

Document downloaded from:

<http://hdl.handle.net/10251/44390>

This paper must be cited as:

Jiao, L.; Balapuwaduge, IAM.; Li, FY.; Pla, V. (2014). On the Performance of Channel Assembling and Fragmentation in Cognitive Radio Networks. *IEEE Transactions on Wireless Communications*. 13(10):5661-5675. doi:10.1109/TWC.2014.2322057.



The final publication is available at

<http://dx.doi.org/10.1109/TWC.2014.2322057>

Copyright Institute of Electrical and Electronics Engineers (IEEE)

# On the Performance of Channel Assembling and Fragmentation in Cognitive Radio Networks

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

**Abstract**—Flexible channel allocation may be applied to multi-channel cognitive radio networks (CRNs) through either channel assembling (CA) or channel fragmentation (CF). While CA allows one secondary user (SU) occupy multiple channels when primary users (PUs) are absent, CF provides finer granularity for channel occupancy by allocating a portion of one channel to an SU flow. In this paper, we investigate the impact of CF together with CA for SU flows by proposing a channel access strategy which activates both CF and CA and correspondingly evaluating its performance. In addition, we also consider a novel scenario where CA is enabled for PU flows. The performance evaluation is conducted based on continuous time Markov chain (CTMC) modeling and simulations. Through mathematical analyses and simulation results, we demonstrate that higher system capacity can be achieved indeed by jointly employing both CA and CF, in comparison with the CA-only strategies. However, this benefit is obtained only under certain conditions which are pointed out in this paper. Furthermore, the theoretical capacity upper bound for SU flows with both CF and CA enabled is derived when PU activities are relatively static compared with SU flows.

**Index Terms**—Multi-channel cognitive radio networks; Channel assembling; Channel fragmentation; Continuous time Markov chain modeling; Performance evaluation.

## I. INTRODUCTION

In cognitive radio networks (CRNs) [1], secondary users (SUs) can access the spectrum which has already been allocated to primary users (PUs) when PUs are inactive. To further enhance spectrum utilization, channel assembling (CA) and channel fragmentation (CF) techniques can be employed. CA means that multiple channels can be assembled by SUs for a single SU flow, instead of using barely one channel per SU flow all the time. This technique has been proposed in many dynamic spectrum access strategies and CRN medium access control (MAC) protocols [2]–[14]. In contrast, a single channel can be shared by multiple SU flows if CF is supported. With CF, an SU flow may be accommodated by a portion of

a channel if the quality of service (QoS) requirement is met [16]–[18], and it can also assemble more than one channel with a non-integer number of channels together with CA [19].

It has been demonstrated in our previous work [13]–[15] that better system performance can be achieved by CA when spectrum adaptation is adopted with proper parameter configuration. In brief, spectrum adaptation embodies two functionalities, i.e., 1) an ongoing SU flow is able to jump to another idle channel if PU appears on the current channel; and 2) the number of assembled channels for an ongoing SU flow can be adapted according to the variation of the channel availability. The capacity<sup>1</sup> upper bound in the quasistationary regime (QSR) for any CA strategy is deduced in [20] and it is demonstrated that this upper bound is achievable by dynamic strategies with appropriately configured parameters.

Recently, CF combined with CA and spectrum adaptation has drawn attention in the research community [19], [20]. However, a common feature of these presented strategies is that each ongoing SU flow occupies *at least one channel*, and CF is introduced in order to achieve a non-integer number of assembled channels for an SU flow. In contrast, the case that an SU flow can be served by *a portion of one channel* is not studied. At the same time, we observed that the effect of CF on real-time SU flows is not investigated in [19], [20]. Therefore, a full scope on the impact of CF on network performance is still hidden in the literature.

The main objective of this study is to make an overall evaluation on the impact of CF to SU performance when both CA and spectrum adaptation are already enabled, given that an SU flow can be accommodated by a portion of one channel. In addition, as more and more emerging wireless and cellular communication systems, like 802.11ac [9] and LTE-A [12], adopt CA, this technique can also be implemented by PUs. Therefore, the impact on SUs when PUs can assemble multiple channels is also of essential interest to study. Considering different impact that CF and CA may have on various types of flows, two SU traffic types, i.e., elastic traffic and real-time traffic, are investigated. For an elastic flow, like file downloading, the service time will be reduced if more channels are assembled for the same flow due to higher data rate, and vice versa. On the contrary, for a real-time flow, like a voice conversation, the service duration will not be affected by the number of assembled channels as long as the basic QoS requirement is satisfied for such flows. Both elastic and

Lei Jiao, Indika A. M. Balapuwaduge, and Frank Y. Li are with the Dept. of Information and Communication Technology, University of Agder (UiA), N-4898 Grimstad, Norway (E-mail: {lei.jiao; indika.balapuwaduge; frank.li}@uia.no).

Vicent Pla is with the Dept. of Communications, Universitat Politècnica de València (UPV), 46022 València, Spain (E-mail: vpla@upv.es).

The cooperation among the authors was supported by the EU Seventh Framework Programme FP7-PEOPLE-IRSES under grant agreement number 247083, project acronym S2EuNet.

The work of L. Jiao was partially supported by the Research Council of Norway through the ECO-boat MOL project under grant number 210426, and by the EU Seventh Framework Programme STREP - FP7-ICT-2013-11 under grant agreement number 619560, project acronym SEMIAH. The work of V. Pla was supported in part by the Ministry of Economy and Competitiveness of Spain under Grant TIN2010-21378-C02-02.

<sup>1</sup>Capacity in [20] is denoted as the number of flows that can be served by the system per time unit, which is different from Shannon capacity. We utilize the same definition for capacity in this work unless otherwise stated.

real-time traffic types are studied in this work.

To investigate system performance for elastic traffic, we first enhance the full adaptation and full sharing (FAFS) strategy proposed in [20]. More specifically, we extend the minimum number and the maximum number of channels that can be utilized by an ongoing SU flow in FAFS into a larger number domain. Then we propose and analyze two different CTMC models for the system under various conditions. The first model is a precise model that corresponds to the case where there is no constraint on the time scales of PU and SU activities. The second model applies to the QSR, i.e., the one in which the PU events occur at a much longer time-scale than the SU events do. The performance of SUs in the QSR is also of essential interest because a comparatively slow PU activity is an important prerequisite for envisaged CRN deployment in real-life [5], [21]. Furthermore, as an extension of the conclusions drawn in [20], we derive a closed-form theoretical capacity upper bound a CRN can achieve in the QSR by employing both CF and CA. To further examine the impact of CF on real-time traffic, similar mathematical models are developed and analyses are performed.

In brief, the main contributions of this paper are outlined as follows:

- A representative spectrum access strategy with *non-integer number* of assembled channels is developed. This strategy does not only accommodate SUs with multiple channels through CA and spectrum adaptation, but also supports SUs which use only *a portion of* one channel for their traffic flows. Its mathematical analysis is significant as it provides theoretical insight on the performance gain that can be obtained by CF together with CA and spectrum adaptation.
- Based on the analysis of the mathematical model, we reveal an important condition under which CF can improve the capacity for elastic traffic of a CRN using a strategy that enables both CA and spectrum adaptation. Furthermore, the closed-form capacity upper bound in the QSR for any strategy when CF and CA are adopted is derived for elastic traffic.
- For real-time traffic, it is demonstrated mathematically that higher capacity can also be achieved if CF is utilized. The corresponding theoretical capacity upper bound in the QSR is also derived.
- The impact of applying CA to PUs on the performance of SUs is also considered and analyzed. It is a more generalized scenario than the previously studied cases in which PU flows would always occupy one channel.

The remainder of this paper is organized as follow. In Sec. II, the related work is summarized. The network scenario and the access strategy are presented in Sec. III. In Sec. IV the impact of CF together with CA for elastic traffic is investigated, followed by Sec. V in which the impact of CF for real-time traffic is examined. Numerical results are presented and discussed in Sec. VI before we conclude the paper in Sec. VII.

## II. RELATED WORK

In what follows, we first summarize the existing work on mathematical modeling of channel access strategies when either CF or CA is supported, and then survey the studies where both CF and CA are enabled.

### A. CF Strategies and Their Mathematical Models

The performance of CRNs with CF has been studied under various scenarios [16]–[18] through CTMCs, where an SU flow can be accommodated by a portion of a channel. Papers [16] and [17] focus mainly on the cases when there are infinite number of users. In those studies, both spectrum handover<sup>2</sup> and non-handover cases are investigated. When spectrum handover is not supported, multiple ongoing SU flows may be preempted by a single PU arrival. Therefore, the system performance is worse than when spectrum handover is enabled. Besides spectrum handover and non-handover cases, channel reservation which makes a tradeoff between forced termination probability and blocking probability is also investigated in [16]. Paper [18] illustrates the performance of CRNs with a finite number of users. The system model in the QSR is also studied in that work.

### B. CA Strategies and Their Mathematical Models

The CA technique has been widely proposed in MAC protocol design for CRNs [2]–[5] and many other communication systems [6]–[12]. The protocol-level studies reveal the feasibility of CA in practice and illustrate the detailed scheduling and signaling process for such a technique. To evaluate the performance of these protocols, simulations and experiments are carried out according to the above literature. Furthermore, Markov chain based mathematical analysis is adopted, e.g., in [2], [3]. In [28], the application of CA to SU packets is studied in-depth for CRNs. The analysis in [28] is performed through discrete time Markov chains.

To investigate the performance of CRNs when CA is applied to SU traffic flows, many channel access schemes have been proposed and their mathematical models are developed [13]–[15], [23]–[27]. Due to its flexibility, elastic traffic is the main focus for many CA strategies [13], [23]–[27]. Heterogeneous traffic types with both elastic and real-time traffic have also been considered in [14], [15], [22]. In our earlier work [14], we consider heterogeneous traffic in a loss system whereas in [15] we introduce queues and scheduling schemes. Different priority-based access mechanisms are proposed for various types of flows in [14]. It is also demonstrated in [14] that spectrum adaptation is an important prerequisite where CA can result in performance improvement compared with the strategies without CA.

### C. Strategies with both CF as well as CA and Their Mathematical Models

Two similar strategies with both CA and CF enabled are proposed independently in [19] and [20] with different scopes.

<sup>2</sup>Spectrum handover means that an ongoing SU flow can hop to another idle channel to maintain the flow if a PU appears on the current channel.

