

Guaranteed Service and Delay-Constrained Capacity of a Multi-Channel Cognitive Secondary Network

Yuehong Gao^{†‡}, Wenting Jiang[†] and Yuming Jiang[‡]

[†]Wireless Theories and Technologies Lab,

Beijing University of Posts and Telecommunications, Beijing China.

[‡]Centre for Quantifiable Quality of Service in Communication Systems,

Norwegian University of Science and Technology, Trondheim Norway.

Email: yhgao@bupt.edu.cn, jwt616@bupt.edu.cn, ymjia@ieee.org

Abstract—In this paper, we consider a multi-channel cognitive radio system serving a primary network and a secondary network, and analyze the quality of service and delay-constrained capacity of the secondary network. Specifically, by assuming that a certain amount of resource is exclusively reserved and used on each channel by the primary network, we derive the traffic transportation capacity that is guaranteed to the secondary network. Based on this, we analyze the traffic delay distribution in the secondary network and derive an upper bound on it, which allows us to further obtain a guaranteed capacity of the secondary network in serving traffic with probabilistic delay requirement. Both numerical and simulation results are presented for an example where the secondary network traffic follows a model taken from 3GPP LTE. The delay distribution, average delay and delay-constrained capacity of the secondary network are compared. The excellent match between analytical results and simulation results validates the theoretical analysis.

I. INTRODUCTION

Cognitive radio is a promising technique for efficiently making use of wireless spectrum [1]. Its fundamental idea is to allow a secondary network to coexist with the primary network in the system, and the secondary network can also access the system resource (or wireless channels) as far as the performance of the primary network is not affected. In this paper, we consider such a cognitive radio system with focus on the quality of service (QoS) performance of the secondary network. Specifically, we analyze the traffic transportation capacity that can be guaranteed to the secondary network, investigate its delay performance, and obtain its capacity in serving traffic with delay requirement.

In the literature, several attempts have been made to conduct performance analysis of the secondary network. Some of them make use of classical queueing theory [2][3]. In order to directly apply existing results, typically $M/G/1$ Priority analysis, Poisson arrival and single channel scenario are assumed with average delay and average queue length as the performance metrics of interest. In some other works, e.g. [4][5], a classic stochastic process analysis technique is used, which establishes its basis on the states of each channel occupancy, i.e. whether a channel is occupied by which network, and uses a Markov chain to model this process. With such Markov chains, the dropping probability and blocking probability are derived. Although multi-channel is considered in these works, it is often assumed that the arrivals (to each channel) form a Poisson

process and the service time (of each channel occupancy) follows some negative exponential distribution in order to ensure the Markov property of the channel occupancy process. In addition, some results on outage/ergodic capacity are available, e.g. [6][7], under various constraints that include power constraints and peak interference power constraints. However, study on the maximum arrival rate under probabilistic delay-constraint (defined as guaranteed delay-constrained capacity in this paper) is very limited.

Furthermore, the problem becomes even challenging when the cognitive radio network is supposed to have multiple parallel channels. For analyzing multi-channel/multi-server systems, another novel approach has been adopted, which lays on the *network calculus* theory [8][9]. For example, in [10][11], service guarantee analysis of multi-server Weighted Fair Queueing and multi-server Round Robin scheduling systems have been respectively studied. However, the considered multi-server scenarios therein do not encompass the priority issue, which is typically inherent in cognitive radio systems. To the best of our knowledge, an analysis of multi-channel cognitive radio system with general traffic model and probabilistic delay requirement is yet to be found, which has motivated the present work.

The objective of this paper is to analyze the quality of service and capacity of a cognitive radio secondary network. Specifically, by assuming that a certain amount of resource is exclusively reserved and used on each channel by the primary network, we first derive the traffic transportation capacity that is guaranteed to the secondary network. Then, we analyze the traffic delay distribution in the secondary network and derive an upper bound on it. This delay distribution bound allows us to further obtain a guaranteed capacity of the secondary network in serving traffic with delay requirement. To validate the analysis, both numerical and simulation results are presented by using a 3GPP LTE scenario as an example. The delay distribution, average delay and capacity of the secondary network are compared and discussed. The comparison shows an excellent match between numerical and simulation results.

The rest is organized as follows. Sec. II describes the considered system model. Sec. III presents a backlog period analysis and derives the guaranteed service provided to the secondary network. Then, delay analysis is conducted in Sec.

