QoS-Aware Joint Access Control and Duty Cycle Control for Machine-to-Machine Communications

Yun Li*, Kok Keong Chai*, Yue Chen *, Jonathan Loo†
*School of Electronic Engineering and Computer Science, Queen Mary University of London
†School of Science and Technology, Middlesex University
{yun.li, michael.chai, yue.chen}@qmul.ac.uk, J.Loo@mdx.ac.uk

***Abstract*—Massive energy constrained devices and various applications imposes new challenges for Machine-to-Machine (M2M) communications to enable Internet of Things (IoT). In this paper, we investigate a QoS-aware joint access control and duty cycle control problem for M2M communications to optimise the overall network performance, including energy efficiency, end-to-end delay, reliability, throughput and fairness. We first model a practical hybrid M2M communication network and define a cost function as the overall network performance indicator. Then, an optimisation problem is formulated for minimisation of long-term aggregated network cost. Further more, we overcome the non-convexity of the cost function and mathematically derive the optimal access control. Finally, we propose a distributed access control followed by a reinforcement learning (RL) based duty cycle control which adapts to various network dynamics without priori network information. Simulation results show that, the proposed joint access control and duty cycle control minimise the network long-term aggregated cost, while achieving fairness among cluster heads with QoS differentiation.**

I. Introduction

Machine-to-machine (M2M) communications aim at providing ubiquitous connectivity between devices without human intervention. With an explosion in the number of devices and various applications, M2M communications are considered as the enabling technology for the realisation of Internet of Things (IoT) [1]. It has been addressed that energy efficiency is critical for M2M communications as most of the devices are battery operated. What's more, concurrent and massive access of devices may cause performance degradation, such as intolerable delay, packet loss, and unfairness. To fulfil the requirements of IoT, the main design challenge for M2M communications is to effectively manage the massive access of energy constrained devices while satisfying different Quality of Service (QoS) requirements [2].

In existing work, throughput maximisation access control for cellular networks and energy efficiency duty cycle control for machine type networks are considered separately. In [2], the optimisation of bit-per-jour capacity under statistical QoS guarantees is achieved via resource allocation and power control. A distributed channel sharing algorithm is proposed in [3] with the aim of maximising the sum weighted data-rate of all devices.

On the other hand, an increasing interest is addressed in duty cycle control to improve the end-to-end network performance in multi-hop networks. Distributed duty cycle controls for multi-hop networks are proposed in [4] and [5]. The end-to-end delay is guaranteed by adapting sleep intervals based on assigned local delay requirement to each single hop. To address the dynamic network conditions and non-perfect information of devices, reinforcement learning (RL) is employed in [6], [7]. More recently, an adaptive optimal duty-cycle algorithm is proposed for non-beacon-enabled IEEE 802.15.4 with the aim of minimising energy consumption while meeting the reliability and delay requirements [8]. In our previous work [9], we investigated a duty cycle control problem for M2M networks as a joint optimisation of energy efficiency, delay and reliability. RL technique is applied to learn the optimal duty cycle control in networks with unavailable network information, various network dynamics, and time-varying traffics.

Recently, a hybrid architecture is proposed by [10] to support M2M communications in 5G systems, where M2M gateways act as protocol translation points between short-range capillary networks and cellular networks. Due to the coexistence of cellular and capillary networks, it is crucial to optimise the overall network performance by simultaneous consider access control and duty cycle control. In this paper, we propose a QoS-aware joint access control and duty cycle control problem. The aim is to optimise the overall networks performance, while achieving fairness among cluster heads with QoS differential. The objective network performance metrics including energy efficiency, end-to-end delay, throughput and packet drop ratio.

The contributions of the paper are summarised as follows. First, we model a practical hybrid M2M network with dynamic traffic generation, different application requirements and device capabilities. Then, we define a cost function as the overall network performance indictor, which take consideration of energy efficiency, end-to-end delay, throughput and packet drop ratio. Next, the formulation of a QoS-aware joint access control and duty cycle control problem to minimise the long-term aggregated cost is given. After that, we overcome the non-convexity of the cost function, and derived the optimal access control. Last but not least, a distributed access control followed by a RL based duty cycle control without priori network information is proposed to ensure the control feasibility under dynamic network conditions is proposed. At convergence, cost minimisation of the network and cost proportional fairness among cluster heads are achieved.

II. System Model

According to [10], base station (BS), M2M gateways, cluster heads and end-devices are four types of devices of networks, as shown in Fig 1. We consider the dominate short-range technology IEEE 802.15.4 in capillary networks.

The network operates at discrete time domain where time is divided into time periods $t = 0, 1, \ldots, T$. Each device is equipped with a single omnidirectional antenna for uplink transmission. The M2M gateways is denoted as $n \in \mathcal{N}$. The cluster heads of gateway n forms the cluster head set $\mathcal{I}_n$ and link set $\mathcal{L}_{i,n}$. The immediate child devices of cluster head i form devices set $\mathcal{C}_i$ and link set $\mathcal{L}_{i,j}$ with cluster head i. As each cluster head may run different applications with each other, θ_i^t is denoted as the priority factor for cluster head i.

