Bandit Framework For Systematic Learning In Wireless Video-Based Face Recognition

Onur Atan, Yiannis Andreopoulos[†], *Member, IEEE,* Cem Tekin, Mihaela van der Schaar, *Fellow, IEEE*

***Abstract*—Video-based object or face recognition services on mobile devices have recently garnered significant attention, given that video cameras are now ubiquitous in all mobile communication devices. In one of the most typical scenarios for such services, each mobile device captures and transmits video frames over wireless to a remote computing cluster (a.k.a. "cloud" computing infrastructure) that performs the heavy-duty video feature extraction and recognition tasks for a large number of mobile devices. A major challenge of such scenarios stems from the highly-varying contention levels in the wireless transmission, as well as the variation in the task-scheduling congestion in the cloud. In order for each device to adapt the transmission, feature extraction and search parameters and maximize its object or face recognition rate under such contention and congestion variability, we propose a systematic learning framework based on *multi-user multi-armed bandits*. The performance loss under two instantiations of the proposed framework is characterized by the derivation of upper bounds for the achievable short-term and long-term loss in the expected recognition rate per face recognition attempt against the "oracle" solution that assumes *a-priori* knowledge of the system performance under every possible setting. Unlike well-known reinforcement learning techniques that exhibit very slow convergence when operating in highly-dynamic environments, the proposed bandit-based systematic learning quickly approaches the optimal transmission and cloud resource allocation policies based on feedback on the experienced dynamics (contention and congestion levels). To validate our approach, time-constrained simulation results are presented via: *(i)* contention-based H.264/AVC video streaming over IEEE 802.11 WLANs and *(ii)* principal-component based face recognition algorithms running under varying congestion levels of a cloud-computing infrastructure. Against state-of-the-art reinforcement learning methods, our framework is shown to provide $17.8\% \sim 44.5\%$ reduction of the number of video frames that must be processed by the cloud for recognition and $11.5\% \sim 36.5\%$ reduction in the video traffic over the WLAN.**

***Index Terms*—multi-armed bandits, learning, face recognition, cloud computing, wireless contention, scheduling congestion**

I. Introduction

MOST of the envisaged applications and services for wearable sensors, smartphones, tablets or portable computers in the next ten years will involve analysis of video streams for event, action, object or user recognition, typically within a remote computing cluster [22], [28], [36], [38], [44], [45]. In this process, they experience time-varying and *a-priori* unknown channel conditions, traffic loads and processing constraints at the remote computing cluster, where the data analysis takes place [4], [16], [23], [30], [34], [38], [44], [45]. Examples of early commercial services in this domain include Google Goggles, Google Glass object recognition, Facebook automatic face tagging [5], Microsoft's Photo Gallery face recognition, as well as technology described in recent publications and patents from Google, Siemens and others[1].

Figure 1 presents an example of such deployments. Video content producers include several types of sensors, mobile phones, as well as other low-end portable devices, that capture, encode and transmit video streams [14] to a remote *computing cluster* (a.k.a. cloud) for recognition purposes. A number of these devices in the same wireless network forms a *wireless cluster*. A cloud-computing cluster is used for analyzing visual data from numerous wireless clusters, as well as for a multitude of other computing tasks unrelated to object or face recognition [28], [34], [44], [45]. Each device uploads its video content and can adapt the encoding bitrate, as well as the number of frames to produce, in order to alleviate the impact of contention in the wireless network. At the same time, the visual analysis performed in the cloud can be adapted to scale the required processing time to alleviate the impact of task scheduling congestion in the cloud. In return, within a predetermined time window, each device receives from the cloud a label that describes the recognized object or face (e.g. the object or person's name), or simply a message that the object or person could not be recognized. In addition, each device or wireless cluster can also receive feedback on the experienced wireless medium access control (MAC) layer contention and the cloud task scheduling congestion conditions. This interaction comprises a *face recognition transaction* between each mobile device and the cloud. The goal of each device is to achieve reliable object or face recognition while minimizing the required wireless transmission and cloud-based processing under highly-varying contention and congestion conditions (respectively).

[†]Corresponding author. O. Atan, C. Tekin and M. van der Schaar are with the Networks, Economics, Communication Systems, Informatics and Multimedia Research Lab, Electrical Engineering Department, University of California, Los Angeles, 56-125B, Engineering IV Building, Box 951594 Los Angeles, CA 90095-1594, USA; tel. +13108255843; fax. +13102068495 email: {oatan, cmtkn}@ucla.edu, mihaela@ee.ucla.edu. Y. Andreopoulos is with the Electronic and Electrical Engineering Department, University College London, Roberts Building, Torrington Place, London, WC1E 7JE, UK; tel. +442076797303; fax. +442073889325; email: i.andreopoulos@ucl.ac.uk. O. Atan, C. Tekin and M. van der Schaar acknowledge the support of the NSF (grant 1016081) and the US Air Force Office of Scientific Research under the DDDAS Program. Y. Andreopoulos acknowledges the support of the UK TSB, projects REVQUAL and VQ-INDEX (#710292).

[1] See "A Google Glass app knows what you're looking at" MIT Tech. Review (Sept. 30, 2013) and EU projects SecurePhone [7], [35] and MoBio [29], [33]. Concerning patents, amongst several others, see the following EU and US patent applications from Google, Siemens, Biometrix Pty and others as an indication of the commercial interest in this area: EP 1580684 B1, US5715325 A, WO 2004029861 A1, US20130219480 A1.

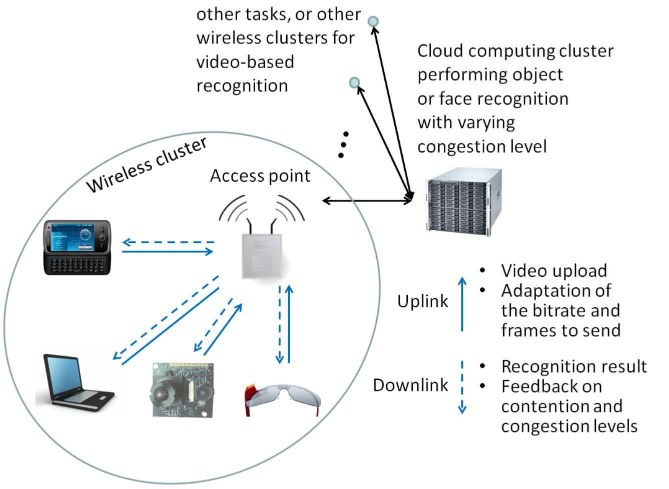

other tasks, or other wireless clusters for video-based recognition
Cloud computing cluster performing object or face recognition with varying congestion level
Wireless cluster
Access point
Uplink
Video upload
Adaptation of the bitrate and frames to send
Downlink
Recognition result
Feedback on contention and congestion levels

Figure 1. Illustration of object or face recognition transactions between mobile devices and a cloud-computing service via adaptive wireless video transport to a remote cloud computing cluster. During each transaction (typically comprising multiple recognition attempts until the person is recognized), the cloud responds with the result for each recognition attempt and each device is given feedback on the wireless contention levels, as well as the congestion levels in the cloud.

A. *Related Work*

Each mobile device of Figure 1 seeks to achieve a certain recognition accuracy rate that is deemed suitable to the application, while minimizing its cost in terms of utilized wireless resources (e.g., MAC superframe transmission opportunities used) and the number of video frames that must be encoded and transmitted. To this end, several approaches have been proposed that are based on reinforcement learning [46] or other methods for resource provisioning and optimization [4], [16], [22], [28], [30], [34]. However, most existing solutions for designing and configuring wireless multimedia applications that offload their processing to the cloud assume that the underlying dynamics (e.g. source and traffic characteristics, channel state transition probabilities, multi-user interactions, cloud congestion, etc.) are either known, or that simple-yet-accurate models of these dynamics can be built [22], [28], [34], [44], [46].

Nevertheless, in practice, this knowledge is not available and models of such complex system dynamics (which include multiple wireless users and the cloud) are very difficult to built and calibrate for specific environments. Hence, despite applying optimization, these solutions tend to result in highly sub-optimal performance since the models they use for the experienced dynamics are not accurate. Hence, reinforcement learning (i.e. learning how to act based on past experience) becomes a vital component in all such wireless multimedia applications with cloud processing. Some of the best-performing online reinforcement learning algorithms are Q-learning [39] and structural-based reinforcement learning [8]–[10]. In these, the goal is to learn the state-value function, which provides a measure of the expected long-term performance (utility) when it is acting optimally in a dynamic environment. It has been proven that online learning algorithms converge to optimal solutions when all the possible system states are visited infinitely often [39].

However, these methods have to learn the state-value function at every possible state. As a result, they incur large memory overheads for storing the state-value function and they are typically slow to adapt to new or dynamically changing environments (i.e., they exhibit a slow convergence rate), especially when the state space is large—as in the considered wireless transmission and recognition problem. These memory and speed-of-learning deficiencies are alleviated in structural-based learning solutions [8]–[10]. Despite this, a key limitation still remains: all these schemes provide only asymptotic bounds for the learning performance—no speed-of-learning guarantees are provided. Nevertheless, in most multimedia analysis and recognition systems, users are interested in both short-term performance and long-term performance. This is because, for example, a user will find it a time-consuming and cumbersome task to train a face recognition app in the mobile device if it requires too many queries and responses to learn to recognize accurately. Therefore, we need algorithms whose performance is adequate even under a modest number of attempts.

One solution is to use multi-armed bandit (MAB) algorithms, for which finite time bounds on the performance can be obtained in addition to the asymptotic convergence results. The fundamental operation of these algorithms involves carefully balancing exploration of actions with highly uncertain performance and exploitation of the action with the highest estimated performance. To do this, most of these algorithms keep an index that weights the *estimated performance* and *uncertainty* of each action and chooses the action with the highest index at each time slot. Then, the indices for the next time slot for all actions are updated based on the feedback received from the chosen action. However, most of the existing work on multi-armed bandits [2], [19] does not take into account the side information (i.e., context) available at each time, which, in this case, is the contention and congestion levels at the wireless network and cloud processing, respectively. These methods utilize all the past observations obtained for a specific transmission setting to estimate the expected performance when using this setting. Hence, they learn fast but are highly sub-optimal for video-based recognition services since congestion and contention are not taken into account. They can be seen as acting blindly, neglecting the current congestion and contention levels when choosing the transmission setting. The side information can be exploited using contextual bandit algorithms [20], [37], where the best action (transmission setting) given the context (side information) is learned online. These methods utilize only a context-dependent history of past observations for a specific transmission setting to estimate its context-dependent performance, but require strong *similarities* between the contexts such that learning can be performed together for a group of contexts. Different from this, our proposed framework learns independently for each context, hence does not require *similarities* between contexts. Moreover, the related literature in contextual bandits is focused on single-user learning over time, rather than multi-user learning, and does not consider the joint effect of the decision of multiple users on the congestion level. On the opposite side, related work in multi-user multi-armed bandits [1], [26] does not take into account the context information and does not consider clustering the action profiles, hence is highly sub-optimal for our context-based recognition framework.