IV. Sec. V presents numerical results and compares with simulation results. Finally, further discussion and concluding remarks are made in Sec. VI and Sec. VII, respectively.

II. THE SYSTEM MODEL

In this paper, we consider a cognitive radio system with multiple independent channels indexed by i ($1 \leq i \leq N$) as shown in Fig. 1. In this system, all channels are slotted¹ and synchronized, where T denotes the slot *time* length. Only at the beginning of a slot, scheduling is made and transmission can start. On each channel i , a certain number of slots are reserved² *periodically* and *exclusively* for the primary network in order to guarantee its service. Here, by *exclusively*, we mean that such slots are never used by the secondary network. When there is traffic, the primary network will always try to use such slots first.

Throughout the rest of this paper, we assume that both the primary network and the number of reserved slots are properly planned such that no additional slots are needed for traffic of the primary network. While this assumption is rather conservative, it *guarantees* a certain amount of service available to the secondary network, which we believe is reasonable and can be expected particularly when the secondary network needs to pay the primary network. Under this channel reservation, each channel becomes an *ON-OFF* process from the viewpoint of the secondary network. With this information, we believe channel sensing will work much more effectively and correctly. Due to this, sensing error is ignored in this paper. Recall that, the objective of the paper is to find the traffic transportation capacity that is guaranteed to the secondary network, which also implies perfect sensing.

We suppose each wireless channel has constant transmission rate C_i . We define R_i as the length (in number of slots) of a reservation period on channel i , and R_i^{on} as the *number* of slots reserved by the primary network in each reservation period on channel i . In addition, $\eta_i = \frac{R_i^{on}}{R_i}$ is called as the active factor of the primary network on channel i . Fig. 1 depicts the aforementioned cognitive radio system, where the traffic generated by the primary and secondary networks are denoted by fl^p and fl^s , respectively.

III. GUARANTEED SERVICE ANALYSIS

By observing the considered system, we notice that every channel provides a deterministic amount of service during any given period. In addition, there also exists an upper bound on the amount of resource reserved for the primary network during any period. Therefore, the amount of service that can be utilized by the secondary network is lower bounded. Intuitively, one may guess the long term average service rate of each channel i , which is available to the secondary network, is $(1 - \eta_i) \cdot C_i$ and hence, the total long term average service rate, which the secondary network can maximally get,

¹One slot is defined as the smallest transmission unit in time domain.

²The considered reservation works in the time domain, but, we would like to emphasize that the analysis can also be applied/extended when the reservation is made in other domains such as in the frequency domain.

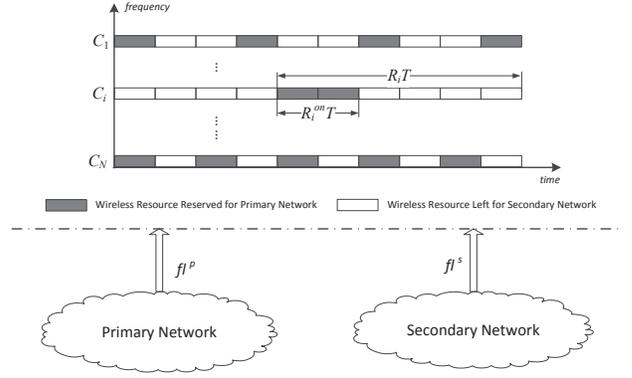


Fig. 1. Considered Cognitive Radio Network

is $\sum_{i=1}^N (1 - \eta_i) \cdot C_i$. In this section, we present results on the amount of service that can be provided to the secondary network, which not only leads to a rigorous validation of the average service rate intuition, but also allows us to view this service on short time scale, which is crucial for QoS analysis of the secondary network.

The following theorem presents the main result, which lays the foundation for later delay analysis of the secondary network.

Theorem 1: (Guaranteed Service.) For the considered cognitive radio system, the amount of service provided to the secondary network during its any backlogged period $(\tau, \tau + t]$, denoted by $W^s(\tau, \tau + t)$, satisfies,

$$C^s \cdot t + U^s \geq W^s(\tau, \tau + t) \geq C^s \cdot (t - L^s)^+$$

where $C^s = \sum_{i=1}^N (1 - \eta_i) C_i$, $L^s = \max_{1 \leq i \leq N} (2R_i^{on} T + 2T)$, $U^s = \sum_{i=1}^N C_i (R_i^{on} + 1) T$, and $x^+ \equiv \max\{x, 0\}$.

