The final publication is available at

http://dx.doi.org/10.1109/TWC.2013.120713.121948

Channel Assembling with Priority-based Queues in Cognitive Radio Networks: Strategies and Performance Evaluation

Indika A. M. Balapuwaduge, *Student Member, IEEE*, Lei Jiao, *Member, IEEE*, Vicent Pla, *Member, IEEE*, and Frank Y. Li, *Senior Member, IEEE*

Abstract—With the implementation of channel assembling (CA) techniques, higher data rate can be achieved for secondary users in multi-channel cognitive radio networks. Recent studies which are based on loss systems show that maximal capacity can be achieved using dynamic CA strategies. However the channel allocation schemes suffer from high blocking and forced termination when primary users become active. In this paper, we propose to introduce queues for secondary users so that those flows that would otherwise be blocked or forcibly terminated could be buffered and possibly served later. More specifically, in a multi-channel network with heterogeneous traffic, two queues are separately allocated to real-time and elastic users and channel access opportunities are distributed between these two queues in a way that real-time services receive higher priority. Two queuing schemes are introduced based on the delay tolerance of interrupted elastic services. Furthermore, continuous time Markov chain models are developed to evaluate the performance of the proposed CA strategy with queues, and the correctness as well as the preciseness of the derived theoretical models are verified through extensive simulations. Numerical results demonstrate that the integration of queues can further increase the capacity of the secondary network and spectrum utilization while decreasing blocking probability and forced termination probability.

Index Terms: Cognitive radio networks, channel assembling, heterogeneous traffic, queuing schemes, CTMC

I. INTRODUCTION

COGNITIVE radio networks (CRNs) have gained increasing popularity nowadays due to its capability of improving spectrum utilization by exploiting the unused spectrum in dynamically changing environments. For example the newly ratified IEEE 802.22 wireless regional area network (WRAN) standard enables license-exempt devices to use TV bands for Internet access, based on the concept of cognitive radio (CR) [1]. In a CRN, a secondary user (SU) opportunistically accesses a spectrum hole, which is a channel licensed to a primary user (PU) but not being used during a particular time in-

terval. Thus, efficient resource allocation in licensed frequency bands by employing dynamic spectrum access (DSA) becomes one of the essential components for CRN design [2]-[8]. Among existing DSA techniques, channel assembling (CA) is a popular approach which targets at increasing the capacity of the secondary network by allowing one SU to aggregate several idle channels if available. CA can be implemented in two ways, depending on the availability of idle channels in the spectrum domain [8]. The idle channels which are located adjacent to each other or contiguous channels could be *bonded* as a single SU channel. On the other hand, channels which are not contiguous can be *aggregated* as a single SU channel [9]. In this study, we use CA to represent both of them.

Different CA strategies [6]-[8] have been proposed in the literature. The dynamic CA strategy proposed in [7] has been designed for heterogeneous traffic environment and spectrum adaptation is implemented by employing channel sharing and spectrum handover. The results illustrate that the proposed dynamic strategy can achieve better performance than the static strategy does. However, the blocking probability and the forced termination probability of the SU services grow rapidly when the arrival rates of PU and SU services increase. Therefore, how to provide low probabilities of blocking and forced termination for SU services despite higher user arrival rates appears as an interesting question. Motivated by this observation, we propose in this paper a queue-based CA strategy considering real-time and elastic traffic in multi-channel CRNs and evaluate the performance of this strategy. The main idea of our strategy is to accommodate those SU services that might be blocked or dropped by the strategy in [7] in a queue so that they can finish their services at a later instant. In brief, the contributions of this paper can be summarized as follows.

1) A queue-based solution is proposed to further reduce the blocking probability and the forced termination probability of the dynamic CA strategy proposed in [7]. We extend the CA strategies proposed in [6], [7] by introducing two queues which are dedicated for allocating real-time and elastic SU services respectively.
2) Consequently, two queuing schemes are proposed with a novel queue scheduling algorithm which can assign different priority levels to heterogeneous traffic.
3) Analytical models are developed using continuous time Markov chains (CTMCs). The performance of the CRN

is evaluated based on these derived mathematical models and the results are further validated by simulations. For performance evaluation, several parameters including capacity of the secondary network and overall spectrum utilization of the system are considered. Numerical results demonstrate that the proposed queuing schemes improve system performance. Furthermore, delay minimization techniques are considered and average total delay per SU service is also evaluated to analyze the tradeoff due to additional queuing delay.

The remainder of this paper is organized as follows. An overview of related work is given in Sec. II, and then the network scenario is presented in Sec. III, together with assumptions. Sec. IV describes the proposed queuing schemes and the dynamic CA strategy. In order to analyze the system performance, CTMC models are developed in Sec. V. In Sec. VI, we present and discuss the main numerical results and thereafter conclude the paper in Sec. VII.

II. RELATED WORK

The majority of the research efforts on improving SU performance focus on efficient spectrum access and accurate spectrum sensing. Designing effective queuing disciplines in CRNs also has gained importance with the proliferation of spectrum access strategies.

In [10], a queuing framework was developed to analyze the packet level performance of opportunistic access by CR users. A bursty traffic arrival pattern and finite buffer size are considered in their proposed queuing model. Since the proposed spectrum scheduling is based on channel quality information (CQI), extra radio resources have to be allocated for CQI transmission. However, the spectrum scheduling scheme in [10] provides less spectrum adaptation flexibility compared with the channel access strategy proposed in this paper. In [11], a CRN was modeled using a queuing model to evaluate service degradation experienced by PUs due to SUs' incorrect spectrum sensing. A notable feature of the model presented in [11] is that each channel in the system is considered as a single server queuing system and the preempted SU services will not be dropped, instead they will enter a queue associated with another channel. In the CRN context, a *preempted* SU service is an SU service which is forcibly terminated due to the arrival of a PU service. Obviously, the network model in [11] is not suitable for delay-critical applications due to long waiting times. In our work, this particular problem is avoided by introducing multiple queues which have different priority levels and finite buffer sizes. Furthermore, in [11] CA is also not considered either.

The queuing model in [5] was designed with two queues. One of the queues is allocated for incoming SUs and the other one is used for allocating bandwidth resources for both PU and SU services. Inside the bandwidth allocation queue (BAQ), both PU and SU requests are buffered. So the first come first served (FCFS) discipline does not apply to BAQ since PU flows always have priority over SU flows. Furthermore, their model is evaluated by access latency and blocking probability of SU services but the effect of queuing on the performance
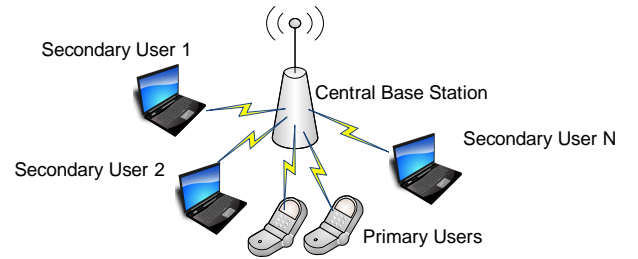

Fig. 1: Network scenario of a CRN with centralized architecture.

of SU capacity and spectrum utilization is not evaluated. In [12], a stochastic fluid queue approach was adopted for delay analysis in CRNs. In that paper, adaptive algorithms were developed for achieving minimum queuing delays in the SU network. To decrease the mean queuing delay, multiple PU channels and multiple SU interfaces were proposed in their queuing model. However, the delay analysis does not differentiate between delay-sensitive SU traffic and elastic SU traffic. In our work, the queuing discipline is designed in a way that it can minimize the waiting time of a real-time SU request in order to maintain its delay-sensitive nature.

The CA strategies proposed in our earlier work [7] provide the basis for the dynamic CA strategy proposed in this paper. In [7], two representative CA strategies were proposed to evaluate the system performance considering heterogeneous SU traffic. However, the CRN considered in [7] is a loss system. That is, new SU requests are simply blocked and the preempted SU services are dropped when spectrum adaptation is not possible in a situation that all the channels in the system are occupied. As an effort to further improve SU performance, the proposed queuing method in this paper can reduce blocking probability and forced termination probability of SU services while increasing the capacity of secondary network and improving spectrum utilization.

III. NETWORK SCENARIO AND ASSUMPTIONS

We consider an infrastructure based CRN consisting of a central base station (CBS) and multiple CR users as shown in Fig. 1. PUs have the full privilege of accessing their designated frequency bands whereas SUs can opportunistically utilize channels which are not occupied by the PUs. Resource allocation for the secondary network is performed by the CBS. Perfect channel sensing capability of the CRN is assumed so that SUs can accurately and promptly detect the appearance of a PU on any channel. That is, the probability of channel sensing errors of the system is considered as negligible [13] which implies that the secondary network capacity to be calculated later in this paper is the ideal capacity. The licensed spectrum band is divided into $M \in \mathbb{Z}^+$ frequency channels and each channel is allocated to a specific PU. Here, $\mathbb{Z}^+$ is the set of positive integers. Even though a PU service occupies only one channel for each service, an SU can assemble several channels in order to complete its service at a higher data rate.

In our analysis, we consider two types of SU flows, i.e., elastic traffic and real-time traffic. Elastic flows such as file transfer and video streaming adjust their transmission rate according to network conditions [14], while real-time applications, such as

video conference and voice over IP, have the same call duration even though the data rate may vary from time to time [15]. For presentation convenience, two abbreviations, *ESU* and *RSU*, are used to represent elastic SU and real-time SU services respectively whereas the term *SU* commonly represents both SU services. In our study, the number of channels allocated to a real-time service is considered to be fixed.

In this paper, we model traffic at the *session level* since modeling at this level captures the dynamics related to the arrival and departure of flows such as flow duration or number of active flows. The number of channels required by real-time traffic, or elastic traffic, is an abstraction at the session level of the QoS requirements at the packet level. For more information on packet level modeling of CA, please refer to [19].

Moreover, the following assumptions are made in order to develop our analytical model.

- The arrivals of both PU and SU services are Poisson processes with arrival rates λ_P, λ_{SE} and λ_{SR} for PU, ESU and RSU services respectively.
- The service times for PU and RSU services, and the volume of information to transfer (measured in bits) for an ESU service are exponentially distributed, with corresponding service rates per channel μ_P, μ_{SR} and μ_{SE} respectively.
- All channels are homogeneous. Thus, the service rate of k assembled channels in a secondary network equals to $k\mu_{SE}$ for elastic traffic.
- The sensing and spectrum adaptation latency is negligible in comparison with the duration between two consecutive service events.

IV. PROPOSED QUEUING SCHEMES AND DYNAMIC CA STRATEGY

To deal with heterogeneous traffic, we propose to deploy separate queues for different traffic types. The queue named as *EQ* is used for elastic traffic while the queue named as *RQ* is dedicated for real-time services. For a queuing discipline, priority levels can be assigned according to the delay tolerance of different services [16]. Since higher priority for channel access is assigned to RSU services when channels are not occupied by PUs, *RQ* is regarded as the *high priority queue* and *EQ* is considered as the *low priority queue* in our model. Moreover, the priority level can also be re-configured according to service requirements. This feature will be discussed in Subsec. IV.C.

A. Queuing Schemes

Two queuing schemes are proposed based on service arrivals to the queues. The main reason that we propose two distinct schemes is due to the consideration of SU traffic characteristics. If the preempted ESU services can be queued, the forced termination probability of ESU services can be reduced, however, at a cost of longer queuing delay for *ESU services*. Therefore, for delay-sensitive applications like real-time traffic, it is better *not* to feed the preempted ESU services back to the EQ. Fig. 2 illustrates the proposed queuing schemes, referred to as Scheme I and Scheme II respectively where λ_{SE}^r denotes the mean arrival rate of the fed back ESU services to the EQ.
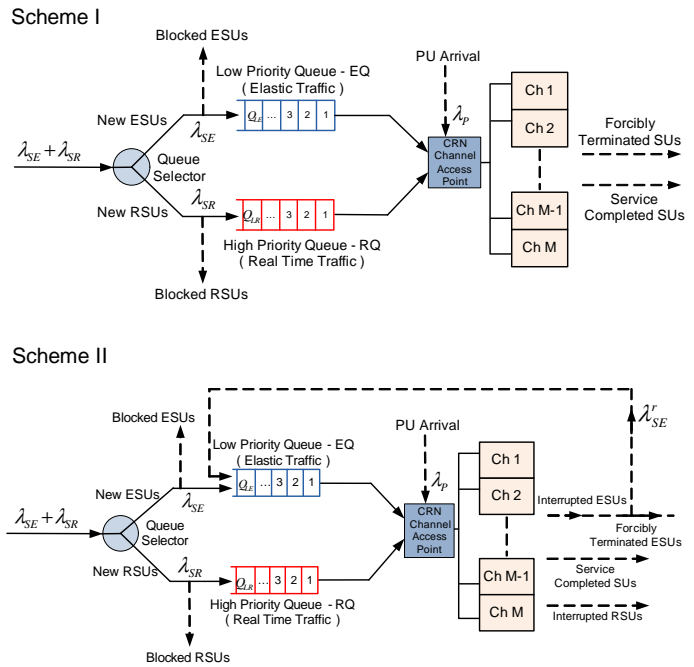

Fig. 2: Illustration of the proposed queuing schemes where the forcibly terminated ESUs may re-join the EQ in Scheme II but not in Scheme I.