B. Paper Contribution

We propose two new multi-armed bandit-based learning algorithms: device-oriented contextual learning and service-oriented contextual learning. Device-oriented contextual bandit algorithm is a single-user bandit-based approach with the use of contextual information. Service-oriented contextual learning algorithm is centralized multi-user bandit-based approach with the use of contextual information. We not only show that our algorithms converge to the optimal action profile that assumes full knowledge of the system parameters, but are also able to quantify at every instance of time how far our algorithms are from this optimal profile. We do this by deriving worst-case performance bounds on our algorithms. Specifically, to measure the performance of our algorithms we use the notion of *regret*, which is the difference between the expected recognition rate the devices obtain per recognition attempt when optimally knowing *a-priori* the exact recognition rate expected for each action (i.e., the complete knowledge benchmark), and the expected recognition rate per attempt that will be achieved following the online learning algorithm. In other words, the notion of regret at the kth recognition transaction is the performance loss due to unknown system parameters. The detailed contributions of the paper are summarized below:

- We propose the use of contextual bandits for mobile devices and prove that the *regret bound*—the maximum loss incurred by the algorithm against the best possible non-cooperative decision that assumes full knowledge of contention and congestion conditions—is logarithmic if users do not collaborate and each would like to maximize their own utility.
- When the cloud congestion depends on the user actions and, therefore, the cloud maximizes the average utility of the users of a wireless cluster, we prove a logarithmic regret bound with respect to the best possible cooperative decision.
- We also achieve much higher learning rate than conventional multi-user multi-armed bandits with grouping the action profiles that lead to the congestion level on the cloud. The proposed contextual bandit framework is general, and can also be used for learning in other wireless video applications that involve offloading of various processing tasks.

A logarithmic regret bound means that the order of the regret is $O(\log k)$. It is known [19] that for most of the MAB problems that assume finite set of contexts, actions and stochastic rewards (i.e., recognition rates in our case), the best order of regret is logarithmic in time, i.e., no learning algorithm can have smaller regret. This implies that the average regret at recognition transaction k, i.e., the regret at k divided by k goes to zero very fast.

Beyond the application scenario of object or face recognition via wireless video transmission and remote server processing, our theoretical framework can be used in many other practical applications, including resource provisioning in cognitive radio networks, wireless sensor networks, etc. Moreover, unlike other learning-based methods, such as Q-learning, we not only provide asymptotic results, but we are able to analytically bound the worst-case short term performance. Table I presents a summary of the different aspects of multi-armed bandit based decision making, highlighting the advantages of this work over other recently-proposed approaches.

Table I
COMPARISON OF PROPOSED APPROACH WITH OTHER WORK ON MULTI-ARMED BANDITS.

	[2], [19]	[20], [24], [37]	[1], [26]	This work
Multi-user	No	No	Yes	Yes
Contextual	No	Yes	No	Yes
Similarity Metric on Context	N/A	Yes	N/A	No
Grouping Joint Profiles	N/A	N/A	No	Yes
Regret	Logarithmic	Sublinear	Logarithmic	Logarithmic

In Section II, we present the detailed system description and the system model under consideration. Sections III and IV present the design and analysis of the proposed multi-armed bandit-based learning algorithms for the distributed (user-based) and centralized (cloud-based) cases, respectively. Section V presents the corresponding simulation results validating our proposals against state-of-the-art learning algorithms from the literature and Section VI concludes the paper.

II. SYSTEM DESCRIPTION AND SYSTEM MODEL

A. Video Capturing and Encoding

Each networked mobile device of Figure 1 involves a video camera capturing several frames that include the object or human face. Each frame can be illuminated via artificial modulation of the light of the camera flash to emulate a light source present in different angles such that the system will not be easily fooled by a photograph or video of the object or person placed in front of the camera (see [3], [7], [17], [32], [33], [35] and footnote 1 for further information on flash illumination variation). Each video frame can be cropped to the object or face area by automated face detection algorithms [5], such as the well-known Viola-Jones classifier for face detection [18] in video frames. Alternatively, the user can be asked to position the mobile device such that its frontal camera places the object or face within a rectangle displayed on the device (smartphone or portable computer) screen prior to the initiation of the video capture. For instance, this approach is followed within the Google Goggles Search App[2]. These cropped areas of the video frames are then compressed using a standard-compliant video codec (such as

[2] http://www.google.com/mobile/goggles/

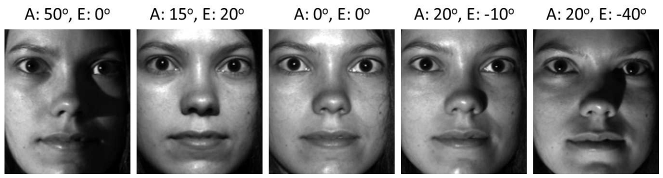


Figure 2. Example of five frames captured under different illumination elevation (E) and azimuth (A) angles from the Yale Face Database B [11], [21].

MPEG/ITU-T H.264/AVC), which is typically realized via a low-power hardware chipset. The created stream, typically comprises a short video of $0.5 \sim 2$s duration, with $5 \sim 30$ frames captured and encoded per second. An example of the facial portion of 5 video frames (from the Yale Face Database B [11], [21]) under different illumination angles is presented in Figure 2.

B. Wireless Transmission

Transmission of the compressed video content can take place in our envisioned system using either wireless local area network (WLAN) infrastructures, such as IEEE 802.11 WLANs [41], or WiMAX/LTE-based cellular networks. A key feature of such networks is that they are simultaneously supporting multiple wireless devices sharing the same spectrum and each device can adapt its transmission parameters (e.g. number of packet retransmissions, modulation and coding schemes, transmit power level etc.) depending on the number of concurrent transmitters. Our solution can learn the behavior of such adaptation mechanisms and, under a time constraint for the transmission of each video, decide on the transmission settings to use (i.e., per device, video encoding bitrate and number of frames to send) in a manner that is agnostic to the specifics of the utilized adaptation of the lower layers of the protocol stack. This is because we only require the existence of a mechanism for obtaining feedback on the current contention level in the wireless transmission. Such a mechanism is indeed supported by all practical deployments of WLANs and 3G/4G networks, e.g., via the use of carrier-sense designs supported by the related standards [6]. Therefore, even though we provide validation results under the assumption of IEEE 802.11 WLANs [41], our proposal is generic and can be applied to a variety of contention-based wireless transmission frameworks.

C. Visual Analysis

The cloud computing cluster processes multiple visual analysis tasks concurrently, possibly in conjunction with the execution of several other services, as shown in Figure 1. Therefore, its task scheduler experiences highly-varying levels of congestion. These levels can be measured in real time [23], but it is generally accepted that it is difficult to anticipate and predict them prior to the actual execution of each task. Therefore, under a time reservation mechanism for each recognition attempt[3], the cloud computing infrastructure may have to adapt the accuracy of its feature matching algorithm, as well as the number of video frames processed, if the available processing cycles do not suffice for the completion of the complete series of processing steps.

Typical scenarios for object or face recognition algorithms consider that the server matches the provided video information to a pre-established database of stored images using an algorithm based on principal component analysis (PCA) [43], classification via ℓ_1 minimization [42], salient-point extraction and matching [27], support vector machines [15], etc. For example, for each video frame, the server computes the feature extraction operation (with a precomputed projection matrix [15], [42], [43] or a predetermined salient-point extraction algorithm [27]) and then sends the extracted features to the distance-calculation routine that retrieves the best match via searching within a large database of such features. When the majority of the video frames are matched to the same object or person in the database, the system classifies the match as successful and the identified object or person is returned as the result.

The exact percentage of video frames that must match the same person can be set *a-priori*, such that the *false positive rate* is substantially reduced, e.g., by experimentally setting this percentage such that accidental identification of the wrong person is extremely unlikely. At the same time, the system must maintain *the false negative rate* under control, i.e., the percentage of times a person is not reliably matched to the correct face in the database and the match is rejected albeit being correct. This occurs when the system identifies the correct person for the majority of video frames, but this majority does not surpass the *a-priori* set percentage. This which may happen due to varying illumination, motion blur or varying distance (or pose) of the person in front of the camera. Evidently, the number of frames processed, the average frame distortion stemming from video encoding, as well as the number of features used for the distance-calculation routine (e.g. number of eigenvectors [15], [43] or salient points [27] used against the training dataset), affect the experienced false negative rate per recognition attempt. Hence, they also affect the number of recognition attempts that are expected to take place until the system recognizes the person and the *face recognition transaction* (comprising multiple recognition attempts) between the mobile device and the cloud is concluded.

If a recognition attempt does not match to the same person for a preset percentage of frames in the video, the device is notified that the recognition attempt was unsuccessful. In this paper we assume that, under appropriate parameter tuning, the only possible responses of such a face recognition transaction are: *(i)* the recognized person identity; *(ii)* a message stating that recognition attempt was unsuccessful. While we shall provide validation results via a face recognition application

[3] Time reservation is the most common way of billing for cloud computing services, such as Amazon WS EC2, and it is therefore natural to focus on time-constrained execution. In addition, each task has a given deadline, which is imposed by the need to provide a recognition result to each device within a few seconds.

mostly suitable to PCA-based methods, or to methods performing classification in sparse representations [42], [43], our learning frameworks can be applied directly under a salient-point extraction based recognition method [27], or under support vector machines (SVM) based methods [15]. In fact, performing the analysis steps in the cloud instead of the mobile device allows for seamless changes in the utilized processing and analysis algorithms, and each mobile device can simply apply our learning framework to adjust its transmission settings to the utilized recognition algorithm used.

D. System Model

Consider M mobile devices, indexed by the set[4] $\mathcal{M} = \{1, 2, \ldots, M\}$. Let $\mathcal{A} = \{a_1, a_2, \ldots, a_S\}$ denote the set of all possible transmission settings (actions) for each mobile device, i.e., all possible video coding bitrates and number of video frames to transmit when a device attempts a recognition action, with S the size of the settings space. In addition, let all devices consider the discrete sets $\mathcal{T}$ and $\mathcal{G}$ to comprise all contention and congestion levels of the wireless medium and cloud-computing infrastructure, respectively. Both $\mathcal{T}$ and $\mathcal{G}$ are discrete sets, since all timeslot and cycle allocation strategies of wireless MAC retransmission mechanisms and cloud schedulers (respectively) operate under a discrete set of states. Importantly, in the proposed systematic learning framework via bandits, we do not utilize any prior information (e.g. training) for the contention and congestion levels and our results apply for arbitrary context variations.

Under this setup, the following events take place sequentially for each *recognition transaction*, k:

1) Each device observes the current wireless contention level, $t(k) \in \mathcal{T}$, and cloud congestion level, $g(k) \in \mathcal{G}$, and selects the bitrate and number of frames to capture, and, within a predetermined deadline, transmits the corresponding H.264/AVC-encoded video to the cloud in order to attempt to recognize the object or face;
2) the cloud decodes the video it received, extracts features out of the decoded video frames, and performs feature matching with the database of available features with search accuracy (i.e. number of features used) that corresponds to its congestion level;
3) each device gets the result from the cloud, which is either the label corresponding to a recognized object or person, or a message stating that the object or face could not be recognized reliably (i.e. "recognition unsuccessful"); based on this result, each device adjusts its expected recognition rate per attempt for a trasmission setting $a \in \mathcal{A}$ it had chosen, i.e $\hat{Y}_{t,g,a}(k)$;
4) in the latter case, the device performs further recognition attempts (going back to Step 1), until a successful result is obtained or the user abandons the recognition transaction; in each attempt, it can select a different setting (in terms of bitrate and number of frames).