Before proving Th. 1, we first discuss how this result can help prove the long term average service rate of the secondary network. Note that the maximum long term average service rate is achieved when there is always traffic to send from the network, and hence can be written as $\lim_{t \rightarrow \infty} \frac{W(\tau, \tau + t)}{t}$. Then, with the first part of Th. 1, we get

$$\lim_{t \rightarrow \infty} \frac{W(\tau, \tau + t)}{t} \leq \lim_{t \rightarrow \infty} \frac{C^s \cdot t + U^s}{t} = C^s$$

and with the second part of Th. 1, we get

$$\lim_{t \rightarrow \infty} \frac{W(\tau, \tau + t)}{t} \geq \lim_{t \rightarrow \infty} \frac{C^s \cdot (t - L^s)^+}{t} = C^s.$$

Summing up, we conclude:

Corollary 1: (Long Term Average Service Rate.) The long-term average service rate that the secondary network can maximally provide is $C^s = \sum_{i=1}^N (1 - \eta_i) C_i$.

A. Proof of Theorem 1

The rest of this section is devoted to the proof of Th. 1. In this paper, we assume that the amount service of a slot is delivered or received by a network when and only when the

slot ends and it is allocated to this network at the start of this slot. The rationale of this assumption is that in packet-switched networks, a packet is considered to be serviced when and only when its last bit has been serviced.

Since all channels are independent with each other, the analysis on each channel is the same. Hence, we start by considering an arbitrary channel indexed with i , and later, the analysis will be extended to the whole system. Consider any backlog time period $(s, t]$ ($0 \leq s \leq t$) for the secondary flow fl^s , which means that there is always traffic waiting to be served in the secondary network during this period. Therefore, we have

$$W_i^s(s, t) = W_i(s, t) - W_i^p(s, t), \quad (1)$$

where $W_i(s, t)$ denotes the total amount of service that can be provided by channel i , $W_i^p(s, t)$ the amount of service that may maximally be occupied by fl^p , and $W_i^s(s, t)$ the amount of service occupied by fl^s . Note that $W_i^p(s, t)$ essentially denotes the amount of service of the reserved slots in $(s, t]$ by the primary network, and due to exclusive reservation, these slots are not used by the secondary network though they may not carry traffic from the primary network.

We first prove the second part of Th. 1.

In Eq. (1), $W_i(s, t)$ can be easily obtained as $W_i(s, t) \geq C_i \cdot (t - s - 2T)^+$. The amount of service provided within two slot $C_i \cdot 2T$ are deducted, because the worse case happens when the time point s (or t) locates just after (or before) a slot starts (or ends), indicating the first slot and the last slot during $(s, t]$ are not complete slots. In addition, $(\cdot)^+$ is due to the fact that the amount of service cannot be negative.

Regarding the specific expression for $W_i^s(s, t)$, there are two scenarios to be analyzed.

- *Scenario 1: The time length $(t - s)$ is no longer than one reservation period, i.e., $t - s \leq R_i T$.*

In this scenario, service reserved for fl^s is upper bounded by $C_i \cdot R_i^{on} T$ as shown in Fig. 2(a). Therefore, we have

$$\begin{aligned} W_i^s(s, t) &= W_i(s, t) - W_i^p(s, t) \\ &\geq [C_i \cdot (t - s - 2T) - C_i \cdot R_i^{on} T]^+ \\ &= C_i \cdot (t - s - 2T - R_i^{on} T)^+. \end{aligned} \quad (2)$$

- *Scenario 2: The time length $(t - s)$ lasts longer than one reservation period, i.e., $t - s > R_i T$.*

In this scenario, let s' denote the start time of the next period just after s , and t' denote the end time of the latest period just before t , as illustrated by Fig. 2(b). Then, the amount of service provided to the secondary network consists of three parts:

$$W_i^s(s, t) = W_i^s(s, s') + W_i^s(s', t') + W_i^s(t', t), \quad (3)$$

where the length of $(s' - s)$ and $(t - t')$ are shorter than one period cycle $R_i T$, and therefore, $W_i^s(s, s')$ and $W_i^s(t', t)$ fall into the range of *Scenario 1*. Then, there hold:

$$W_i^s(s, s') \geq C_i \cdot (s' - s - R_i^{on} T - T)^+ \quad (4)$$

$$W_i^s(t', t) \geq C_i \cdot (t - t' - R_i^{on} T - T)^+. \quad (5)$$

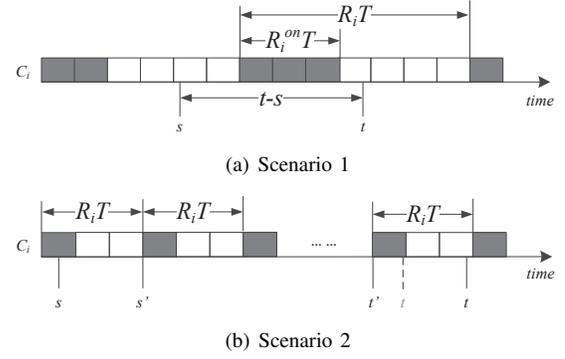


Fig. 2. Illustration of Two Scenarios

Note that, only one slot length is deducted in (4) and (5) compared with (2), because s' and t' are at the edge of a slot.

Intuitively from the definition, we know the time length between s' and t' is integer times of one period $R_i T$, and hence, the service left for the secondary network during $[s', t']$ can be obtained deterministically as

$$W_i^s(s', t') = \frac{t' - s'}{R_i T} \cdot C_i \cdot (R_i T - R_i^{on} T). \quad (6)$$

Combining these scenarios together, we have

$$\begin{aligned} W_i^s(s, t) &= W_i^s(s, s') + W_i^s(s', t') + W_i^s(t', t) \\ &\geq C_i \cdot (s' - s - R_i^{on} T - T)^+ \\ &\quad + C_i \cdot (1 - \frac{R_i^{on}}{R_i})(t' - s') \\ &\quad + C_i \cdot (t - t' - R_i^{on} T - T)^+ \\ &\geq (1 - \eta_i) C_i \cdot (t - s - 2R_i^{on} T - 2T)^+. \end{aligned}$$

Considering the fact of $R_i^{on} \geq 1$ and $\eta_i \leq 1$, the result of *Scenario 1* and *Scenario 2* can be further merged as

$$\begin{aligned} W_i^s(s, t) &\geq (1 - \eta_i) C_i \cdot (t - s - 2R_i^{on} T - 2T)^+ \\ &\triangleq C_i^s \cdot (t - s - L_i^s)^+ \end{aligned}$$

for any backlog period $(s, t]$.

Then, a lower bound on the amount of service provided by the whole system to the secondary network can be obtained by making a summation as

$$\begin{aligned} W^s(s, t) &= \sum_{i=1}^{i=N} W_i^s(s, t) \\ &\geq \sum_{i=1}^{i=N} C_i^s (t - s - L_i^s)^+ \\ &\geq (\sum_{i=1}^{i=N} C_i^s) \cdot (t - s - \max_{1 \leq i \leq N} L_i^s)^+ \\ &\triangleq C^s \cdot (t - s - L^s)^+ = \beta^s (t - s), \end{aligned} \quad (7)$$

which ends the proof of the second part.

For the first part, the prove follows similarly. Particularly, it can be easily verified that $W_i(s, t) \leq C_i \cdot (t - s + T)$.

In addition, in any time interval $(s, t]$, the number of reserved slots for the primary network is not smaller than $\lfloor \frac{t-s}{R_i T} \rfloor R_i^{on}$ and hence the corresponding time length not shorter than $(\frac{t-s}{R_i T} - 1)^+ \cdot (R_i^{on} T)$. We then have

$$W_i^p(s, t) \geq (\frac{t-s}{R_i T} - 1)^+ \cdot (R_i^{on} T) C_i$$

and hence

$$W_i^s(s, t) \leq (1 - \eta_i) C_i \cdot (t - s) + C_i \cdot T \cdot (R_i^{on} + 1)$$

with which the first part is proved by summing up all channels' service.

IV. DELAY DISTRIBUTION ANALYSIS

Th. 1 fundamentally indicates the amount of service that can be guaranteed for the secondary network. With this, the following theorem presents that, if the traffic arrival process of the secondary network is stochastically bounded, the traffic delay (including queueing delay and transmission time) in the secondary network is probabilistically upper bounded.

