisy, the enactment of the algorithms is contingent on the "local" behavior of the function around its global optimal communicated in terms of the quantity of near-optimal states phlegmatic with some geometric as presented in figure 4. In case the indigenous unevenness of the utility is recognized formerly solitary can enterprise identical effective optimizations.

The stochastic MAB problem remains a significant typical for studying the regrets, exploration, and exploitation adjustment in reinforcement learning as well as observed figure 5. While numerous procedures for the problem are well-understood hypothetically, empirical authorization of their efficiency is generally limited.

## 4-11- The several issues in wireless transportations

The model the problem as that MAB delinquent where numerous issues in wireless transportations obligate been treated by the use of multi-armed bandits, a decision-making individual has to choose a subsection of schedules of unidentified recompenses to activity the recompense terminated time. Importantly a MAB method is appropriate for our problematics since an umBS outflow an inadequate customary of beams instantaneously.

The Algorithms matches the issues of mmWave UAV communication on numerous facades are critically identified for example (a) the model distinguishes perpetual structures like edifices, and recurrently congested extents owing to transitory obstructions for example base stations or structure situates haunted by huge automobiles exhausting operational knowledge; (b) the model regulate traffic designs to have structure aptitude maximization by providing grander coverage (like in this case distribution of additional *bms*) in situations with heftier traffics.

## 4-12- The significance of Mobile Base station

This is significant since wave BS may communicate instantaneously terminated an inadequate number of *bm*. This restraint is subject to the hardware traits, the millimeter Wave frequency sparsity, and the beam form practice lastly (b) on the list it conjectures traffics from the perspective (like the automobile's path of the entrance) and chooses the superlative beams. The mainstreams of boulevards have dissimilar road traffic arrays unbiased by the interval of the diurnal. Although inferring these arrangements are obtainable of the latitude of this paper, the model strategies algorithms to recognize and absorb from such configurations.

### 4-13- The effect of user speed vehicle

On the focus on animate everyday traffic configuration: The higher the speed of vehicle as per the UCB, the lower the computation rate, where the higher the speed of vehicle, the lower optimum computation done per the random and optimum estimation per the emblematic circulation arrangement. The 2D-FML algorithm slightly performs close to the 3D-FML model presented due to the fact that 3D models perform better than the 2D the figures of demonstration illustrated since in areas of hard to reach increases connectivity, and mobility conditions. Because of a regular of consequence, Googles' distinctive traffics do not confiscation the instantaneous deviations in overcrowding designs that are perceptible in conscious traffic tale. Therefore, the model depicted the tentative sentient traffic accounts for Google's positioning within a period of two days and half an hour.

## 5- Conclusions

This paper presented a first semi-online algorithm that selects the best beams emitted from the UAV since there are gaining popularity nowadays in social networking and service delivery like now commercial UAVs are becoming smarter. The paper stresses that the applicability of 5G research is in its initial stages whereby few countries have embraced it due to international technological conflicts. The paper clearly demonstrated that it is applicable to use UAVs as a base station in hard to reach areas by the use of reinforcement learning as per the details.

Furthermore, the paper depicted that UAV's advancement will lead to increased illegal surveillance easier, fly into private property, and get video, photos, or both that are not possible for camera-wielding beings that cannot mainly be controlled on the web. The papers did not consider an international policy on the height at which UAV can fly, never consider any category of UAVs. Also, it used a geometric dataset that is not of real-time. In the future, the proposed model is to be subjected to other algorithms besides those ones that have been used in perspective to beam orthogonality, location recording, and situation of co-located automobiles, interference and amount of selected beams.

## References


[1] R. Torkamani and R. A. Sadeghzadeh, "Wavelet-based Bayesian Algorithm for Distributed Compressed Sensing.", Journal of Information Systems and Telecommunication, Vol. 7, No. 2, April-June 2019.

[2] M. Jaderyan and H. Khotanlou, "SGF (Semantic Graphs Fusion): A Knowledge-based Representation of Textual Resources for Text Mining Applications." , Jour. of Information Systems and Tele.mmunication, Vol. 7, No. 1, January-March 2019.