Each recognition transaction is therefore expected to comprise several attempts. Furthermore, each attempt is carried out under a time constraint for both the wireless transmission and the feature extraction and matching in the cloud. Therefore, depending on the contention and congestion levels, a varying number of frames will be transmitted and processed for each recognition attempt, which will affect the recognition rate per attempt, as well as the number of recognition attempts expected to be required by the recognition transaction in order to ensure that the system recognizes with a certain accuracy rate (e.g., 90% recognition accuracy).

In this paper, we propose two models for the derivation of a bandit-based systematic learning framework that, under given time constraints and recognition accuracy rate required by an application, lead to significantly decreased resource consumption in the wireless network and the cloud infrastructure against other state-of-the-art learning methods.

In the first model, illustrated in Figure 3(a) and termed as the *device-oriented model*, devices strive to systematically learn the best transmission setting to maximize their own recognition rate per attempt under given contention level in the wireless medium and congestion level in the cloud. Therefore, the reward for this case is the recognition result at each time step. For this case, we assume that both the wireless access point and the cloud infrastructure serve many more requests than the ones from a given cluster of devices (as illustrated in Figure 1). Therefore, both contention and congestion levels vary randomly and are not affected by the settings used by each device. This makes the devices completely independent.

In the second model, illustrated in Figure 3(b) and termed as the *service-oriented model*, the cloud systematically learns the best action profile that maximizes the cluster's average recognition rate per attempt under given contention in the wireless transmission. For this model, we assume that, while the wireless contention level remains independent of the decisions made by the devices, the cloud congestion level varies depending on the actions taken at the devices. For example, this corresponds to the scenario where a virtual machine instance is allocated on dedicated hardware in the cloud and serves solely a given wireless cluster of devices.

For the device-oriented model, all devices use the contention level of the wireless medium and the congestion level in the cloud as contexts for the bandit-based systematic learning framework. For the service-oriented model, the cloud uses the wireless contention level as context and the recognition rate per attempt, as well as the cloud congestion level, depends on the aggregate actions of all devices, which are represented by vector $\mathbf{a}(k)$.

III. Distributed, Device-oriented, Bandit Learning Algorithm

We propose a device-oriented learning framework, where the mobile devices select their own transmission settings (actions) and learn through their interaction with the wireless medium and the cloud, assuming that the wireless contention

[4] Notations: Uppercase letters indicate system settings; lowercase letters indicate variables and functions; uppercase calligraphic letters indicate sets, e.g., $\mathcal{T}$, with their cardinality indicated by $|\mathcal{T}|$; α indicates a mobile device's transmission settings based on an optimization or learning framework; $a \leftarrow b$ assigns value b to a; $\hat{x}$ indicates an estimate of variable x; Pr$\{\mathcal{E}\}$ denotes the probability of occurrence of event $\mathcal{E}$; and $E[\cdot]$ is the statistical expectation operator.

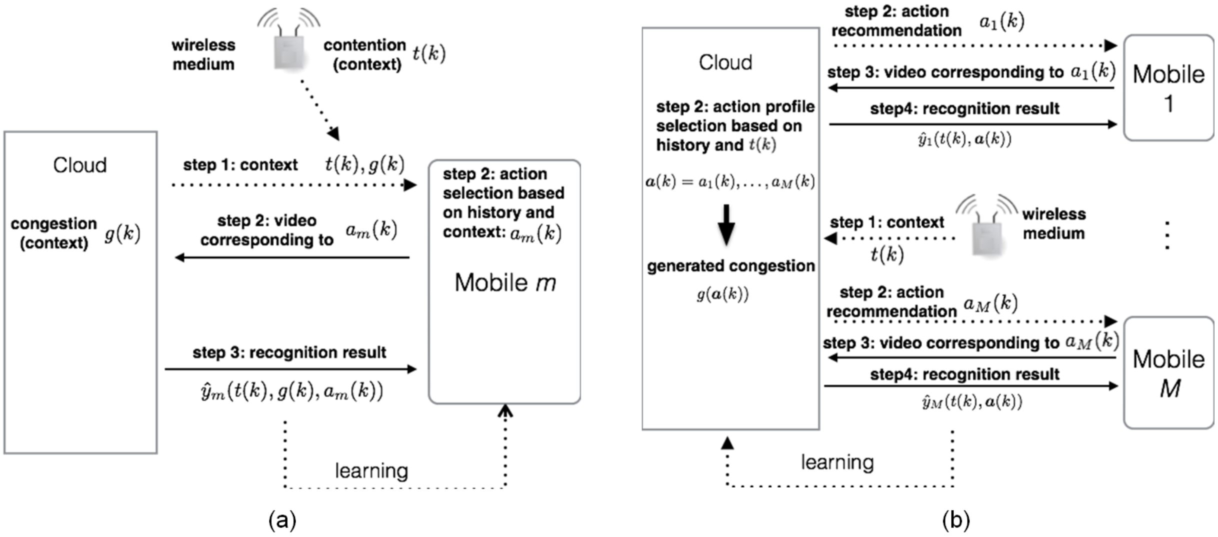


Figure 3. (a) device-oriented model; (b) service-oriented model.

and cloud congestion levels vary independently of the actions of each device in the same wireless cluster.

Let $\eta_m(t\left(k\right), g\left(k\right), a)$ be the expected recognition rate of an attempt of the mth device with transmission settings a, given the contention and congestion levels $t\left(k\right)$ and $g\left(k\right)$ at the kth recognition transaction, respectively. The goal of each device is to explore the transmission settings in $\mathcal{A}$ and learn the expected recognition rate $\eta \in \left(0,1\right)$ depending on $g\left(k\right)$ and $t(k)$. It can then anticipate how many attempts it will require on average, in order to receive a recognition result with a predetermined recognition accuracy rate (e.g., 90%, which is generally deemed adequate for general face matching applications [5], [15], but higher rates may be needed for other services that are partially based on face recognition [29]). We will determine the performance of each learning algorithm in comparison to the optimal solution that selects the best transmission setting a^* for the mth mobile device, i.e., the setting that yields the lowest number of expected attempts to receive a recognition result under the same recognition accuracy rate. The optimal solution for the kth recognition transaction is given by

$$a_m^*\left(t\left(k\right), g\left(k\right)\right) = \arg\max_{\forall a \in \mathcal{A}} \left\{\eta_m\left(t\left(k\right), g\left(k\right), a\right)\right\} \tag{1}$$

and it is defined as the oracle solution, since it assumes that all conditions for each case are precisely known beforehand. We now define the regret of the algorithm as a performance measure.

Definition 1 (Regret for the mth Device in the Device-oriented Model). The regret after k iterations (recognition transactions) is the expected loss against the optimal solution of (1), which is incurred due to unknown system dynamics. For the mth device, the regret of learning algorithm α that selects the setting a_l at each transaction l, $1 \leq l \leq k$, with respect to the best action is given by

$$\begin{aligned} R_m(k) &= \sum_{l=1}^{k} \eta_m\left(t\left(l\right), g\left(l\right), a^*\right) \\ &- E\left[\sum_{l=1}^{k} Y_m\left(t(l), g(l), a(l)\right)\right], \end{aligned} \tag{2}$$

where $Y \in \{0,1\}$ is a binary random variable modeling the recognition result received from the cloud under transmission setting $a(k)$ for the mth device. ■

The regret gives the rate of convergence of the expected recognition rate of each algorithm, under systematic learning aiming towards the value of optimal solution, given by (1). It is therefore essential in quantifying the expected performance of a learning algorithm. Specifically, providing upper bounds on the regret after k recognition transactions can characterize: *(i)* whether a learning algorithm can approach the optimal recognition rate and *(ii)* at what speed this can take place.

A. Proposed Device-oriented Contextual Learning Algorithm

At any recognition transaction k, the mobile device can be in one of the two following stages: *(i) exploration stage*, where the mobile device chooses an arbitrary transmission setting to update the estimated recognition rate per attempt given the contention and the congestion levels at the cloud; and *(ii) exploitation stage*, where mobile devices select the transmission setting that yields the highest estimated recognition rate given the network contention level and the congestion level in the cloud[5]. In order to determine the stage of the algorithm, we need to keep the number of times each transmission setting has been selected for each congestion and contention level at the cloud. Let $N_{t,g,a}(k)$ be the number of times transmission setting a has been selected until the kth recognition transaction

[5] all the parameters defined in this subsection are different for each mobile device $m \in \mathcal{M}$. However, for notational brevity and given we are considering an arbitrary device here, we will refrain from using the subscript m in the notations.

by a mobile device under contention and congestion levels t and g. At the kth transaction, each device is given the levels $t(k)$ and $g(k)$ and checks[6]: $t \leftarrow t(k), g \leftarrow g(k), \forall a: N_{t,g,a}(k)$, to identify whether there exists a transmission setting that should be explored. Let us define $\mathcal{S}_{t,g}(k)$ as the set of transmission settings that need to be explored at the kth transaction

$$\mathcal{S}_{t,g}(k) = \left\{t \leftarrow t(k), g \leftarrow g(k), \forall a \in \mathcal{A}: N_{t,g,a}(k) \le c(k)\right\}, \tag{3}$$

where $c(k)$ is a deterministic control function that is monotonically increasing in k. Function $c(k)$ can be interpreted as the minimum number of exploration steps required by the algorithm such that the deviation probability of the sample mean estimate of the expected reward of setting a decays at rate k^{-b} for some $b \ge 1$ [37]. In practice, the control function $c(k)$ guarantees that the sample mean of the recognition rate for each device's attempts is high enough to be used for the exploitation stage of the learning process.

Each mobile device estimates the recognition rate of its transmission setting at a specific contention and congestion level based on the recognition attempts it observed for that particular setting so far. Therefore, let $\mathcal{X}_{t,g,a}(k)$ be the set of all recognition results (i.e., set of all "rewards") obtained by the mobile device until the kth recognition transaction when selecting transmission setting a under under contention and congestion levels t and g. In addition, let $\hat{\alpha}(k)$ be the (estimated) best transmission setting at the kth transaction based on the estimated recognition rates for contexts $t \leftarrow t(k)$ and $g \leftarrow g(k)$:

$$\hat{\alpha}(k) \in \arg\max_{\forall a \in \mathcal{A}} \left\{\hat{Y}_{t,g,a}(k)\right\}, \tag{4}$$

where $\hat{Y}_{t,g,a}(k)$ is the sample mean of the obtained recognition results in $\mathcal{X}_{t,g,a}(k)$, i.e.,

$$\hat{Y}_{t,g,a}(k) = \sum_{\forall Y(t,g,a) \in \mathcal{X}_{t,g,a}(k)} \frac{Y(t,g,a)}{|\mathcal{X}_{t,g,a}(k)|}, \tag{5}$$

with $Y(t,g,a) \in \{0,1\}$ each recognition result (or reward) obtained by the attempts of each device ($\forall t,g,a$: 0 for no recognition and 1 for successful recognition). We do not assume the uniqueness of $\hat{\alpha}(k)$. Indeed, if more than one setting maximizes (4), then the mobile device m chooses any arbitrary setting from that set.