1) Scheme I: Queuing scheme without feedback loop: In Scheme I, only the new services are allowed to enter the queuing system. Correspondingly, new ESU services will occupy waiting positions in the EQ while new RSU services will enter the RQ. The main feature of this scheme is that the preempted SU services will not be buffered in the queue even though they are forced to terminate.

2) Scheme II: Queuing scheme with feedback loop: In Scheme II, the interrupted ESU services due to PU arrivals, which cannot find any idle channel in the system by means of channel adaptation, are also inserted at the end of the EQ together with the new ESU arrivals. However, if there are not enough waiting rooms in the EQ, the preempted ESUs are forced to terminate. Other than this feed back feature, the rest of the functions of Scheme II are the same as in Scheme I. It is worth mentioning that, Scheme I is not a particular case of Scheme II as λ_{SE}^r is not a parameter of the system which can be set endogenously to zero or to any other value. Therefore, two separate CTMC analyses are performed since the behavior of both schemes is qualitatively different.

It is worth mentioning that if the transition includes the feedback feature, the resulting arrival process is not Poisson anymore. In other words, the fed back flow prevents the aggregated arrival stream to the queue (fresh arrivals plus feedback sessions) from being of Poisson process (i.e., exponentially distributed inter-arrival times). However, our model does not rely on such an assumption. We need only to assume that the exogenous arrivals are Poisson processes. The former assumption (i.e., the aggregated arrival process is Poisson) might be necessary if, for instance, we model the buffer and the channel by separate and independent Markov chains (MCs). This may be useful to reduce the dimension of the MCs we are analyzing, but dimensionality is not an issue in

our case. In this paper, we assume that the service times at each queue are exponentially distributed and the *original external* arrival process follows Poisson process. The analytical results to be presented later are valid under this assumption.

In our model, each queue is considered to hold homogeneous traffic, i.e., the EQ processes only elastic traffic and the RQ processes only real-time applications, although the SU arrival process is heterogeneous. For each queue, the FCFS order is applied, i.e., the first user in the queue is the first service that is to be processed. Moreover, when channel access opportunities appear, those opportunities are divided between the EQ and the RQ according to the assigned priority levels [15], [16]. For example, assuming that 6 channels become idle, 4 channels are allocated to the real-time services in the RQ and 2 channels are allocated to the elastic traffic in the EQ, if the ratio of priority levels between the RQ and the EQ is given as 2:1. More details about the queue scheduling algorithm will be presented in Subsec. IV.C. Note that the queues proposed in our model are different from the *priority queuing discipline*[1].

B. Dynamic$(a, W, V, Q_{LE}, Q_{LR})$ Channel Access Strategy

The channel access strategy which is utilized for the CRN studied in this paper is enhanced from the dynamic CA strategy proposed in [7], i.e., $D(a, W, V)$. Both spectrum handover and adjustable channel assembling are used in $D(a, W, V)$. Two strategies, dynamic and static are studied in-depth in [7], and the results show that the dynamic CA strategy achieves much better performance than the static one. Therefore, in this paper, we improve the $D(a, W, V)$ strategy by integrating the priority-based queuing system presented above. The DSA strategy proposed in this paper is denoted herein as $Dynamic(a, W, V, Q_{LE}, Q_{LR})$ where W and V denote the lower and upper bounds of the number of assembled channels for an ESU service and $a \in \mathbb{Z}^+$ represents the number of channels assembled by an RSU service. The maximum queue sizes of EQ and RQ are denoted as Q_{LE} and Q_{LR} respectively.

Consider a CR network with M licensed channels, where SUs can access the channels when PUs are inactive. Depending on the channel status and the user activities, an ESU service can dynamically adjust the number of assembled channels. An ESU service can assemble several channels to increase its service rate [8]. However for an RSU service, the number of aggregated channels is always fixed as a. The states when modeling the system are represented by the numbers of channels occupied by various users and the numbers of users in the queues. State transitions are triggered by one of the following four events.

1) PU Arrivals: When a PU arrives to a channel which is currently in the idle state, it starts its transmission and the secondary network does not suffer any channel state variation. However, when a PU arrives to a busy channel which is currently occupied by an ESU service with k_1 assembled

channels, where $W \leq k_1 \leq V$, the interrupted ESU service has to vacate the channel immediately. The ESU service with the maximum number of assembled channels, k_2, where $k_2 > k_1$, will donate a channel for the interrupted ESU service. If there is no such ESU service which has $k_2 > k_1$ assembled channels, then the interrupted ESU service will reduce the number of its assembled channels by one, as long as the remaining number of the aggregated channels is still greater than or equals to W. On the other hand, if an RSU service is interrupted when there is no idle channel in the system, the ESU service with the maximum number of channels will donate a channel to the RSU service as long as it has at least W assembled channels after donation. Otherwise, the interrupted RSU service will be forced to terminate. An important feature that should be mentioned here is that RSU services do not donate channels to other SU services no matter which arrival or departure event occurs. Moreover, when the interrupted ESU service has exactly W channels and there is no other ESU service which has more than W assembled channels, it is forced to terminate according to Scheme I. If Scheme II is applied to the queuing process, the interrupted ESU service is temporarily suspended and it is buffered inside the EQ as long as the queue is not full. In the worst case, i.e., if the EQ is full, the corresponding ESU service is forced to terminate. In Scheme II, if an RSU service is interrupted upon a PU arrival when all the channels in the system are occupied, it is forced to terminate unless there is at least one ESU service which has more than W assembled channels. As mentioned earlier, the interrupted RSU services will not be buffered into the RQ. The channels which become idle due to forced termination of SU services are allocated to the awaiting SU services in the queues by using the queue scheduling algorithm explained in Subsec. IV.C.

2) PU Departures: As a result of a PU departure, the corresponding channel becomes idle. When an idle channel appears, the priority of accessing that channel is given to the services waiting at the RQ. According to the FCFS discipline, if $a = 1$, the RSU service in the front end of the RQ will receive the chance for service commencement on the idle channel. Upon a PU departure, if $a > 1$, none of the real-time services in the RQ will be commenced unless there are $a - 1 > 0$ idle channels already in the system due to earlier departed PU/SU services. As mentioned before, a PU service always occupies a single channel in the system. Thus, only one channel becomes idle due to a PU departure. In the definition of real-time services in Sec. III, we mention that the number of channels assembled by an RSU service is fixed, as a. Therefore in order to commence one of the queued RSU services, exactly a channels are required. Consequently, upon a PU departure, one of the queued RSU can be commenced if there are already $a-1$ idle channels in the system due to earlier PU/SU departures. Otherwise none of the real-time services in the RQ will be commenced. ESU services in the EQ will receive the opportunity in a similar way. If none of the queued SUs utilize the idle channel, the ongoing ESU service with the minimum aggregated channels, i.e., k_{min}, where $k_{min} < V$, can assemble the newly idle channel.

---

[1]According to the priority queuing discipline, each service request is assigned a priority, and the service request with the highest priority will be served first, regardless of the order of arrival, whereas FCFS in our model applies to each traffic type.

Algorithm 1 : Pseudo code for the queue scheduling of *Dynamic* $(a, W, V, Q_{LE}, Q_{LR})$ strategy upon PU/SU departures and PU arrivals.

1:    **if** $n = 2k$ where, $k \in \mathbb{Z}^+$ % Lines 1 and 4 check whether n is even or odd

2:        $n_r = \lfloor \frac{\beta}{\beta+1} n \rfloor$ % If n is even, n_r and n_e are calculated as in lines 2 and 3 respectively.

3:        $n_e = n - n_r$

4:    **elseif** $n = (2k + 1)$ where, $k \in \mathbb{Z}^+,\ k > 0$

5:        $n_r = \lceil \frac{\beta}{\beta+1} n \rceil$ % If n is odd, n_r and n_e are calculated as in lines 5 and 6 respectively.

6:        $n_e = n - n_r$

7:    **end**

8:    **if** $j_{eq} \geq 1$

9:        $n_{eq,\ max} = V \cdot j_{eq}$ % To commence all queued ESUs with V channels, $V \cdot j_{eq}$ channels are required.

10:      $n_{eq,\ min} = W$

11:   **else**

12:      $n_{eq,\ max} = 0$ % If the EQ is empty, none of the channels need to be allocated.

13:      $n_{eq,\ min} = 0$

14:   **end**

15:   $n_{rq,\ max} = a \cdot j_{rq}$ % Even though $n_{rq,\ max}$ channels are required to commence all queued RSU services,

16:   **if** $n_{rq,\ max} \geq n_r$ only n_R channels will be assigned.

17:      $n_R = \lfloor \frac{n_r}{a} \rfloor \cdot a$ % n_R is the product of a and $\lfloor \frac{n_r}{a} \rfloor$ since a commenced RSU occupies a channels.

18:   **else**

19:      $n_R = n_{rq,\ max}$

20:   **end**

21:   $n_e = n - n_R$ % After assigning channels to the RQ, the rest are allocated to the EQ.

22:   **if** $n_{eq,\ max} > n_e$ % Even though $n_{eq,\ max}$ channels are required to commence all queued ESU services,

23:      **if** $n_e \geq n_{eq,\ min}$ only n_E channels will be assigned.

24:        $n_E = n_e$

25:      **else**   % If the number of available channels, n_e is lower than the minimum requirement,

26:        $n_E = 0$ none of the queued ESUs will be commenced.

27:      **end**

28:   **else**

29:      $n_E = n_{eq,\ max}$ % If condition in line 22 fails, all queued ESUs can be commenced.

30:   **end**

31:   $n_Q = n_R + n_E$

32:   $n_O = n - n_Q$ % Once channels are allocated for both queues, the remaining is allocated for ongoing ESUs.

---

3) SU Arrivals: Upon a new ESU service arrival, the service can be commenced if it can assemble at least W idle channels in the system. Similarly, a new RSU service can be commenced if it can assemble a idle channels in the system. If there are not enough idle channels for the newly arrived SU service, the ongoing ESU service which has the maximum number of channels will donate channels to the new user. After donation, the ongoing ESU service should be able to keep at least W channels, i.e., the lower bound of CA. If the one with the maximum number of assembled channels cannot provide W channels to the new ESU request or a channels to the new RSU request by itself, the next ESU with the second maximum number of aggregated channels will donate its channels, and so on. If the ongoing ESU services collectively cannot provide the required number of channels, then the new service is put in the corresponding queue (i.e., ESU services are directed to the EQ and RSUs are directed to the RQ). However, if the queue is fully occupied, then the new request is blocked. One of the important aspects herein is that ESU services are not put in the RQ even at a situation when the EQ is fully occupied and the RQ is empty. Similar rules apply to RSU services. Moreover, the ongoing ESU connections are not preempted by RSU arrivals.

4) SU Departures: Due to an SU departure, k, where $W \leq k \leq V$, channels become idle. The opportunity of accessing those channels will be shared between awaiting

RSU and ESU services in the queues by utilizing the queue scheduling algorithm explained in the next subsection. By varying the priority factor β, the number of allocated channels for each queue can be adjusted. If one of the queues is empty, then all the idle channels are allocated to the services awaiting in the other queue. If there are still idle channels even after allocating channels to the queued SU services, then the ongoing ESU service which has the minimum number of aggregated channels receives the chance to upgrade its service up to V channels. Furthermore, if the ESU service with the minimum number of channels has aggregated V channels after adjustment and there are still vacant channels, the other ESU services will occupy the remaining ones following the same principle. However, the ongoing real-time SU services will not assemble newly idle channels.

C. Queue Scheduling Algorithm

The allocation process of the n newly idle channels among the services in the EQ and the RQ is specified in Algorithm I. This algorithm applies to the events of PU arrival, PU departure and SU departure. Let $\beta: 1$ be the ratio between priority levels assigned to the RQ and the EQ respectively where, $\beta \in \mathbb{Q}^+$ and $\beta \geq 1$. Here, $\mathbb{Q}^+$ represents the set of positive rational numbers. According to the QoS requirements of elastic and real-time traffic, a proper value for β can be determined. The constraint $\beta \geq 1$ is made in order to assign

higher priority to the RSU services in the RQ. As we already mentioned, priority levels (i.e., β) can be set according to the delay tolerance of each service type. To obtain the value of β we can first specify the QoS objectives we intend to meet (e.g., an upper bound for the mean delay of RSUs). Once this is done, we could follow any (or a combination of both) of the following general paths: a) use our mathematical expressions to find (e.g., iteratively) the minimum value of β so that the QoS objective is met; b) implement an adaptive scheme to adjust β online by increasing/decreasing β values based on the measurement of the target performance parameter and its comparison with the objective [24], [25].