[3] N. Sharma, A. S. Arora, A. P. Singh, and J. Singh, "The Role of Infrared Thermal Imaging in Road Patrolling Using Unmanned Aerial Vehicles," in Unmanned Aerial Vehicle: Applications in Agriculture and Environment, Springer, 2020, pp. 143–157.

[4] C. Qu, W. Gai, M. Zhong, and J. Zhang, "A novel reinforcement learning-based grey wolf optimizer algorithm for unmanned aerial vehicles (UAVs) path planning," Appl. Soft Comput., p. 106099, 2020.

[5] F. Al-Turjman, H. Zahmatkesh, and R. Daboul, "Optimized Unmanned Aerial Vehicles Deployment for Static and Mobile Targets' Monitoring," Comput. Commun., vol. 149, pp. 27–35, 2020.

[6] Y.-F. Liu, X. Nie, J.-S. Fan, and X.-G. Liu, "Image-based crack assessment of bridge piers using unmanned aerial vehicles and three-dimensional scene reconstruction," Comput.-Aided Civ. Infrastruct. Eng., 2020

[7] S. Lee et al., "Intelligent traffic control for autonomous vehicle systems based on machine learning," Expert Syst. Appl., vol. 144, p. 113074, 2020.

[8] E. dos Santos Moreira, R. M. P. Vanni, D. L. Função, and C. A. C. Marcondes, "A Context-Aware Commu. Link for Unmanned Aerial Vehicles," in 2010 Sixth Advanced International Conference on Telecommunications, 2010, pp. 497–502.

[9] D. He, S. Chan, and M. Guizani, "Communication security of unmanned aerial vehicles," IEEE Wirel. Commun., vol. 24, no. 4, pp. 134–139, 2016.

[10]A. Abdessameud and A. Tayebi, "Formation control of VTOL unmanned aerial vehicles with communication delays," Automatica, vol. 4, no. 11, pp. 2383–2394, 2011.

[11]W. Gaofeng, G. Xiaoguang, Z. Kun, and F. Xiaowei, "Multi-objective Placement of Unmanned Aerial Vehicles as Communication Relays Based on Clustering Method," in 2019 Chinese Control And Decision Conference (CCDC), 2019, pp. 1462–1467.

[12]H. Nawaz, H. M. Ali, and M. H. Mahar, "Swarm of Unmanned Aerial Vehicles Communication Using 802.11 g and 802.11 n," IJCSNS, vol. 19, no. 4, p. 289, 2019.

[13]G. S. Voronkov, E. A. Smirnova, and I. V. Kuznetsov, "The method for synthesis of the coordinated group DPCM codec for unmanned aerial vehicles communication systems," in 2019 International Conference on Electrotechnical Complexes and Systems (ICOECS), 2019, pp. 1–4.

[14]H. Harmon, "The Use of Unmanned Aerial Vehicles in Public Safety Communication and Optimization of Coverage (Case Study-Cook Islands)," PhD Thesis, Auckland University of Technology, 2017

[15]S. Fekri-Ershad and F. Tajeripour, "Multi-resolution and noise-resistant surface defect detection approach using new

version of local binary patterns," Appl. Artif. Intell., vol. 31, no. 5–6, pp. 395–410, 2017.

[16] S. Fekri-Ershad and F. Tajeripour, "Impulse-Noise resistant color-texture classification approach using hybrid color local binary patterns and pullback–leibler divergence," The Computer Journal, vol. 60, no. 11, pp. 1633–1648, 2017.

[17]J. A. Adams et al., "Cognitive task analysis for developing unmanned aerial vehicle wilderness search support," Journal of cognitive engineering and decision making, vol. 3, no. 1, pp. 1–26, 2009.

[18]H. Ayaz et al., "Monitoring expertise development during simulated UAV piloting tasks using optical brain imaging," in 2012 IEEE Aerospace Conference, 2012, pp. 1–11.

[19]S. O. Murphy, C. Sreenan, and K. N. Brown, "Autonomous Unmanned Aerial Vehicle for Search and Rescue Using Software Defined Radio," in 2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring), 2019, pp. 1–6.

[20]H. El-Sayed, M. Chaqfa, S. Zeadally, and D. Putkal, "A Traffic-Aware Approach for Enabling Unmanned Aerial Vehicles (UAVs) in Smart City Scenarios*," IEEE Access, pp. 1–1, 2019.

