Abstract—Multimedia applications have been widely supported by cognitive radio networks in which high bandwidth can be provided by fully utilizing radio spectrum. However, the research community lacks a theoretical modeling and analysis for multimedia system performance over cognitive radio networks. In this paper, a novel Markov-Chain based model of centralized cognitive radio network (CCRN) consisting of APs and user equipments (UEs) equipped with several antennas (one antenna for channel sensing, the others for data communication) is proposed. In the proposed model, AP will scan the idle primary channels (PCs) that are not used by primary users (PUs), and then transmit data through selected idle channels. Each selected channel in CCRN will be divided into many sub-time slots and some of them are used as the signalling control and the other slots are used only for data transmissions. In addition, three types of communications are considered and an appropriate number of time slots are pre-assigned to ensure QoS of multiple types of multimedia traffic. We used a multi-dimension Markov chain to model time slot allocation and proposed a general channel allocation model for multimedia traffic over CCRN. We validated our model through both theoretical analysis and simulations.

Index Terms—Channel allocation, cognitive radio, Markov Chain, multimedia.

I. INTRODUCTION

EW research engages in studying multimedia transmission over cognitive radio networks. In [1], the authors basically proposed a content-aware multimedia delivery scheme and formulate the spectrum allocation problem as an auction game. In [2], the authors studied a dynamic channel selection to satisfy the rate requirements and delay deadlines of heterogeneous multimedia users through dynamic channel selection. In [3], a joint source-channel coding approach is proposed for distributed multimedia transmission over second users. The fountain code is studied because of its efficiency in error correction capability over lossy wireless channel. The study in [3] also shows that the trade-offs between link reliability, spectral efficiency and coding overheads for multimedia transmission. In [4], the authors specially developed a cross-layer optimization approach to optimize the video quality. The quality optimization is formed as a mixed integer nonlinear programming (MINLP) problem. A distributed algorithm is also developed to solve the MINLP problem. In [5], the authors presented a game-theoretic solution for wireless resource management for delay-sensitive multimedia applications. The authors in [7] proposed an approach that jointly optimizes multimedia intra refreshing rate and lower layer access strategy for high quality multimedia transmission. The optimization is formed as a partially observable Markov decision process. All these approaches above basically consider both multimedia features and channel selection for cognitive radio networks. In [8], the video transmission system is formulated as a switching control dynamic Markovian game. The control scheme based on the game can improve video transmission PSNR quality. In [9], the authors studied voice service over cognitive radio networks and an analytical model is proposed to estimate the capacity for the secondary users. In [10], the authors proposed an open cognitive architecture that can support efficient multi-service operations. In [11], the voice capacity of the CR system based on the theory of effective bandwidth (EB) is derived. The authors show that the simulation results match the analytical results well. A call admission control policy for QoS provisioning in CR networks is also developed. The authors in [13]–[15] studied the principle of resource allocations and multimedia scheduling services in wireless networks. An analysis of channel allocation scheme for wireless cellular networks is given in [16]. In [17], a spectrum sharing scheme across multiple service providers via cognitive radio nodes is proposed. The authors in [18], [19] have studied the multimedia transmission over cognitive radio networks in a cross-layer manner. The works shows the significance of joint design of multimedia coding and channel allocation of cognitive radio networks. In [20], the authors proposed an end-to-end system to optimize the user-perceived video quality at the receiver end under the constraint of packet delay bound in CCRN. Some recent related studies on cognitive radio network have been conducted in [21], [22].