Although a certain number of channels is initially allocated to the RSU services in the RQ or the ESU services in the EQ, the maximum number of required channels to reach the upper bound of CA for all RSU or ESU services may be different. For example, even though n_r channels can be allocated to RSU services in the queue by lines 2 and 5 in Algorithm I, the required maximum number of channels for them is $n_{rq,\,max}$ as mentioned in line 15. The number of actually allocated channels for RSU services, i.e., n_R, is calculated by lines 17 and 19. The following notations indicate the parameters used in Algorithm I.

n	Number of newly idle channels,
n_r, n_e	Number of channels available for RSU services in the RQ and ESU services in the EQ respectively,
$n_{rq,\,max}$	Maximum number of channels required by all the RSU services in the RQ,
$n_{eq,\,max}$	Maximum number of channels required by all the ESU services in the EQ,
$n_{eq,\,min}$	Minimum number of channels required by all the ESU services in the EQ,
n_E, n_R	Number of channels actually used by ESU and RSU services in the EQ and the RQ respectively,
n_O	Number of channels to be allocated to ongoing ESUs,
j_{eq}	Current queue length of the EQ, where $j_{eq} \leq Q_{LE}$,
j_{rq}	Current queue length of the RQ, where $j_{rq} \leq Q_{LR}$.

V. CTMC ANALYSIS

In this section, CTMCs are developed to model the proposed dynamic channel access strategy with queuing schemes. Let $\boldsymbol{x} = \{j_W, j_{W+1}, \cdots, j_k, \cdots, j_V, g_a, j_{pu}, j_{eq}, j_{rq}\}$ be the general state representation of the system and $\pi(\boldsymbol{x})$ be the steady state probability of being in state $\boldsymbol{x}$. Here, j_k denotes the number of ESU services with k aggregated channels, and $W \leq k \leq V$, g_a is the number of RSUs with a aggregated channels and j_{pu} denotes the number of PU services in the system.

The set of feasible states of the system is denoted as $\mathcal{S} = \{\boldsymbol{x} \mid j_W, j_{W+1}, \cdots j_V, g_a, j_{pu}, j_{eq}, j_{rq} \geq 0; b(\boldsymbol{x}) \leq M, j_{eq} \leq Q_{LE}, j_{rq} \leq Q_{LR}; \sum_{i=1}^{V-W} ij_{W+i} < W$, if $j_{eq} > 0; \sum_{i=1}^{V-W} ij_{W+i} < a$, if $j_{rq} > 0\}$, where $b(\boldsymbol{x})$ is the total number of utilized channels at state $\boldsymbol{x}$.

$$b(\boldsymbol{x}) = j_{pu} + ag_a + \sum_{k=W}^{V} kj_k. \tag{1}$$

Condition $\sum_{i=1}^{V-W} ij_{W+i} < W$, if $j_{eq} > 0$ means that if EQ is not empty, the summation of maximum number of channels which can be donated by the already commenced ESU services is always lower than W. Similarly, $\sum_{i=1}^{V-W} ij_{W+i} < a$, if $j_{rq} > 0$ declares that the summation of maximum number of channels which can be donated by the already commenced ESU services is always lower than a if RQ is not empty. The stationary probabilities, $\pi(\boldsymbol{x})$ can be calculated from the global balance equations and the normalization equation, which are given as

$$\boldsymbol{\pi Q} = 0, \quad \sum_{\boldsymbol{x} \in \mathcal{S}} \pi(\boldsymbol{x}) = 1, \tag{2}$$

where $\boldsymbol{\pi}$ is the steady state probability vector and $\boldsymbol{Q}$ denotes the transition rate matrix.

The detailed state transition tables can be found in Tables I-IV. Those Tables summarize the state transitions associated with different events with different conditions. The parameters, n, n_R and n_E in Tables I-IV, have the same meanings as described in *Algorithm I*. Moreover DP and AR indicate a departure event and an arrival event respectively. ESU_k denotes an ESU service with k aggregated channels. Each row entry in those tables is corresponding to a particular transition from a state, i, to another state, j, where $i, j \in \mathcal{S}$, due to PU/SU arrivals or departures.

The total transition rate from state i to state j, namely q_{ij}, is the summation of transition rates from state i to state j considering all possible user activities. Once we determine the q_{ij} for all $i, j (i \neq j) \in \mathcal{S}$, the diagonal elements in $\boldsymbol{Q}$, i.e., q_{ii}, $i \in \mathcal{S}$ are found as $q_{ii} = -\sum_{j \in \mathcal{S}, j \neq i} q_{ij}$. When the steady state probabilities are determined from (2), the performance of the CRN can be evaluated with respect to different parameters. The derivations of mathematical expressions for those parameters are presented in the following subsections.

A. Analysis of Scheme I

In this subsection we derive equations to evaluate the system performance of Scheme I.

1) Secondary network capacity: In this study, *capacity* indicates the rate of service completions, i.e., the average number of service completions per time unit in a secondary network. Let ρ_1 and ρ_2 be the capacity of ESU and RSU services respectively. Then, we obtain that

$$\rho_1 = \sum_{\boldsymbol{x} \in \mathcal{S}} \sum_{k=W}^{V} kj_k \mu_{SE} \pi(\boldsymbol{x}), \tag{3}$$

$$\rho_2 = \sum_{\boldsymbol{x} \in \mathcal{S}} ag_a \mu_{SR} \pi(\boldsymbol{x}). \tag{4}$$

2) Spectrum utilization: In our analysis, we calculate the spectrum utilization [17] of the system as the average number of utilized channels over the total number of channels. In state $\boldsymbol{x} \in \mathcal{S}$, a total number of $b(\boldsymbol{x})$ channels out of M are utilized. Consequently, the spectrum utilization of the CRN, U_{CRN}, can be expressed as

$$U_{CRN} = \sum_{\boldsymbol{x} \in \mathcal{S}} \pi(\boldsymbol{x}) \frac{b(\boldsymbol{x})}{M}. \tag{5}$$

TABLE I: Transitions from a generic state $\boldsymbol{x} = \{j_W, j_{W+1}, \cdots, j_k, \cdots, j_V, g_a, j_{pu}, j_{eq}, j_{rq}\}$ of $Dynamic(a, W, V, Q_{LE}, Q_{LR})$ upon a PU arrival in Scheme I

Activity	Dest. State	Trans. rate	Conditions
PU AR. A vacant channel exists.	$(j_W, \cdots, j_k, \cdots, j_V, g_a, j_{pu} + 1, j_{eq}, j_{rq})$	λ_P	$M - b(\boldsymbol{x}) > 0$.
PU AR. An ESU_k is interrupted and reduces its channels.	$(j_W, \cdots, j_{k-1} + 1, j_k - 1, \cdots, j_V, g_a, j_{pu} + 1, j_{eq}, j_{rq})$	$\frac{kj_k}{M - j_{pu}}\lambda_P$	$V > W; b(\boldsymbol{x}) = M; j_k > 0; k > W$.
PU AR. ESU_W is forced terminated. No spectrum adaptation.	$(j_W - 1, \cdots, j_k, \cdots, j_V, g_a, j_{pu} + 1, j_{eq}, j_{rq})$	$\frac{Wj_W}{M - j_{pu}}\lambda_P$	$j_W \geq 1; b(\boldsymbol{x}) = M; W = V$ or $W = 1, j_k = 0, W + 1 \leq k \leq V$. Or $j_W = 1; b(\boldsymbol{x}) = M; j_k = 0, W + 1 \leq k \leq V$.
PU AR. ESU_W is forced terminated. RSU in RQ uses all the vacant channels.	$(j_W - 1, \cdots, j_k, \cdots, j_V, g_a + 1, j_{pu} + 1, j_{eq}, j_{rq} - 1)$	$\frac{Wj_W}{M - j_{pu}}\lambda_P$	$j_W \geq 1; b(\boldsymbol{x}) = M; j_k = 0, \forall k > W; j_{rq} > 0; W > 1; n_R = a = W - 1$.
$\cdots$	$\cdots$	$\cdots$	$\cdots$
PU AR. ESU_W is forced terminated and another ongoing ESU_W uses all the vacant channels.	$(j_W - 2, \cdots, j_l + 1, \cdots, j_V, g_a, j_{pu} + 1, j_{eq}, j_{rq})$	$\frac{Wj_W}{M - j_{pu}}\lambda_P$	$j_W > 1; b(\boldsymbol{x}) = M; j_{rq} = 0$ or $n_R = 0; j_k = 0, \forall k > W; W > 1; l = 2W - 1 \leq V$.
$\cdots$	$\cdots$	$\cdots$	$\cdots$
PU AR. ESU_W is forced terminated and other all ESUs uses all the vacant channels and achieve V.	$(0, 0, \cdots, j_V + q, g_a, j_{pu} + 1, j_{eq}, j_{rq})$	$\frac{Wj_W}{M - j_{pu}}\lambda_P$	$j_W > 1; b(\boldsymbol{x}) = M; j_{rq} = 0$ or $n_R = 0; j_k = 0, \forall k > W; V > W > 1; q = j_W - 1; W - 1 \geq (V - W)(j_W - 1)$.
PU AR. RSU is interrupted and ESU_k with max. channels, k donates a channel.	$(j_W, \cdots, j_{k-1} + 1, j_k - 1, \cdots, j_V, g_a, j_{pu} + 1, j_{eq}, j_{rq})$	$\frac{ag_a}{M - j_{pu}}\lambda_P$	$V > W; b(\boldsymbol{x}) = M; g_a > 0; k > W; k = max\{r | j_r > 0, W + 1 \leq r \leq V\}$.
PU AR. RSU is forced terminated. No spectrum adaptation.	$(j_W, \cdots, j_k, \cdots, j_V, g_a - 1, j_{pu} + 1, j_{eq}, j_{rq})$	$\frac{ag_a}{M - j_{pu}}\lambda_P$	$g_a > 0; b(\boldsymbol{x}) = M; a = 1, j_k = 0, W + 1 \leq k \leq V$. Or $g_a > 0; b(\boldsymbol{x}) = M; W = V$.
PU AR. RSU is forced terminated. An ESU in EQ uses all the vacant channels[†].	$(j_W, \cdots, j_k, \cdots, j_h + 1, \cdots, j_V, g_a - 1, j_{pu} + 1, j_{eq} - 1, j_{rq})$	$\frac{ag_a}{M - j_{pu}}\lambda_P$	$g_a > 0; b(\boldsymbol{x}) = M; j_k = 0, \forall k > W; a \geq W + 1; n_E = h = a - 1; j_{eq} = 1, W \leq a - 1 \leq V$ or $j_{eq} > 1, W \leq a - 1 < 2W$.
$\cdots$	$\cdots$	$\cdots$	$\cdots$
PU AR. RSU is forced terminated. An ESU_W uses all the vacant channels.	$(j_W - 1, \cdots, j_k + 1, \cdots, j_V, g_a - 1, j_{pu} + 1, j_{eq}, j_{rq})$	$\frac{ag_a}{M - j_{pu}}\lambda_P$	$g_a > 0; b(\boldsymbol{x}) = M; a > 1, j_k = 0, W + 1 \leq k \leq V; j_W > 0, k = a + W - 1 \leq V; V > W$.
$\cdots$	$\cdots$	$\cdots$	$\cdots$
PU AR. RSU is interrupted and ESU_W is forced terminated. All ESUs uses the vacant channels and achieve V.	$(0, 0, \cdots, j_V + q, g_a - 1, j_{pu} + 1, j_{eq}, j_{rq})$	$\frac{ag_a}{M - j_{pu}}\lambda_P$	$g_a > 0; b(\boldsymbol{x}) = M; a > 1, j_k = 0, W + 1 \leq k \leq V; q = j_W; a - 1 \geq (V - W)j_W; V > W$.

[†]Upon a PU arrival, an ongoing RSU service is forced to terminate when all channels are occupied in the system and there is no more than one ESU service which has more than W channels. Those two conditions are represented by $b(\boldsymbol{x}) = M$ and $j_k = 0, \forall k > W$. After forced termination, only $a - 1$ channels are released in the system. Thus, no queued RSU services could acquire the idle channels since exactly a number of channels are required for a new RSU commencement. Then, $n_E = a - 1$. If there is only one queued ESU in the EQ, all $a - 1$ channels can be allocated for that ESU if $W \leq a - 1 \leq V$. This is because that, if $W > a - 1$, service cannot be commenced and if $a - 1 > V$, $V - (a - 1)$ number of channels are still idle. On the other hand, if there are more than one queued ESU services, only one can be commenced if $(a - 1) < 2W$. The reason is that at least $2 \times W$ number of channels are needed to commence two ESU services.