Theorem 2: (Delay Distribution.) For the considered cognitive radio system, if the amount of traffic of the secondary network $A^s(s, t)$ is stochastically bounded by an arrival function $\alpha^s(t) \in F^3$ and a probability distribution function $f(x) \in \bar{F}^4$, i.e., there holds

$$P\{\sup_{0 \leq s \leq t} \{A^s(s, t) - \alpha^s(t - s)\} > x\} \leq f(x), \quad (8)$$

then the system delay for any traffic from the secondary network is probabilistically upper bounded by

$$P\{d^s > h(\alpha^s(t) + x, \beta^s(t - L^\sigma))\} \leq f(x), \quad (9)$$

where $\beta^s(t) \equiv C^s \cdot (t - L^s)^+$, $h(\alpha^s(t) + x, \beta^s(t - L^\sigma))$ is the maximum horizontal distance between $\alpha^s(t) + x$ and $\beta^s(t - L^\sigma)$, and $L^\sigma = \sigma_{max}^s / C^s$ is the latency of serving the largest unit of traffic denoted as σ_{max}^s .

Here we would like to remark the difference between L^s and L^σ in Th. 2. L^s is given in Th. 1, denoting the latency term if the service would have been defined using the Latency Rate server model [12]. However, L^σ is a time length related to serving the largest traffic unit, such as maximum length packet or maximum length file, in the secondary network. This difference is clearly seen from the example given in the next section. In addition, we would like to highlight that the literature has proved that a lot types of traffic satisfy (8) and extensive discussion on this can be found from e.g. [9][13].

Define delay-constrained capacity as the maximum long term traffic rate that can be supported by a network under delay constraint (D, ϵ) , denoted by $C_{(D, \epsilon)}$. Specifically, for the secondary network, the delay-constrained capacity $C_{(D, \epsilon)}^s$ is defined as

$$C_{(D, \epsilon)}^s \equiv \max \lim_{t \rightarrow \infty} \frac{A(\tau, \tau + t)}{t} \quad \text{such that} \quad P\{d^s > D\} = \epsilon.$$

³ F : the set of non-negative wide-sensing increasing functions

⁴ \bar{F} : the set of non-negative wide-sensing decreasing functions

With Th. 2, the following result is immediately obtained:

Corollary 2: (Guaranteed Delay-Constrained Capacity.)

It is guaranteed that the delay-constrained capacity of the secondary network is not smaller than $\max \lim_{t \rightarrow \infty} \frac{\alpha(t)}{t}$, where $x = f^{-1}(\epsilon)$ is the inverse function of ϵ and $\alpha(t)$ satisfies

$$P\{h(\alpha^s(t) + x, \beta^s(t - L^\sigma)) \leq D\} \leq \epsilon.$$

A. Proof of Theorem 2

Consider any traffic unit σ_j^s that arrives at the secondary network at time t . There exists a time point $0 \leq t_0 \leq t$ which is the start of the backlog period containing time t . We can always find such t_0 , because at least the arrival of itself will start the backlog period. Then, the system delay can be expressed as

$$d_j^s = \inf\{\tau : A^s(t_0, t) \leq A_{out}^s(t_0, t + \tau)\}. \quad (10)$$

We can prove that for any $x \geq 0$, if $d_j^s > x$, there must be $A^s(t_0, t) > A_{out}^s(t_0, t + x)$, since otherwise if $A^s(t_0, t) \leq A_{out}^s(t_0, t + x)$, then $d_j^s \leq x$ should hold, which will contradict the condition $d_j^s > x$. To sum up, event $d_j^s > x$ implies event $A^s(t_0, t) > A_{out}^s(t_0, t + x)$. Therefore, it holds:

$$P\{d_j^s > x\} \leq P\{A^s(t_0, t) > A_{out}^s(t_0, t + x)\}. \quad (11)$$

Note that $A_{out}^s(t_0, t + x) \leq W^s(t_0, t + x)$, because the system may be busy severing previous traffic unit $\sigma_k^s \leq \sigma_{max}^s$ at time t_0 , which is the last one sent to the system and empties the buffer before time t_0 . Therefore, we have

$$W^s(t_0, t + x) - \sigma_{max}^s \leq A^s(t_0, t + x) \leq W^s(t_0, t + x). \quad (12)$$

Then, the following steps hold:

$$\begin{aligned} & P\{A^s(t_0, t) > A_{out}^s(t_0, t + x)\} \\ & \leq P\{A^s(t_0, t) > W^s(t_0, t + x) - \sigma_{max}^s\} \\ & \leq P\{A^s(t_0, t) > \beta^s(t + x - t_0) - \sigma_{max}^s\} \\ & \leq P\{A^s(t_0, t) > \beta^s(t + x - t_0) - \frac{\sigma_{max}^s}{C^s}\} \\ & = P\{A^s(t_0, t) - \alpha^s(t - t_0) > \\ & \quad \beta^s(t + x - t_0 - L^\sigma) - \alpha^s(t - t_0)\}, \end{aligned}$$