Given levels $t \leftarrow t(k)$ and $g \leftarrow g(k)$, if $\mathcal{S}_{t,g}(k) \ne \varnothing$, then there exists at least one transmission setting that must be explored, and the mobile device chooses an arbitrary setting in this set. If, however, $\mathcal{S}_{t,g}(k) = \varnothing$, then all transmission settings have been explored sufficiently, and the mobile device will select (i.e., exploit) the setting that yields the highest estimated recognition rate per attempt.

In order to define the minimum suboptimality gap that provides an indication of the performance difference between the best setting and the next-best setting that can be selected, we need to define suboptimality gap of any setting for each congestion and contention levels at the cloud side.

```
Algorithm 1: Device-Oriented Contextual Learning
Input: $c(k)$; sets: $\mathcal{A}$, $\mathcal{G}$, $\mathcal{T}$
Initialization:
  $\forall t \in \mathcal{T}, \forall g \in \mathcal{G}\ \forall a \in \mathcal{A}: N_{t,g,a} = 0$ & $\hat{Y}_{t,g,a} = 0$; $k = 1$
Repeat
  Get contention and congestion levels $t \leftarrow t(k)$, $g \leftarrow g(k)$
  If $\exists a \in \mathcal{A}$ s.t. $N_{t,g,a}(k) \le c(k)$ // exploration stage
    Choose setting $a$
    Receive recognition rate $Y(t,g,a)$ after multiple attempts
    Update($N_{t,g,a}(k)$,$\hat{Y}_{t,g,a}$,$Y(t,g,a)$)
  Else // exploitation stage
    Find $\hat{\alpha}(k) \in \arg\max_{\forall a \in \mathcal{A}} \hat{Y}_{t,g,a}$
      Receive recognition rate $Y(t,g,\alpha(k))$ after multiple attempts
    Update($N_{t,g,\hat{\alpha}(k)}(k)$,$\hat{Y}_{t,g,\hat{\alpha}(k)}$,$Y(t,g,\alpha(k))$)
  End If
  $k \leftarrow k + 1$
End
Update($n$,$\hat{Y}$,$Y$):   $\hat{Y} \leftarrow \frac{n\hat{Y}+Y}{n+1}$; $n \leftarrow n + 1$
```

Definition 2 (Suboptimality Gap and Minimum Suboptimality Gap). Let $\Delta_{t,g}(a^-) \triangleq \eta(t,g,a^*) - \eta(t,g,a^-)$ be the suboptimality gap of any transmission setting a^-, with $a^- \in \mathcal{A} \backslash a^*(t,g)$, and its corresponding optimal setting $a^*(t,g)$ given by (1). We now define the minimum suboptimality gap $\Delta_{\min}$ as the minimum difference between the expected recognition rate of the best transmission setting and second-best transmission setting, i.e., $\forall t \in \mathcal{T}$, $\forall g \in \mathcal{G}$, $\forall a^- \in \mathcal{A} \backslash a^*(t,g)$:

$$\Delta_{\min} \triangleq \min_{\forall t,g,a^-} \Delta_{t,g}(a^-). \tag{6}$$

■

The suboptimality gap defines the performance difference between the best transmission setting and other transmission settings. Due to the existence of the suboptimality gap, if our algorithm can form good-enough estimates of the expected recognition rate per attempt, then it will almost always choose the setting with highest true recognition rate in exploitation phases.

The proposed algorithm for device-oriented contextual learning is given in Algorithm 1. Below we present a Lemma that characterizes the conditions under which this algorithm achieves the optimal performance.

Lemma 1 (Condition for Optimal Exploitation of Algorithm 1). $\forall a \in \mathcal{A}$, $\forall t \in \mathcal{T}$, $\forall g \in \mathcal{G}$: if

$$\left|\hat{Y}_{t,g,a}(k) - \eta(t,g,a)\right| < \frac{1}{2}\Delta_{\min}, \tag{7}$$

then the optimized transmission setting given in (4) is $a^*(t,g)$ given in (1).

Proof: See Appendix A. ■

[6] with $a \leftarrow b$ the operator that assigns b to variable a

Lemma 1 proves that, under accurate-enough estimates, Algorithm 1 will select the optimal transmission setting in the exploitations. We will use this to bound the suboptimal transmission setting selection in the exploitations in the analysis that follows.

B. Analysis

There are two components of the regret in contextual learning via Algorithm 1: $R_e(k)$, i.e., the regret due to the explorations, and $R_s(k)$, i.e., the regret due to suboptimal action selection in the exploitations. Since the expected rewards are bounded in $(0,1)$, it is sufficient to bound the number of times that device chooses a suboptimal setting. In the following lemmas, we will derive bounds for $R_e(K)$ and $R_s(K)$. .

Lemma 2 (Regret Bound for Exploitations). For any recognition transaction $l \le k$, if we set: $c(l) = 4\frac{b\ln l}{\Delta_{\min}^2}$ with $b > \frac{1}{2}$, then the expected regret due to suboptimal setting selection in exploitation steps performed until recognition transaction k is upper bounded by

$$E\left[R_s\left(k\right)\right] \le 2S|\mathcal{G}|\left|\mathcal{T}\right|H_k^{(2b)}, \tag{8}$$

where $H_k^{(2b)}$ is the Generalized Harmonic Number [13]: $H_k^{(2b)} = \sum_{l=1}^{k}\frac{1}{l^{2b}}$.

Proof: See Appendix A. ■

In Lemma 2 we proved that, when $b > \frac{1}{2}$, the expected number of times a mobile device selects a suboptimal transmission setting in exploitation phases is bounded by a constant term that is independent of the recognition transaction k. This means that the regret in exploitation phases does not diverge to infinity as k goes to infinity. In other words, the regret due to exploitation phases is $O\left(1\right)$. In the next lemma, we bound the regret due to explorations.

Lemma 3 (Regret Bound for Explorations). For any recognition transaction $l \le k$, if we set: $c(l) = 4\frac{b\ln l}{\Delta_{\min}^2}$ with $b > \frac{1}{2}$, then the expected regret due to the explorations performed until recognition transaction k is upper bounded by

$$E[R_e(k)] \le |\mathcal{G}|\left|\mathcal{T}\right|S\left(1 + c(k)\right). \tag{9}$$

Proof: See Appendix A.

■

Theorem 1. *Under the conditions of Lemmas 2 and 3, the total expected regret due to to explorations and exploitations until recognition attempt k is upper-bounded by*

$$E\left[R\left(k\right)\right] \le |\mathcal{G}|\left|\mathcal{T}\right|S\left(1 + 4\frac{b\ln k}{\Delta_{\min}^2} + 2H_k^{(2b)}\right). \tag{10}$$

Proof: We have: $E\left[R\left(k\right)\right] = E\left[R_e\left(k\right)\right] + E\left[R_s\left(k\right)\right]$, which, from Lemmas 2 and 3, leads to the desired result.

■

We proved that Algorithm 1 achieves logarithmic regret. Moreover, for $b > \frac{1}{2}$, $H_k^{(2b)}$ is finite as $k \to \infty$. Therefore, the expected averaged regret goes to zero, i.e,

$$\lim_{k\to\infty}\frac{R(k)}{k} = 0. \tag{11}$$

An interesting question is whether the logarithmic (with respect to attempts performed) regret is the best that can be achieved. It is shown by Lai and Robbins [19] that, for the non-contextual standard multi-armed bandit problem, the logaritmic in time l value of $c(l)$ is the smallest possible amount of exploration which guarantees that the expected number of suboptimal action selections in exploitations are sub-logaritmic, and the slowest growth rate of regret is logarithmic in the number of attempts for any learning algorithm. Since non-contextual multi-armed bandit problem is a special case of the problem we consider in this paper, our order-of-regret in Theorem 1 matches the lower bound, and, hence, it is tight. Given that logarithmic regret, $O\left(\ln k\right)$, is the lowest possible regret that can be achieved by any function $c(k)$ [20], [25], the average recognition rate of an attempt of each mobile device will converge to the recognition rate of the oracle solution defined in (1).

IV. Centralized, Cloud-based, Bandit Learning

In the previous section we proposed a device-oriented learning approach, where the mobile devices select their own transmission settings (actions) and learn through their interaction with the cloud, assuming that the cloud congestion level varies independently of the actions of each device in the wireless cluster. However, if we assume that the devices' actions affect the cloud's congestion level (under, for example, a dedicated hardware instance in the cloud for a given wireless cluster), if many mobile devices send large volumes of video frames to the cloud, they will all experience low recognition rate per attempt due to the high congestion caused in the cloud. Hence, the algorithm proposed in Section III may not lead to the optimal solution for this case, since the recognition rate of each attempt of a mobile device is inherently affected by the settings chosen by the other devices of the same cluster.

To address this case, in this section we take a service-oriented approach, where mobile devices follow the suggestions of the cloud for their transmission settings. Thus, as illustrated in Figure 3(b), it is the cloud that learns which joint action profile, $\mathbf{a}\left(k\right) = \left[a_1\left(k\right), a_2\left(k\right), \ldots, a_M\left(k\right)\right]$, should be used by the M mobile devices based on the contention level, $t\left(k\right)$, at each attempt k.

The recognition rate per attempt for this case depends on: *(i)* the transmission settings and *(ii)* the contention level in the wireless medium. Let $g(\mathbf{a})$ be the congestion caused by the mobile devices when they select transmission settings $\mathbf{a}$. The feature-matching complexity used at the cloud depends on the settings of the mobile devices, since the cloud uses different number of features for each congestion level. Let $\mathcal{H}$ be a partition of all the joint action profiles $\mathcal{A}^M$, where each element is a subset of joint action profiles that include the same settings with different permutations. We assume that different permutations of action profile correspond to the same congestion level g, i.e $g\left(\boldsymbol{a}\right) = g\,\forall\boldsymbol{a} \in h\,\forall h \in \mathcal{H}$. Then, we have $|\mathcal{H}| = \binom{M+S-1}{S-1} = \frac{(M+S-1)!}{(S-1)!M!}$.