3) Blocking probability: An incoming ESU service will be blocked if the following three conditions are met simultaneously: there are not enough idle channels to commence the service; there is no possibility of channel sharing with ongoing ESU services; and the EQ is fully occupied. Similar conditions apply to the calculation of the blocking probability for incoming RSU services except the fact that they need exactly a channels to commence a service. Let P_{b1} and P_{b2} be the blocking probability of ESU and RSU services respectively, then we have that,

$$P_{b1} = \sum_{\substack{\boldsymbol{x} \in \mathcal{S}, \\ M - b(\boldsymbol{x}) + \sum_{k=W+1}^{V}(k-W)j_k < W, \\ j_{eq} = Q_{LE}}} \pi(\boldsymbol{x}), \quad (6)$$

$$P_{b2} = \sum_{\substack{\boldsymbol{x} \in \mathcal{S}, \\ M - b(\boldsymbol{x}) + \sum_{k=W+1}^{V}(k-W)j_k < a, \\ j_{rq} = Q_{LR}}} \pi(\boldsymbol{x}). \quad (7)$$

4) Mean PU interruption rate of ESUs: The mean PU interruption rate of ESUs, R_{int}, can be expressed as the preemption rate of ESUs when there are no ways of accessing an idle channel, upon a PU arrival. That is,

$$R_{int} = \sum_{\substack{\boldsymbol{x} \in \mathcal{S}, \\ M = b(\boldsymbol{x}), \ j_{pu} < M, \ j_W > 0, \ j_n = 0; \ \forall n > W}} \frac{\lambda_P W j_W}{M - j_{pu}} \pi(\boldsymbol{x}). \quad (8)$$

5) Forced termination probability of ESU service: The forced termination probability is the probability that an active

TABLE II: Transitions from a generic state $\boldsymbol{x} = \{j_W, j_{W+1}, \cdots, j_k, \cdots, j_V, g_a, j_{pu}, j_{eq}, j_{rq}\}$ of $Dynamic(a, W, V, Q_{LE}, Q_{LR})$ upon a PU arrival in Scheme II

Activity	Dest. State	Trans. rate	Conditions
PU AR. A vacant channel exists.	$(j_W, \cdots, j_k, \cdots, j_V, g_a,$ $j_{pu} + 1, j_{eq}, j_{rq})$	λ_P	$M - b(\boldsymbol{x}) > 0.$
PU AR. An ESU service with k channels reduces its channels.	$(j_W, \cdots, j_{k-1} + 1, j_k - 1,$ $\cdots, j_V, g_a, j_{pu} + 1, j_{eq}, j_{rq})$	$\frac{k j_k}{M - j_{pu}} \lambda_P$	$V > W; M = b(\boldsymbol{x}); j_k > 0;$ $k > W.$
PU AR. The interrupted ESU_W is put into the queue EQ and no spectrum adaptation	$(j_W - 1, \cdots, j_k, \cdots, j_V,$ $g_a, j_{pu} + 1, j_{eq} + 1, j_{rq})$	$\frac{W j_W}{M - j_{pu}} \lambda_P$	$j_W > 0; b(\boldsymbol{x}) = M; j_{eq} < Q_{LE};$ $W = V$ or $W = 1, j_k = 0,$ $W + 1 \leq k \leq V.$ Or $j_W = 1;$ $j_k = 0, W + 1 \leq k \leq V,$ $b(\boldsymbol{x}) = M.$
PU AR. ESU_W is put into the queue, EQ. Another ESU_W uses the idle channels[‡].	$(j_W - 2, \cdots, j_l + 1, \cdots, j_V, g_a,$ $j_V, g_a, j_{pu} + 1, j_{eq} + 1, j_{rq})$	$\frac{W j_W}{M - j_{pu}} \lambda_P$	$j_W > 1; b(\boldsymbol{x}) = M; j_k = 0,$ $\forall k > W; W > 1;$ $l = 2W - 1 \leq V; j_{eq} < Q_{LE}.$
$\cdots$	$\cdots$	$\cdots$	$\cdots$
PU AR. ESU_W is put into the queue, EQ and other all ESUs uses the idle channels and achieve V	$(0, 0, \cdots, j_V + q, g_a,$ $j_{pu} + 1, j_{eq} + 1, j_{rq})$	$\frac{W j_W}{M - j_{pu}} \lambda_P$	$j_W > 1; b(\boldsymbol{x}) = M; j_k = 0, \forall k > W;$ $V > W > 1; q = j_W - 1; j_{eq} < Q_{LE};$ $W - 1 \geq (V - W)(j_W - 1).$
PU AR. The interrupted ESU_W is forced terminated and no spectrum adaptation	$(j_W - 1, \cdots, j_k, \cdots, j_V,$ $g_a, j_{pu} + 1, j_{eq} + 1, j_{rq})$	$\frac{W j_W}{M - j_{pu}} \lambda_P$	$j_W > 0; b(\boldsymbol{x}) = M; j_{eq} = Q_{LE};$ $W = V$ or $W = 1, j_k = 0,$ $W + 1 \leq k \leq V.$ Or $j_W = 1;$ $b(\boldsymbol{x}) = M, j_k = 0, W + 1 \leq k \leq V.$
PU AR. ESU_W is forced terminated, another ESU_W uses the idle channels.	$(j_W - 2, \cdots, j_l + 1, \cdots,$ $j_V, g_a, j_{pu} + 1, j_{eq} + 1, j_{rq})$	$\frac{W j_W}{M - j_{pu}} \lambda_P$	$j_W > 1; b(\boldsymbol{x}) = M; j_k = 0, \forall k > W;$ $W > 1; l = 2W - 1 \leq V; j_{eq} = Q_{LE}.$
$\cdots$	$\cdots$	$\cdots$	$\cdots$
PU AR. ESU_W is forced terminated and other all ESUs uses the idle channel and achieve V.	$(0, 0, \cdots, j_V + q, g_a,$ $j_{pu} + 1, j_{eq} + 1, j_{rq})$	$\frac{W j_W}{M - j_{pu}} \lambda_P$	$j_W > 1; b(\boldsymbol{x}) = M; j_k = 0, \forall k > W;$ $V > W > 1; q = j_W - 1; j_{eq} = Q_{LE};$ $W - 1 \geq (V - W)(j_W - 1).$
PU AR. RSU is interrupted and ESU_k with maximum channels, k donates a channel.	$(j_W, \cdots, j_{k-1} + 1, j_k - 1,$ $\cdots, j_V, g_a, j_{pu} + 1, j_{eq}, j_{rq})$	$\frac{a g_a}{M - j_{pu}} \lambda_P$	$V > W; M = b(\boldsymbol{x}); g_a > 0; k > W;$ $k = max\{r | j_r > 0, W + 1 \leq r \leq V\}.$
PU AR. RSU is forced terminated, No spectrum adaptation.	$(j_W, \cdots, j_k, \cdots, j_V,$ $g_a - 1, j_{pu} + 1, j_{eq}, j_{rq})$	$\frac{a g_a}{M - j_{pu}} \lambda_P$	$g_a > 0; b(\boldsymbol{x}) = M; a = 1, j_k = 0,$ $W + 1 \leq k \leq V.$ Or $g_a > 0; b(\boldsymbol{x}) = M; W = V.$
PU AR. RSU is forced terminated. An ESU in EQ uses all the vacant channels.	$(j_W, \cdots, j_k, \cdots, j_h + 1,$ $\cdots, j_V, g_a - 1, j_{pu} + 1,$ $j_{eq} - 1, j_{rq})$	$\frac{a g_a}{M - j_{pu}} \lambda_P$	$g_a > 0; b(\boldsymbol{x}) = M; j_k = 0, \forall k > W;$ $a \geq W + 1; n_E = h = a - 1;$ $j_{eq} = 1, W \leq a - 1 \geq V$ or $j_{eq} > 1, W \leq a - 1 < 2W.$
$\cdots$	$\cdots$	$\cdots$	$\cdots$
PU AR. RSU is forced terminated. An ESU_W uses all the vacant channels.	$(j_W - 1, \cdots, j_k + 1, \cdots,$ $j_V, g_a - 1, j_{pu} + 1, j_{eq}, j_{rq})$	$\frac{a g_a}{M - j_{pu}} \lambda_P$	$g_a > 0; b(\boldsymbol{x}) = M; a > 1, j_k = 0,$ $W + 1 \leq k \leq V; j_W > 0,$ $k = a + W - 1 \leq V; V > W.$
$\cdots$	$\cdots$	$\cdots$	$\cdots$
PU AR. RSU is interrupted and ESU_W is forced terminated. All ESUs uses the vacant channels and achieve V.	$(0, 0, \cdots, j_V + q, g_a - 1,$ $j_{pu} + 1, j_{eq}, j_{rq})$	$\frac{a g_a}{M - j_{pu}} \lambda_P$	$g_a > 0; b(\boldsymbol{x}) = M; a > 1, j_k = 0,$ $W + 1 \leq k \leq V; q = j_W > 0,$ $a - 1 \geq (V - W) j_W; V > W.$

[‡]Upon a PU arrival, an ongoing ESU service which has W channels (ESU_W) is put into the EQ when all channels are occupied in the system and there are not more than one ESU services which have more than W channels. Furthermore the queue EQ should not be fully occupied. Those three conditions are represented by $b(\boldsymbol{x}) = M$, $j_k = 0, \forall k > W$ and $j_{eq} < Q_{LE}$. Then, one of the remaining ESU_W services can aggregate up to $2W - 1$ by adding newly released $W - 1$ channels.

session is forced to terminate before the communication is regularly finished [18]. In Scheme I, if a PU preempts an ESU service which has exactly W assembled channels when there is no idle channel or channel sharing opportunity with another ongoing ESU service, then the ESU service is forced to terminate. Therefore, the interrupted ESU service has to terminate its communication before its service is finished. Correspondingly the forced termination probability of ESU services, P_{f1}, can be expressed as the mean forced termination

rate of ESU services, R_{f1}, divided by the mean admitted ESU rate, $\lambda_{SE}^* = \lambda_{SE}(1 - P_{b1})$. Furthermore, $R_{f1} = R_{int}$ since there is no feedback from the interrupted ESU services into the queue, EQ. Therefore, the forced termination probability of ESU services becomes,

$$P_{f1} = \frac{R_{f1}}{\lambda_{SE}^*} = \sum_{\substack{\boldsymbol{x} \in \mathcal{S}, \\ M = b(\boldsymbol{x}), \, j_{pu} < M, \\ j_W > 0, \, j_n = 0; \, \forall n > W}} \frac{\lambda_P W j_W}{(M - j_{pu}) \lambda_{SE}^*} \pi(\boldsymbol{x}). \quad (9)$$

TABLE III: Transitions from a generic state $\boldsymbol{x} = \{j_W, j_{W+1}, \cdots, j_k, \cdots, j_V, g_a, j_{pu}, j_{eq}, j_{rq}\}$ of $Dynamic(a, W, V, Q_{LE}, Q_{LR})$ upon a PU departure with transition rate $j_{pu}\mu_P$ and RSU departure with transition rate $ag_a\mu_{SR}$