Both strategies require that all ongoing SU flows equally share the available channels, implying that CF is necessary for such strategies in order to achieve a non-integer number of assembled channels. In [20], the capacity upper bound of any strategy employing CA techniques is derived in the QSR. In order to reduce the dimension of the Markov chain models, CF is enabled therein as a supplementary technique. However, we did not explore in [20] the benefit of enabling CF in addition to CA. The study in [19] instead focuses on analyzing the strategy itself, and its mathematical analysis as well as simulations are performed based on an implicit assumption that each SU flow requires *at least one channel*. In fact, as shown later in this work, we can find a CA-only strategy that is equivalent to the strategies modeled in [19] [20] in the condition that at least one channel is needed for each flow. The property which allows accommodating an SU flow by *a portion of a channel* results in a fundamental difference in the system performance. This point is not revealed earlier, by neither [19] nor [20].

In short, although CF and CA are proposed and analyzed for CRNs, it is still necessary 1) *to identify under which circumstances CF can bring further benefit to a system in addition to CA*, and 2) *to explore how much benefit CF can bring to the system concerning the parameters of interests*. In the rest of this paper, we intend to provide the answers to the above two questions by proposing and analyzing a new strategy. Since spectrum adaptation is important for improving system performance [14], spectrum adaptation is enabled together with CF and CA by default.

### III. SYSTEM CONFIGURATIONS AND ACCESS STRATEGIES

#### A. Network Scenario and System Configurations

Consider a CRN with multiple channels. Assume that there are two types of radios, PUs and SUs, operating in the same frequency band. PUs have preemptive privilege for spectrum access and can acquire the channels being used by SUs at any time. There is no cooperation between PUs and SUs.

Without loss of generality, we define the minimum spectrum requirement for a single PU flow as one channel (a unit channel for the system). The total spectrum band consists of  $M \in \mathbb{Z}^+$  channels for PUs, where  $\mathbb{Z}^+$  denotes the set of positive *integer numbers*. Denote the number of channels that a PU flow assembles as  $G$  and assume that the maximum number of channels a PU flow can utilize is  $H$ . According to the above configuration, it is obvious that  $1 \leq G \leq H \leq M$  holds, where  $G, H \in \mathbb{R}^+$ , and  $\mathbb{R}^+$  denotes the set of positive *real numbers*. If  $H = 1$  holds, only one channel is utilized for each PU flow. On the other hand, one PU flow may occupy multiple channels given  $H > 1$ .

SUs have the freedom to transmit a flow using a portion of one channel by utilizing CF, or in multiple channels by adopting CA. The assembled multiple channels can be either neighboring to or separated from each other in the spectrum domain. Denote by  $W$  and  $V \in \mathbb{R}^+$  as the minimum number and the maximum number of channels that a single SU flow can utilize respectively, and  $R \in \mathbb{R}^+$  is the number of channels that an SU flow assembles. Obviously,  $0 < W \leq R \leq V \leq M$  holds. Note that  $W$  and  $V$  in this study are considered as

*positive real numbers*, which is different from [20] where  $W$  and  $V$  are positive integer numbers.

Assume that SUs can detect PU activities with high enough spectrum sensing accuracy and the effect of sensing failure at the flow level is ignorable. In our analysis, it is assumed that there is a protocol running among the SUs to support CA, CF, and spectrum adaptation. We further assume that the SU flows are independent of each other. Once accepted, the flow level service will not be terminated due to channel variations with the help of advanced physical and MAC layer techniques, like link adaptation and adaptive modulation, hybrid automatic repeat request with soft combining, etc. Given the above assumptions, we regard the capacity obtained from the theoretical analyses as ideal compared with the results based on more realistic conditions.

#### B. PU Channel Access Strategy

It is assumed that PUs can adaptively adjust the number of channels that a flow assembles according to channel availability. More specifically, PUs operate in the following manner. The ongoing PU flows will 1) always utilize as many channels as they are allowed to; 2) always *equally* share<sup>3</sup> the available channels; and 3) always share their occupied channels with a new PU arrival. In more details, if one or more channels become idle due to a departure of a PU flow, the remaining ongoing PU flows will equally share the newly available channels, up to  $H$  channels that are assembled for each flow. Upon the arrival of a new PU flow, it will be allowed to commence if and only if the number of channels per PU flow is not lower than one after spectrum sharing. Note that if  $H = 1$  holds, CA is not supported.

#### C. SU Channel Access Strategy

Consider an extended full adaptation and full sharing strategy for SUs, denoted as EFAFS ( $W \leq R \leq V$ ). In this strategy, ongoing SU flows have the same properties 1) – 3) as discussed in the previous subsection for PUs. Moreover, whenever channels become idle due to a PU or SU flow completion or a forced SU flow termination, these channels will be equally shared by ongoing SU flows, up to  $V$  channels per flow. Upon an SU flow arrival, it will be commenced if the number of channels per SU flow is not lower than  $W$  after sharing with ongoing SU flows. However, due to their inferiority in the network, the ongoing SU flows will reduce the number of occupied channels and possibly continue their flows only when at least  $W$  channels per each ongoing SU flow are kept upon a PU arrival. In the extreme case, one or multiple ongoing SU flows may be forced to terminate if and only if the average number of assembled channels is lower than  $W$  for ongoing SU flows after a PU appearance.

There exist several techniques which allow equally sharing of a channel among flows. In the frequency domain, orthogonal frequency division multiple access (OFDMA) [29] can be

<sup>3</sup>Note that an ongoing PU flow must be able to assemble a non-integer number of channels. Therefore, CF is also allowed. For PU flows,  $G \geq 1$  always holds since we have defined the minimum spectrum requirement for a single PU flow as one channel.

TABLE I

TRANSITIONS FROM A GENERIC STATE  $\mathbf{x} = (i, j)$  FOR EFABS. IN THIS TABLE, PU AND SU REPRESENT A PU FLOW AND AN SU FLOW RESPECTIVELY.

Activity	Destination state	Transition rate	Conditions
PU arrives when no SU exists	$(i + 1, 0)$	$\lambda_P$	$i < M$ and $j = 0$
PU arrives when at least one SU exists. No SU forced termination happens	$(i + 1, j)$	$\lambda_P$	$M - (i + 1)H \geq jW$ , $j > 0$ , and $i < M$
PU arrives when at least one SU exists. SU forced termination happens	$(i + 1, \lfloor \frac{\max(0, M - (i + 1)H)}{W} \rfloor)$	$\lambda_P$	$M - (i + 1)H < jW$ , $i < M$ , and $j > 0$
PU departure	$(i - 1, j)$	$\min(M, iH)\mu_P$	$i > 0$
SU arrival	$(i, j + 1)$	$\lambda_S$	$M - iH \geq (j + 1)W$
SU departure	$(i, j - 1)$	$\min(M - iH, jV)\mu_S$	$j > 0$

employed by facilitating ongoing SU flows to share a channel, as illustrated in Fig. 1. Alternatively, multiple SU flows may also share a channel in the time domain, or in both time and frequency domains in a hybrid manner. However, in real-life, the granularity of channel sharing is constrained by hardware limitation. In other words, it might be impossible to achieve an absolutely equal channel share among all ongoing SU flows. In this work, for the clarity of our theoretical analysis, we assume that the difference of the amount of bandwidth allocated to various flows is negligible.

In OFDMA, different sub-channels may have different characteristics, resulting in distinct service rates. However, to evaluate the impact of different parameters for each sub-channel, the states of each sub-channel must be tracked, leading to a probably non-manageable size of state space in the Markov model when the number of sub-channels is substantially large. Furthermore, to avoid adjacent channel interference between neighboring users/flows, guard-bands need to be utilized. The requirements of guard-bands may be different due to various physical layer techniques employed for CA [30]. As elaborated in [8], [10], optimization on channel utilization can also be carried out when guard-bands are considered for CA. The variety of sub-channels and guard-bands may result in heterogeneity in channel parameters before or after CF and CA. Additionally, the transmission power may also vary based on different system configurations. To make our analysis tractable and comparable with previous studies, we assume that all channels are homogeneous with identical transmission power, and the impact of guard-band is ignored when CF and CA are employed. In other words, the average data rate for a flow is assumed to be linearly proportional to the number of its occupied channels in our model.

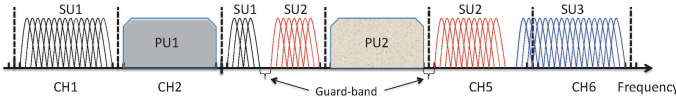


Fig. 1. Illustration of SUs that occupy non-integer number of channels without overlap with PUs by employing OFDMA.

Note that this strategy is similar to the one proposed in [19], [20], i.e., FAFS. The main differences are, however, that: 1) in EFABS,  $W, V \in \mathbb{R}^+$  whereas in FAFS,  $W, V \in \mathbb{Z}^+$ ; and 2) in EFABS, the scenario in which CA applies to PUs is investigated whereas this scenario was not considered in FAFS. In the following two sections, we present the mathematical analysis of the proposed strategy for elastic traffic and real-time traffic respectively.

#### IV. IMPACT OF CF AND CA ON ELASTIC SU FLOWS

Unless otherwise stated, PU and SU flows in this section refer to elastic flows. Following the common practice for Markov chain modeling, we assume that the arrivals of both SU and PU flows are Poisson processes with rates  $\lambda_S$  and  $\lambda_P$  respectively. Likewise, the service times are exponentially distributed with service rates  $\mu_S$  and  $\mu_P$  in *one* channel respectively. With homogeneous channels, the service rate of an SU flow with  $R$  channels is  $R\mu_S$ , and similarly  $G\mu_P$  for a PU flow with  $G$  channels.

There are four steps in our analysis. The precise analytical model for EFABS performance is derived in the first subsection. In the second subsection, the equivalence of EFABS and a strategy without CF, i.e., dynamic fully adjustable (DFA) [13], is demonstrated under certain conditions. The purpose of introducing DFA and comparing it with EFABS is to reveal the fact that CF cannot always improve system performance if the system parameters are not properly adjusted and configured. In the third subsection, we analyze the conditions that CF can enhance the system performance and derive further the theoretical capacity upper bound in the QSR. In the last subsection, we provide deeper insight into our mathematical analysis considering its limitations in real-life networks.

##### A. The Precise Model for EFABS with Elastic Traffic

Let  $i$  be the number of ongoing PU flows and  $j$  be the number of ongoing SU flows. The state in the Markov chain of EFABS is expressed as  $\mathbf{x} = (i, j)$ . The transitions from state  $(i, j)$  to other possible states are shown in Table I. We denote by  $b(\mathbf{x})$  the total number of occupied channels at state  $\mathbf{x}$ , given as  $b(\mathbf{x}) = iH + \min(M - iH, jV)$ . Given concrete values of  $M, H, W$  and  $V$ , the feasible states of the CTMC model can be obtained, as  $\mathcal{S} = \{(i, j) | iH + Vj < M\} \cup \{\mathbf{x} | b(\mathbf{x}) = M\}$ .