Let $\eta_m(t(k), \mathbf{a}(k)) \equiv \mu(t(k), g(\mathbf{a}(k)), a_m(k))$ be the expected recognition rate of an attempt of the mth device at the kth recognition transaction, where $\mu : \mathcal{T} \times \mathcal{G} \times \mathcal{A} \to (0,1)$ is the expected recognition rate function that depends on contention, congestion and the mobile device's transmission setting. The goal of cloud is to find best transmission settings for all devices to maximize the average recognition rate per attempt of all M devices. Because the different permutations of joint action profile will lead to the same congestion level, let $\eta(t(k), h(k))$ be the expected recognition rate per attempt at kth recognition transaction of all M devices selecting any action profile $\boldsymbol{a} \in h$

$$\eta(t(k), h(k)) = \frac{1}{M} \sum_{m=1}^{M} \eta_m(t(k), \mathbf{a}(k)) \; \forall \boldsymbol{a} \in h. \qquad (12)$$

The goal of the cloud is to explore action profiles in $\mathcal{H}$ and learn the expected average recognition rates per attempt, $\eta \in (0,1)$, depending on the congestion level $t(k)$. We now define the benchmark solution which is computed under the full knowledge of the recognition rates per attempt. The benchmark solution for the contention level is given by

$$h^*(t(k)) = \arg\max_{\forall h \in \mathcal{H}} \{\eta(t(k), h(k))\} \qquad (13)$$

and it is defined as the oracle solution, since it assumes that all conditions are precisely known beforehand. We now define the regret of the algorithm as a performance measure.

Let $\hat{Z}(t,h) \in (0,1)$ be a random variable modeling the average recognition rate for an attempt of all M devices under contention level t and transmission settings $\mathbf{a} \in h$ chosen for all devices. At the kth recognition transaction

$$\hat{Z}(t(k), h(k)) = \frac{1}{M} \sum_{m=1}^{M} \hat{Y}_m(t(k), h(k)), \qquad (14)$$

where $\hat{Y}_m = \{0,1\}$ is the binary random variable modeling the recognition results for the mth device.

Definition 3 (The Regret for the Service-oriented Model). The regret for the service-oriented model is the loss due to unknown recognition rates per attempt obtained via each setting. For the cloud, the regret of learning algorithm α that selects the any action profile $\mathbf{a}(t(l)) \in h$ at each recognition transaction l, $1 \le l \le k$, with respect to the best action is given by

$$R_{\text{cloud}}(k) = \sum_{l=1}^{k} \eta(t(l), h^*) - E\left[\sum_{l=1}^{k} \hat{Z}(t(l), \alpha(l))\right]. \qquad (15)$$

■

The regret gives a measure of the different in performance between our learning algorithm and the oracle solution defined in (13).

A. Service-Oriented Contextual Learning

In this section, we will propose the service-oriented contextual learning for the cloud, which tries to find the best profile of transmission settings for each congestion level. The proposed algorithm is similar to the Algorithm 1 defined in the Section 3 in the sense that it also balances exploration with exploitation. However, there are important differences between the two algorithms. First of all, in service-oriented learning, it is the cloud that makes the decisions, i.e., all mobile devices simply obey to the transmission setting suggested to them by the cloud. Secondly, in this case the cloud takes into account the aggregate of the recognition rates of all devices and the contention level in the wireless medium, but not its own congestion level, as this is indirectly controlled by the settings decided for each device. In contrast, in the previous section, a mobile device takes *both* the contention level *and* congestion levels as contexts, and selects a transmission setting that will maximize its own estimated recognition rate per attempt.

At any recognition transaction k, the service-oriented learning algorithm can be in one of two phases: *(i)* exploration step, in which the cloud chooses an arbitrary transmission setting in $\mathcal{H}$ depending on the contention level t and updates the estimated recognition rate (per attempt) of any transmission action profile $h \in \mathcal{H}$; *(ii)* exploitation step, in which the cloud selects any transmission action profile that yields the highest-expected average recognition rate per attempt.

Let $\mathcal{X}_{t,h}(k)$ be the recognition accuracies collected until recognition transaction k by selecting all possible transmission settings in h given contention level t. The cloud selects a transmission-action profile that yields the highest estimated average recognition rate. Let $\hat{\alpha}(k)$ be the (estimated) best transmission setting for all mobile devices for context $t \leftarrow t(k)$, i.e,

$$\hat{\alpha}(k) \in \arg\max_{\forall h \in \mathcal{H}} \left\{\hat{Z}_{t,h}(k)\right\}, \qquad (16)$$

where $\hat{Z}_{t,h}(k)$ is the estimated sample mean of the elements in $\mathcal{X}_{t,h}(k)$. Explicitly,

$$\hat{Z}_{t,h}(k) = \sum_{Z_{t,h} \in \mathcal{X}_{t,h}(k)} \frac{Z_{t,h}}{|\mathcal{X}_{t,h}(k)|} \qquad (17)$$

with $Z_{t,h} \in \{0, ..., M\}$ the sum of each recognition result (or reward) obtained by all M devices (per device: 0 for no recognition and 1 for successful recognition). Once the transmission settings have been selected for the mobile devices, the cloud will randomly select among the joint action profiles $\boldsymbol{a} \in h$ to the devices to send their requests. The reason the cloud randomizes in h is *fairness*: since the optimal profile will include some transmission settings that correspond to less frames and lower encoding bitrates, some devices will be punished at a particular attempt; therefore, the cloud randomizes in h each time to ensure no single device is penalized more than the others.

To differentiate between the exploration and exploitation steps, the cloud needs to keep track of the number of times a particular vector of settings in h, has been chosen for each contention level. Let $N_{t,h}(k)$ be the number of times the cloud selected any transmission action profile $\mathbf{a} \in h$ until the kth recognition transaction, given the contention level t. For each recognition transaction k, the cloud receives the contention level, $t \leftarrow t(k)$ and checks whether the following set is empty

Algorithm 2: Service-Oriented Contextual Learning

```
Input: $c(k)$; sets: $\mathcal{H}$,$\mathcal{T}$
Initialization:
  $\forall t \in \mathcal{T}$ , $\forall h \in \mathcal{H}:\ N_{t,h} = 0$ & $\hat{Z}_{t,h} = 0$; $k = 1$
Repeat
  Get contention level $t \leftarrow t(k)$
  If $\exists h \in \mathcal{H}$ s.t. $N_{t,h}(k) \leq c(k)$
    Choose setting $h$ and randomly choose joint setting $\boldsymbol{a} \in h$
    Receive recognition rate $Z_{t,h}$ after multiple attempts
    Update($N_{t,h}(k)$,$\hat{Z}_{t,h}$,$Z_{t,h}$)
  Else
    Find $\hat{\alpha}(k) \in \arg\max_{\forall h \in \mathcal{H}} \hat{Z}_{t,h}$
    Randomly choose joint setting$\hat{\alpha}(k) \in h$
    Recommend the $m$th element to the $m$th device
    Receive recognition rate $Z_{t,\hat{\alpha}(k)}$ after multiple attempts
    Update($N_{t,\hat{\alpha}(k)}(k)$,$\hat{Z}_{t,\hat{\alpha}(k)}$,$Z_{t,\hat{\alpha}(k)}$)
  End If
  $k \leftarrow k + 1$
End
Update($n$,$\hat{Z}$,$Z$):   $\hat{Z} \leftarrow \frac{n\hat{Z}+Z}{n+1}$; $n \leftarrow n + 1$
```

$$\mathcal{S}_t = \{t \leftarrow t(k),\ \forall h \in \mathcal{H}:\ N_{t,h}(k) \leq c(k)\},$$

where $c(k)$ is defined as for (3) of the previous section. When $\mathcal{S}_t \neq \varnothing$, the cloud selects an arbitrary transmission setting from this set and collects the recognition rates for all devices. If $\mathcal{S}_t = \varnothing$, this means that all the transmission action profiles are explored sufficiently.

Definition 4 (Suboptimality Gap and Minimum Suboptimality Gap). Let $\Delta_t(h^-) \triangleq \eta(t, h^*) - \eta(t, h^-)$ be the suboptimality gap of any transmission setting h^-, with $h^- \in \mathcal{H}\backslash h^*$, and its corresponding optimal setting $h^*(t)$ given by (13). We now define the minimum suboptimality gap, $\Delta_{\min}$, as the minimum difference between the expected recognition rate per attempt of the best profile and second-best profile, i.e., $\forall t \in \mathcal{T}$, $\forall h^- \in \mathcal{H}\backslash h^*$:

$$\Delta_{\min} \triangleq \min_{\forall t, h^-} \{\Delta_t(h^-)\}. \tag{18}$$

■

The proposed algorithm for service-oriented contextual learning is given in Algorithm 2. Below we present a Lemma that characterizes the conditions under which this algorithm achieves the optimal performance.

Lemma 4. $\forall h \in \mathcal{H}$, $\forall t \in \mathcal{T}$, if

$$\left|\hat{Z}_{t,h}(k) - \eta(t,h)\right| < \frac{1}{2}\Delta_{\min}, \tag{19}$$

then the optimized transmission setting given in is $h^*(t)$ given in (16) is oracle solution given in (13).

Proof: The proof follows the one of Lemma 1. ■

B. Analysis

There are two components of the regret in service-oriented contextual learning. The first one is $R_e(k)$, i.e. the regret due to the explorations and $R_s(k)$, i.e., the regret due to suboptimal profile selection in the exploitations. Since the rewards are bounded in $(0, M)$, it is sufficient to bound the number of times that device selects an suboptimal action. In the following lemmas, we will bound $R_e(k)$ and $R_s(k)$ separately.

Lemma 5. For any recognition transaction $l \leq k$, if we set: $c(l) = 4\frac{b \ln l}{\Delta^2_{\min}}$ with $b > \frac{1}{2}$, then the expected regret due to suboptimal setting selection in exploitation steps performed until recognition transaction k is upper bounded by

$$E[R_s(k)] \leq 2\binom{M+S-1}{S-1}|\mathcal{T}|H_k^{(2b)}, \tag{20}$$

where $H_k^{(2b)}$ is the Generalized Harmonic Number [13].

Proof: See Appendix A. ■

With this lemma, we proved that the regret for suboptimal settings' selection in exploitations is finite for $b > \frac{1}{2}$.

Lemma 6. For any recognition transaction $l \leq k$, if we set: $c(l) = 4\frac{b \ln l}{\Delta^2_{\min}}$ for some $b > \frac{1}{2}$, then the expected regret due to the explorations is upper bounded by

$$E[R_e(k)] \leq |\mathcal{T}|\binom{M+S-1}{S-1}(1 + c(k)). \tag{21}$$

Proof: See Appendix A.

■

Theorem 2. *Under the conditions of Lemmas 5 and 6, the total expected regret due to to explorations and exploitations until* recognition transaction k *is upper-bounded by*

$$E[R_{cloud}(k)] \leq M|\mathcal{T}|\binom{M+S-1}{S-1}\left(1 + 4\frac{b \ln k}{\Delta^2_{\min}} + 2H_k^{(2b)}\right). \tag{22}$$

Proof: We have: $E[R_{\text{cloud}}(k)] \leq M(E[R_e(k)] + E[R_s(k)])$. From Lemmas 5 and 6, and due to the reward being bounded by M, this leads to the desired result.

■

We notice that only the constants are different between the regret bounds of Theorems 1 and 2. In addition, the regret bound of Theorem 2 depends on the number of mobile devices.

C. Discussion

Our analysis is also valid when the feedback is arriving with some delay and the correct recognition results (rewards) are not always revealed. The algorithm keeps the results produced by the classification and updates the rewards whenever the correct recognition results are revealed. This will add some extra lag in learning process, however, the asymptotic regret for both Algorithm 1 and 2 will still be valid and the expected total reward will still converge to the value of the optimal solution.