Assume all generated packets are available at the beginning of each time period t. Each device maintains one queue. The new arrived packets will be dropped if the queue length reaches its maximum q_i^{max}. The change of queue length of cluster head i is given as

$$q_i^{t+1} = \min\left([q_i^t + r_i^t - f_i^t + g_i^t]^+, q_i^{max}\right), \tag{1}$$

where $0 \le t \le T-1$, $[\cdot]^+ = \max(0, \cdot)$, g_i^t is the number of packets being generated by device i in time period t; f_i^t is the number of packet transmitted by cluster head i in time period t; and r_i^t is the number of packets received by cluster head i in time period t.

To fit into the practical scenarios, we adopt an empirical dual-slope propagation model of path loss with distance, Nakagami frequency-flat small-scale fading, and lognormal shadowing [**?**]. The overall channel propagation loss is

$$\begin{aligned} L_{c,dB} =& L_{0,dB} + X_{s,dB} + X_{f,dB} \\ &+ \begin{cases} 10n_0 \log_{10}(d) & d \le d_1, \\ 10n_0 \log_{10}(d_1) + 10n_1 \log_{10}(\frac{d}{d_1}) & d > d_1, \end{cases} \end{aligned} \tag{2}$$

where X_s dB is a zero mean Gaussian random variable with standard deviation σ_s, d is the distance between the sender and receiver, $X_{f,dB} = 10\log(X_f)$ and X_f is a unit-mean gamma-distributed random variable with variance $1/m$ (m is the Nakagami fading parameter).

We assume that the fading and shadowing are constant during each time period. The condition for the successful transmission is that the received signal power $P_{rec,i}^t$ is above the sensitivity threshold $P_{sens,i}^t$ of the device. The received power $P_{rec,i}^t$ of device i is the sum of the conducted power to the transmit antenna, the path loss due to channel propagation, the transmit and receive antenna gains. We denote the successful transmission probability $\rho_{i,j}^t$ as,

$$\rho_{i,j}^t = \begin{cases} 1 & P_{rec,i}^t \ge P_{sens,i}^t \\ 0 & P_{rec,i}^t < P_{sens,i}^t. \end{cases} \tag{3}$$

Due to the duty cycle mechanism, the effective 802.15.4 link rate is $R_{i,n}^t = 250kbps \times \textit{duty cycle}$ and the cellular link rate of gateway is $R_{n,b} = Blog_2(1 + SNR_{n,b})$.

For IEEE 802.15.4 (2011) beacon-enabled mode in capillary networks, the duration between two consecutive beacons is called Beacon Interval (BI), while the duration of an active period is called Superframe Duration (SD). More specifically, $BI = aBaseSuperFrameDuration \times 2^{BO}$ and $SD = aBaseSuperFrameDuration \times 2^{SO}$, where Beacon Order ($BO$) and Superframe Order ($SO$) $0 \le SO \le BO \le 14$ are two integers. $aBaseSuperFrameDuration = 15.36ms$ at 2.4 GHz with 250 kbps bandwidth.

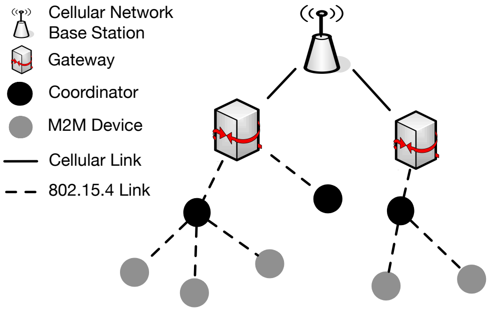


Fig. 1: Network model.

The duty cycle is defined as the ratio of the active portion over each time period, thus $Duty\ Cycle = SD/BI = 2^{SO-BO}$. For multi-hop networks, each cluster head divides its BI into two superframes, named incoming superframe and outgoing superframe [11]. The cluster head i receives the beacon from its parent M2M gateway in the incoming superframe, and transmits its beacon to its child devices in the outgoing superframe. While the incoming superframe duty cycle is decided by the parent device and is enclosed in the received beacon, each M2M gateway/router can control its outgoing superframe duty cycle (refered as duty cycle in this paper). To simplify the synchronisation, we assume all routers have same BO. Thus, the duty cycle control of device i is achieved by setting its outgoing SO.

III. Problem Formulation

A. Cost Function Design

We define the transmitting energy cost $E_t(f_i^t)$, receiving energy cost $E_r(r_i^t)$, idle listening energy cost $E_l(r_i^t)$, and end-to-end delay cost $D(r_i^t)$ of the device i all in terms of number of packets, thus the costs of energy consumption and end-to-end delay are additive. As the energy consumption of ACK packets transmission exists only when the cluster head receives packets. Thus a fixed ACK transmission energy cost A is introduced along with the receiving energy consumption. The specific definition of each cost is given as:

$$E_r(r_i^t) = \begin{cases} A + c_r \cdot \frac{r_i^t}{q_i^{max} \cdot l_i} & \text{if} \quad r_i^t > 0, \\ 0 & \text{if} \quad r_i^t = 0. \end{cases} \tag{4}$$