[21]M. Meng et al., "BeamRaster: A Practical Fast Massive MU-MIMO System With Pre-Computed Precoders," IEEE Transactions on Mobile Computing, vol. 18, no. 5, pp. 1014–1027, May 2019.

[22]W. Shafik and S. A. Mostafavi, "Knowledge Engineering on Internet of Things through Reinforcement Learning," Int. J. Comput. Appl., vol. 975, p. 8887.

[23]S. M. Matinkhah, W. Shafik, and M. Ghasemzadeh, "Emerging Artificial Intelligence Application: Reinforcement Learning Issues on Current Internet of Things," in 2019 16th international Conference in information knowledge and Technology (ikt2019), p. 2019.

[24]W. Shafik, S. M. Matinkhah, and M. Ghasemazade, "Fog-Mobile Edge Performance Evaluation and Analysis on Internet of Things," J. Adv. Res. Mob. Comput., vol. 1, no. 3.

[25]Bonald, T., Borst, S., Hegde, N. and Proutiére, A., 2004. Wireless data performance in multi-cell scenarios. ACM SIGMETRICS Performance Evaluation Review, 32(1), pp.378-380.

[26]Jose, J., Ashikhmin, A., Marzetta, T.L. and Vishwanath, S., 2009, June. Pilot contamination problem in multi-cell TDD systems. In 2009 IEEE International Symposium on Information Theory (pp. 2184-2188). IEEE.

[27]Kazi, B.U., and Wainer, G., 2020. Coordinated multi-cell cooperation with a user-centric dynamic coordination station. Computer Networks, 166, p.106948.

[28] A. Asadi, S. Müller, G. H. Sim, A. Klein, and M. Hollick, "FML: Fast machine learning for 5G mmWave vehicular communications," in IEEE INFOCOM 2018-IEEE Conference on Computer Communications, 2018, pp. 1961–1969.

[29]M. M. Ferdaus, S. G. Anavatti, M. Pratama, and M. A. Garratt, "Towards the use of fuzzy logic systems in rotary wing unmanned aerial vehicle: a review," Artif. Intell. Rev., vol. 53, no. 1, pp. 257–290, 2020.

[30]W. Shafik, M. Matinkhah, M. Asadi, Z. Ahmadi, and Z. Hadiyan, "A Study on Internet of Things Performance Evaluation," J. Commun. Technol. Electron. Comput. Sci., vol. 28, pp. 1–19, 2020.

**Wasswa Shafik** received a B.Sc. degree in Computer Engineering from Ndejje University Kampala, Uganda in 2016. He is currently pursuing his MSc degree in Computer Engineering (Networks Option) at Yazd University, Iran. He is a research member of the Intelligent Connectivity Research laboratory at Yazd University. His research interests include Smart Grid, Smart Cities, Cyber Security, 5G & Beyond, and Machine Learning.

**S. Mojtaba Matinkhah** is an assistant professor of Computer Engineering at Yazd University, Iran. He received his B.Sc. degree in software engineering from the University of Isfahan, Iran, an M.Sc. degree in Computer Security from Tehran Polytechnic, and his Ph.D. in Computer Networks from Tehran Polytechnic. He was a scholar visitor of Mississippi State University, the USA in 2011. He is currently the head of the Intelligent Connectivity Research laboratory in Yazd. His main interests are 5G, IoT, Smart Grids, post-quantum security, Cloud-Fog-Edge commuting.

**Mohammad Ghasemzadeh** is an Associate Professor of Computer science and Engineering, working at Yazd University in Iran, since 1996. Received his BSc degree in computer science and Engineering from Shiraz University, Shiraz, Iran in 1989, the MSc degree in artificial intelligence and robotics from Amirkabir University of Technology (Tehran Polytechnic) in 1995. For his PhD research program, he worked at the University of Trier and at the University of Potsdam both in Germany, from Feb. 2002 to Feb. 2006. He completed his PhD degree in "Theoretical Computer Science", in Nov. 2005. He also spent his sabbatical as guest/postdoc researcher at HPI in Germany, from Feb. to Nov. 2016. His main interests are algorithms, intelligent systems, machine learning, big-data analytics and soft computing.