where $\alpha^s(t)$ is a non-negative wide-sensing increasing function, and L^σ is the latency term introduced by serving σ_{max}^s . When $\lim_{t \rightarrow \infty} \frac{\alpha^s(t)}{t} < \lim_{t \rightarrow \infty} \frac{\beta^s(t)}{t}$ holds, there exists a maximum horizontal distance between $\alpha^s(t) + y$ and $\beta^s(t - L^\sigma)$ for $\forall y > 0$, defined as:

$$\begin{aligned} & h(\alpha^s(t) + y, \beta^s(t - L^\sigma)) \\ & = \sup_{t \geq 0} \{\inf\{\zeta \geq 0 : \alpha^s(t) + y \leq \beta^s(t - L^\sigma + \zeta)\}\}. \end{aligned} \quad (13)$$

By setting $x = h(\alpha^s(t) + y, \beta^s(t - L^\sigma))$ and by definition (13), we have:

$$\begin{aligned} & P\{d_j^s > h(\alpha^s(t) + y, \beta^s(t - L^\sigma))\} \\ & \leq P\{A^s(t_0, t) - \alpha^s(t - t_0) > y\} \\ & \leq P\{\sup_{0 \leq t_0 \leq t} \{A^s(t_0, t) - \alpha^s(t - t_0)\} > y\}. \end{aligned} \quad (14)$$

where the $\sup_{0 \leq t_0 \leq t} \{\cdot\}$ in (14) is used to remove the randomness of t_0 .

When the arrival process $A^s(s, t)$ is stochastically bounded as defined in (8), there holds

$$P\{d^s > h(\alpha^s(t) + y, \beta^s(t - L^\sigma))\} \leq f(x), \quad (15)$$

which ends the proof.

V. NUMERICAL AND SIMULATION RESULTS

In this section, we provide both numerical results and simulation results for the uplink of a FDD LTE system. As specified by [14], the uplink transmissions are organized into radio frames with duration of 10 ms, which is employed as the smallest transmission unit here, i.e., $T = 10$ ms. The number of slots in a reservation period (i.e., R_i) is set to 10 frames. The active factor η_i varies within the range of $[0.1, 0.2, \dots, 0.9]$. Along the frequency axis, the system is “grided” into Resource Blocks (RB), and hereafter, each RB is considered as a single channel (i.e., $N = N^{RB}$), which varies depending on the system bandwidth within $\{15, 25, 50, 75, 100\}$ channels. Each RB in frequency domain contains 12 sub-carriers and the channel rate C_i is 224 kbps under 1/3 coding rate and 16QAM modulation.

The traffic generated by the secondary network is considered to follow a non-full buffer FTP model as suggested by [15]. Specifically, file is the concerned traffic unit with fixed length of $\sigma^s = 4Mbits$ indicating bursty traffic. The bursty file is divided into small packets to fill each slot when transmitted on channels. The arrival process of files is supposed to be a Poisson process⁵. The *delay* to be presented later is defined as the time length between the arrival of a file and the end point of the slot containing the last bit of this file. This arrival process is a *compound* Poisson process. It can be proved that the considered process is stochastically bounded by Eq (8) with the following setting [16][17]:

$$\begin{aligned} a^s(t) &= \frac{\lambda^s}{\theta} (e^{\sigma^s \theta} - 1) t \\ f(x) &= e^{-\theta x} \end{aligned}$$

where $\theta > 0$ is a free parameter and can be used to optimize the results presented later.

By applying Th. 1 and the aforementioned configurations into Th. 2, the probabilistic delay distribution bound can be expressed as:

$$P\{d^s > x\} \leq f(C^s(x - L^s - L^\sigma)^+), \quad (16)$$

subject to

$$\frac{\lambda^s}{\theta} (e^{\sigma^s \theta} - 1) \leq C^s, \quad (17)$$

where $C^s = (1 - \eta_i)NC_i$, $L^s = 2(R_i^{\text{on}} + 1)T$ and $L^\sigma = \sigma^s/C^s$.

Fig.3 compares the upper bound of delay distribution probabilities obtained by the theoretical analysis with simulation

results, where the number of channels is 50, the average file arrival rate from the secondary network is 0.5 files per second and the active factor of the primary network is 0.1 and 0.5, respectively. Though a gap exists between the theoretical and corresponding simulation results, the probabilities obtained by the theoretical analysis are close to and in the same order of magnitude as the simulation results.