Activity	Dest. State	Conditions
PU DP. An RSU in RQ uses the vacant channel§.	$(j_W, \cdots, j_V, g_a+1, j_{pu}-1, j_{eq}, j_{rq}-1)$	$j_{rq} > 0;\ j_{pu} > 0;\ a = 1$ or $a > 1, M - b(\boldsymbol{x}) = a - 1.$
PU DP. An ESU in EQ uses the vacant channel.	$(j_W+1, \cdots, j_V, g_a, j_{pu}-1, j_{eq}-1, j_{rq})$	$j_{rq} = 0$ or $a > 1, M - b(\boldsymbol{x}) < a - 1;$ $j_{eq} > 0; j_{pu} > 0;\ W = 1$ or $W > 1,$ $M - b(\boldsymbol{x}) = W - 1.$
PU DP. An ongoing ESU_k uses the vacant channel.	$(j_W, \cdots, j_k-1, j_{k+1}+1, \cdots, j_V, g_a,$ $j_{pu}-1, j_{eq}, j_{rq})$	$j_{rq} = 0$ or $a > 1, M - b(\boldsymbol{x}) < a - 1;$ $j_{eq} > 0;$ or $W > 1, M - b(\boldsymbol{x}) < W - 1,$ $j_{pu} > 0;\ j_k > 0,\ V > 1;$ $k = min\{r | j_r > 0, W \le r < V\}.$
PU DP. No SUs uses the vacant channel.	$(j_W, \cdots, j_V, g_a, j_{pu}-1, j_{eq}, j_{rq})$	$j_{rq} = 0$ or $a > 1, M - b(\boldsymbol{x}) < a - 1;$ $j_{eq} > 0;$ or $W > 1, M - b(\boldsymbol{x}) < W - 1,$ $j_k = 0, W \le k < V$ or $W = V; j_{pu} > 0.$
RSU DP. An RSU in RQ uses all the vacant channels.	$(j_W, \cdots, j_V, g_a, j_{pu}, j_{eq}, j_{rq}-1)$	$j_{rq} > 0;\ g_a > 0;\ n_R = n = a.$
RSU DP. An ESU in EQ uses all the vacant channels.	$(j_W, \cdots, j_a+1, \cdots, j_V, g_a-1, j_{pu},$ $j_{eq}-1, j_{rq})$	$j_{rq} = 0$ or $n_R = 0;\ g_a > 0;$ $n_E = n = a; j_{eq} = 1,\ W \le a \le V$ or $j_{eq} > 1,\ W \le a < 2W.$
RSU DP. Two ESUs in EQ uses all the vacant channels.	$(j_W, \cdots, j_h+1, \cdots, j_l+1, \cdots, j_V,$ $g_a-1, j_{pu}, j_{eq}-2, j_{rq})$	$j_{rq} = 0$ or $n_R = 0;\ g_a > 0;$ $n_E = n = a; j_{eq} = 2,\ 2W \le a \le 2V,$ $a = h + l,\ W \le h, l \le V$ or $j_{eq} > 2,$ $2W \le a < 3W,\ a = h + l,\ W \le h, l \le V.$
$\cdots$	$\cdots$	$\cdots$
RSU DP. An ongoing ESU with the minimum no. of channels, h uses all the vacant channels.	$(j_W, \cdots, j_h-1, \cdots, j_l+1, \cdots, j_V,$ $g_a-1, j_{pu}, j_{eq}, j_{rq})$	$j_{rq} = 0$ or $n_R = 0;\ j_{eq} = 0$ or $n_E = 0;$ $g_a > 0;\ h = min\{r | j_r > 0, W \le r < V\};$ $l = a + h \le V;\ V > W.$
$\cdots$	$\cdots$	$\cdots$
RSU DP. All other ESUs use the vacant channels and reach the upper bound V.	$(0, \cdots, 0, \cdots, j_V+q, g_a-1, j_{pu}, j_{eq}, j_{rq})$	$j_{rq} = 0$ or $n_R = 0;\ j_{eq} = 0$ or $n_E = 0;$ $g_a > 0;\ q = \sum_{m=W}^{V-1} j_m;$ $a \ge \sum_{m=W}^{V-1} (V - m) j_m;\ V > W.$
RSU DP. Other SUs cannot use the vacant channels.	$(j_W, \cdots, j_V, g_a-1, j_{pu}, j_{eq}, j_{rq})$	$j_{rq} = 0$ or $n_R = 0;\ j_{eq} = 0$ or $n_E = 0;$ $g_a > 0; V > W,\ j_m = 0, \forall m < V$ or $W = V.$

§Once a PU service is completed, one channel becomes idle. This channel may be assigned to an RSU service in the RQ if one of the following conditions is satisfied, given that the RQ is not empty. That is, a) the required number of channels for a new RSU service commencement is $a = 1$; b) the required number of channels for a new RSU service commencement is greater than 1, but there are already $a - 1$ idle channels in the system due to previous PU/SU departures. Therefore, the system's current state $\boldsymbol{x} = \{j_W, j_{W+1}, \cdots, j_k, \cdots, j_V, g_a, j_{pu}, j_{eq}, j_{rq}\}$ transits to a new state $\boldsymbol{x} = \{j_W, j_{W+1}, \cdots, j_k, \cdots, j_V, g_a+1, j_{pu}-1, j_{eq}, j_{rq}-1\}$ with rate $j_{pu}\mu_P$.

6) Forced termination probability of RSU: A forced termination of an RSU service occurs when a PU service preempts an RSU service and the preempted RSU service cannot find an idle channel in the system, nor is it possible to share a channel with other ongoing ESU services. The forced termination probability of RSUs, P_{f2}, can be expressed as the mean forced termination rate of RSUs, R_{f2}, divided by the mean admitted RSU rate, $\lambda_{SR}^* = \lambda_{SR}(1 - P_{b2})$. Correspondingly the forced termination probability of RSU services becomes, $P_{f2} = \frac{R_{f2}}{\lambda_{SR}^*}$. Therefore,

$$P_{f2} = \sum_{\substack{\boldsymbol{x} \in \mathcal{S}, \\ M = b(\boldsymbol{x}), g_a > 0, \\ j_n = 0;\ \forall n > W}} \frac{\lambda_P a g_a}{(M - j_{pu})\lambda_{SR}^*} \pi(\boldsymbol{x}). \quad (10)$$

7) Mean queue length: The queue length depends on the arrival rates of both PUs and SUs as well as their service completion rates. Recall that j_{eq} and j_{rq} are the corresponding queue lengths of the EQ and the RQ respectively at each state $\boldsymbol{x} \in \mathcal{S}$. Therefore, the mean queue length of ESU services, L_{EQ}, and the mean queue length of RSU services, L_{RQ}, are given by,

$$L_{EQ} = \sum_{\boldsymbol{x} \in \mathcal{S}} \pi(\boldsymbol{x}) j_{eq}, \quad (11)$$

$$L_{RQ} = \sum_{\boldsymbol{x} \in \mathcal{S}} \pi(\boldsymbol{x}) j_{rq}. \quad (12)$$

8) Average total delay of SU services: In our model, the average total delay of an SU service equals to the summation of average waiting time of SU services in the queue plus the average transmission time of an ongoing SU service. The total delay is averaged over the number of SU services. Prior to presenting the mathematical analysis, the following notations are introduced.

N_{SE}^s and N_{SR}^s	:	Average number of ongoing ESU and RSU services in the system,
T_{SE}^s and T_{SR}^s	:	Average transmission time of ongoing ESU and RSU services in the system,
T_{SE}^q and T_{SR}^q	:	Average waiting time of ESU and RSU services in the queues,
T_{SE}^t and T_{SR}^t	:	Average total delay of ESUs and RSUs.

TABLE IV: Transitions from a generic state $\boldsymbol{x} = \{j_W, j_{W+1}, \cdots, j_k, \cdots, j_V, g_a, j_{pu}, j_{eq}, j_{rq}\}$ of $Dynamic(a, W, V, Q_{LE}, Q_{LR})$ upon an ESU_k departure with transition rate $kj_k\mu_{SE}$, an ESU arrival with transition rate λ_{SE} and RSU arrival with transition rate λ_{SR}

Activity	Dest. State	Conditions
ESU_k DP. An RSU in RQ uses all the vacant channels.	$(j_w, \cdots, j_k - 1, \cdots, j_V, g_a + 1, j_{pu}, j_{eq}, j_{rq} - 1)$	$j_{rq} \geq 1$; $j_k > 0$; $k = a$; $n_R = k$.
ESU_k DP. Two RSUs in RQ uses all the vacant channels.	$(j_W, \cdots, j_k - 1, \cdots, j_V, g_a + 2, j_{pu}, j_{eq}, j_{rq} - 2)$	$j_{rq} \geq 2$; $j_k > 0$; $k = 2a$; $n_R = k$.
$\cdots$	$\cdots$	$\cdots$
ESU_k DP. An RSU in RQ and an ESU in EQ uses all the vacant channels.	$(j_W, \cdots, j_k - 1, \cdots, j_h + 1, \cdots j_V, g_a + 1, j_{pu}, j_{eq} - 1, j_{rq} - 1)$	$j_k > 0$; $k = a + h$; $j_{rq} \geq 1$; $n_R = a$; $n_E = h$; $j_{eq} = 1$, $W \leq h \leq V$ or $j_{eq} > 1$, $W \leq h < 2W$.
$\cdots$	$\cdots$	$\cdots$
ESU_k DP. An ESU in EQ uses all the vacant channels.	$(j_W, \cdots, j_k, \cdots, j_V, g_a, j_{pu} j_{eq} - 1, j_{rq})$	$j_{rq} = 0$ or $n_R = 0$; $j_k > 0$; $n_E = n = k$; $j_{eq} = 1$, $W \leq k \leq V$ or $j_{eq} > 1$, $W \leq k < 2W$.
ESU_k DP. Two ESUs in EQ uses all the vacant channels.	$(j_W, \cdots, j_k - 1, \cdots, j_h + 1, \cdots, j_l + 1, \cdots, j_V, g_a, j_{pu}, j_{eq} - 2, j_{rq})$	$j_{rq} = 0$ or $n_R = 0$; $j_k > 0$; $n_E = n = k$; $j_{eq} = 2$, $2W \leq k \leq 2V$, $k = h + l$, $W \leq h, l \leq V$ or $j_{eq} > 2$, $2W \leq k < 3W$, $k = h + l$, $W \leq h, l \leq V$.
$\cdots$	$\cdots$	$\cdots$
ESU_k DP. An ongoing ESU with the minimum no. of channels, h uses all the vacant channels.	$(j_W, \cdots, j_h - 1, \cdots, j_k - 1, \cdots, j_l + 1, \cdots, j_V, g_a, j_{pu}, j_{eq}, j_{rq})$	$j_{rq} = 0$ or $n_R = 0$; $j_{eq} = 0$ or $n_E = 0$; $j_k > 0$; $h = min\{r | j_r > 0, W \leq r < V\}$; $l = k + h \leq V$; $V > W$.
$\cdots$	$\cdots$	$\cdots$
ESU_k DP. All other ESUs use the vacant channels and reach the upper bound V.	$(0, \cdots, 0, \cdots, j_V + q, g_a, j_{pu}, j_{eq}, j_{rq})$	$j_{rq} = 0$ or $n_R = 0$; $j_{eq} = 0$ or $n_E = 0$; $j_k > 0$; $q = \sum_{m=W}^{V-1} j_m - 1$; $k \geq \sum_{m=W}^{V-1}(V - m)j_m - (V - k)$; $V > W$.
ESU_k DP. Other SUs cannot use the vacant channels.	$(j_W, \cdots, j_k - 1, \cdots, j_V, g_a, j_{pu}, j_{eq}, j_{rq})$	$j_{rq} = 0$ or $n_R = 0$; $j_{eq} = 0$ or $n_E = 0$; $j_k > 0$; $V > W$, $j_m = 0, \forall m < V$ or $W = V$.
ESU AR. Enough idle channels exist.	$(j_W, \cdots, j_k + 1, \cdots, j_V, g_a, j_{pu}, j_{eq}, j_{rq})$	$k = min\{M - b(\boldsymbol{x}), V\} \geq W$.
ESU AR. The ESU with the maximum no. of channels, m donate W channels.	$(j_W + 1, \cdots, j_n + 1, \cdots, j_m - 1, \cdots, j_V, g_a, j_{pu}, j_{eq}, j_{rq})$	$V > W$; $m = max\{r | j_r > 0, W + 1 \leq r \leq V\}$; $n = m - [W - (M - b(\boldsymbol{x}))]$; $W \leq n < m$.
ESU AR. Two ESUs with m and h channels collectively donate W channels.	$(j_W + 2, \cdots, j_n + 1, \cdots, j_h - 1, \cdots, j_m - 1, \cdots, J_V, g_a, j_{pu}, j_{eq}, j_{rq})$	$V > W$; $m = max\{r | j_r > 0, W + 1 \leq r \leq V\}$; $h = max\{r | j_r > 0, W + 1 \leq r \leq m - 1\}$, $j_m = 1$ or $h = m$, $j_m > 1$; $n = h + m - [2W - (M - b(\boldsymbol{x}))]$, $W \leq n < h$.
$\cdots$	$\cdots$	$\cdots$
ESU AR. The new ESU request is put into the queue, EQ[¶].	$(j_W, \cdots, j_k, \cdots, j_V, g_a, j_{pu}, j_{eq} + 1, j_{rq})$	$M - b(\boldsymbol{x}) + \sum_{m=W+1}^{V}(m - W)j_m < W$; $j_{eq} < Q_{LE}$.
RSU AR. Enough idle channels exist.	$(j_W, \cdots, j_k, \cdots, j_V, g_a + 1, j_{pu}, j_{eq}, j_{rq})$	$M - b(\boldsymbol{x}) \geq a$.
RSU AR. The ESU with the maximum no. of channels, m donate a channels.	$(j_W, \cdots, j_n + 1, \cdots, j_m - 1, \cdots, j_V, g_a + 1, j_{pu}, j_{eq}, j_{rq})$	$V > W$; $m = max\{r | j_r > 0, W + 1 \leq r \leq V\}$; $n = m - [a - (M - b(\boldsymbol{x}))]$, $W \leq n < m$.
RSU AR. Two ESUs with m and h channels collectively donate a channels.	$(j_W + 1, \cdots, j_n + 1, \cdots, j_h - 1, \cdots, j_m - 1, \cdots, J_V, g_a + 1, j_{pu}, j_{eq}, j_{rq})$	$V > W$; $m = max\{r | j_r > 0, W + 1 \leq r \leq V\}$; $h = max\{r | j_r > 0, W + 1 \leq r \leq m - 1\}$, $j_m = 1$ or $h = m$, $j_m > 1$; $n = h + m - [(W + a) - (M - b(\boldsymbol{x}))]$, $W \leq n < h$.
$\cdots$	$\cdots$	$\cdots$
RSU AR. The new ESU request is put into the queue, RQ.	$(j_W, \cdots, j_k, \cdots, j_V, g_a, j_{pu}, j_{eq}, j_{rq} + 1)$	$M - b(\boldsymbol{x}) + \sum_{m=W+1}^{V}(m - W)j_m < a$; $j_{rq} < Q_{LR}$.