Based on the state transitions presented in Table I, transition rate matrix  $\mathbf{Q}$  can be established, with each element representing the transition rate from one state to another. To calculate the steady state probability,  $\pi(\mathbf{x})$ , the global balance equations and the normalization equation can be established, as  $\pi\mathbf{Q} = \mathbf{0}$ , and  $\sum_{\mathbf{x}} \pi(\mathbf{x}) = 1$ , where  $\pi$  is the steady state probability vector. When the steady state probability,  $\pi(\mathbf{x})$ , is obtained, the other system parameters can be calculated.

The capacity of the secondary network,  $\rho_e$ , is defined as the average number of SU flow completions per time unit [16]. It can be calculated by summing up the products of state probabilities and their corresponding SU service rates, as

$$\rho_e = \sum_{i: M-iH > 0} \left[ \frac{(M-iH)}{W} \right] \sum_{j=0}^{\lfloor \frac{(M-iH)}{W} \rfloor} \min(M-iH, Vj) \mu_S \pi(i, j). \quad (1)$$

The blocking probability,  $P_{be}$ , which refers to the probability that an SU flow arrival is blocked and lost due to insufficient resources, is given by

$$P_{be} = \sum_{\substack{\mathbf{x} \in \mathcal{S}, \\ M-iH < (j+1)W}} \pi(\mathbf{x}). \quad (2)$$

The forced termination probability,  $P_{fe}$ , which represents the fraction of the forced terminations over those commenced SU flows, is given by

$$P_{fe} = \frac{R_{fe}}{\lambda_S^*} = \frac{\lambda_P}{\lambda_S^*} \sum_{\substack{\mathbf{x} \in \mathcal{S}, j > 0, \\ M-(i+1)H < Wj}} \left( j - \left\lfloor \frac{\max(0, M-(i+1)H)}{W} \right\rfloor \right) \pi(\mathbf{x}), \quad (3)$$

where  $R_{fe}$  is the forced termination rate and  $\lambda_S^* = (1 - P_{be})\lambda_S$ . From the definition of  $\lambda_S^*$ , we observe that the blocked SU flows are not taken into account for calculating  $P_{fe}$ . Note that the mathematical model developed in [19], [20] becomes a special case of the above model as  $W, V \in \mathbb{Z}^+$ , and  $H = 1$  hold in [19], [20].

### B. Equivalence between EFAPS and DFA

In this part, we illustrate the equivalence between EFAPS ( $W \leq R \leq V$ ) and DFA ( $W' \leq R' \leq V'$ ) under certain conditions. These two strategies follow similar principles but are quite distinct, as EFAPS ( $W \leq R \leq V$ ) requires both CF and CA, but DFA ( $W' \leq R' \leq V'$ ) requires CA only. Notation DFA ( $W' \leq R' \leq V'$ ) refers to the dynamic fully adjustable strategy, where  $W', V'$  and  $R'$  have the same meanings as  $W, V$  and  $R$  in the EFAPS, but belong to the set of positive integers, i.e.,  $\mathbb{Z}^+$ . More detailed description on DFA and its mathematical modeling can be found in [13].

The purpose of illustrating the equivalence is to figure out the circumstances where CF cannot get further benefit if CA and spectrum adaptation have already been enabled in the system. The comprehension on this point will lead us to the conditions under which the system performance can be further improved by employing CF additionally. Note that when DFA was developed in [13], CA was not considered for PUs. Therefore, we consider  $H = 1$  in this subsection by default unless otherwise states.

*Proposition 1:* Given  $W = W'$  and  $V = V'$ , EFAPS ( $W \leq R \leq V$ ) is equivalent to DFA ( $W' \leq R' \leq V'$ ) with respect to capacity, blocking probability, and forced termination probability.

*Proof:* Refer to Appendix A.

Note that in the proof of *Proposition 1*, there is no constraint on the time scale of PU and SU activities. Therefore, the equivalence of these two strategies holds *in general* for any time scale rather than just in the QSR.

When  $W$  and  $V$  are integer numbers, EFAPS is actually the same as the strategies in [19]<sup>4</sup> and [20]. *Proposition 1* tells us that even though CF is enabled in [19], [20], i.e.,  $R \in \mathbb{R}^+$ , the network *cannot* achieve benefit from CF compared with a CA-only strategy. Whether  $W, V \in \mathbb{R}^+$  can yield further benefit or not will be examined in the following subsection.

Note that such an equivalence still holds for  $H > 1$  if the PU access strategies are identical for EFAPS and DFA and  $H \in \mathbb{Z}^+$  holds. The proof of equivalence is comparatively similar to the one for *Proposition 1* with a revised model representing CA of PUs. However, if  $H$  is a non-integer number, the equivalence does not hold because the DFA that can only utilize integer number of channels cannot utilize channel resources as much as EFAPS does.

### C. The Advantages of CF and its Limit

In this subsection, we will elucidate the circumstance that CF together with CA can lead to better system performance than CA-only does, and then reveal the maximum theoretical capacity when CA and CF are enabled in the QSR.

1) *When CF can improve system performance:* The key point for possible system performance improvement by CF is to *utilize the values of  $W$  and  $V$  that cannot be achieved without CF*. Both  $W$  and  $V$  may have impact on system performance if they are selected from a larger number set. For  $W$ , after the extension from  $\mathbb{Z}^+$  to  $\mathbb{R}^+$ , a larger number of SU flows can be potentially supported by the system. For example, given  $W < 1$ , meaning that an SU flow can be supported by a portion of a channel, capacity improvement can be observed (to be illustrated later in the numerical results) as more SU flows can be accommodated. For  $V$ , after extending it from  $\mathbb{Z}^+$  to  $\mathbb{R}^+$ , the maximum service rate for an SU flow is not constrained to integer numbers, resulting in higher channel utilization. For example, consider a system that has  $M = 4$  channels in total and can support up to 3.5 channels for a single SU flow. With the same value of  $W$  in EFAPS, it is straightforward that the system capacity will become higher if  $V = 3.5$  instead of restricting  $V$  to an integer, i.e.,  $V = 3$ , since more channels are utilized by a single SU flow. In what follows, we will elaborate a mathematical interpretation for the reason of the increased capacity in the QSR, given that  $W, V \in \mathbb{R}^+$ .

2) *Capacity analysis in the QSR with CA and CF:* In the QSR, the distribution of SU flows reaches equilibrium between two consecutive PU events. In other words, PU flows are relatively static compared with SU flows. Therefore, when  $i$  PU flows exist and  $M - iH > 0$  holds, there are  $M - iH$  channels that are in a sense *dedicated* to SU flows. In the QSR, the number of ongoing SU flows with  $M - iH$  dedicated channels can be modeled by a birth and death process (BDP), as shown in Fig. 2, where  $Q = M - iH$ ,  $I = \left\lfloor \frac{Q}{W} \right\rfloor$  and

<sup>4</sup>It is worth mentioning that in [19], the service rate of SU flows is a sum of two rates, which represents the minimum of two exponential random variables. In fact, the equivalence still holds if DFA has the same exponential distribution as in [19]. Its proof has the same principle as the proof of *Proposition 1*, only with an increased notational complexity in the SU service rate.

$C = \lfloor \frac{Q}{V} \rfloor$ . In this BDP, the service rate of the  $r$ th state is  $rV\mu_S$  when  $r \leq C$  and  $Q\mu_S$  when  $r > C$ .

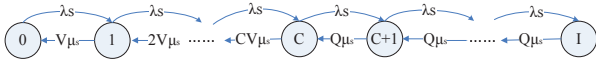


Fig. 2. The BDP for SU flows given  $M - iH$  dedicated channels.

According to the proof in [20], the capacity for such a network can be improved by *either increasing the service rate or enlarging the chain length of the BDP for SU flows*. Obviously, if  $W < 1$ , the value of  $I$  will increase compared with the case of  $W = 1$ . In other words, the chain length is increased with  $W < 1$  for any given  $i$ . By utilizing *Proposition 2* in [20], we can conclude that higher capacity is achieved with  $W < 1$  when both CF and CA are enabled<sup>5</sup>. Similarly, re-visit the second example mentioned in the second paragraph of this subsection. By utilizing *Proposition 1* in [20], it is obvious that the system capacity will be higher when  $V = 3.5$  compared with the case when  $V = 3$  since the service rates increase in the former case<sup>6</sup>. With CA-only,  $W$  and  $V$  cannot be a fractional value. Consequently, we understand mathematically the reasons why CF can result in higher capacity than the system enabling only CA, since the chain length can be extended and the service rate can be increased.

Knowing that CF together with CA can result in higher capacity in the QSR, the next question is how much capacity a system can achieve at most with both CF and CA.

3) *Capacity upper bound in the QSR with CA and CF*: For any strategy with CA, the capacity upper bound in the QSR equals to the capacity of FAFS given  $W = 1$  and  $V = M$  in the same regime [20]. To study the capacity upper bound with EFAFS, we need to check how much the service rate as well as the chain length can be further extended.

As mentioned earlier, the system service rate increases when  $V$  becomes larger. However, the service rate cannot be further improved through CF because  $V = M$  is already required for achieving the capacity upper bound with a CA-only strategy and  $V$  cannot be greater than  $M$ . On the other hand, the chain length can be extended with a reduced value of  $W$  by utilizing CF when  $W < 1$ . Ideally, if an SU flow can be accommodated by an infinitely small portion of a channel, i.e.,  $W \rightarrow 0$ , the capacity in the QSR will achieve its maximum possible value. In reality, however,  $W \rightarrow 0$  is unrealistic since each flow has its minimal QoS requirement and  $W$  cannot be arbitrarily small considering hardware limitation. More discussions on the granularity of  $W$  will be resumed later.

*Proposition 2*: With  $W, V \in \mathbb{R}^+$  and CA and CF enabled, the maximum theoretical achievable capacity for elastic SU flows in the QSR is given by

$$\rho_{qel} = \sum_{i: M-iH > 0} \pi(i) \min((M-iH)\mu_S, \lambda_S), \quad (4)$$

<sup>5</sup>Proposition 2 in [20] approves that the capacity of a BDP will increase if the chain length becomes longer with correspondingly non-decreasing service rates.