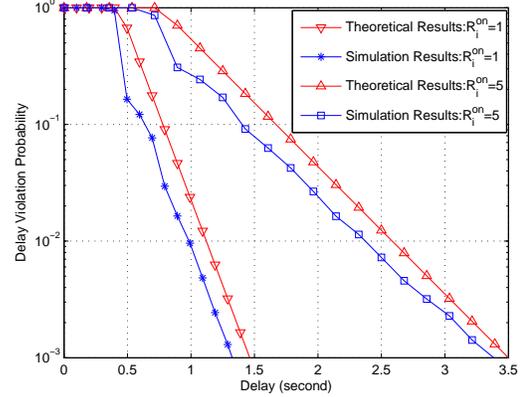


Fig. 3. Delay Distribution Probability

Fig.4 plots the average delay under different configurations. Firstly, the influence of R_i^{on} is investigated. It is obvious that the average delay of secondary files increases when more resource is reserved for the primary network. In addition, the average delay goes to infinity when $R_i^{\text{on}} = 9$, which is not included in the figure. Similar trend can be found when the average file arrival rate from the secondary network increases. On the contrary, more channels (equivalent to larger bandwidth) will guarantee better delay requirements. In addition, the theoretical results locate close to the simulation results, especially when the system is heavily loaded. Those heavily loaded points are usually utilized to find the guaranteed capacity for admission control.

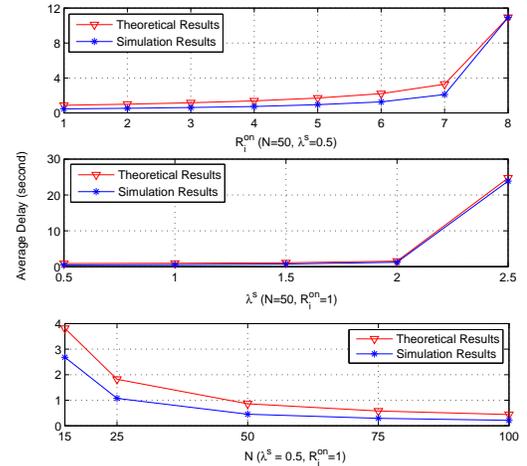


Fig. 4. Average Delay

⁵Note that, we are talking about the service demand of flow arrival (not packet level) that has Poisson property.

Finally, Fig. 5 compares the guaranteed capacity under the constraint that the delay exceeding 3 seconds has a probability less than 1%. The figure shows that the theoretical results match well with the simulation results, which validates the effectiveness of the presented analysis.

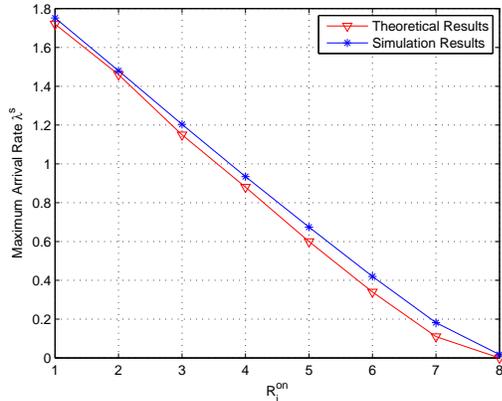


Fig. 5. Delay-Constrained Capacity of Secondary Network

VI. DISCUSSION

We would like to further discuss several key issues in this work. First, the theoretical analysis here finds its root in the area of network calculus [8][9]. The guaranteed service $\beta^s(t)$ in Th. 1 is indeed the so-called *Latency-Rate Service Curve* [12], which is an important type of service curve in network calculus. In addition, in characterizing the stochastic arrival process of the secondary network for delay analysis, the definition in (8) is known as the *virtual-centric-backlog* stochastic arrival curve in stochastic network calculus [9], and $\sup\{\cdot\}$ in (8) cannot be omitted. Importantly, while at a first glance, one might think to apply the leftover service property in network calculus to find the service guaranteed to the secondary network, we stress that this cannot be done easily. The fundamental reason is that the available network calculus leftover service property is applicable only to *single server* systems. Furthermore, while much of the existing multi-channel analysis literature (e.g. [10][11]) is also based on network calculus, the considered systems therein do not match with the considered cognitive radio system. All these have motivated the present work, which also imply the challenge in the analysis.