[¶]When the system cannot provide the minimum number of required channels for a new ESU, it will be kept in the EQ if the queue is not full, i.e., with the condition $j_{eq} < Q_{LE}$. $M - b(\boldsymbol{x})$ indicates the number of idle channels and $\sum_{m=W+1}^{V}(m - W)j_m$ indicates the total number of channels that ongoing ESUs can provide.

Thereafter, we have the following expressions.

$$N_{SE}^s = \sum_{\boldsymbol{x} \in \mathcal{S}} \sum_{k=W}^{V} \pi(\boldsymbol{x})kj_k \quad \text{and} \quad N_{SR}^s = \sum_{\boldsymbol{x} \in \mathcal{S}} \pi(\boldsymbol{x})ag_a.$$

Using Little's law [20], the average total delay of ESU services, T_{SE}^t, is obtained as,

$$T_{SE}^t = T_{SE}^q + T_{SE}^s = \frac{L_{EQ}}{\lambda_{SE}^*} + \frac{N_{SE}^s}{\lambda_{SE}^*}. \quad (13)$$

Similarly, the average total delay of RSU services can be expressed as,

$$T_{SR}^t = T_{SR}^q + T_{SR}^s = \frac{L_{RQ}}{\lambda_{SR}^*} + \frac{N_{SR}^s}{\lambda_{SR}^*}. \quad (14)$$

B. Analysis of Scheme II

From the queuing scheme analysis presented earlier, it is straightforward to conclude that the derived mathematical equations for calculating the capacity, spectrum utilization and

the blocking probability of the secondary network, the average PU interruption rate and the expressions for determining the mean queue length apply to both schemes. Furthermore, the equations for determining the forced termination probability and the mean delay of RSU services are also the same in both schemes. However, the forced termination probability and the average delay of an ESU service have to be analyzed separately for the two cases since the preempted ESU services in Scheme II is fed back in the EQ instead of direct service termination.

1) Forced termination probability of ESU services: In Scheme II, a preempted ESU service is forced to terminate if all the waiting positions of the EQ are occupied by other ESU services. The forced termination probability of ESU services in Scheme II, P'_{f1}, can therefore be expressed as the mean forced termination rate of ESU services, R'_{f1}, divided by the mean admitted ESU rate, i.e., λ^*_{SE}. Mathematically, the forced termination probability of ESU services is obtained by,

$$P'_{f1} = \frac{R'_{f1}}{\lambda^*_{SE}} = \sum_{\substack{\boldsymbol{x} \in \mathcal{S}, \\ M=b(\boldsymbol{x}),\ j_{pu}<M, \\ j_W>0,\ j_n=0;\ \forall n>W,\ j_{eq}=Q_{LE}}} \frac{\lambda_P W j_W}{(M - j_{pu})\lambda^*_{SE}} \pi(\boldsymbol{x}).$$

(15)

2) Average delay of ESU services: In Scheme II, the service arrival rate at the EQ is higher than that of Scheme I due to the admission of interrupted ESU services. The parameters used to calculate the average delay of ESU services in Scheme II are the same as in Scheme I. The mean arrival rate of the preempted ESU services at the EQ, equals $\lambda^r_{SE} = R_{int} - R'_{f1}$. Using Little's law, the average waiting time of ESU services, T^q_{SE}, and the average transmission time of ESU services (each time they gain access to the channels), T^s_{SE}, can be expressed as,

$$T^q_{SE} = L_{EQ}/(\lambda^*_{SE}+\lambda^r_{SE}) \quad \text{and} \quad T^s_{SE} = N^s_{SE}/(\lambda^*_{SE}+\lambda^r_{SE}).$$

Let U_r be the average number of times that an ongoing ESU service returns to the EQ after PU interruption. Then,

$$U_r = \frac{\lambda^r_{SE}}{\lambda^*_{SE}} = \frac{R_{int} - R'_{f1}}{\lambda_{SE}(1 - P_{b1})}.$$

(16)

Note that the average total delay of ESU services is

$$T^t_{SE} = (L_{EQ} + N^s_{SE})/\lambda^*_{SE}.$$

Denoting the total arrival rate as $\lambda^t_{SE} = \lambda^*_{SE}+\lambda^r_{SE}$, we obtain finally

$$T^t_{SE} = \frac{L_{EQ} + N^s_{SE}}{\lambda^t_{SE}} \cdot \frac{\lambda^t_{SE}}{\lambda^*_{SE}} = \frac{L_{EQ} + N^s_{SE}}{\lambda^t_{SE}} \cdot \frac{\lambda^*_{SE} + \lambda^r_{SE}}{\lambda^*_{SE}}$$
$$= (T^q_{SE} + T^s_{SE})(1 + U_r).$$

(17)

VI. NUMERICAL RESULTS AND DISCUSSIONS

In this section, numerical results are presented to illustrate the performance of the proposed CA strategy with queues. We use MATLAB to solve the Markov model and to run simulations.

A CRN with 6 channels (i.e., $M = 6$) in the allocated spectrum band is configured and we evaluate the performance of this CRN with respect to the parameters discussed in Sec. V.
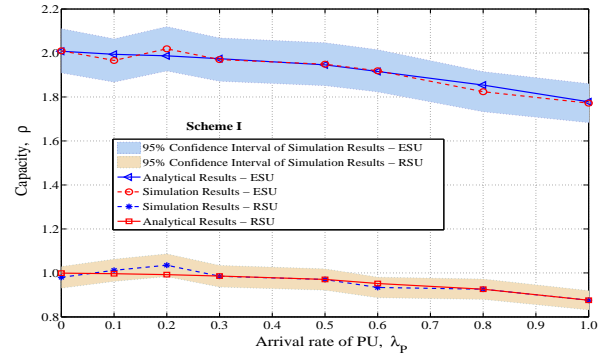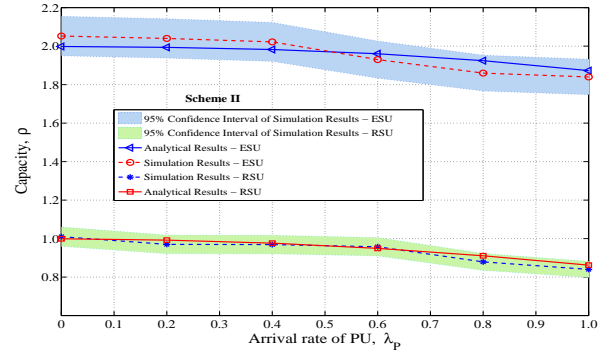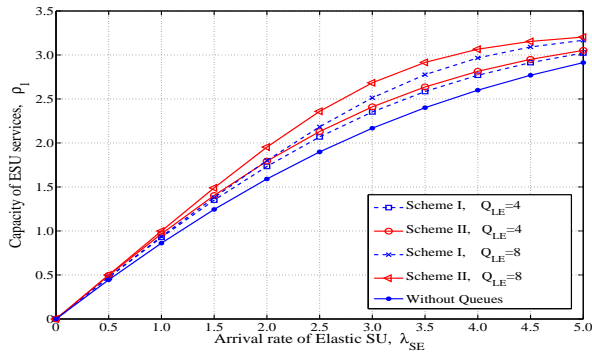

Fig. 3: SU capacity - Scheme I.

Fig. 4: SU capacity - Scheme II.

The default values for PU/SU arrivals and departures are set as $\lambda_P = 1.0$, $\lambda_{SE} = 2.0$, $\lambda_{SR} = 1.0$, $\mu_P = 0.5$, $\mu_{SE} = 1.0$ and $\mu_{SR} = 1.0$. For each result in our performance evaluation, one of these parameters[2] varies at a time while the others are kept the same as the default values. The configurable parameters of the dynamic spectrum access strategy used in the system, i.e., Dynamic$(a, W, V, Q_{LE}, Q_{LR})$ are fixed as $W = 1$, $V = 3$ and $a = 1$. Unless otherwise stated, the maximum lengths of the queues are configured as $Q_{LE} = 4$ and $Q_{LR} = 2$ respectively and the priority factor of the queue scheduling algorithm is fixed as $\beta = 2$.

A. Model Validation

To validate the correctness and the preciseness of the mathematical analysis, we simulate PU and SU events based on the proposed channel access strategy and compare the analytical results with the simulation results. Fig. 3 and Fig. 4 depict the variation of the secondary network capacity as a function of PU arrival rate for Scheme I and Scheme II respectively. When more PU services arrive, the capacity of both ESU and RSU services decreases. As shown in those figures, the analytical results lie within the 95% confidence intervals of the simulation results, meaning that 95% of the observed confidence intervals will hold the true value of the parameter. Therefore, we ascertain that the results of the analytical model precisely coincide with the obtained simulation results. Many

[2]The unit of these parameters could be session per unit of time. Given concrete values for these parameters, e.g., 0.5 sessions/sec and flow length in bits, the capacity could be expressed in kbps or Mbps.

Fig. 5: SU Capacity as a function of λ_{SE}.

Fig. 6: Spectrum utilization as a function of λ_{SE}.

empirical studies have found that the log-normal distribution is a more realistic model for service times of both real-time and elastic traffic [21]-[23]. However, in our analysis, for the sake of mathematical tractability we have assumed that service times are exponentially distributed. The capacity performance under log-normal service time distribution is validated in Subsec. VI.F by means of computer simulations.

B. Performance Evaluation

In this subsection, we use the derived analytical models in Sec. V to evaluate the performance of the $Dynamic(a, W, V, Q_{LE}, Q_{LR})$ CA strategy in terms of secondary network capacity, spectrum utilization, blocking probability, forced termination probability and average total delay.

1) Secondary network capacity: Fig. 5 depicts the capacity of the ESU services as the ESU arrival rate, λ_{SE}, varies. With higher arrival rate of ESU services, the capacity of ESU services initially increases almost linearly. However, the increasing rate of the ESU capacity flattens out after $\lambda_{SE} = 3.0$. The reason is as follows. Initially, the capacity increases with a high rate since more ESU services are accommodated in the system. With even higher λ_{SE}, almost all channels in the system are fully utilized. That is, the network is close to saturation and the achieved capacity increases moderately when injecting more SU traffic. Another important observation is that the capacity increases due to the introduction of queues and it can be further increased by using a larger buffer. That is, the capacity corresponding to the queue size of eight, i.e., $Q_{LE} = 8$, is always higher than that of $Q_{LE} = 4$. This is because that a large queue size allows to store more service requests, leading to better channel access opportunities for ESUs even though no channels were available upon their arrivals.

When comparing the capacity achieved by the two queuing schemes, it is obvious that Scheme II obtains higher capacity than Scheme I does. This is because that some of the interrupted ESU services are forced to terminate in Scheme I even though the EQ is empty. These services are not considered as service completions, and correspondingly are not counted for capacity calculations. In Scheme II, however, those interrupted services may have chances to finish their services later, since they are queued and still have opportunity to access system resources. Furthermore, it is worth mentioning that the curve marked as *Without Queues* means the capacity achieved by the
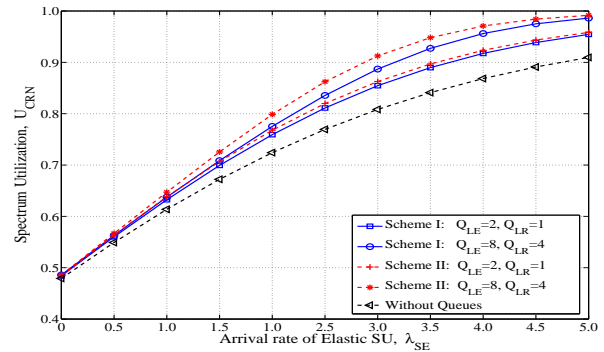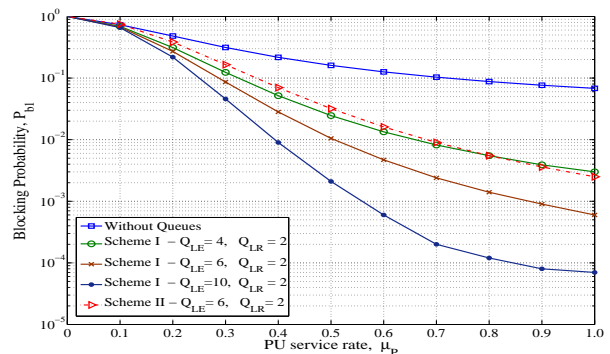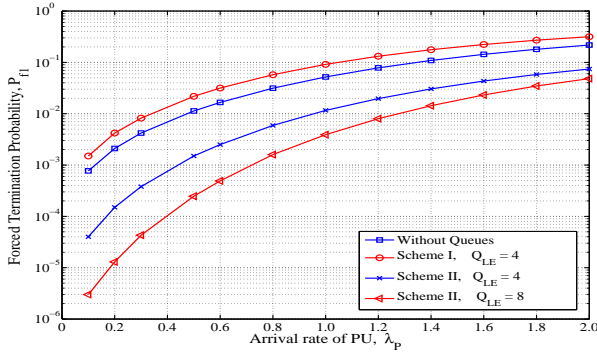

Fig. 7: Blocking probability as a function of PU service rate.