<sup>6</sup>Proposition 1 in [20] reveals the fact that the capacity of a BDP will increase if the service rate increases for any state.

where

$$\pi(0) = \left[ \sum_{i=0}^F \left( \frac{\lambda_P}{H\mu_P} \right)^i \frac{1}{i!} + \sum_{i=F+1}^M \left( \frac{\lambda_P}{M\mu_P} \right)^i \frac{M^F}{H^F F!} \right]^{-1}, \quad (5)$$

$$\pi(i) = \begin{cases} \left( \frac{\lambda_P}{H\mu_P} \right)^i \frac{\pi(0)}{i!}, & \forall 1 \leq i \leq F, \\ \left( \frac{\lambda_P}{M\mu_P} \right)^i \frac{M^F}{H^F} \frac{\pi(0)}{F!}, & \forall F < i \leq M, \end{cases} \quad (6)$$

and  $F = \lfloor \frac{M}{H} \rfloor$ .

*Proof*: Refer to Appendix B.

From (4), we observe that for any given  $M - iH > 0$  channels, the maximum capacity equals to  $\min((M - iH)\mu_S, \lambda_S)$ . The reason is as follows. With an infinitely long chain as well as the same service rate and arrival rate for each state, meaning that the network is saturated, the accomplished flows per time unit equal to the service rate if the arrival rate is greater than the total service rate. Otherwise, the number of flows that can be finished by the system per time unit equals to the offered load.

#### D. Further Discussions on the Derived Upper Bound

If the secondary network is saturated for any occupancy level of PUs (i.e.,  $\lambda_S \geq M\mu_S$ ), the maximum capacity in (4) becomes

$$\rho_{qel} = (M - E[N_{PU}])\mu_S, \quad (7)$$

where  $N_{PU}$  is a random variable representing the number of channels that PUs occupy in the system and  $E[N_{PU}]$  is its expected value. This result tells us that all the unused capacity of the primary network could be fully utilized, if there were no lower limit about the data rate that SUs' elastic flows require. This must be regarded as a theoretical limit that cannot be fulfilled in practice. By letting  $W$  go to zero and without applying admission control to elastic traffic (i.e., an unlimited number of elastic flows can be initiated), a minimum service rate per flow cannot be guaranteed as a consequence. Although elastic traffic has a certain level of tolerance on low bandwidth, flow abandonment due to an extremely low service rate may occur because of service timeout or human impatience. Therefore, a sharp drop of capacity may happen if  $W$  decreases below a certain threshold. Flow drop has negative impact on efficiency because certain amount of capacity is wasted by non-completed flows. The necessity of applying admission control for elastic traffic to prevent non-efficient utilization of resources has been explained in [31].

Considering system stability in reality, a minimum value of  $W$  should be enforced unless  $\lambda_S < \min\{M - iH : i \geq 0, M - iH > 0\}\mu_S$  is guaranteed. Note that  $\min\{M - iH : i \geq 0, M - iH > 0\}\mu_S$  is the lowest rate that the SUs in the system will receive. Although  $\lambda_S < (M - E[N_{PU}])\mu_S$  might result in a stable system as a whole, the arrival rate of elastic flows may be greater than its completion rate when the number of active PU flows surpasses  $E[N_{PU}]$ . Especially in the QSR, such a *local instability* situation, i.e., the period that the arrival rate is greater than the service rate, may last for an exceedingly long time [32], since we are assuming that the PU activities change at a much lower rate than for SUs. Therefore, admission control may still be necessary even though the system can be considered as stable in the long run.

## V. IMPACT OF CF AND CA ON REAL-TIME SU FLOWS

In this section, we analyze the impact of CF on real-time flows. Unless otherwise stated, SU flows in this section refer to real-time flows. Since the service duration of a real-time SU flow will not be affected by the number of assembled channels, applying CA for real-time flows cannot improve SU capacity. Correspondingly, the remaining alternative to enhance capacity is to accommodate a larger number of flows in the system, by utilizing a portion of a channel through CF.

### A. Precise and QSR Models for Real-time Traffic

Consider a strategy where each real-time SU flow occupies  $W \in \mathbb{R}^+$  channels. The value of  $W$  decides the number of SU flows that the system can accommodate for any given number of  $i$  PU flows, as  $\lfloor (M - iH)/W \rfloor$ . Denote the service rate of the SU real-time flows by  $\mu'_S$  and the arrival rate by  $\lambda'_S$ , while the arrival and service rates for PUs are the same as the ones discussed in the previous section.

The precise model of a CRNs with real-time SU flows has the same transitions as those illustrated in Table I except that we need to substitute  $\lambda_S$  and  $\min(M - Hi, jV)\mu_S$  by  $\lambda'_S$  and  $j\mu'_S$  respectively. Note that the service rate of a state is related to the number of ongoing SU flows in that state rather than the number of channels that those flows assemble. The capacity for real-time flows,  $\rho_r$ , is therefore given by

$$\rho_r = \sum_{i: M-iH>0} \sum_{j=0}^{\lfloor (M-iH)/W \rfloor} j\mu'_S \pi(i, j), \quad (8)$$

while the blocking probability and the forced termination probability have the same expression as in (2) and (3).

Let us further consider the capacity of real-time traffic in the QSR. In this regime, for any given  $M - iH > 0$ , the behavior of SU flows can be modeled by a BDP with arrival rate  $\lambda'_S$  and service rate  $m\mu'_S$  for a state with  $m$  flows, and the chain length is given by  $\lfloor (M - iH)/W \rfloor + 1$ . Similar to the analysis in [20], the capacity expression for real-time traffic in the QSR can be expressed by

$$\begin{aligned} \rho_{qr} &= \sum_{i: M-iH>0} \pi(i)(1 - \pi(I|i))\lambda'_S \\ &= \sum_{i: M-iH>0} \pi(i) \left[ 1 - \left( \sum_{j=0}^I \frac{I!}{j!} \left( \frac{\mu'_S}{\lambda'_S} \right)^{I-j} \right)^{-1} \right] \lambda'_S. \end{aligned} \quad (9)$$

Despite practical constraints on  $W$  as discussed earlier, it is of theoretical interest to study the case when  $W$  approaches to zero. As  $W$  becomes smaller, the bandwidth occupied by each SU flow reduces. With the same bandwidth provided by a channel, the number of SU flows that a channel could accommodate becomes larger as  $W$  decreases. If the real-time flow can be accommodated by an infinitely small fraction of a channel, a channel can actually support the offered traffic. Mathematically, as  $W$  diminishes,  $I$  becomes a larger number, and  $\left( \sum_{j=0}^I \frac{I!}{j!} \left( \frac{\mu'_S}{\lambda'_S} \right)^{I-j} \right)^{-1}$  in (9) approaches zero as  $I$  increases. Therefore, the capacity approaches  $\lambda'_S$  for any given  $M - iH > 0$  channels.

Note that for real-time traffic, reducing the value of  $W$  while keeping other parameters constant leads to lighter traffic load offered to the system, as the required data rate by each SU flows is reduced. In the next subsection, we study the case where  $W$  decreases but the total offered load is kept constant.

### B. The Advantage of CF and Its Limits

In order to study the effect of diminishing  $W$ , we let  $W$  approach zero while keeping the offered load constant. Let  $c$  denote the average data rate that can be obtained from one channel for real-time traffic. Then,  $R_o = \lambda'_S W c / \mu'_S$  represents the total offered load measured in bits per time unit. Thus, if the arrival rate is varied with  $W$  as  $\lambda'_S = \lambda_0 / W$ , where  $\lambda_0$  is a given constant, the offered load in bits per time unit becomes a constant, i.e.,  $R_o = \lambda_0 c / \mu'_S$ , for any value of  $W$ .

As mentioned above, considering real-time traffic that requires an arbitrarily small amount of resources is not physically sound. Our purpose in this subsection, however, is to study from a theoretical point of view the capacity for SUs when the bandwidth of one PU channel is very large compared with the bandwidth required by an SU flow, while CF is enabled.

In this case, the BDP that models real-time SUs in the QSR becomes an Erlang-B system. The capacity in bits per second when there are  $M - iH > 0$  channels (i.e., there are  $i$  PU flows) for SU flows can be expressed as

$$R_c(i) = \left( 1 - B \left( \left\lfloor \frac{M - iH}{W} \right\rfloor, \frac{\lambda_0 / \mu'_S}{W} \right) \right) R_o, \quad (10)$$

where  $B(x, a)$  represents the Erlang loss function for  $x$  trunks with traffic intensity  $a$ .

It is known that when  $x \rightarrow \infty$  [33],

$$B(x, \alpha x)^{-1} = \frac{\alpha}{\alpha - 1} - \frac{\alpha}{(\alpha - 1)^3} \frac{1}{x} + o\left(\frac{1}{x}\right), \quad (11)$$

$$B(x, x + \beta\sqrt{x})^{-1} = a_0(\beta)\sqrt{x} + a_1(\beta) + \frac{a_2(\beta)}{\sqrt{x}} + o\left(\frac{1}{\sqrt{x}}\right). \quad (12)$$

Having these results in mind and taking the limit  $W \rightarrow 0$  in (10), it yields

$$R_c(i) = \begin{cases} R_o & \text{if } \lambda_0 \leq (M - iH)\mu'_S, \\ (M - iH)c & \text{if } \lambda_0 > (M - iH)\mu'_S, \end{cases} \quad (13)$$

or equivalently,  $R_c(i) = \min(R_o, (M - iH)c)$ .

Now, we use the distribution of the number of PUs in the system,  $\pi(i)$ , to obtain the unconditioned value of the capacity for SUs as

$$R_c = \sum_{i: M-iH>0} \pi(i) R_c(i) = \sum_{i: M-iH>0} \pi(i) \min(R_o, (M - iH)c). \quad (14)$$

In particular, if  $\lambda_0 \geq M\mu'_S$ ,  $R_c = (M - E[N_{\text{PU}}])c$ , which has an analogous interpretation to that given for (7).

### C. Further Discussions

We have investigated the impact of CF and CA in ideal configurations. In the following, we discuss practical limitations that may affect the accuracy of our analytical models.



In reality, the bit rate achieved by a flow may not be linearly proportional to the number of channels that it occupies when guard-bands are considered. This aspect has different effects on real-time and elastic traffic in our analytical model. For real-time traffic, the service rate is determined by the number of flows rather than the bit rate of each flow, as long as the basic QoS requirement is met. Therefore, if the number of real-time flows is proportional to the number of channels, the non-linearity will not influence our analytical model. However, for elastic traffic, the flow service rate may be influenced by the heterogeneity of channels due to guard-bands. For example, in LTE, 10% of the spectrum is required as guard-band for a channel with bandwidth between 3 MHz and 20 MHz, and 23% for a channel with bandwidth 1.4 MHz [34]. In this example, the bit rate of multiple assembled channels is not proportional to the number of channels even if the same power and noise level apply to all channels. As a result, our model can be regarded as an ideal analysis without considering guard-bands.