$$E_f(f_i^t) = c_f \cdot \frac{f_i^t}{q_i^{max} \cdot l_i}, \tag{5}$$

$$E_l(f_i^t, r_i^t) = c_l \cdot \frac{[f_i^t - g_i^t - q_i^t - r_i^t]^+}{q_i^{max} \cdot l_i}, \tag{6}$$

$$D(f_i^t, r_i^t) = c_d \cdot \frac{[q_i^t + r_i^t + g_i^t - f_i^t]^+}{q_i^{max} \cdot l_i}, \tag{7}$$

where c_f, c_r, c_l and c_d are the cost coefficients of transmitting, receiving, idle listening and delay, respectively and $A = c_f \cdot J$. Note that $c_r < c_l$, as if c_r were greater than c_l, it would never be optimal to receive new packet at the last period and possibly in earlier periods. Different device buffer size q_i^{max} and device type l_i have also been integrated into the problem formulation to differentiate gateway, cluster heads and child devices. With the operation $[\cdot]^+$, the packet drop ratio is also considered in the designed costs. The expected weighted-sum joint-cost function for device i at time period t is

$$\begin{aligned} & C_i(f_i^t, r_i^t) \qquad (8) \\ & = \mathbb{E}\bigg\{\alpha\bigg(E_f(f_i^t) + E_r(r_i^t) + E_l(f_i^t, r_i^t)\bigg) + \beta D(f_i^t, r_i^t)\bigg\}, \end{aligned}$$

where α and β ($\alpha+\beta = 1$) are the weighting factors of energy efficiency and end-to-end delay for different applications.

B. Access Control Problem

The aim of our access control is to minimised the long-term aggregated costs. Based on the modelled M2M network and cost function, the access control problem is formulated as:

$$P1 : \min \sum_{t=0}^{T} \sum_{i \in \mathcal{I}_n} \theta_i^t C_i(f_i^t, r_i^t) \qquad (9)$$

$$s.t. \quad \sum_{i \in \mathcal{I}_n} f_i^t \leq R_{n,b}^t, \qquad (10)$$

$$0 \leq f_i^t \leq R_l^t, \qquad (11)$$

$$r_i^t \leq \min(R_l^t, q_i^{max}), \qquad (12)$$

where $l \in \mathcal{L}_{i,n}$. The objective of (10) is to minimise the long-term aggregated cost. The constraint (10) - (12) are the link capacity constraints, which state the total transmitted packets of $i \in \mathcal{I}_n$ within each time period should be no more than the transmission link capacity or the number of available packets.

Similar to the network utility maximisation (NUM) framework [13], P1 can be decompose into two distributed sub-problem. Taking different link characteristic, device function and hardware limitations into consideration, the main objective of cluster heads is to achieve joint optimisation of energy efficiency, end-to-end delay, and reliability, while that of the gateways is to maximise the throughput while achieving fairness allocation among cluster heads. We decomposed P1 into the following P2 and P3,

$$P2 : \max - \sum_{t=0}^{T} \sum_{i \in \mathcal{I}_n} \theta_i^t \bigg(C_i(\frac{p_{i,n}^t}{p_i^t}, r_i^t) - p_{i,n}^t \bigg) \qquad (13)$$

$$s.t. \quad p_{i,n}^t \geq p_i^t \geq 0, \qquad (14)$$

$$r_i^t \leq \min(R_l^t, q_i^{max}), \qquad (15)$$

where p_i^t is regarded as a bid price (per packet), the charged price $p_{i,n}^t = p_i^t \cdot f_i^t, i \in \mathcal{I}$ and $f_i^t = p_{i,n}^t / p_i^t$.

Suppose the gateway n knows its revenue vector $P_n = (p_{i,n}^t, i \in \mathcal{I}_n)$, and the priority factor of cluster head θ_i^t. We offset f_i^t by +1 to ensure the positiveness of the cost, thus

$$P3 : \max \sum_{t=0}^{T} \sum_{l \in \mathcal{I}_n} p_{i,n}^t \theta_i^t \log(1 + f_i^t) \qquad (16)$$

$$s.t. \quad \sum_{i \in \mathcal{I}_n} f_i^t \leq R_{n,b}^t, \qquad (17)$$

$$0 \leq f_i^t \leq R_l^t. \qquad (18)$$

Under the decomposition, the cost function C_i of cluster head is not required by the gateway, and only appears in the optimisation problem P2 faced by cluster head i. Then the solution of P3 is the gateway control with the aim of fairly access control between the gateway and cluster heads with different QoS requirements. By solving the access control distributively, significant reduction of overhead and complexity is achieved compared to the centralised control [3].