$D(a, W, V)$ strategy proposed in [7]. *This result demonstrates that the $Dynamic(a, W, V, Q_{LE}, Q_{LR})$ strategy outperforms its predecessor, $D(a, W, V)$, in terms of achieved network capacity.*

2) Spectrum utilization: From Fig. 6, we observe that spectrum utilization is improved using our proposed queue-based dynamic CA strategy. When the system is operated without queues, spectrum utilization shows lower values in contrast with the system with queues. With higher λ_{SE} the improvement of spectrum utilization is more significant. For example, with the queue sizes for the low and high priority queues in the system configured as $Q_{LE} = 8$ and $Q_{LR} = 4$ respectively, and $\lambda_{SE} = 5$, the spectrum utilization is improved by 4%, compared with the system configured as $Q_{LE} = 2$ and $Q_{LR} = 1$. Again, with larger queue size, more SUs can be accommodated in the queues. Therefore, when a channel becomes idle, it is more likely that it will be utilized by an SU. To improve spectrum utilization, it is required to reduce the time period in which the channels stay in idle state. Larger queue size can fulfil this requirement since more user requests are allowed. Furthermore, the numerical results reveal that under identical conditions, the CRN with Scheme II has the highest spectrum utilization in comparison with Scheme I and *the system without queue*. The reason for non-zero spectrum utilization at $\lambda_{SE} = 0$ is because of the baseline utilization due to PU and RSU services.

3) Blocking probability: The ESU blocking probability as a function of PU service rate is plotted in Fig. 7. From this figure, we observe that a continuous descent in the blocking probability of new user requests as the service rate of PUs

Fig. 8: Forced termination probability as a function of λ_P.

Fig. 9: Average total delay as a function of SU arrival rates.

Fig. 10: P_{b1}, P_{f1} and T_{SE}^t as a function of λ_{SE}.

increases while keeping λ_P constant. When the PU service rate is low, fewer PUs finish their services per unit of time. Then a majority of the channels are occupied by respective licensed users and the probability of blocking a new user request is therefore high. Since the probability that the system stays in the busy state decreases with higher μ_P, the probability of blocking a new ESU service becomes lower. As shown in Fig. 7, the blocking probability of ESU services decreases using the proposed queuing schemes since the blocked services due to insufficient channels are buffered into the queues until they are offered with the required number of channels. For this reason, the blocking probability is significantly reduced with the integration of queues. With larger queue size, more SUs can be queued instead of being blocked, resulting in lower blocking probability.

Furthermore, the blocking probability of ESU services in Scheme II is higher than that of Scheme I for the same configuration. The reason for this behavior can be explained as follows. In Scheme II the preempted ESU services are also allowed to be queued in the EQ. In other words, the low priority queue in Scheme II is allocated for both new arrivals and interrupted services. Thus, the waiting positions of the EQ are more likely to be occupied by ESU services in Scheme II. For this reason, the chance of finding a queue position becomes lower when a new ESU service arrives. Therefore, the new ESU service requests are blocked and consequently the blocking probability in Scheme II is higher than that of Scheme I.

4) Forced termination probability: The forced termination probability is illustrated in Fig. 8 as PU arrival rate varies. Obviously the forced termination probability becomes higher as PUs become more active in all cases, including the system without queues. However, the slope of the forced termination probability curve becomes flatter as λ_P grows. The reason is as follows. When λ_P is comparatively low, only a small number of channels in the system are occupied by PUs. As λ_P increases, more ongoing ESU services will be forced to terminate. However, when λ_P further increases, almost all the channels are occupied by PUs. Therefore there are fewer ongoing SUs to be interrupted. According to the numerical results depicted in Fig. 8, the forced termination probability of the queuing system with Scheme I is always higher than that of the system without queues. This result reveals that Scheme I of the proposed queuing scheme cannot provide
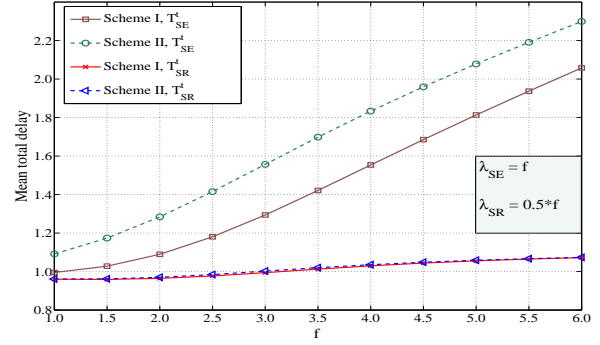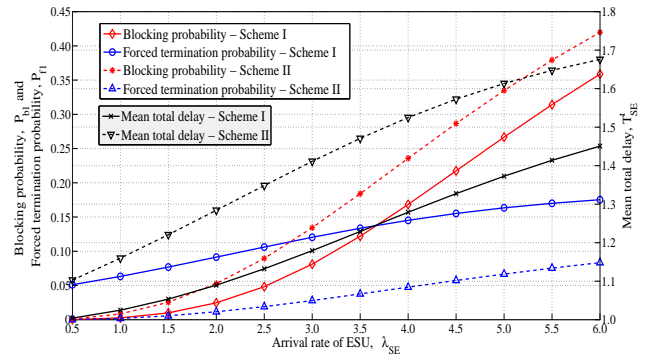
better performance in terms of forced termination probability. In Scheme I when queues are integrated, the would-be-blocked ESU services by $D(a, W, V)$ are not rejected but instead buffered in the queue in order to provide channel access when it is available. This implies that more service interruptions occur due to PU arrivals since there are more ESUs in the system. As a consequence, a higher forced termination probability is observed in Scheme I. On the other hand, even though ESU services in Scheme II also encounter more interruptions due to more active PU activities, the interrupted ESU services are kept and fed back into the EQ instead of being terminated. Hence, the forced termination probability of ESU services reduces in Scheme II.

C. Average total delay

From Fig. 9, we observe a significant increase of delay in ESU services when Scheme II is employed. The reason is that in Scheme II, the EQ is more likely to be occupied than in Scheme I since the interrupted ESU services are also buffered. However, there is no significant difference in RSU delay between these two schemes since Scheme II does not differ from Scheme I in terms of channel access for RSU services. Furthermore, it is clear that the reduced queue length of the RQ shows a noticeable contribution to maintain a controlled delay on RSU services.

D. Allowable Maximum Arrival Rate for a Required Grade of Service

In this subsection we offer an approach to analyze the maximum ESU arrival rate that can be allowed by the CRN
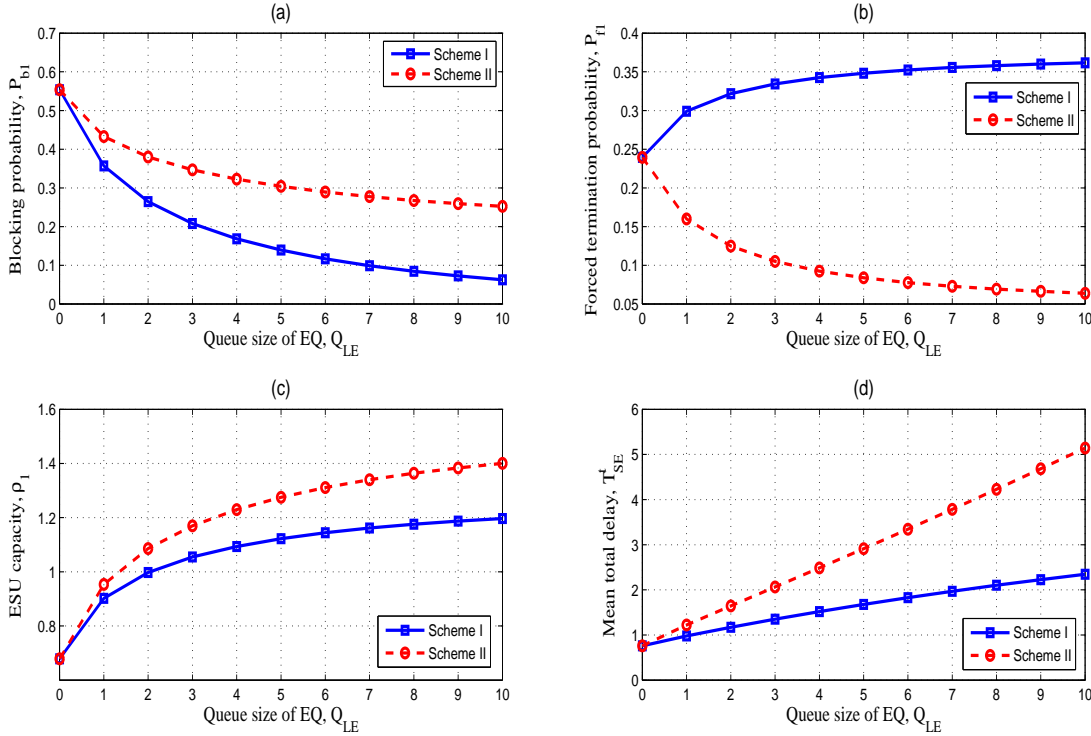
Fig. 11: Performance measures as a function of the queue size of EQ.

under different grade of service (GoS) requirements. In particular, we focus on three performance measures, i.e., blocking probability, forced termination probability and average total delay. Those performance metrics are plotted against the ESU arrival rate as shown in Fig. 10 for both schemes. Different GoS requirements may apply to different traffic types of SUs in a CRN. For example, consider a CRN with the following GoS requirements. Blocking probability of new ESU services, $P_{b1} \leq 0.25$, forced termination probability of ongoing ESU services, $P_{f1} \leq 0.15$ and average total delay of an ESU service, $T_{SE}^t \leq 1.2$ time units. In order to satisfy all these three requirements in the proposed model, the maximum ESU arrival rate that the system with Scheme II can support is $\lambda_{SE} = 1.3$ session per time unit. However, under the same GoS requirement, the system can support ESU arrival rates up to $\lambda_{SE} = 3.2$ sessions per time unit if Scheme I is employed. This is because the average waiting time of an ESU service in Scheme II is always higher than in Scheme I due to the additional arrivals of the interrupted ESU services to the EQ.

On the other hand, if the above mentioned CRN has a stringent requirement for forced termination probability to protect ongoing ESU services, i.e., $P_{f1} < 0.05$, Scheme I cannot support ESUs with an arrival rate of greater than $\lambda_{SE} = 0.5$ sessions per time unit. However, Scheme II can support for a maximum ESU arrival rate of $\lambda_{SE} = 1.3$ sessions per time unit while still satisfying the GoS requirement. The reason can be explained as follows. In this set of GoS requirement, the forced termination probability constraint is more dominant than the other constraints. For a given forced termination constraint, the maximum allowable ESU arrival rate of Scheme I must be

lower than that of Scheme II because more forced terminations happen in Scheme I. The results presented in this subsection indicate that an appropriate queuing scheme may apply to a specific type of service depending on the requirements of the application.

E. Effect of Queue Length on Performance Measures

In the proposed model, queue lengths play a significant role on system performance. In order to illustrate this effect, the behavior of four performance measures is observed in Fig. 11 by varying the queue size of the EQ while keeping the queue size of the RQ constant. In addition, the arrival rates are set as $\lambda_{SE} = \lambda_{SR} = \lambda_P = 2.0$. As shown in Fig. 11(a), the blocking probability of ESU services decreases significantly with larger queue size. Furthermore, Scheme I achieves low blocking probability than Scheme II does. However, only Scheme II shows performance enhancement in terms of forced termination probability as illustrated in Fig. 11(b). Even though both schemes improve the capacity when queue size increases, as depicted in Fig. 11(c), Scheme II exhibits approximately 15% higher SU capacity in comparison with Scheme I when $Q_{LE} > 4$. The main cost incurred by the proposed queuing schemes is the additional waiting time of SU services, i.e., the queuing delay as part of the overall delay of services. To illustrate the effect of the queue size on the overall delay, we present the numerical results shown in Fig. 11(d). As expected, longer average delay occurs to ESU services with a larger buffer size. Moreover, since the preempted ESUs are buffered in the queue in Scheme II, the
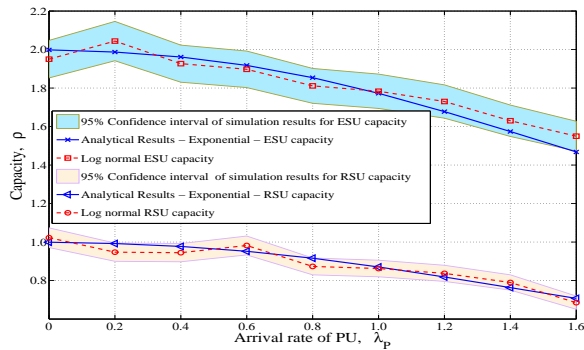
Fig. 12: Capacity as a function of λ_P with log-normal distribution with the same mean value and variance as exponential distribution.
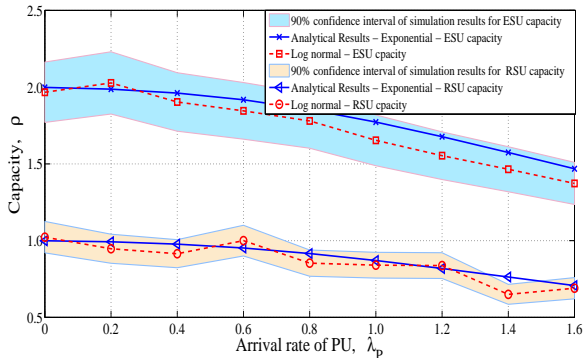

Fig. 13: Capacity as a function of λ_P using log-normal distribution with the same mean value but a different variance from the exponential distribution.

average waiting time of an ESU service increases faster than in Scheme I. Jointly considering the results in Fig. 11(b) and Fig. 11(d), we observe that *there is a tradeoff between the average delay and the forced termination probability*. That is, even though Scheme II contributes to a significant reduction of the forced termination probability for ESU services, the benefit is achieved at a cost of relatively high total delay. Please note that the results with $Q_{LE} = 0$ for all four parameters coincide with the results reported in [7] for both schemes.