Moreover, packet drop due to sensing failure or channel fading may affect flow transmissions. Different types of flows have various tolerance on packet drop. For example, a real-time voice conversation may tolerate up to 3% of packet loss [35]. Furthermore, retransmission schemes at the MAC layer may also help reducing packet drop. Therefore, at the flow level, packet loss may not necessarily cause a termination of a traffic flow. Indeed, a real-time traffic flow will not terminate due to packet loss although its QoS will degrade. For elastic traffic, however, the service time will be longer if packet retransmission happens, resulting a smaller  $\mu_S$ . On the other hand, if the percentage of packet loss is large enough to cause a flow termination, the termination probability will increase, and consequently, the capacity will decrease.

In reality, due to hardware constraints for instance the bandwidth of the bandpass filter, the span of the spectrum that an SU can assemble is limited. Therefore, the value of  $V$  cannot be arbitrarily large. For example, if two idle channels are too far away in the frequency domain, they may not be assembled for a single SU flow, resulting in decreased capacity for SUs. In real-life system where CA is adopted, e.g., LTE-A, up to 100 MHz can be utilized by a single user [12].

Another practical consideration is that the arrival rates and service rates can be time-variant. Correspondingly, one may decide when and how many channels would be assembled/fragmented for SU traffic to adapt to such dynamics on the fly. For this purpose, a measurement-based estimation on arrival/service rates can be performed so that the measured values can be fed into our analytical models in a timely fashion. Based on the output of our model, the system can adjust the number of channels to be assembled/fragmented accordingly.

Furthermore, it is worth mentioning that CF and CA are not the only approaches that can improve system performance. Other techniques, e.g., adding queues to SU flows, can also enhance the capacity. Anyhow, studying the impact of other techniques on capacity is beyond the scope of this paper.

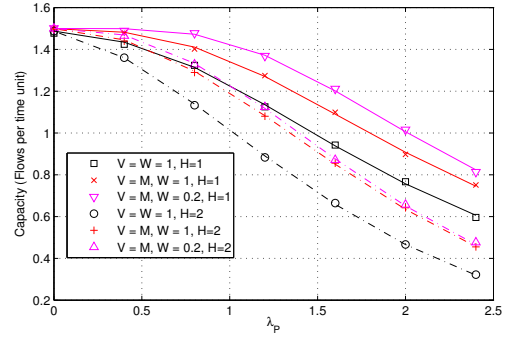


Fig. 3. Capacity as a function of  $\lambda_P$  for elastic traffic.

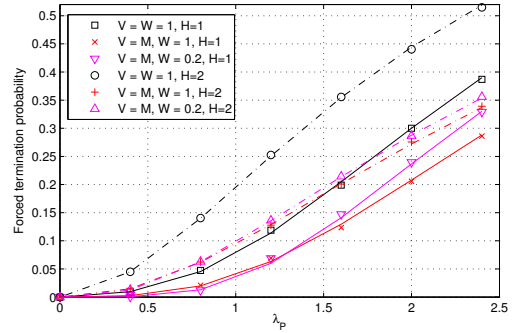


Fig. 4. Forced termination probability as a function of  $\lambda_P$  for elastic traffic.

## VI. NUMERICAL RESULTS AND DISCUSSIONS

To illustrate the impact of CF and CA on system performance, the numerical results based on our mathematical analyses and simulations are presented in this section for elastic traffic and real-time traffic consecutively. To validate the developed model, we also present simulation results with various traffic distributions based on real-life observations from [36], [37].

Discrete-event simulations are carried out following the structure proposed in [38] through Matlab. Each round of the simulation is performed for 20000 time units. We also simulate the system with the duration of 50% time units longer and the results are similar. Therefore, the duration of simulation is regarded as long enough. For an ongoing elastic flow with  $R$  channels assembled, its service rate in the simulation is  $R$  times higher than a flow with only one channel. On the contrary, for a real-time flow, its service rate is a constant, no matter how many channels it utilizes. During the simulation period, the events of both PUs and SUs are checked, processed, and tracked. At the end of simulation, the statistics of the system can be calculated. For example, the capacity is obtained based on the total number of successful SU services averaged by the total simulated time units.

In order to verify the mathematical models, the simulation results together with the analytical results are plotted in Figs. 3-5, 8-11, 14 and 15. From these curves, we can observe that the simulation results precisely coincide with the analytical results. Therefore, the correctness and the preciseness of the mathematical analysis are validated. For the ease of illustration clarity, only the numerical results from mathematic analyses

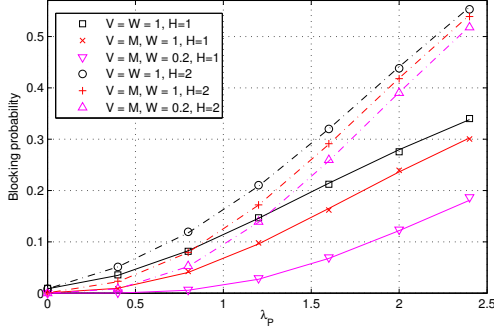


Fig. 5. Blocking probability as a function of  $\lambda_P$  for elastic traffic.

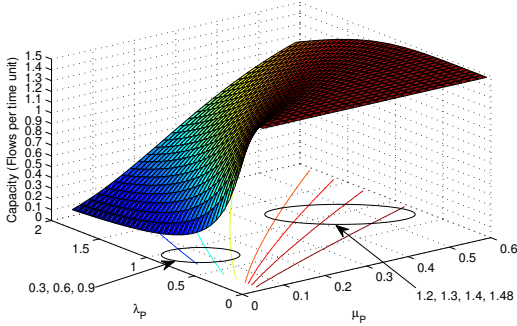


Fig. 6. Capacity as a function of  $\lambda_P$  and  $\mu_P$  for elastic traffic ( $V = M, W = 0.2, H = 1$ ).

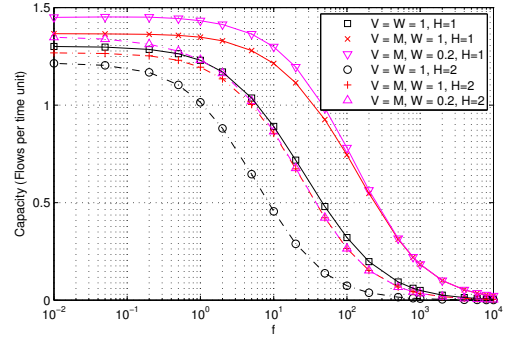


Fig. 7. Capacity as a function of  $f$  for elastic traffic.

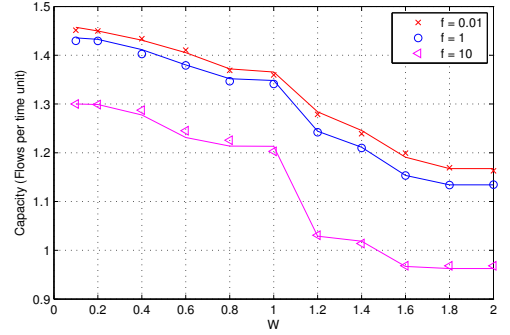


Fig. 8. Capacity as a function of  $W$  for elastic traffic.

are depicted in Figs. 6, 7, 12, and 13.

#### A. Performance Evaluation for Elastic Traffic

For elastic traffic, we investigate  $\rho_e$ ,  $P_{fe}$  and  $P_{be}$  as a function of  $\lambda_P$  with various values of  $W$  and  $V$ . The rest of parameters are configured as  $\lambda_S = 1.5$ ,  $\mu_S = 0.82$ ,  $\mu_P = 0.5$ , and  $M = 6$ . The rates of SU flows offer moderate traffic load. The value of  $\mu_P$  is selected as 0.5 because it is at the same order of the service rate of SU, meaning that the dynamics of PU and SU flows are comparatively similar. The corresponding results are shown in Figs. 3-5. Other cases when the dynamics of PUs vary are illustrated in Figs. 7 and 8 later. The curve with  $W = V = 1$  represents the strategy without CF and CA. The curve of  $W = 1, V = M$ , and  $H = 1$  also represents the results from the strategy proposed in [19], [20] where both CF and CA are utilized, but at least one channel is required for each flow. For  $W = 1, V = M$ , and  $H = 1$ , we have checked DFA [13] with the same configuration and find that these two curves fit exactly with one another. Therefore, Proposition 1 is validated.

As  $\lambda_P$  increases, indicating that more PU flows arrive per time unit, the capacity of SUs decreases while the blocking and the forced termination probabilities become higher. In Fig. 3, higher capacity is observed if  $V$  increases with the help of CA, given the same value of  $W$ . The capacity becomes even higher with a smaller  $W$  value with the help of CF in addition to CA. By selecting a portion of a channel for a flow, the capacity gain is obvious compared with [19], [20] where at least one channel is required. This is mainly because more SU flows

can be accommodated since the blocking probability becomes lower as  $W$  decreases, which can be observed from Fig. 5. The forced termination probability plotted in Fig. 4 grows higher for the case with  $W = 0.2$  than the case with  $W = 1$  when  $\lambda_P$  is relatively large. This is because more than one SU flow can be potentially preempted by a PU appearance in the former case. Therefore, when the arrival rate of PU becomes higher, the forced termination probability of the strategy with CF and CA increases faster compared with the strategy that allows CA only. However, if the arrival of PU is relatively low, the overall system performance is further improved if CF is adopted together with CA.

Comparing the group of curves with  $H = 1$  and  $H = 2$ , the SU performance with  $H = 2$  is worse than their counterparts in  $H = 1$ . It means that with the option of enabling CA for PUs, the system resource is utilized more efficiently by PUs, resulting in an overall performance degradation of SUs.

To illustrate the impact of the service rate together with the arrival rate of PUs, it is depicted in Fig. 6 the capacity as a function of both  $\lambda_P$  and  $\mu_P$  with contour lines. As can be observed, when  $\mu_P$  increases, indicating a shorter service time for PU flows on average, the SU capacity becomes higher. Note that the curve for  $V = M, W = 0.2, H = 1$  in Fig. 3 is integrated in the surface in Fig. 6 (given  $\mu_P = 0.5$ ).