C. Duty Cycle Control

IEEE 802.15.4 beacon-enabled mode applies slotted carrier-sense multiple access with slotted collision avoidance (CSMA/CA) for frame transmission. We assume devices need to perform two clear channel assessments (CCAs) before frame transmission. The beacon transmission duration is D_{bcn}. For each cluster head, the total frame transmission duration is given as $PD = SD - D_{bcn} = \sum_{j=1}^{J} \lceil D_j \rceil + \lceil \delta + D_{ACK} \rceil$, where D_j is the frame transmission duration of child device $j \in \mathcal{C}_i$ and $\rho_{i,j}^t = 1$, δ and D_{ACK} are waiting time and transmission duration of the ACK packet, respectively. Then the number of packets that can be received by i is $r_i^t = \sum_{j=1}^{J} \lfloor b \cdot D_j / D_p \rfloor$, where D_p is transmission duration per packet and b is the throughput limitation coefficient [12], which shows impact of the backoff and contention during CSMA/CA transmission. Then, the relationship between the minimum SO and r_i^t can be presented as

$$SO(r_i^t) = \left\lceil \log_2(\left\lceil \frac{r_i^t D_p}{b} \right\rceil + \lceil \delta + D_{ACK} + D_{bcn} \rceil) \right\rceil. \quad (19)$$

IV. Joint Access Control and Duty Cycle Control

Using duality, the distributed iterative algorithm to the P3 can be achieved. At convergence, the cost proportional fairness is achieved. Then based on the optimal result f_i^{t*} of P3, P2 is the cluster head control with the aim of minimising the long-term aggregated costs. Thus the objectives of optimising the overall network performance, including energy efficiency, end-to-end delay, reliability, throughput and fairness can be achieved by a QoS-aware joint access control and duty cycle control.

A. Gateway Control

Under the strict concavity on $\log(1+f_i^t)$, there always exists a unique optimal solution f_i^t to the maximisation problem P3. The optimal solution f_i^t can be obtained by looking for a *saddle-point* in the following Lagrangian form:

$$\begin{aligned} & L(f_i^t, p_{i,n}^t; \mu_i^t) \qquad (20) \\ & = \sum_{t=0}^{T} \bigg(\sum_{i \in \mathcal{I}_n} p_{i,n}^t \log(1 + f_i^t) - \sum_{i \in \mathcal{I}_i} \mu_i^t (f_i^t - R_{i,n}^t) \bigg), \end{aligned}$$

where μ_i^t is Lagrange multipliers, $\frac{\partial L(\cdot)}{\partial f_i^t} = \frac{p_{i,n}^t}{f_i^t} - \sum_{i \in \mathcal{I}_i} \mu_i^t$.

The access control algorithm is implemented at each gateway to adapt transmission based on the feedback charged price $p_{i,n}^t$ of its child cluster heads. With the the bid price p_i^t of each cluster head $i \in \mathcal{I}_n$, the unique optimum to the P3 is

$$p_{i,n}^t = p_i^t \cdot f_i^t, \quad f_i^t = \frac{p_{i,n}^t}{\sum_{i \in \mathcal{I}_i} \mu_i^t}. \tag{21}$$

B. Cluster Head Control

The bid price algorithm at cluster head i is operated to get p_i^t depending on the QoS requirement and severity of buffered queue length. If η is the feedback control parameter, at time period $t+1$ each cluster head updates its bid price p_i^t as,

$$p_i^{t+1} = p_i^t \cdot \left(1 + \eta \theta_i^t \operatorname{sign}\left[f_i^t - q_i^t - r_i^t\right]^+\right) \tag{22}$$

The above equation indicate that if the aggregated number of packets exceeds the maximum buffer size q_i^{max}, the bid price will be increased; otherwise it will be decreased.

Based on the results of the P3 at gateway, each cluster head need to solve its local P2. Then, the duty cycle control at cluster head is achieved according to (19). However, it is not trivial to solve $P2$ as function $C_i(f_i^t, r_i^t)$ is not a convex function due to the $[\cdot]^+$ operation of limited buffer size. However, it has been proved by Scarf that an optimal $multi{-}period(s, S)$ solution exists, if $C_i(f_i^t, r_i^t)$ is $A-convex$ function [14].

Definition 1. *The real-valued function f is an $A-convex$ function, if $A \geq 0$, for all $z \geq 0, b > 0$, f satisfies the $A-convexity$ property*

$$A + f(z+y) \geq f(y) + z\left(\frac{f(y) - f(y-b)}{b}\right). \tag{23}$$

Now we will give sufficient conditions to the $A{-}convexity$ of the cost-to-go function $C_i(f_i^t, r_i^t)$. To reduce the number of notations in the equations, we denote $m_i^t = q_i^t + r_i^t$ and $i^t = f_i^t - g_i^t$. If $\delta(0) = 0$, $\delta(r_i^t) = 1$ for $r_i^t > 0$, based on (4) - (8), we have

$$C_i(f_i^t, r_i^t) = \min_{\pi_i \in \mathfrak{D}} \mathbb{E}\left\{A\delta(r_i^t) + W(m_i^t)\right\} - \frac{\alpha c_r \cdot q_i^t}{q_i^{max}}, \tag{24}$$

where $W(m_i^t, i^t) = \alpha E_f(f_i^t) + \alpha E_r(r_i^t) + \alpha c_l \cdot \frac{[i^t - m_i^t]^+}{q_i^{max}} + \beta c_d \cdot \frac{[m_i^t - i^t]^+}{q_i^{max}} + U([m_i^t - i^t]^+)$.