Since the set of device transmission settings grows combinatorially with the number of concurrent devices in the wireless cluster, the cloud incurs certain complexity overhead for storing and adapting the estimated transmission rates for all these settings in comparison to the device-oriented model, where each device only stores adapts the estimated transmission rates for its own transmission settings. Therefore the complexity of the cloud-oriented model is greater than the complexity of the device-oriented model.

Specifically, at each exploitation step, each algorithm needs to pick best setting among the possible settings it has available. For the device-oriented learning, a device has S actions; therefore, its complexity is of order $O(S)$. For the service-oriented learning, the cloud has S^M action profiles, but only $\binom{M+S-1}{S-1}$ of them are distinct in terms of the congestion they generate and, therefore, it only needs to keep estimates for the distinct ones. Thus, the complexity for service-oriented case is of order $O\binom{M+S-1}{S-1}$. However, this will not be a problem in practice, since the computational and memory resources of the cloud are significantly higher than the resources of the mobile devices.

V. Numerical Results

For each algorithm under consideration, we present simulation results with respect to the average recognition attempts required per recognition transaction[7], as well as the average bitstream size per recognition transaction. Given that there are several parameters that vary in our system (contention and congestion levels and training and testing subsets), we repeat each experiment 100 times with random training and testing subsets per person and present the average results. Therefore, our results correspond to a mean-based analysis of performance instead of best or worst case analysis. This relates to the expected performance of such a system that would be assessed prior to cumbersome deployment and testing in the field.

A. General Setup

Our simulation environment comprises mobile devices connected via an IEEE 802.11 WLAN to a computing cluster, i.e., cloud computing service. Videos of human faces are produced by random images of persons taken from the extended Yale Face Database B (39 cropped faces of human subjects under varying illumination). While this is not a large-scale dataset and it is not as specific to mobile systems as other datasets (e.g., see [29], [35]), it corresponds to a scenario where users within a group (e.g. a residential or office environment) would be recognized automatically. In addition, it resembles our assumption in Section II that the camera flash is modulated to emulate light coming from different angles (Figure 2) such that person recognition remains robust to counterfeit measures (e.g., spoofing attacks by placing a photo or video of a person in front of the camera of the mobile device [3], [17], [32]). Each video to be recognized comprises up to 30 images from the same person and it is compressed to a wide range of bitrates via the H.264/AVC codec (x264 codec, crf $\in \{4, 14, 24, 34, 44, 51\}$). The 2D-PCA algorithm [43] is used at the cloud side for face recognition from each decoded video frame (with the required training done offline as per the 2D-PCA setup [43]) and the cloud-computing server being able to use a varying number of features (eigenvectors) for the projection and matching process of 2D-PCA, depending on its congestion level. For all simulations, we have set a time window of two seconds per recognition attempt, which was separated to one second for capturing, encoding and transmission and one second for processing on the cloud.

B. Wireless Transmission

The time-constrained transmission limits the number of video frames received by the cloud under varying WLAN contention levels at the MAC layer, as delay is increased under MAC-layer contention due to the backoff and retransmissions of IEEE 802.11 WLAN standards. In our simulations, we generated wireless contention via the well-known backoff and retransmission mechanism of the Distributed Coordination Function (DCF) of such networks under the default settings of the DCF simulator of Bianchi's method [6].

C. Face Recognition on the Cloud

In terms of the recognition algorithm, in order to declare this video as "recognized" by the cloud while at the same time substantially reducing the possibility of false positives under the utilized setup, more than 80% of the received video frames have to match to the same person in the database. However, because of varying congestion in the cloud: *(i)* only a limited number of the received video frames is actually used by 2D-PCA and *(ii)* the utilized number of eigenvectors, d, used for the distance calculation during the matching stage [43] is chosen from $d \in \{2, 4, 6, 10\}$, according to the cloud congestion level, thereby affecting the recognition rate per attempt. In our simulations, for the device-oriented learning of Algorithm 1 [Figure 3(a)], we generated random congestion levels at each recognition transaction. Instead, for the service-oriented learning of Algorithm 2 [Figure 3(b)], the generated congestion level was analogous to the volume of video frames received by all M devices.

In order to ensure that all methods are compared on an equal basis, we report the average number of attempts per recognition transaction, as well as the average video traffic transported per recognition transaction by each method. By setting the recognition accuracy rate of the system to 0.9 (90%), the average number of attempts per transaction (and their corresponding bitstream size) is calculated based on the (per-attempt) empirically-derived recognition rate, $\mu(k)$, with $\mu(k) = \hat{Y}_{t,g,a}(k)$ of (5) for the device-oriented framework and $\mu(k) = \hat{Z}_{t,\mathbf{a}}(k)$ of (17) for the service-oriented framework. Specifically, by denoting the average attempts of the kth recognition transaction by "$\mathrm{rec}(k)$" and recalling that each attempt is independent of the previous ones, we have

[7] given that the system may reply that it is unable to recognize reliably if a substantial number of frames is not matched to the same person

$$\begin{aligned} 1-[1-\mu(k)]^{rec(k)} &= 0.9 \\ \Longleftrightarrow rec(k) &= \log_{1-\mu(k)} 0.1. \end{aligned} \tag{23}$$

We remark that recognition accuracy rate of 0.9 is deemed as adequate for real-world face recognition applications [5], [15] and, following (23), our results can be derived for arbitrary recognition accuracy rates.

D. Results

Considering first the single-device case, Figure 4(a) presents the average number of attempts per recognition transaction k, by: *(i)* our method (with and without using the cloud congestion information as context); *(ii)* the optimal setting of (1) that assumes full system knowledge (oracle); *(iii)* Q-learning [39]. The results indicate that, after 250 transactions (each transaction comprises the attempts listed), our algorithm approaches the oracle bound and, for the same recognition accuracy rate per transaction (90%), incurs less attempts in comparison to Q-learning, reaching a reduction of up to 30%. In addition, Figure 4(b) presents the corresponding video bitstream size transported per transaction by each method. Our approach allows up to 20% reduction in the video traffic over the IEEE 802.11 contention-based MAC in comparison to Q-learning and approaches the oracle bound as the number of recognition transactions increases.

Figure 5 presents the results for the service-oriented learning via Algorithm 2. We again observe that our approach outperforms Q-learning and it also approaches the number of attempts required by the oracle bound. Extending these results to more mobile devices, it can be seen from Fig. 6 (b) that although Q-learning is better initially then the proposed method (because of the fact that the number time slots spend in explorations which has a large contribution to the regret decrease as time progresses and), the proposed methods rate of exploitation increases as time goes on and hence it starts performing much better than Q-learning under the same contexts (especially in terms of the number of attempts per recognition transaction) by efficiently allocating transmission settings (actions) to the mobile users that maximize the average recognition rate per attempt under wireless contention, which minimize the expected number of required attempts following (23). We can observe from the Fig. 6(a) that our approach converges to oracle bound with a higher rate than Q-learning when the number of users increase since our proposed method considers clustering the action profiles to reduce the exploration rate.

Based on these results, we see that the proposed systematic learning framework based on multi-user multi-armed bandits is able to achieve a high performance, i.e., recognition transactions with the minimum number of attempts, in dynamically changing and unknown environments. Our algorithms are able to learn significantly faster than existing reinforcement learning solutions. Moreover, our results show that the proposed context-based MAB solutions are significantly outperforming conventional MAB schemes by exploiting the available side-information and therefore lead to lower bandwidth usage and faster recognition as they require less attempts. Finally, while Q-learning appears to reach a saturation point in the average number of attempts to achieve a certain recognition accuracy rate (thereby implying fewer faces are recognized successfully per attempt), the simulations with the proposed bandit-based learning show that, given enough recognition transactions k, the performance bounds of the oracle method will be approached.

VI. Conclusions

We propose a contextual bandit framework for learning contention and congestion conditions in object or face recognition via wireless mobile streaming and cloud-based processing. Analytic results show that our framework converges to the value of the oracle solution (i.e., the solution that assumes full knowledge of congestion and contention conditions). Simulations within a cloud-based face recognition system demonstrate it outperforms Q-learning, as well as context-free bandit-based learning, as it quickly adjusts to contention and congestion conditions. Therefore, our analysis and results demonstrate the importance of using contexts in multi-user bandit-based learning methods, as well as their efficacy within wireless transmission and cloud-based processing environments. Beyond the proposed application for face recognition via cloud-based processing of wireless streams, our proposed learning framework can be applied in a variety of other scenarios where fast learning under uncertain conditions is essential, such as multi-user wireless video streaming systems and, in general, multimedia signal processing systems where decisions need to be made in environments with unknown dynamics.

Appendix A

Lemma 1.

Proof: We have

$$\eta_m(t,g,a) - \hat{Y}_{t,g,a}(k) \le \frac{1}{2}\Delta_{\min} \tag{24}$$

and, for any suboptimal a^-, i.e., $a^- \in \mathcal{A} \backslash a^*(t,g)$:

$$\hat{Y}_{t,g,a^*}(k) - \eta_m(t,g,a^*) < \frac{1}{2}\Delta_{\min}. \tag{25}$$

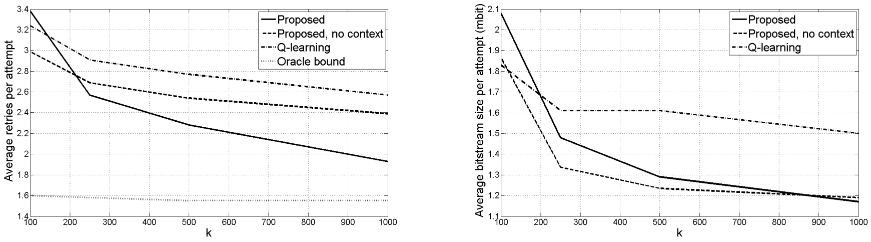
Combining the last two inequalities with the fact that $\Delta_{t,g}(a^-) \ge \Delta_{\min}$ leads to: $\hat{Y}_{t,g,a^*}(k) - \hat{Y}_{t,g,a^-}(k) > 0$, which leads to the desired result. ■

Lemma 2.