In Fig. 7, we plot the capacity as a function of  $f$ , where  $\lambda_S = 1.5$ ,  $\mu_S = 0.82$ ,  $\lambda_P = f$ ,  $\mu_P = 0.5f$ , and  $M = 6$ . Parameter  $f$  is a scaler introduced to describe the dynamics of the system. The larger the value of  $f$  is, the farther the system is away from the QSR. Again, in the QSR, it is validated that the capacity of DFA is exactly the same as the one achieved

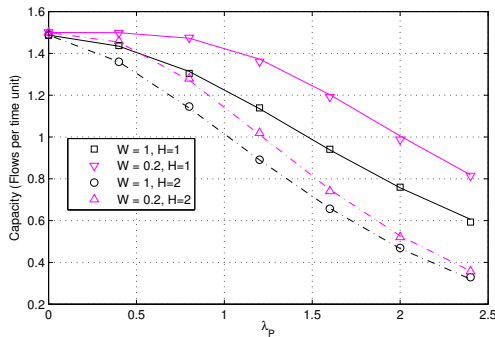


Fig. 9. Capacity as a function of  $\lambda_P$  for real-time traffic.

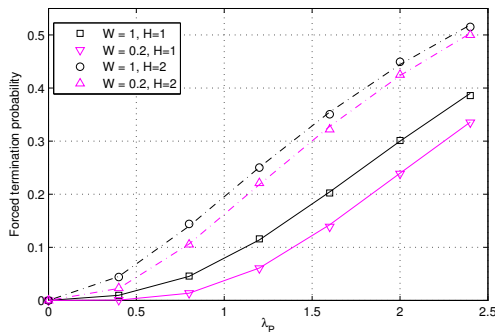


Fig. 10. Forced termination probability as a function of  $\lambda_P$  for real-time traffic.

in EFAS with the same configuration, i.e.,  $V = M$ ,  $W = 1$ , and  $H = 1$ . Furthermore, in the QSR, it is obvious in Fig. 7 that with a smaller value of  $W$ , higher capacity is achieved by EFAS. Therefore, the mathematical analysis in Sec. IV is validated.

Fig. 8 illustrates the capacity as a function of  $W$ , given  $H = 1$ ,  $V = M = 6$ ,  $\lambda_S = 1.5$ ,  $\mu_S = 0.82$ ,  $\lambda_P = f$ , and  $\mu_P = 0.5f$ . As  $W$  increases, the capacity has a decreasing trend because the number of simultaneous ongoing SU flows is reduced. Again, the capacity is higher as PUs become less active, which can be observed by comparing the curve with  $f = 0.01$  and  $f = 10$ . When  $W \rightarrow 0$ , the theoretical value of capacity given by (4) is 1.4572 in the QSR. Note that the curves in this figure are not smooth because different values of  $W$  involve various integer numbers of SU flows. For instance, there is an obvious drop when  $W$  increases from 1 to 1.2. The reason is that the maximum number of SU flows that the system can accommodate is reduced by one for all states given a specific number of PU flows, resulting in a higher blocking probability, and consequently, a descent in the capacity.

### B. Performance Evaluation for Real-time Traffic

For real-time flows, we investigate  $\rho_r$ ,  $P_{fr}$  and  $P_{br}$  as a function of  $\lambda_P$  with different  $W$  values, given  $\lambda'_S = 1.5$ ,  $\mu'_S = 0.82$ ,  $\mu_P = 0.5$ , and  $M = 6$ . The results are shown in Figs. 9-11. As can be observed from these figures, the capacity has a descending trend, and the forced termination and the blocking probabilities increase when PUs arrive more frequently. Similar to the elastic traffic case, the capacity becomes

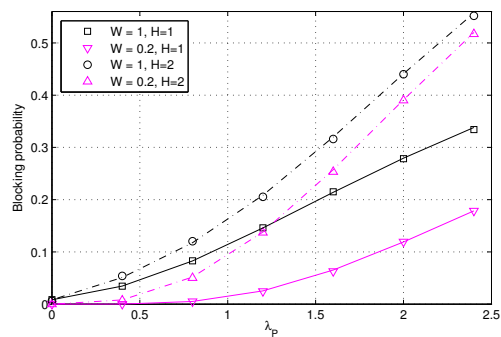


Fig. 11. Blocking probability as a function of  $\lambda_P$  for real-time traffic.

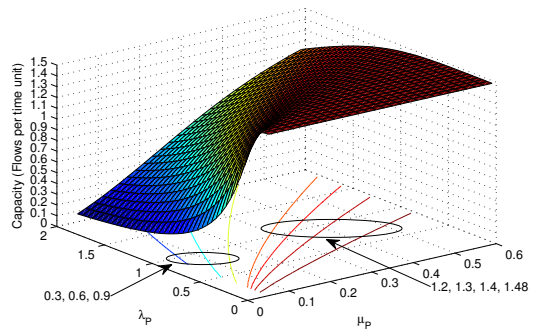


Fig. 12. Capacity as a function of  $\lambda_P$  and  $\mu_P$  for real-time traffic ( $W = 0.2$ ,  $H = 1$ ).

higher as  $W$  decreases. Note that the forced termination and the blocking probabilities decrease when  $W$  becomes smaller. In other words, the higher the number of SU flows in the system, the better the overall system performance. The reason for this result is that for real-time flows, we offer lighter traffic load to the system by reducing  $W$  if we keep all the other parameters constant. Fig. 12 depicts the capacity as a function of both  $\lambda_P$  and  $\mu_P$ . The results are similar to those observed in Fig. 6.

Similar to Fig. 7, we plot in Fig. 13 the capacity as a function of  $f$ , with  $\lambda'_S = 1.5$ ,  $\mu'_S = 0.82$ ,  $\lambda_P = f$ ,  $\mu_P = 0.5f$ , and  $M = 6$ . A similar capacity trend as the one for elastic traffic is observed for real-time traffic as a function of  $f$ . Again, it is clear that with a smaller value of  $W$ , higher capacity is achieved in the QSR due to the fact that a larger number of SU flows are accommodated.

Figs. 14 and 15 illustrate the capacity as a function of  $W$  for real-time traffic with a constant offered load. The system is configured as  $H = 1$ ,  $M = 6$ ,  $\lambda_o = 1.5$ ,  $\lambda'_S = \lambda_o/W$ ,  $\mu'_S = 0.82$ ,  $\lambda_P = f$ ,  $\mu_P = 0.5f$ , and  $c = 2$  Mbps. Therefore, the constant offered load is  $R_o = 3.6585$  Mbps. Fig. 14 demonstrates the capacity in terms of SU flows per time unit, i.e.,  $\rho_r$ , while Fig. 15 depicts the capacity in terms of Mbps. The capacity in Mbps is calculated by  $\rho_r W c / \mu'_S$ . From both figures, we can observe that the capacity is generally higher when PU activities become static, i.e., as  $f$  decreases. As can be observed from Fig. 14, the number of completed SU flows in each time unit increases dramatically as  $W$  decreases, indicating that more flows can be served by the system when

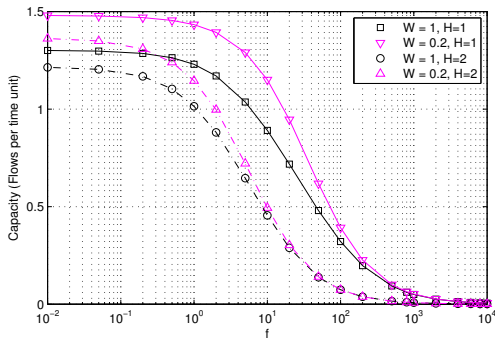


Fig. 13. Capacity as a function of  $f$  for real-time traffic.

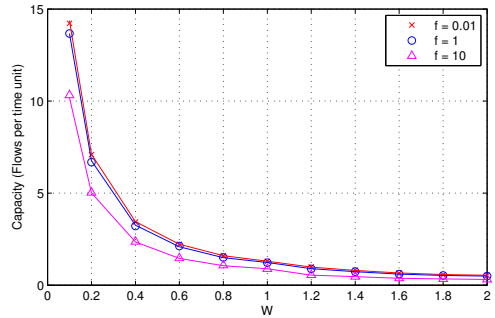


Fig. 14. Capacity as a function of  $W$  for real-time traffic.

$W$  is smaller with the same intensity of injected traffic load. A similar trend can also be observed from Fig. 15. It is interesting that the case with  $W = 0.8$  has a larger value of  $\rho_r$ , i.e., the number of SU flow completions per time unit, but lower capacity in Mbps compared with the case with  $W = 1$ . This result can be interpreted as follows. Although  $W = 0.8$  can result in a larger average number of flows, the data rate in Mbps for each flow reduces correspondingly (because a portion of a channel is used). Note that the capacity in Mbps is a product of the number of flows and the data rate for each flow. Therefore, if the increased number of flows cannot compensate the loss of data rate in each flow as a product, a sudden jump in capacity happens. Note that the capacity upper bound in Mbps in the QSR is  $R_c = 3.5542$  Mbps according to (14), and the curve with  $f = 0.01$  is approaching this value as  $W$  decreases.

### C. Performance for Traffic Patterns with Various Distributions

The traffic patterns in reality may not follow Poisson arrival process and exponential distribution. This observation motivates us to further investigate the applicability and the preciseness of our Markov chain analyses for other distributions. The results presented in this subsection is obtained from computer simulation with distributions based on empirical traffic observations [36], [37].

We consider three cases with various traffic patterns according to [36], [37]. In the first two cases, the arrival processes of PU and SU flows still follow the Poisson process, but the service time is modeled as a log-normal random variable. In case one, both the mean values and the variances of the log-

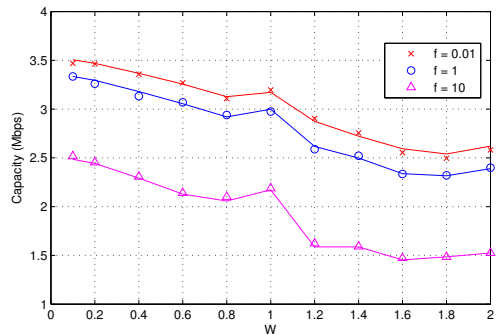


Fig. 15. Capacity as a function of  $W$  for real-time traffic in Mbps.

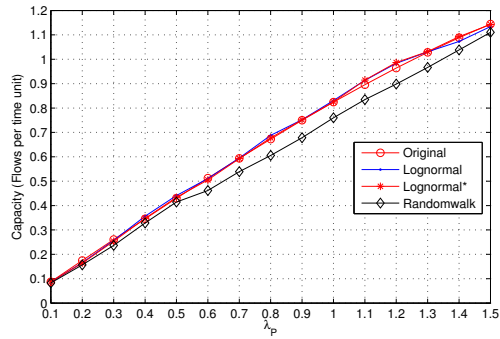


Fig. 16. Capacity as a function of  $\lambda_S$  when various distributions are utilized.

normal distributions equal to those of the corresponding exponential distributions, labeled as Lognormal. In case two, the variance values of the log-normal distributions are greater than those of the exponential distributions while the mean values are kept the same, labeled as Lognormal\*. More specifically, we adopt a squared coefficient of variation  $SCV=4.618$  [37] ( $SCV = \text{variance}/\text{mean}^2$ ). The reason for adopting larger variance values in Lognormal\* is that high variability has been observed in flow sizes in modern data networks [37]. In case three, a random walk model for PU activities [36] is adopted while the patterns for the SU flows are the same as in case one, labeled as Randomwalk. The results from the original distributions adopted in our mathematical models are also depicted, labeled as Original.