Lemma 1. *According to (13), if $W(m_i^t)$ is an $A-convex$ function, so is $C_i(f_i^t, r_i^t)$.*

Proof. Please refer to Appendix A □

To show the property between $C_i([m_i^t - n_i^t]^+)$ and $W(m_i^t)$, we rewrite (13) as

$$W(m_i^t) = \alpha\left(E_f(f_i^t) + E_r(r_i^t)\right) + \frac{\beta c_d \cdot [m_i^t - n_i^t]^+}{q_i^{max} \cdot l_i} + R(m_i^t),$$

where $R(m_i^t) = \alpha c_l \cdot \frac{[n_i^t - m_i^t]^+}{q_i^{max} \cdot l_i} + C_i([m_i^t - n_i^t]^+)$. The $A-convexity$ of $W(m_i^t)$ holds if the $A{-}convexity$ of $W(m_i^{t+1})$ implies $A-convexity$ of $R(m_i^t)$.

Lemma 2. *According to (20), if $W(m_i^{t+1})$ is an $A-convex$ function, so is $R(m_i^t)$.*

Proof. Please refer to Appendix B □

Theorem 1. *If function $C_i(f_i^t, r_i^t)$ is* A-convex*, the optimal duty cycle control is a multi-period policy: when the queue length q_i^t is smaller than the T_i^t, SO_i^t is set based on (19), otherwise, SO equals to zero:*

$$SO_i^{t^*} = \begin{cases} SO(r_i^{t^*}), & \text{if } q_i^t < T_i^t, \\ 0, & \text{if } q_i^t \geq T_i^t, \end{cases} \tag{25}$$

where $r_i^{t^*}$ is the optimal number of packets the device should received at each time period [9].

To ensure the feasibility of the duty cycle control under various network dynamics, a Q-learning algorithm is applied. For a given policy π, a Q-value represent the expected discounted cost when executing action r_i^t at state q_i^t and then following policy π thereafter, and it is defined as $Q_\pi(r_i^t) = C_i + \delta \min Q_\pi(r_i^{t+1})$, where δ is the discount factor. Then given the learning rate γ, the device will learn the optimal duty cycle by updating its estimation of $Q(r_i^t)$ as

$$Q^{t+1}(r_i^t) = Q^t(r_i^t) + \gamma\left\{C_i^t + \min Q^{t+1}(r_i^t) - Q^t(r_i^t)\right\}. \tag{26}$$

where the learning rate $\gamma \in (0, 1]$ specifies how far the estimate of $Q(r_i^t)$ is adjusted.

V. Simulation Results and Analysis

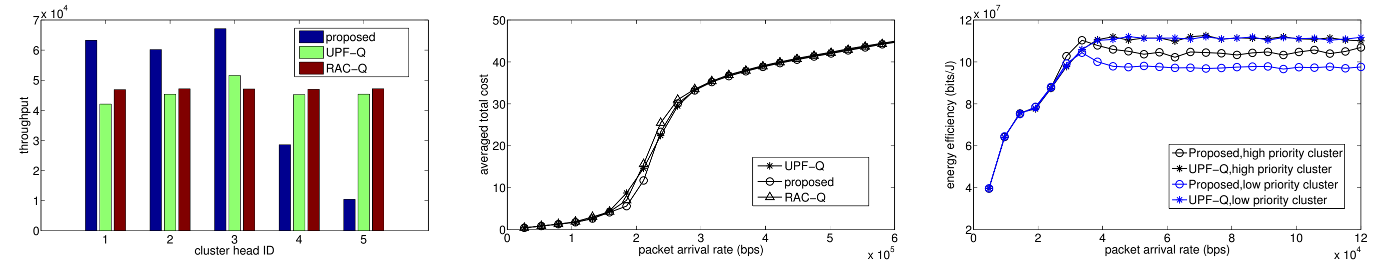
In this section, the performance of proposed QoS-aware joint access control and duty cycle control is evaluated in a two hop cluster-tree network with Matlab. Simulation parameters are given in TABLE I. Device energy consumption parameters are based on CC2420 data sheet [15] and MAC layer parameters are based on IEEE 802.15.4.

We applied ON/OFF traffic model in the simulation. When the device is active (ON), the distribution of the traffic generation follows Poisson distribution. When the device is inactive (OFF), it is idle and does not generates any packets. The service rate f_i^t follows poisson distribution and the number of observation time periods t is 100. The length of each time period BI is 0.49s ($BO = 5$). The results are the averaged values of 1000 runs. The maximum queue length of cluster heads is 50 packets and that of the child devices is 20 packets.