Proof: Let $W(l)$ be the event that proposed algorithm chooses a suboptimal setting in the exploitation stage at recognition transaction l, $l \le k$. Then the expected regret until recognition transaction k is expressed as

$$E[R_s(k)] = \sum_{l=1}^{k} \Pr\{W(l)\}. \tag{26}$$

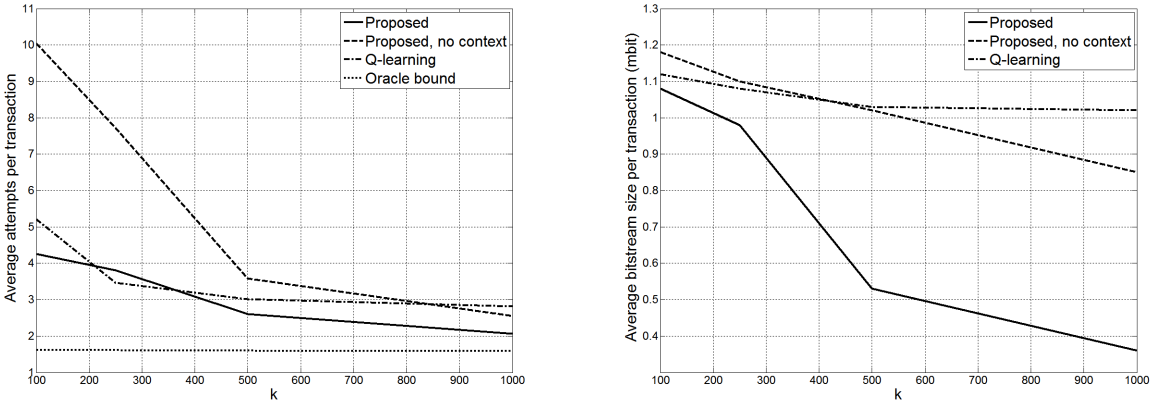
The occurrence of event $W(l)$ depends on: *(i)* the suboptimal setting selection, i.e $\forall l: \ \hat{\alpha}(l) \notin a^*(t,g)$ and *(ii)* the occurrence of exploitation stages in the proposed algorithm, i.e, $N_{t,g,a}(l) \ge c(l)$, given the contention and congestion



(a) Average attempts per transaction k, $M = 1$

(b) Average bitstream size, $M = 1$

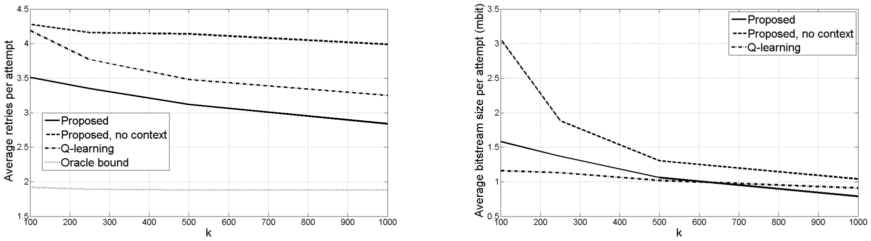
Figure 4. Device-oriented model. Per recognition transaction, k, and under recognition accuracy rate of 0.9, we present: (a) Average attempts and (b) average bitstream size with the 2D-PCA algorithm and a single device ($M = 1$).



(a) Average attempts per transaction k, $M = 3$

(b) Average bitstream size, $M = 3$

Figure 5. Service-oriented model. Per recognition transaction, k, and under recognition accuracy rate of 0.9, we present: (a) Average attempts and (b) average bitstream size with the 2D-PCA algorithm and with $M = 3$.



(a) Average attempts per transaction k, $M = 5$

(b) Average bitstream size, $M = 5$

Figure 6. Service-oriented model. Per recognition transaction, k, and under recognition accuracy rate of 0.9, we present: (a) Average attempts and (b) average bitstream size with the 2D-PCA algorithm and with $M = 5$.

levels $t \leftarrow t(l)$ and $g \leftarrow g(l)$ respectively, for all settings a. Therefore, for all contention and congestion levels, (26) is upper bounded by the sum of the probabilities of concurrent occurrence of these two events. Explicitly,

$$E[R_s(k)] \le \sum_{l=1}^{k}\sum_{g\in\mathcal{G}}\sum_{t\in\mathcal{T}}\sum_{a\in\mathcal{A}} \Pr\left\{\hat{\alpha}(k)\notin a^{*}(t,g),\; N_{t,g,a}(l) \ge 4\frac{b\ln l}{\Delta_{\min}^{2}}\right\}. \quad (27)$$

Using Lemma 1, we can rewrite (27) as

$$E[R_s(k)] \le \sum_{l=1}^{k}\sum_{g\in\mathcal{G}}\sum_{t\in\mathcal{T}}\sum_{a\in\mathcal{A}} \left\{\Pr\left|\hat{Y}_{t,g,a}(l) - \eta_m(t,g,a)\right| \ge \frac{1}{2}\Delta_{\min},\; N_{t,g,a}(l) \ge 4\frac{b\ln l}{\Delta_{\min}^{2}}\right\}. \quad (28)$$

Given that $N_{t,g,a}(l) \ge 4\frac{b\ln l}{\Delta_{\min}^{2}}$ in the second condition of the probability of (28), $\Delta_{\min} \ge 2\sqrt{\frac{b\ln l}{N_{t,g,a}(l)}}$. Therefore,

$$\Pr\left\{\left|\hat{Y}_{t,g,a}(l) - \eta_m(t,g,a)\right| \ge \frac{1}{2}\Delta_{\min},\; N_{t,g,a}(l) \ge 4\frac{b\ln l}{\Delta_{\min}^{2}}\right\}$$

$$\leq \quad \Pr\left\{\left|\hat{Y}_{t,g,a}(l) - \eta_m(t,g,a)\right| \geq \sqrt{\frac{b\ln l}{N_{t,g,a}(l)}}, \right.$$
$$\left. N_{t,g,a}(l) \geq 4\frac{b\ln l}{\Delta^2_{\min}}\right\}.$$

Using the Chernoff-Hoeffding bound, the last probability can be upper bounded by $2\exp\left(-2b\ln l\right)$. Replacing this upper bound on (28), we reach

$$E[R_s(k)] \leq \sum_{l=1}^{k}\sum_{g\in\mathcal{G}}\sum_{t\in\mathcal{T}}\sum_{a\in\mathcal{A}} 2\exp\left(-2b\ln l\right). \qquad (29)$$

The last expression is upper bounded by the desired result. ■

Lemma 3.

Proof: At any recognition transaction l, $l \leq k$, at most $c(l) + 1$ exploration steps have been experienced for each contention and congestion levels $g \leftarrow g(l)$ and $t \leftarrow t(l)$. Then

$$\begin{aligned} E\left[R_e(k)\right] \quad &\leq \quad \sum_{g\in\mathcal{G}}\sum_{t\in\mathcal{T}}\sum_{a\in\mathcal{A}}\left(4\frac{b\ln k}{\Delta^2_{\min}} + 1\right) \qquad (30) \\ &= \quad |\mathcal{G}|\,|\mathcal{T}|\,S\left(1 + c\left(k\right)\right). \end{aligned}$$

■

Lemma 5.

Proof: Let $W(l)$ be the event that proposed algorithm chooses a suboptimal joint action profile in the exploitation stage at recognition transaction l, $l \leq k$. Then the expected regret for this case is expressed as

$$E\left[R_s\left(k\right)\right] = \sum_{l=1}^{k}\Pr\{W(l)\}. \qquad (31)$$

The occurrence of event $W(l)$ depends on: *(i)* the suboptimal profile selection, i.e $\hat{\alpha}(l) \notin \mathbf{a}^*(t)$ and *(ii)* the occurrence of the exploitation stages in the proposed algorithm, i.e $N_{t,\mathbf{a}}(l) \geq c(l)$ given contention $t \leftarrow t(l)$ for all settings $\mathbf{a}$. Therefore, for all congestion levels, (31) will be upper bounded by the sum of the probabilities of these two events occurring simultaneously. Explicitly

$$E\left[R_s\left(k\right)\right] \quad \leq \quad \sum_{l=1}^{k}\sum_{t\in\mathcal{T}}\sum_{h\in\mathcal{H}}\Pr\left\{\alpha(l) \notin h^*(t)\right., \qquad (32)$$
$$\left. N_{t,h}(l) \geq 4\frac{b\ln l}{\Delta^2_{\min}}\right\}.$$

Via Lemma 4, we can rewrite (32) as

$$E\left[R_s\left(l\right)\right] \quad \leq \quad \sum_{l=1}^{k}\sum_{t\in\mathcal{T}}\sum_{h\in\mathcal{H}}\Pr\left\{\left|\hat{Z}_{t,h}(l) - \eta\left(t,h^-\right)\right| \leq \frac{1}{2}\Delta_{\min}, \right.$$
$$\left. N_{t,h}(l) \geq 4\frac{b\ln l}{\Delta^2_{\min}}\right\}. \qquad (33)$$

Given that $N_{t,\mathbf{a}}(l) \geq 4\frac{b\ln l}{\Delta^2_{\min}}$, we have $\Delta_{\min} \geq 2\sqrt{\frac{b\ln l}{N_{t,\mathbf{a}}(l)}}$. Therefore the probability in the summation of (33) is upper-bounded by

$$\Pr\left\{\left|\hat{Z}_{t,h}(l) - \eta\left(t,h\right)\right| \geq \sqrt{\frac{b\ln l}{N_{t,h}(l)}}, N_{t,h}(l) \geq 4\frac{b\ln l}{\Delta^2_{\min}}\right\}$$
$$\leq 2\exp(-2b\ln l).$$

The last result follows from using Chernoff-Hoeffding bound. Then, we have

$$E\left[R_s(k)\right] \leq 2\sum_{l=1}^{k}\sum_{t\in\mathcal{T}}\sum_{h\in\mathcal{H}}\exp(-2b\ln l). \qquad (34)$$

Since $|\mathcal{H}| = \binom{M+S-1}{S-1}$, (34) is upper bounded by $2\binom{M+S-1}{S-1}|\mathcal{T}|\sum_{l=1}^{k} l^{-2b}$, with the sum being $H_k^{(2b)}$. ■

Lemma 6.

Proof: At any recognition transaction l, $l \leq k$, at most $c(l)+1$ exploration steps have taken place for each contention level $t \leftarrow t\left(l\right)$. Then,

$$\begin{aligned} E\left[R_e(k)\right] \quad &\leq \quad \sum_{t\in\mathcal{T}}\sum_{h\in\mathcal{H}}\left(\frac{b\ln k}{\Delta^2_{\min}} + 1\right) \qquad (35) \\ &= \quad |\mathcal{T}|\binom{M+S-1}{S-1}\left(1 + \frac{b\ln k}{\Delta^2_{\min}}\right). \end{aligned}$$

■

References

[1] A. Anandkumar, N. Michael, and A. Tang. Opportunistic spectrum access with multiple players: Learning under competition. In *Proc. of IEEE INFOCOM*, 2010.

[2] P. Auer, N. Cesa-Bianchi, and P. Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47:235–256, 2002.

[3] W. Bao, H. Li, N. Li, and W Jiang. A liveness detection method for face recognition based on optical flow field. In *Proc. IEEE Int. Conf. on Image Anal. and Signal Process., 2009, IASP 2009*, pages 233–236. IEEE, 2009.

[4] S. Barbarossa, S. Sardellitti, and P. Di Lorenzo. Computation offloading for mobile cloud computing based on wide cross-layer optimization. In *IEEE Future Network and Mobile Summit (FutureNetworkSummit), 2013*, pages 1–10. IEEE, 2013.

[5] B. C. Becker and E. G Ortiz. Evaluation of face recognition techniques for application to facebook. In *8th IEEE Internat. Conf. on Automatic Face & Gesture Recognition, 2008. FG'08.*, pages 1–6. IEEE, 2008.

[6] G. Bianchi. Performance analysis of the ieee 802.11 distributed coordination function. *IEEE J. Select. Areas in Commun.*, 18(3):535–547, 2000.