Fig. 16 illustrates the capacity of EFAPS configured as  $W = 0.2$ ,  $M = V = 6$  with elastic SU traffic as a function of  $\lambda_S$ . From this figure, we can observe that the results of two log-normal types are quite close to the analytical one. This result confirms the fact that the analytical results are not sensitive to the type of distribution for service time as long as Poisson arrival is held. Furthermore, the results under the random walk model exhibit similar results to a large extent. The simulation results with other parameter configurations, although not plotted here explicitly, have also similar properties. These observations demonstrate that the Markov chain analysis can be adopted as a robust reference model for analyzing the performance of CA and CF in CRNs.

In more details, Fig. 16 is obtained given that  $\lambda_P = 0.5$ ,  $\mu_P = 0.15601$ , and  $\mu_S = 0.5$  in the original distributions for PU and SU flows. The configuration of PUs gives three

PU-occupied channels on average, which is equal to the average number of PU-occupied channels in the random walk model [36]. Then the average time interval between two PU events in the random walk model is computed as  $1/(2\mu_P E) = 1.0683$ , where  $E = 3$  is the average number of PU-occupied channels.

## VII. CONCLUSIONS

In this work, we investigate the impact of CF technique thoroughly when CA and spectrum adaptation are already enabled for SU flows. Furthermore, the performance of SUs when CA is adopted by PUs is also studied. Based on the mathematical analyses, we reveal that to improve system performance for elastic flows through CF in addition to CA, it is essential to admit a flow with the smallest-possible portion of a channel, given that the QoS requirement is satisfied. For real-time traffic, potential performance improvement can also be expected by CF as CF can increase the total number of SU flows. Furthermore, in the QSR, we derive the theoretical capacity upper bound for both elastic and real-time traffic types in closed-forms when CF and CA techniques are utilized. The results from our analytical model apply also to various kinds of traffic distributions.

### APPENDIX A PROOF OF PROPOSITION 1

*Statement:* The procedure of this proof is similar to what is used for proving Proposition 1 in [14]. However, the proof herein is different from the one used in [14] with respect to two aspects: 1) the following proof has no constraint on the time scale of PU activities while the proof in [14] applies only to the QSR; and 2) the following proof establishes the equivalence between DFA and EFAPS when  $W = W'$  and  $V = V'$  only. On the other hand, the proof in [14] establishes the equivalence among various strategies in the QSR as long as the two features mentioned in the paragraph after Proposition 1 in [14] are met. Explicitly, these two features are: 1) ongoing SU flows will always occupy as many idle channels as they are able to; and 2) if there are fewer than  $W$  idle channels upon an SU arrival, the ongoing SU flows will donate their occupied channels to the newcomer, as long as each SU flow has at least  $W$  channels after donation.

*Proof:* The state of DFA is expressed by  $\mathbf{x}' = (i, j_{W'}, \dots, j_{V'})$ , where  $i$  is the number of PU flows, and  $j_k$  is the number of SU flows that assemble  $k = W', W' + 1, \dots, V'$  channels [13]. In order to illustrate the equivalence, we transfer the CTMC of DFA to a CTMC which is exactly the same as the one based on the EFAPS strategy.

To illustrate the equivalence of those two processes, we establish a one-to-one correspondence from a multiple dimension space,  $(i, j_{W'}, \dots, j_{V'})$ , to a three dimension space,  $(i, r, l)$ , for DFA. Consider all states with  $i$  PU flows and  $r$  SU flows in DFA, where  $r = \sum_{k=W'}^{V'} j_k$ . Let  $l$  represent the index of a particular state among all the states with  $i$  PU flows and  $r$  SU flows. Let  $L(i, r)$  be the number of states that have  $r$  SU flows given  $i$  PU flows. Therefore, we have  $l \in \{1, \dots, L(i, r)\}$ . Thus, using state  $(i, r, l)$  to represent a

specific state  $(i, j_{W'}, \dots, j_{V'})$ , we can establish one-to-one mapping from  $(i, j_{W'}, \dots, j_{V'})$  to  $(i, r, l)$  in DFA.

Let  $\pi'(i, r, l)$  be the steady probability of the  $l$ th state which has  $i$  PU flows and  $r$  SU flows, and  $\xi(i, r, l)$  be an indicator of the existence of the state  $(i, r, l)$ . If state  $(i, j_{W'}, \dots, j_{V'})$  which is represented by state  $(i, r, l)$  in the new notation exists,  $\xi(i, r, l) = 1$ . Otherwise,  $\xi(i, r, l) = 0$ . Similarly, we introduce another indicator for the event of forced termination,  $\phi(i, r)$ , where  $\phi(i, r) = 1$  if  $i + rW' = M$  holds, and  $\phi(i, r) = 0$  otherwise.

For a general state  $\mathbf{x}'$ , i.e., the  $l$ th state with  $r$  SU flows and  $i$  PU flows, the balance equation can be expressed as shown in (15). The notations used in (15) are explained as follows.  $m_1$  to  $m_5$  represent the state indices of the destination states upon different events transferred from state  $(i, r, l)$ . In more details,  $m_1$  denotes the index of the destination state within the state set with  $i$  PU flows and  $r + 1$  SU flows, and state  $(i, r, l)$  is transferred to  $(i, r + 1, m_1)$  upon an SU flow arrival.  $m_2$  denotes the index of the destination state within the state set with  $i$  PU flows and  $r - 1$  SU flows upon an SU flow departure. Similarly,  $m_3$  denotes the index of the destination state within the state set with  $i - 1$  PU flows and  $r$  SU flows upon a PU flow departure. When a PU flow arrives, an ongoing SU flow will either reduce the number of assembled channels and continue its flow as long as the lower bound is met or become terminated. Which event will happen depends on the current state of the system. Correspondingly, in (15), there are two potential destinations for a general state  $(i, r, l)$  upon a PU arrival, i.e.,  $(i + 1, r, m_4)$  or  $(i + 1, r - 1, m_5)$ , representing a destination state with or without forced termination. Note that according to DFA, these two events will not happen simultaneously.

Similarly, at the bottom half of (15),  $n_1$ ,  $n_2$ , and  $n_3$  represent the state indexes of source states from which  $(i, r, l)$  is transferred due to an SU flow arrival, an SU flow departure, and a PU flow departure respectively.  $n_4$  and  $n_5$  represent the source state indexes when a PU flow arrives with and without forced termination respectively.

If we sum up all these equations of states with  $r$  SU flows which have the same  $i$ , the left-hand side of (15) can be expressed as a common factor of outgoing rates by the sum of the probabilities of the states with  $r$  SU flows given  $i$  PU flows. Similarly, on the right-hand side, the transitions representing SU arrivals, SU departures, PU arrivals, and PU departures share the common factor  $\lambda_S$ ,  $rV'\mu_S$ ,  $\lambda_P$ , and  $i\mu_P$  respectively. For example, considering the states when  $0 < i < M$  and  $0 < r < \lfloor \frac{M-i}{W'} \rfloor$ , we have (16).

Similar expressions apply to the states with other values of  $i$  and  $r$  in the system, and the only difference is that some of the state transitions may not be feasible due to the non-existence of source or destination states.

Note that if we consider the sum of the state probabilities for all the states with the same number of PU and SU flows as a whole, the balance equation expressions in DFA are exactly the same as the balance equations according to the transitions described in Table I, given that  $W = W'$  and  $V = V'$ . Furthermore, the calculation of  $\rho_e$ ,  $P_{fe}$ , and  $P_{be}$  does not need any modification if we consider the sum of the probabilities of

$$\begin{aligned}
& [\lambda_S \xi(i, r+1, m_1) + \min\{rV', M-i\} \mu_S \xi(i, r-1, m_2) + i \mu_P \xi(i-1, r, m_3) + \\
& \lambda_P (\xi(i+1, r-1, m_4) \phi(i, r) + \xi(i+1, r, m_5) (1 - \phi(i, r)))] \pi'(i, r, l) \xi(i, r, l) \\
& = \lambda_S \pi'(i, r-1, n_1) \xi(i, r-1, n_1) + \min\{(r+1)V', M-i\} \mu_S \pi'(i, r+1, n_2) \xi(i, r+1, n_2) \\
& + (i+1) \mu_P \pi'(i+1, r, n_3) \xi(i+1, r, n_3) \\
& + \lambda_P \pi'(i-1, r+1, n_4) \xi(i-1, r+1, n_4) \phi(i-1, r+1) + \lambda_P \pi'(i-1, r, n_5) \xi(i-1, r, n_5),
\end{aligned} \tag{15}$$

$$\begin{aligned}
& [\lambda_S + \lambda_P + \min\{rV', M-i\} \mu_S + i \mu_P] \sum_{l=1}^{L(i,r)} \pi'(i, r, l) = \lambda_S \sum_{n_1=1}^{L(i,r-1)} \pi'(i, r-1, n_1) + \lambda_P \sum_{n_3=1}^{L(i-1,r)} \pi'(i-1, r, n_3) \\
& + \min\{(r+1)V', M-i\} \mu_S \sum_{n_2=1}^{L(i,r+1)} \pi'(i, r+1, n_2) + (i+1) \mu_P \sum_{n_4=1}^{L(i+1,r)} \pi'(i+1, r, n_4).
\end{aligned} \tag{16}$$

states as a whole in DFA. Therefore, we conclude that these two strategies are equivalent. ■

## APPENDIX B PROOF OF PROPOSITION 2

*Proof:* Consider a CRN with  $i$  PU flows where  $M - iH \geq 0$  holds.