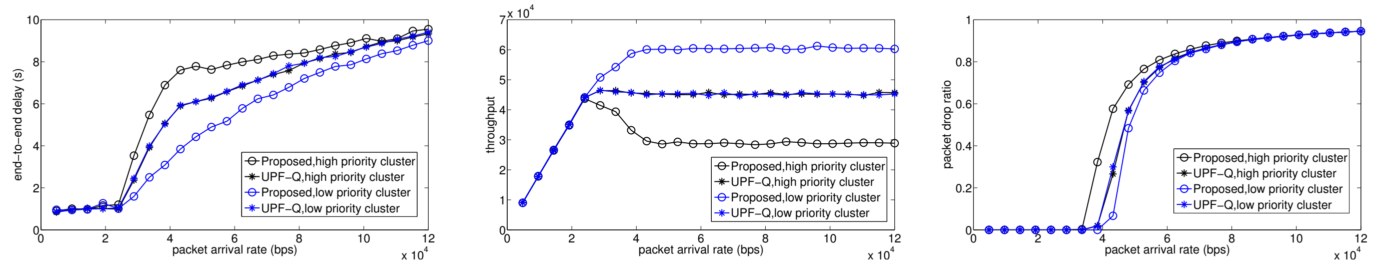
In order to evaluate the performance of the proposed control, in terms of the energy efficiency, end-to-end delay, throughput, packet drop ratio and cost, we consider a hybrid M2M network formed by one BS, one gateway, and 5 cluster heads of two types QoS priorities are connected to the gateway as example. We denoted the 5 cluster heads with cluster head

TABLE I: Simulation Parameters

Parameter	Value	Parameter	Value
frequency	2.4 GHz	α	0.2
data rate	250kbps	β	0.4
transmit power	36.5 mw	packet size	100 bytes
receive power	41.4 mw	CCA size	8 symbols
idle listen power	41.4 mw	ACK packet size	10 symbols
sleep power	0.042 mw	unit backoff period	20 symbols
learning rate	0.9	discount factor	0.5



(a) Throughput value vs. cluster head ID (b) Averaged network cost vs. arrival rate (c) Energy efficiency vs. arrival rate



(d) End-to-end delay vs. arrival rate (e) Throughput vs. arrival rate (f) Packet drop ratio vs. arrival rate

Fig. 2: Performance comparison of the proposed joint control with UPF-Q and RAC-Q.
i) UPF-Q is the utility proportional fairness (UPF) access control proposed in [16] with the Q-learning based duty cycle control proposed in [9]. ii) RAC-Q is a random access control with the Q-learning based duty cycle control proposed in [9]

ID [1,2,3,4,5], respectively. The number of child devices of each cluster head is [6,6,9,6,6]. Among the 5 cluster heads, cluster 1, 2 and 3 have high QoS priority with a large θ_i, while 4 and 5 have low QoS priority with a small θ_i.

Fig. 2(a) shows the throughput of different cluster heads. The throughput of cluster head 3 is higher than the others due to it has more end devices, which will generate more traffic. It can be seen that the throughput of UPF-Q and RAC-Q are evenly distributed among all cluster heads without QoS differentiation, while the proposed control achieved higher throughput for cluster heads with high QoS priority.

Fig. 2(b) shows the averaged total cost of all devices in the network. The averaged cost of the proposed control is no larger than the other two compared controls, and for the packet arrival rate interval [150 - 300] kbps, the proposed control has lower cost than the other two compared controls.

We can conclude from Fig. 2(a) and Fig. 2(b) that the proposed QoS-aware access control and duty cycle control minimised aggregated costs of the network with QoS differentiation among cluster heads.

Fig. 2(c) - (f) show the performance of proposed control energy efficiency, end-to-end delay, throughput, and packet drop ratio. With the focus on QoS differentiation the performance of two representative cluster heads: cluster head 2 with high QoS priority and cluster head 4 with low QoS priority are presented. We can see that the cluster head with high priority has higher throughput, lower end-to-end delay and packet drop ratio compared to the cluster head with low QoS priority. As the trade-off, cluster head with high priority has lower energy efficiency compared to that of the cluster head with low QoS priority.

Similar results can be found among other cluster heads with different QoS priorities. We omit the results due to space limitation.

VI. Conclusion

In this paper, we solved a QoS-aware joint access control and duty cycle control problem for M2M networks. Based on theoretical analysis, the access control is decomposed into a distributive gateway control and a cluster head control, followed by a Q-learning based duty cycle control. The proposed distributed access control reduce the overhead and complexity compared to the centralised control significantly and the Q-learning based duty cycle control ensure the flexibility under various network dynamic without priori networks information. In simulation, a typical M2M communication scenario has been investigated and analysed thoroughly in terms of energy efficiency, end-to-end delay, packet drop ratio, throughput and cost. Simulation results shown that the proposed QoS-aware access control and duty cycle control minimised network aggregated costs while achieving fairness among cluster heads with QoS differentiation.