[7] H. Bredin, A. Miguel, I. H. Witten, and G. Chollet. Detecting replay attacks in audiovisual identity verification. In *Proc. IEEE Int. Conf. Acoust., Speech and Signal Process., 2006, ICASSP 2006*, volume 1, pages I–I. IEEE, 2006.

[8] F. Fu and M. van der Schaar. Decomposition principles and online learning in cross-layer optimization for delay-sensitive applications. *IEEE Trans. Signal Process.*, 58(3):1401–1415, 2010.

[9] F. Fu and M. van der Schaar. Structure-aware stochastic control for transmission scheduling. *IEEE Trans. Veh. Tech.*, 61(9):3931–3945, 2010.

[10] F. Fu and M. van der Schaar. Structural solutions for dynamic scheduling in wireless multimedia transmission. *IEEE Trans. Circ. Syst. for Video Techol.*, 22(5):727–739, 2012.

[11] A. S. Georghiades, P. N. Belhumeur, and D. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. on Pat. Anal. and Mach. Intel.*, 23(6):643–660, 2001.

[12] B. Girod, V. Chandrasekhar, D. M. Chen, N.-M. Cheung, R. Grzeszczuk, Y. Reznik, G. Takacs, S. S. Tsai, and R. Vedantham. Mobile visual search. *IEEE Signal Processing Magazine*, 28(4):61–76, 2011.

[13] Knuth D. E. Graham, R. L. and O. Patashnik, editors. *Concrete Mathematics: A Foundation for Computer Science, 2nd ed.* Reading, MA: Addison-Wesley, 1994.

[14] C. Greco, M. Cagnazzo, and B. Pesquet-Popescu. Low-latency video streaming with congestion control in mobile ad-hoc networks. *IEEE Trans. on Multimedia*, 14(4):1337–1350, 2012.

[15] E. Gumus, N. Kilic, A. Sertbas, and O. N. Ucan. Evaluation of face recognition techniques using PCA, wavelets and SVM. *Elsevier Expert Syst. with Appl.*, 37(9):6404–6408, 2010.

[16] Y. Im, C. Joe-Wong, S. Ha, S. Sen, T. Kwon, and M. Chiang. Amuser: Empowering users for cost aware offloading with throughput delay tradeoffs. In *Proc. IEEE INFOCOM), 2013*. IEEE, 2013.

[17] H.-K. Jee, S.-U. Jung, and J.-H. Yoo. Liveness detection for embedded face recognition system. *Int. J. of Biomedical Sciences*, 1(4), 2006.

[18] M. Jones and P. Viola. Fast multi-view face detection. *Mitsubishi Electric Research Lab TR-20003-96*, 3:14, 2003.

[19] T. Lai and H. Robbins. Asymptotically efficient adaptive allocation rules. *Adv. Appl. Math*, 6, 1985.

[20] J. Langford and T. Zhang. The epoch-greedy algorithm for contextual multi-armed bandits. *Adv. in Neural Informat. Process. Syst.*, 20:1096–1103, 2007.

[21] K.-C. Lee, J. Ho, and D. Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Trans. on Pat. Anal. and Mach. Intel.*, 27(5):684–698, 2005.

[22] V. C. M. Leung, M. Chen, M. Guizani, and B. Vucetic. Cloud-assisted mobile computing and pervasive services. *IEEE Network*, page 4, 2013.

[23] A. Li, X. Yang, S. Kandula, and M. Zhang. Cloudcmp: comparing public cloud providers. In *Proc. 10th ACM SIGCOMM Conf. on Internet Meas.*, pages 1–14. ACM, 2010.

[24] L. Li, Langford J. Chu, W. and, and R. E. Schapire. A contextual-bandit approach to personalized news article recommendation. *Proc. of the 19th Internat. Conf. on World Wide Web*, pages 661–670, 2010.

[25] H. Liu, K. Liu, and Q. Zhao. Learning in a changing world: Restless multiarmed bandit with unknown dynamics. *IEEE Trans. on Information Theory*, 59:1902 1916, 2013.

[26] K. Liu and Q. Zhao. Distributed learning in multi-armed bandit with multiple players. *IEEE Trans. on Signal Processing*, 58:5667–5681, 2010.

[27] D. G. Lowe. Object recognition from local scale-invariant features. In *Proc. 7th Internat. Conf. Computer vision, 1999*, volume 2, pages 1150–1157. Ieee, 1999.

[28] Xiaoqiang M., Yuan Z., Lei Z., Haiyang W., and Limei P. When mobile terminals meet the cloud: computation offloading as the bridge. *IEEE Network*, 27(5):28–33, 2013.

[29] S. Marcel, C. McCool, C. Atanasoaei, F. Tarsetti, J. Pesán, P. Matejka, J. Cernocky, M. Helistekangas, and M. Turtinen. MOBIO: Mobile biometric face and speaker authentication. In *Proc. IEEE Conf. Comput. Vision and Pat. Rec.*, San Francisco, CA, USA, 2010.

[30] D. Miao, W. Zhu, C. Luo, and C. W. Chen. Resource allocation for cloud-based free viewpoint video rendering for mobile phones. In *Proc. of the 19th ACM Internat. Conf. on Multimedia*, pages 1237–1240. ACM, 2011.

[31] G. Pan, L. Sun, Z. Wu, and S. Lao. Eyeblink-based anti-spoofing in face recognition from a generic webcamera. In *Proc. IEEE Int. Conf. on Computer Vision, 2007, ICCV 2007*, pages 1–8. IEEE, 2007.

[32] G. Pan, Z. Wu, and L. Sun. Liveness detection for face recognition. *J. Recent Adv. in Face Recogn.*, pages 109–124, 2008.

[33] N. Poh, C. H. Chan, J. Kittler, S. Marcel, C. McCool, E. A. Rúa, J. L. A. Castro, M. Villegas, R. Paredes, V. Struc, et al. An evaluation of video-to-video face verification. *IEEE Trans. Inf. Forens. and Sec.*, 5(4):781–801, 2010.

[34] Shaolei R. and M. van der Schaar. Efficient resource provisioning and rate selection for stream mining in a community cloud. *IEEE Trans. on Multimedia*, 15(4):723–734, 2013.

[35] H. Sellahewa and S. A. Jassim. Wavelet-based face verification for constrained platforms. In *SPIE Proc. Defense and Secur. Conf.*, pages 173–183. International Society for Optics and Photonics, 2005.

[36] D. Siewiorek. Generation smartphone. *IEEE Spectrum*, 49(9):54–58, 2012.

[37] A. Slivkins. Contextual bandits with similarity information. In *24th Annual Conference on Learning Theory (COLT)*, 2011.

[38] T. Soyata, R. Muraleedharan, C. Funai, M. Kwon, and W. Heinzelman. Cloud-vision: Real-time face recognition using a mobile-cloudlet-cloud acceleration architecture. In *2012 IEEE Symposium on Computers and Communications (ISCC)*, pages 59–66. IEEE, 2012.

[39] Barto A. Sutton, R., editor. *Reinforcement learning, an introduction.* Cambridge: MIT Press/Bradford Books, 1998.

[40] C. Tekin and M. Liu. Online learning in decentralized multi-user spectrum access with synchronized explorations. In *Proc. IEEE MILCOM*, 2012.

[41] M. van Der Schaar and S. Shankar. Cross-layer wireless multimedia transmission: challenges, principles, and new paradigms. *IEEE Wireless Communications Mag.*, 12(4):50–58, 2005.

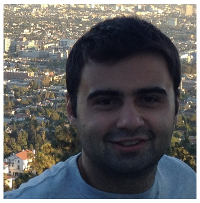
[42] J. Wright, A. Y Yang, A. Ganesh, S. S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Trans. on Pattern Anal. and Machine Intell.*, 31(2):210–227, 2009.

[43] J. Yang, D. Zhang, A. F. Frangi, and J.-Y. Yang. Two-dimensional pca: a new approach to appearance-based face representation and recognition. *IEEE Trans. on Pattern Anal. and Machine Intell.*, 26(1):131–137, 2004.

[44] Weiwen Z., Yonggang W., Jun W., and Hui L. Toward a unified elastic computing platform for smartphones with cloud support. *IEEE Network*, 27(5):34–40, 2013.

[45] Wenwu Z., Chong L., Jianfeng W., and Shipeng L. Multimedia cloud computing. *IEEE Signal Proces. Mag.*, 28(3):59–69, 2011.

[46] X. Zhu, C. Lany, and M. van der Schaar. Low-complexity reinforcement learning for delay-sensitive compression in networked video stream mining. In *IEEE Internat. Conf. on Multimedia and Expo (ICME), 2013*, pages 1–6. IEEE, 2013.



Onur Atan received B.Sc degree in Electrical Engineering from Bilkent University, Ankara, Turkey in 2013. He is currently pursuing the PhD degree in Electrical Engineering at University of California, Los Angeles. His research interests include online learning and multi-armed bandit problems.

Yiannis Andreopoulos (M'00) is Senior Lecturer at the Electronic and Electrical Engineering Department of University College London (UK). His research interests are in wireless sensor networks, error-tolerant computing and multimedia systems. He received the 2007 "Most-Cited Paper" award from the Elsevier EURASIP Signal Processing: Image Communication journal and a best-paper award from the 2009 IEEE Workshop on Signal Processing Systems. Dr. Andreopoulos was Special Sessions Co-chair of the 10th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2009) and Programme Co-chair of the 18th International Conference on Multimedia Modeling (MMM 2012). He is an Associate editor of the IEEE Transactions on Multimedia, the IEEE Signal Processing Letters and the Elsevier Image and Vision Computing journal.

Cem Tekin is a Postdoctoral Scholar at University of California, Los Angeles. He received the B.Sc. degree in electrical and electronics engineering from the Middle East Technical University, Ankara, Turkey, in 2008, the M.S.E. degree in electrical engineering: systems, M.S. degree in mathematics, PhD degree in electrical engineering: systems from the University of Michigan, Ann Arbor, in 2010, 2011 and 2013, respectively. His research interests include machine learning, multi-armed bandit problems, data mining, cognitive radio and networks. He received the University of Michigan Electrical Engineering Departmental Fellowship in 2008, and the Fred W. Ellersick award for the best paper in MILCOM 2009.

Mihaela van der Schaar [F'10] is Chancellor's Professor of Electrical Engineering at University of California, Los Angeles. Her research interests include network economics and game theory, network science, online learning, real-time stream mining, etc. She was a Distinguished Lecturer of the Communications Society for 2011- 2012, the Editor in Chief of IEEE TRANS. ON MULTIMEDIA and a member of the Editorial Board of the IEEE JOURNAL ON SELECTED TOPICS IN SIGNAL PROCESSING and IEEE JOURNAL ON EMERGING AND SELECTED TOPICS IN CIRCUITS AND SYSTEMS. She received an NSF CAREER Award, the Best Paper Award from IEEE TRANS. ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the Okawa Foundation Award, the IBM Faculty Award, the Most Cited Paper Award from EURASIP: IMAGE COMMUNICATIONS JOURNAL, the Gamenets Conference Best Paper Award and the IEEE Circuits and Systems Society Darlington Award Best Paper Award. For more information about her research visit: http://medianetlab.ee.ucla.edu/