VII. CONCLUSION

In this paper, performance analysis of a multi-channel cognitive radio secondary network is conducted. In order to ensure a certain level of service guarantee in the cognitive radio system, it is assumed that some amount of resource is reserved for the primary network. With this assumption, we derive the guaranteed amount of service that can be provided to the secondary network. Then, an upper bound on delay distribution probability in the secondary network is obtained if its traffic arrival process is stochastically bounded. In addition,

a delay constrained capacity of the secondary network is derived. Both numerical and simulation results are presented and discussed by considering an LTE parameter setting. Specifically, delay distribution probabilities, average delay and capacity are compared, which shows a good match between the analytical results and the simulation results, indicating the effectiveness of the theoretical analysis.

We stress that in order to move forward in multi-channel cognitive radio analysis, several assumptions are made. While this will certainly restrict the application of results in this paper, the work still sheds light on how multi-channel cognitive radio systems may be analyzed. For more complex scenarios, such as fading channel and random primary traffic arrival, more effort is needed, which are our on-going work.

REFERENCES

- [1] J. I. Mitola and G. J. Maguire, "Cognitive radio: making software radios more personal," *IEEE Personal Communications*, vol. 6, pp. 13–18, Aug. 1999.
- [2] I. Suliman and J. Lehtomaki, "Queueing analysis of opportunistic access in cognitive radios," in *Proc. 2nd International Workshop on Cognitive Radio and Advanced Spectrum Management*, Aalborg, Denmark, May 2009, pp. 153–157.
- [3] C. Ghosh, S. Pagadarai, D. Agrawal, and A. Wyglinski, "Queueing theory representation and modeling of spectrum occupancy employing radio frequency measurements," in *IEEE 70th Vehicular Technology Conference Fall (VTC 2009-Fall)*, Sep. 2009, pp. 1–5.
- [4] Y. R. Kondareddy, N. Andrews, and P. Agrawal, "On the capacity of secondary users in a cognitive radio network," in *IEEE Sarnoff Symposium*, Princeton, USA, Mar. 2009, pp. 1–5.
- [5] S. Kannappa and M. Saquib, "Performance analysis of a cognitive network with dynamic spectrum assignment to secondary users," in *IEEE International Conference on Communications (ICC)*, May 2010, pp. 1–5.
- [6] X. Kang, Y.-C. Liang, A. Nallanathan, H. Garg, and R. Zhang, "Optimal power allocation for fading channels in cognitive radio networks: Ergodic capacity and outage capacity," *IEEE Transactions on Wireless Communications*, vol. 8, no. 2, pp. 940–950, Feb. 2009.
- [7] S. Stotas and A. Nallanathan, "Enhancing the capacity of spectrum sharing cognitive radio networks," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 8, pp. 3768–3779, Oct. 2011.
- [8] J.-Y. L. Boudec and P. Thiran, *Network Calculus: A Theory of Deterministic Queueing Systems for the Internet*. Springer, 2001.
- [9] Y. Jiang and Y. Liu, *Stochastic Network Calculus*. Springer, 2008.
- [10] J. M. Blanquer and B. Özden, "Fair queuing for aggregated multiple links," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 31, pp. 189–197, Aug. 2001.
- [11] H. Xiao and Y. Jiang, "Analysis of multi-server round robin scheduling disciplines," *IEICE Transactions on Communications*, vol. E87-B, pp. 3593–3602, Dec. 2004.
- [12] D. Stiliadis and A. Varma, "Latency rate servers: a general model for analysis of traffic scheduling algorithms," *IEEE/ACM Trans. Networking*, vol. 6, no. 5, pp. 611–624, Oct. 1998.
- [13] Y. Jiang and et al., "Fundamental calculus on generalized stochastically bounded bursty traffic for communication networks," *Computer Networks*, vol. 53, pp. 2011–2021, 2009.
- [14] 3GPP, "Evolved universal terrestrial radio access (E-UTRA): Physical channels and modulation (release 8)," Tech. Rep., Mar. 2008.
- [15] —, "Evolved universal terrestrial radio access (E-UTRA): Further advancements for E-UTRA physical layer aspects (release 9)," Tech. Rep., Mar. 2010.
- [16] F. Kelly, "Notes on effective bandwidths," in *Stochastic Networks: Theory and Applications*, Royal Statistical Society Lecture Notes Series, 4. Oxford University Press, 1996.
- [17] Y. Jiang, "A note on applying stochastic network calculus," Tech. Rep., 2010. [Online]. Available: <http://q2s.ntnu.no/~jiang/publications/note-on-applying-snetcal-v20100501.PDF>