Appendix A
Proof of Lemma 1

Proof. Based on the definition of $A-convex$, we need to show the following equation holds for all $z \geq 0, b > 0$,

$$A + C_i(q_i^t + z) \geq C_i(f_i^t, r_i^t) + z\left(\frac{C_i(f_i^t, r_i^t) - C_i(q_i^t - b)}{b}\right). \quad (27)$$

As $A > 0$ and $W(m_i^t)$ is $A-convex$, we denote $T_i^t = m_i^{t*} = \arg\min_{m_i^t \in \Re} W(m_i^t)$. Based on (8), the aggregated cost at time period t is

$$C_i(q_i^t) = \begin{cases} A + W(T_i^t) - \frac{\alpha \cdot c_r \cdot q_i^t}{q_i^{max} \cdot l_i} & q_i^t < t_i^t, \\ W(q_i^t) - \frac{\alpha \cdot c_r \cdot q_i^t}{q_i^{max} \cdot l_i} & q_i^t \geq t_i^t, \end{cases} \quad (28)$$

We distinguish three cases to show $C_i(q_i^t)$ is $A-convex$:

Case 1: $q_i^t \geq t_i^t$. If $q_i^t - b \geq t_i^t$, then function $C_i(q_i^t)$ is the sum of a $A - convex$ function and a linear function. Hence, $C_i(q_i^t)$ is $A - convex$ and (27) holds. If $q_i^t - b < q_i^t$, in view of (28), we can write (27) as

$$A + W(q_i^t + z) \geq W(q_i^t) + z\left(\frac{W(q_i^t) - W(t_i^t)}{b}\right). \quad (29)$$

1) If $C_i(q_i^t) \geq C_i(t_i^t)$, then by $A - convexity$ of $W(m_i^t)$,

$$A + W(q_i^t + z) \geq W(q_i^t) + z\left(\frac{W(q_i^t) - W(t_i^t)}{q_i^t - t_i^t}\right)$$
$$\geq W(q_i^t) + z\left(\frac{W(q_i^t) - W(t_i^t)}{b}\right).$$

2) If $C_i(q_i^t) < C_i(t_i^t)$, then

$$A + W(q_i^t + z) \geq A + W(T_i^t) = W(t_i^t) > W(q_i^t)$$
$$\geq W(q_i^t) + z\left(\frac{W(q_i^t) - W(t_i^t)}{b}\right).$$

So for this case, (29) and hence (27) hold.

Case 2: $m_i^t \leq m_i^t + z \leq t_i^t$. In this region, in the view of (25), the function $C_i(q_i^t)$ is linear hence (24) holds.

Case 3: $m_i^t \leq t_i^t \leq m_i^t + z$. For this case, we can write (24) as $A + C_i(q_i^t + z) \geq C_i(t_i^t)$ which holds by the definition of t_i^t. Thus the $A - convexity$ of $C_i(q_i^t)$ is proved given the $A - convexity$ of $W(m_i^t)$. □

Appendix B
Proof of Lemma 2

Proof. We distinguish four cases to show $R(m_i^t)$ is an $A - convex$ function:

Case 1: $0 \leq m_i^{t+1} - b < m_i^{t+1} \leq m_i^{t+1} + z, A - convexity$ of $R(m_i^t)$ follows that of $W(m_i^{t+1})$.

Case 2: $m_i^{t+1} - b < m_i^{t+1} \leq m_i^{t+1} + z \leq 0$: in this region, $R(m_i^t)$ is linear and hence $A - convex$.

Case 3: $m_i^{t+1} - b < m_i^{t+1} \leq 0 \leq m_i^{t+1} + z$: for simplicity, we denote $x = m_i^{t+1} + z$ in this region.

1) $0 < t_i^{t+1} \leq x$*:* $A + C_i(x) \geq A - x \cdot \frac{\alpha c_l \cdot q_i^{t+1}}{q_i^{max} \cdot l_i} + C_i(T_i^{t+1}) = C_i(0) - x \cdot \frac{\alpha c_l \cdot q_i^{t+1}}{q_i^{max} \cdot l_i}$.

2) $t_i^{t+1} \leq 0 \leq x$ *and* $0 \leq x \leq t_i^{t+1}$*:* $A + C_i(x) = 2A - x \cdot \frac{\alpha c_r \cdot q_i^{t+1}}{q_i^{max} \cdot l_i} + C_i(T_i^{t+1}) \geq C_i(0) - x \cdot \frac{\alpha c_l \cdot q_i^{t+1}}{q_i^{max} \cdot l_i}$;

Thus $A + C_i(x) \geq C_i(0) - x \cdot \frac{\alpha c_l \cdot q_i^{t+1}}{q_i^{max} \cdot l_i}$ in this case. According to the definition of $R(m_i^t)$, the $A - convexity$ of $R(m_i^t)$ holds.

Case 4: $m_i^{t+1} - b < 0 < m_i^{t+1} \leq m_i^{t+1} + z$. Then, $0 < m_i^{t+1} < b$.

1) If $\frac{R(m_i^{t+1}) - R(0)}{m_i^{t+1}} \geq \frac{R(m_i^{t+1}) - R(m_i^{t+1} - b)}{b}$, thus

$$A + R(m_i^{t+1} + z) \geq R(m_i^{t+1}) + z\frac{R(m_i^{t+1}) - R(0)}{m_i^{t+1}}$$
$$\geq R(m_i^{t+1}) + z\frac{R(m_i^{t+1}) - R(m_i^{t+1} - b)}{b},$$

2) If $\frac{R(m_i^{t+1}) - R(0)}{m_i^{t+1}} < \frac{R(m_i^{t+1}) - R(m_i^{t+1} - b)}{b}$, then we have

$$R(m_i^{t+1}) - R(0) < \frac{m_i^{t+1}}{b}\left(R(m_i^{t+1}) - R(m_i^{t+1} - b)\right)$$
$$= \frac{m_i^{t+1}}{b}\left(R(m_i^{t+1}) - R(0) + \frac{\alpha c_l \cdot q_i^{t+1}}{q_i^{max} \cdot l_i}(m_i^{t+1} - b)\right).$$