F. Performance Evaluation under Log-normal Service Time Distributions

Finally we illustrate the secondary network capacity of the CRN with Scheme I with log-normal service time distribution in Fig. 12. For comparison, we include the corresponding analytical results of the original exponential distribution in the same figure. The capacity curve is generated by simulations based on the log-normal distribution with the same mean value and the variance as used in the original exponential distribution, while the arrival process is still kept as a Poisson process. From this figure, we can observe that the results with log-normal service time distribution are quite close to the analytical results which are based on exponential distribution. This observation reveals that the results obtained in the proposed model are not sensitive to the type of service distributions. Furthermore, we perform the same test by re-configuring the variance of log-normal distributions 20% higher than the variance of the original exponential distributions, while

the mean value is kept the same. The obtained results are illustrated in Fig. 13. Even though the results under the log-normal service time distribution with different variance do not precisely coincide with the analytical output compared with the results depicted in Fig. 12, the theoretical model still exhibits a good match with the simulation results. Thus, we claim that the mathematical model proposed in this paper suits well for evaluating the performance of CA strategies in CRNs with queues.

VII. CONCLUSIONS

In this paper, we have proposed a queue-based channel assembling strategy for multi-channel CRNs as well as the analytical framework for performance evaluation of such networks. An existing dynamic channel assembling strategy is improved by introducing two queuing schemes which consist of separate queues allocated for different traffic types with different priorities. The developed analytical models are validated by extensive simulations. We evaluated the performance of the system in terms of different parameters and showed that the proposed schemes outperform the existing ones in terms of secondary network's capacity, blocking probability and spectrum utilization. With the increase of queue size, the blocking probability of CR users could be further decreased, however at a cost of longer total delay. Therefore, the queue size of the RQ is kept as small as possible in order to reduce the delay experienced by real-time services and to achieve higher capacity for elastic services. Since elastic traffic services benefit more from low blocking probability and high capacity rather than overall delay, the EQ is designed with comparatively large queue size. The proposed two queuing schemes are different in terms of how the interrupted elastic SU services are dealt with. Since Scheme II significantly reduces forced terminations of ESU services, it is recommended if PUs are more active in a CRN. Considering delay performance of these two schemes, Scheme I is more appropriate for delay-critical applications like video streaming and web TV while Scheme II suits better for applications like interactive web browsing which tolerate longer delays.

REFERENCES

[1] C. R. Stevenson, G. Chouinard, Z. Lei, W. Hu, S. J. Shellhammer, and W. Caldwell, "IEEE 802.22: The First Cognitive Radio Wireless Regional Area Network Standard", *IEEE Communications Magazine*, vol. 47, no. 1, pp. 130-138, Jan. 2009.

[2] E. W. M. Wong, and C. H. Foh, "Analysis of Cognitive Radio Spectrum Access with Finite User Population", *IEEE Communications Letters*, vol. 13, no. 5, May 2009.

[3] H. Al-Mahdi, M. A. Kalil, F. Liers, and A. M. Thiel, "Increasing Spectrum Capacity for Ad Hoc Networks using Cognitive Radios", *IEEE Communications Letters*, vol. 13, no. 9, Sep. 2009.

[4] G. D. Nguyen, and S. Kompella, "Channel Sharing in Cognitive Radio Networks", *in Proc. IEEE Military Communications Conference (MILCOM)*, pp. 2268-2273, San Jose, CA, USA, Nov. 2010.

[5] P. Kaur, A. Khosla, and M. Uddin, "Markovian Queuing Model for Dynamic Spectrum Allocation in Centralized Architecture for Cognitive Radios", *IACSIT International Journal of Engineering and Technology*, vol. 3, no. 1, Feb. 2011.

[6] L. Jiao, F. Y. Li, and V. Pla, "Dynamic Channel Aggregation Strategies in Cognitive Radio Networks with Spectrum Adaptation", *in Proc. IEEE GLOBECOM*, Houston, TX, USA, Dec. 2011.

[7] L. Jiao, F. Y. Li, and V. Pla, "Modeling and Performance Analysis of Channel Assembling in Multi-channel Cognitive Radio Networks with Spectrum Adaptation", *IEEE Transactions on Vehicular Technology*, vol. 61, no. 6, pp. 2686-2697, Jul. 2012.

[8] L. Jiao, V. Pla, and F. Y. Li, "Analysis on Channel Bonding/Aggregation for Multi-channel Cognitive Radio Network", *in Proc. European Wireless (EW)*, Lucca, Italy, Apr. 2010.

[9] B. Gao, Y. Yang, J. M. Park, "Channel Aggregation in Cognitive Radio Networks with Practical Considerations", *in Proc. IEEE ICC*, Kyoto, Japan, Jun. 2011.

[10] M. M. Rashid, M. J. Hossain, E. Hossain, and V. K. Bhargava, "Opportunistic Spectrum Scheduling for Multiuser Cognitive Radio: A Queuing Analysis", *IEEE Transactions on Wireless Communications*, vol. 8, no. 10, pp. 5259-5269, Oct. 2009.

[11] S. Tang, and B. L. Mark, "Performance of a Cognitive Radio Network with Tolerable Service Degradation", *in Proc. Int. Workshop on Design of Reliable Communications Networks (DRCN'09)*, Alexandria, VA, USA, Oct. 2009.

[12] S. Wang, J. Zhang, and L. Tong, "A Characterization of Delay Performance of Cognitive Medium Access", *IEEE Transactions on Wireless Communications*, vol. 11, no. 2, pp. 800-809, Feb. 2012.

[13] A. A. El-Sherif, and K. J. R Liu, "Joint Design of Spectrum Sensing and Channel Access in Cognitive Radio Networks", *IEEE Transactions on Wireless Communications*, vol. 10, no. 6, pp. 1743-1753, Jun. 2011.

[14] H. Yaiche, R. R. Mazumdar, and C. Rosenberg, "A Game Theoretic Framework for Bandwidth Allocation and Pricing in Broadband Networks", *IEEE/ACM Transactions on Networking*, vol. 8, no. 5, pp. 667-678, Oct. 2000.

[15] M. H. Cheung, A. H. Mohsenian-Rad, V. W. S. Wong, and R. Schober, "Random Access for Elastic and Inelastic Traffic in WLANs", *IEEE Transactions on Wireless Communications*, vol. 9, no. 6, pp. 1861-1866, Jun. 2010.

[16] K. Wu, and W. Liao, "On Service Differentiation for Multimedia Traffic in Multi-Hop Wireless Networks", *IEEE Transactions on Wireless Communications*, vol. 8, no. 5, pp. 2464-2472 , May 2009.

[17] W. Han, J. Li, Z. Tian, and Y. Zhang, "Dynamic Sensing Strategies for Efficient Spectrum Utilization in Cognitive Radio Networks", *IEEE Transactions on Wireless Communications*, vol. 10, no. 11, pp. 3644-3655, Nov. 2011.

[18] Q. Huang, S. Xu, and X. Jin, "Performance Evaluation of Secondary Users in Dynamic Spectrum Access System", *in Proc. IEEE PIMRC*, Toronto, Canada, Sep. 2011.

[19] S. Joshi, P. Pawełczak, D. Čabrić, and J. Villasenor, "When Channel Bonding is Beneficial for Opportunistic Spectrum Access Networks," *IEEE Transactions on Wireless Communications*, vol. 11, no. 11, pp. 3942-3956, Nov. 2012.

[20] V. B. Iversen, "Teletraffic Engineering and Network Planning", DTU (Technical University of Denmark) Course 34330, May 2011, available at: http://oldwww.com.dtu.dk/education/34340/telenook.pdf.

[21] D. Willkomm, S. Machiraju, J. Bolot, and A. Wolisz, "Primary Users Behavior in Cellular Networks and Implications for Dynamic Spectrum Access," *IEEE Communications Magazine*, vol. 47, no. 3, pp. 88-95, Mar. 2009.

[22] F. Barcelo and J. Jordan, "Channel Holding Time Distribution in Public Telephony Systems", *IEEE Transactions on Vehicular Technology*, vol. 49, no. 5, pp. 1615-1625, Sep. 2000.

[23] G. -F. Zhao, Q. Shan, S. Xiao, and C. Xu, "Modeling Web Browsing on Mobile Internet," *IEEE Communications Letters*, vol. 15, no. 10, pp. 1081-1083, Oct. 2011.

[24] F. Wu, N. Singh, N. Vaidya, and G. Chen, "On Adaptive-width Channel Allocation in Non-cooperative, Multi-radio Wireless Networks", *in Proc. IEEE INFOCOM*, Shanghai, China, Apr. 2011.

[25] M. G. Khoshkholgh, K. Navaie, and H. Yanikomeroglu, "Adaptive Multiple Time-Scale Power Allocation for Spectrum Sharing in DS-CDMA Networks", *in Proc. IEEE International Workshop on Towards Cognition in Wireless Networks (CogNet) Collocated with IEEE ICC*, Beijing, China, May 2008.

Indika A. M. Balapuwaduge (S'12) received his B.Sc. Eng. degree with 1st class honors from University of Ruhuna, Sri Lanka, in 2008 and M.Sc. degree in Information and Communication Technology (ICT) from University of Agder (UiA), Norway in 2012. His Master thesis was awarded as the Best Master's thesis in ICT at UiA. He was a technical product engineer at Huawei technologies, Sri Lanka, from Oct. 2008 to Aug. 2009 and he worked as a lecturer at the Department of Electrical and Information Engineering, University of Ruhuna from Aug 2009 to Aug 2010. Indika's current research interest covers various areas of mobile and wireless communication, including reliability analysis, spectrum management, modeling and performance evaluation of communication networks especially cognitive radio networks.

Lei Jiao (S'08, M'13) received his BE and ME degrees in Telecommunication Engineering, and Communication and Information System from Hunan University and Shandong University, China respectively in 2005 and 2008. He received his Ph.D. degree in Information and Communication Technology from University of Agder (UiA), Norway in 2012. He is now working at the Department of Information and Communication Technology, University of Agder as a postdoc researcher. His research interests include cognitive radio networks, wireless sensor networks, and artificial intelligence.

Vicent Pla received the M.E. and Ph.D. degrees in Telecommunication Engineering from the Universitat Politècnica de València (UPV), Spain. He was the recipient of the 3rd National Award for Excellence in Academic Performance in the Telecommunications Engineering degree in 1999, bestowed by the government of Spain, and in 2006 he received the UPV Outstanding Ph.D. Thesis Award. He is currently an Associate Professor at UPV. During the last years he has been an active participant in several national and European research projects, such as the Euro-NGI, FGI and NF Networks of Excellence. His research interests lie primarily in the areas of teletraffic and performance evaluation of communication networks. In these areas he has published numerous papers in refereed journals and conference proceedings.

Frank Y. Li (S'99, M'03, SM'09) holds a Ph.D. degree from the Norwegian University of Science and Technology (NTNU). He worked as a Senior Researcher at UniK - University Graduate Center, University of Oslo before joining the Department of Information and Communication Technology, University of Agder (UiA) in August 2007 where he is currently a Professor. During the past few years, he has been an active participant in several Norwegian and EU FP6/FP7 research projects. He is listed as a *Lead Scientist* by the European Commission DG RTD Unit A.03 - Evaluation and Monitoring of Programmes in Nov. 2007. Dr. Li's research interests include MAC mechanisms and routing protocols in 4G and beyond mobile systems and wireless networks, mesh and ad hoc networks; wireless sensor network; D2D communication; cooperative communication; cognitive radio networks; green wireless communications; QoS, resource management and traffic engineering in wired and wireless IP-based networks; analysis, simulation and performance evaluation of communication protocols and networks.