When  $V = M$ , the BDP that models SUs in the QSR is that of an  $M/M/1/(1+I)$  system. For any given  $i$  PU flows, the capacity for SUs can be expressed as

$$\rho' = (1 - \pi_0)(M - iH)\mu_S. \tag{17}$$

Let  $r_i = \frac{\lambda_S}{(M - iH)\mu_S}$ . It is straightforward to obtain that

$$\pi_0 = \begin{cases} \frac{1-r_i}{1-r_i^{I+1}} & \text{if } r_i \neq 1, \\ \frac{1}{I+1} & \text{if } r_i = 1. \end{cases} \tag{18}$$

Since  $I = \left\lfloor \frac{M - iH}{W} \right\rfloor$ ,  $W \rightarrow 0$  is equivalent to  $I \rightarrow \infty$ . By taking the limit in (18), it leads to

$$\lim_{I \rightarrow \infty} \pi_0 = \begin{cases} 1 - r_i & \text{if } r_i < 1, \\ 0 & \text{if } r_i \geq 1. \end{cases} \tag{19}$$

Finally, substituting (19) into (17) yields

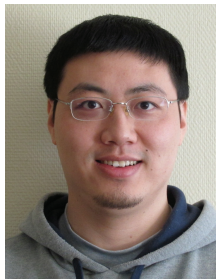
$$\lim_{I \rightarrow \infty} \rho' = \begin{cases} \lambda_S & \text{if } r_i < 1, \\ (M - iH)\mu_S & \text{if } r_i \geq 1, \end{cases} \tag{20}$$

or equivalently,  $\rho' = \min(\lambda_S, (M - iH)\mu_S)$ . Considering the state probabilities of PU flows, (4) holds. ■

## REFERENCES

- [1] Y. C. Liang, K. C. Chen, G. Y. Li, and P. Mahönen, "Cognitive radio networking and communications: An overview," *IEEE Trans. Veh. Technol.*, vol. 60, no. 7, pp. 3386–3407, Sep. 2011.
- [2] H. Su and X. Zhang, "Cross-layer based opportunistic MAC protocols for QoS provisionings over cognitive radio mobile wireless networks," *IEEE J. Select. Areas Commun.*, vol. 26, no. 1, pp. 118–129, Jan. 2008.
- [3] J. Jia, Q. Zhang, and X. Shen, "HC-MAC: A hardware-constrained cognitive MAC for efficient spectrum management," *IEEE J. Select. Areas Commun.*, vol. 26, no. 1, pp. 106–117, Jan. 2008.
- [4] H. A. B. Salameh, M. Krunz, and D. Manzi, "Spectrum bonding and aggregation with guard-band awareness in cognitive radio networks," *IEEE Trans. Mob. Comput.*, vol. 13, no. 3, pp. 569–581, Mar. 2014.
- [5] *IEEE Standard for Wireless Regional Area Networks Part 22: Cognitive Wireless RAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Policies and Procedures for Operation in the TV Bands*, IEEE Standard, Jul. 2011.
- [6] K. Tan, J. Fang, Y. Zhang, S. Chen, L. Shi, J. Zhang, and Y. Zhang, "Fine-grained channel access in wireless LAN," in *Proc. ACM SIGCOMM*, New Delhi, India, Aug. 2010.
- [7] M. Y. Arslan, K. Pelechrinis, I. Broustis, S. V. Krishnamurthy, S. Adepalli, and K. Papagiannaki, "Auto-configuration of IEEE 802.11n WLANs," in *Proc. ACM CoNEXT*, Philadelphia, PA, USA, Nov. 2010.
- [8] L. Deek, E. Garcia-Villegas, E. Belding, S. Lee, and K. Almeroth, "Intelligent channel bonding in 802.11n WLANs," *IEEE Trans. Mob. Comput.*, earlier access available in IEEEXplore, Dec. 2013, DOI: 10.1109/TMC.2013.73.
- [9] M. Park, "IEEE 802.11ac: Dynamic bandwidth channel access," in *Proc. IEEE ICC*, Tokyo, Japan, Jun. 2011.
- [10] G. S. Uyanik, M. A. Rahman, and M. Krunz, "Optimal guard-band-aware channel assignment with bonding and aggregation in multi-channel systems," in *Proc. IEEE GLOBECOM*, Atlanta, GA, USA, Dec. 2013.
- [11] Y. Xiao, C. Yuen, P. D. Francesco, and L. A. DaSilva, "Dynamic spectrum scheduling for carrier aggregation: A game theoretic approach," in *Proc. IEEE ICC*, Budapest, Hungary, Jun. 2013.
- [12] S. Parkvall, A. Furuskär, and E. Dahlman, "Evolution of LTE toward IMT-Advanced," *IEEE Communications Magazine*, vol. 49, no. 2, pp. 84–91, Feb. 2011.
- [13] 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.
- [14] —, "Modeling and performance analysis of channel assembling in multi-channel cognitive radio networks with spectrum adaptation," *IEEE Trans. Veh. Technol.*, vol. 61, no. 6, pp. 2686–2697, Jul. 2012.
- [15] I. A. M. Balapuwaduge, L. Jiao, V. Pla, and F. Y. Li, "Channel assembling with priority-based queues in cognitive radio networks: Strategies and performance evaluation," *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 630–645, Feb. 2014.
- [16] X. R. Zhu, L. F. Shen, and T. S. P. Yum, "Analysis of cognitive radio spectrum access with optimal channel reservation," *IEEE Commun. Lett.*, vol. 11, no. 4, pp. 304–306, Apr. 2007.
- [17] J. Martinez-Bauset, V. Pla, and D. Pacheco-Paramo, "Comments on 'analysis of cognitive radio spectrum access with optimal channel reservation'," *IEEE Commun. Lett.*, vol. 13, no. 10, pp. 739, Oct. 2009.
- [18] E. W. M. Wong and C. H. Foh, "Analysis of cognitive radio spectrum access with finite user population," *IEEE Commun. Lett.*, vol. 13, no. 5, pp. 294–296, May 2009.
- [19] L. Li, S. Zhang, K. Wang, and W. Zhou, "Combined channel aggregation and fragmentation strategy in cognitive radio networks," in *arXiv:1203.4913v2*, Jun. 2012.
- [20] L. Jiao, E. Song, V. Pla and F. Y. Li, "Capacity upper bound of channel assembling in cognitive radio networks with quasistationary primary user activities," *IEEE Trans. Veh. Technol.*, vol. 62, no. 4, pp. 1849–1855, May 2013.

- [21] C. Politis, "Managing the radio spectrum: the role of cognitive radio in future spectrum use," *IEEE Vehicular Technology Magazine*, vol. 4, no. 1, pp. 20–26, Mar. 2009.
- [22] L. Yu, T. Jiang, P. Guo, Y. Cao, D. Qu, and P. Gao, "Improving achievable traffic load of secondary users under GoS constraints in cognitive wireless networks," in *Proc. IEEE GLOBECOM*, Houston, TX, USA, Dec. 2011.
- [23] J. Lee and J. So, "Channel aggregation schemes for cognitive radio networks," *IEICE Trans. Commun.*, vol. E95-B, no. 5, pp. 1802–1809, May 2012.
- [24] T. M. N. Ngatched, S. Dong, A. S. Alfa, and J. Cai, "Performance analysis of cognitive radio networks with channel assembling and imperfect sensing," in *Proc. IEEE ICC*, Ottawa, Canada, Jun. 2012.
- [25] S. M. Kannappa and M. Saquid, "Performance analysis of a cognitive network with dynamic spectrum assignment to secondary users," in *Proc. IEEE ICC*, Cape Town, South Africa, May 2010.
- [26] T. Jiang, H. Wang, and S. Leng, "Channel allocation and reallocation for cognitive radio networks," *Wirel. Commun., Mob. Comput.*, vol.13, 1073–1081, 2013.
- [27] J. Martinez-Bauset, A. Popescu, V. Pla, and A. Popescu, "Cognitive radio networks with elastic traffic," in *Proc. 8th Euro-NF Conference on Next Generation Internet (NGI)*, Karlskrona, Sweden, Jun. 2012.
- [28] S. Joshi, P. Pawelczak, D. Čabrić, and J. Villaseñor, "When channel bonding is beneficial for opportunistic spectrum access networks," *IEEE Trans. Wireless Commun.*, vol. 11, no. 11, pp. 3942–3956, Nov. 2012.
- [29] D. Wang, H. Minn, and N. Al-Dhahir, "A distributed opportunistic access scheme and its application to OFDMA systems," *IEEE Trans. Commun.*, vol. 57, no. 3, pp. 738–746, Mar. 2009.
- [30] C. Cordeiro and M. Ghosh, "Channel bonding versus channel aggregation," *IEEE P802.22 WG*, available online at: <https://mentor.ieee.org/802.22/dcn/06/2022-06-0108-00-0000-bondingvs-aggregation.ppt>, Jul. 2006.
- [31] T. Bonald and J. W. Roberts, "Congestion at flow level and the impact of user behaviour," *Computer Networks*, vol. 42, no. 4, pp. 521–536, Jul. 2003.
- [32] F. Delcoigne, A. Proutière, and G. Régnié, "Modeling integration of streaming and data traffic," *Performance Evaluation*, vol. 55, no. 3–4, pp. 185–209, Feb. 2004.
- [33] D. L. Jagerman, "Some properties of the Erlang loss function," *Bell System Tech. J.*, vol. 53, no. 3, pp. 525–552, Mar. 1974.
- [34] H. Holma and A. Toskala, *LTE for UMTS: Evolution to LTE-advanced*. Wiley, 2011.
- [35] ITU-T, *End-user multimedia QoS categories*, ITU-T Recommendation G.1010, Nov. 2001.
- [36] D. Willkomm, S. Machiraju, J. Bolot, and A. Wolisz, "Primary users behavior in cellular networks and implications for dynamic spectrum access," *IEEE Commun. Mag.*, vol. 47, no. 3, pp. 88–95, Mar. 2009.
- [37] P. Barford and M. Crovella, "Generating representative web workloads for network and server performance evaluation," in *Proc. ACM SIGMETRICS*, Madison, WI, USA, Jul. 1998.
- [38] W. J. Stewart, *Probability, Markov chains, queues and simulation: The mathematical basis of performance modeling*. Princeton, NJ: Princeton University Press, 2009.



**Lei Jiao** (S'08-M'12) received his B.E. and M.E. 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.



**Indika A. M. Balapuwaduge** (S'12) received his B.Sc. Eng. degree with 1<sup>st</sup> 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.



**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 3G/4G and beyond mobile systems and wireless networks, mesh and ad hoc networks; wireless sensor network; cooperative communications; 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.



**Vicent Pla** received the M.E. and Ph.D. degrees in Telecommunication Engineering from the Universitat Politècnica de València (UPV), Spain. In December 1999, he joined the Department of Communications at the UPV, where he is currently an Associate Professor. His research interests lie primarily in the areas of teletraffic and performance evaluation of communication networks. During the past few years, most of his research activity has focused on resource management in wireless networks. In these areas he has published numerous papers in refereed journals and conference proceedings. He has also been an active participant in several national and European research projects.