Since $b > m_i^{t+1}$, $R(m_i^{t+1}) - R(0) < -\frac{\alpha \cdot c_l \cdot q_i^{t+1}}{q_i^{max} \cdot l_i} m_i^{t+1}$. Then we have

$$R(m_i^{t+1}) + z\frac{R(m_i^{t+1}) - R(m_i^{t+1} - b)}{b}$$
$$= R(m_i^{t+1}) - z\frac{\alpha c_l \cdot q_i^{t+1}}{q_i^{max} \cdot l_i} \quad (30)$$
$$< R(0) - (m_i^{t+1} + z)\frac{\alpha c_l \cdot q_i^{t+1}}{q_i^{max} \cdot l_i} \leq A + R(m_i^{t+1} + z).$$

Hence, $R(m_i^{t+1})$ is $A - convex$ for all cases. □

References

[1] K. Zheng, F. Hu, W. Wang, W. Xiang and M. Dohler "Resource Radio Allocation in LTE-advanced Cellular Networks with M2M Communications," *Communications Magazine, IEEE*, vol.50, no.7, pp.184-192, 2012.

[2] A. Aijaz, M. Tshangini, M.R. Nakhai, X. Chu and A.H. Aghvami " Results Energy-Efficient Uplink Resource Allocation in LTE Networks With M2M/H2H Co-Existence Under Statistical QoS Guarantees," *Communications, IEEE Transactions on* 2014; vol. 62, no. 7, pp. 2353 - 2365.

[3] S. Bayat, Y. Li, Z. Han, M. Dohler and B. Vucetic "Distributed Massive Wireless Access for Cellular Machine-to-Machine Communication," in *Proc. of Communications (ICC), 2014 IEEE International Conference on*, Sydney, NSW, June 2014.

[4] X. Wang, X. Wang, G. Xing and Y. Yao, "Dynamic Duty Cycle Control for End-to-End Delay Guarantees in Wireless Sensor Networks," in *Proc. IEEE International Workshop on Quality of Service, IWQoS*, Beijing, China, Jun. 2010.

[5] H. Byun and J. Yu, "Adaptive Duty Cycle Control with Queue Management in Wireless Sensor Networks," *IEEE Transactions on Mobile Computing*, vol. 12, no. 6, pp. 1214-1224, Jun. 2013.

[6] Z. Liu, I. Elhanany, "RL-MAC: a Reinforcement Learning based MAC Protocol for Wireless Sensor Networks", *International Journal of Sensor Networks* vol. 1, pp.117124, 2006.

[7] R. Alberola and D. Pesch, "Duty Cycle Learning Algorithm (DCLA) for IEEE 802.15.4 Beacon-enabled Wireless Sensor Networks," *Journal of Ad Hoc Networks*, vol. 10, no. 4, pp. 664–679, 2012.

[8] P. Park, S. C. Ergen, C. Fischione and A. S. Vincentelli, "Duty-Cycle Oprimization for IEEE 802.15.4 Wireless Sensor Networks," *ACM Transactions on Sensor Networks,* vol.10, no.1, pp.12:1–12:32, 2013.

[9] Y. Li, K. K. Chai, Y. Chen and J. Loo "Smart Duty Cycle Control with Reinforcement Learning for Machine to Machine Communications," in *Communications Workshops (ICC), 2014 IEEE International Conference on*, London, UK, June 2015.

[10] ETSI, ETSI TR 102 935: "Machine-to-Machine communications (M2M); Applicability of M2M Architecture to Smart Grid Networks; Impact of Smart Grids on M2M platform", September 2012.

[11] *IEEE std. 802.15.4, Part. 15.4: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks (LR-WPANs)*, IEEE Std., 2011.

[12] T. R. Park, T. H. Kim, J. Y. Choi, S. Choi and W.H. Kwon, "Throughput and energy consumption analysis of IEEE 802.15.4 slotted CSMA/CA," *Electronics Letters*, vol. 41, no. 18, 2005.

[13] F. P. Kelly, A. Maulloo, and D. Tan, "Rate control for communication networks: Shadow prices, proportional fairness and stability," *J. Oper. Res. Sot.*, vol. 49, no. 3, pp. 237-252, Mar. 1998.

[14] H. Scarf, "The Optimality of (s,S) Policies for the Dynamic Inventory Problem," in *Proc. of the 1st Stanford Symposium on Mathematical Methods in the Social Sciences*, Stanford, CA, 1960.

[15] CC2420 2.4 GHz IEEE 802.15.4 Compliant and ZigBee Ready RF Transceiver, online: http://www.ti.com/product/cc2420

[16] J. Jin, A. Sridharan, B. Krishnamachari, and M. Palaniswami, Handling inelastic traffic in multi-hop wireless networks, *Selected Areas in Communications, IEEE Journal on*, vol. 28, no. 7, pp. 1105-1115, 2010.