However, in the CCRN, each communication node is usually equipped with multiple antennas and these antennas are dy-
The number of working antennas may be dynamically changed according to the arrival of PUs. Most of existing works ignore the practical and real configurations. To be best of our knowledge, there are no existing research works that model an infrastructure based cognitive radio networks for multimedia transmission. A theoretical analysis model for this network is strongly required to evaluate multimedia system performance in CCRN. In this paper, we studied an infrastructure based cognitive radio networks in which both access point (AP) and primary user base station (PBS) co-exist in the networks and the AP is acted as the center of slot allocation and data relay. Fig. 1 shows the proposed infrastructure which includes three types of communications: between UEs within CCRN, from UEs in CCRN to the nodes outside CCRN, and from nodes outside CCRN to the UEs in CCRN. As described in Fig. 1, the PBS manages all the PCs that can be freely used by the nodes of primary user equipment (PUE), while some idle PCs can be used by other second users in the CCRN. The CCRN consists of access points (AP) that connect user equipment (UEs). AP communicates with other nodes outside CCRN via either wire or wireless connections. All APs and UEs are equipped with multiple antennas.

In the infrastructure based cognitive radio networks, free channels of a PU are scanned and reused by the CCRN. AP decides what channels could be re-used and maintains the resource allocation and data relay. For example, we can assume that TDMA based protocol is applied. Each frame lasts 10 ms consisting of 40 OFDM signals, and each signal (signal slot) lasts 250 $\mu$s. Each channel reused by AP will be divided into multiple time slots and be reallocated to SUs. Two types of frames will be also classified due to the needs of signalling and data exchange between the AP and UEs. The signalling from AP to UE includes the synchronizing, cognition information broadcast and allocation table broadcasting. The signalling from UE to AP includes upper layer application QoS requirements which will be sent by UE to AP via a fixed slot. In the paper, we focus on the time slots allocation and assume the QoS requirement information always

**TABLE I
SYMBOLS AND NOTATIONS**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Notations</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_s$</td>
<td>Number of time slots for each channel used by CCRN</td>
</tr>
<tr>
<td>$N_h$</td>
<td>Number of time slots for signalling per channel</td>
</tr>
<tr>
<td>$N_c$</td>
<td>Time slot number for data transmission per channel: $N_c = N_s - N_h$</td>
</tr>
<tr>
<td>$N_f$</td>
<td>Idle time slots per channel in CCRN, which are not used by any user.</td>
</tr>
<tr>
<td>$N_a$</td>
<td>Number of available slots per channel.</td>
</tr>
<tr>
<td>$D$</td>
<td>Number of multimedia traffic types</td>
</tr>
<tr>
<td>$d_{1,2}$</td>
<td>The minimum and maximum slots required for a call.</td>
</tr>
<tr>
<td>$\lambda^p$</td>
<td>The arrival rate of PUs</td>
</tr>
<tr>
<td>$\lambda^{a,b,c,k}$</td>
<td>The arrival rate of SU of type $(a,b,c,k)$</td>
</tr>
<tr>
<td>$\alpha^p$</td>
<td>The service rate of PUs.</td>
</tr>
<tr>
<td>$\alpha^a$</td>
<td>The service rate of SU if one slot is used.</td>
</tr>
</tbody>
</table>

**TABLE II
TIME SLOT DEFINITION**

<table>
<thead>
<tr>
<th>Index</th>
<th>Types</th>
<th>from</th>
<th>to</th>
<th>source</th>
<th>destination</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Uplink (control slot)</td>
<td>UE1</td>
<td>AP</td>
<td>UE1</td>
<td>AP</td>
</tr>
<tr>
<td>2</td>
<td>Downlink (control slot)</td>
<td>AP</td>
<td>UE1</td>
<td>AP</td>
<td>UE1</td>
</tr>
<tr>
<td>3</td>
<td>Uplink (Data slot)</td>
<td>UE1</td>
<td>AP</td>
<td>UE1</td>
<td>UE2</td>
</tr>
<tr>
<td>4</td>
<td>Downlink (Data slot)</td>
<td>AP</td>
<td>UE2</td>
<td>UE1</td>
<td>UE2</td>
</tr>
<tr>
<td>5</td>
<td>Uplink (Data slot)</td>
<td>UE1</td>
<td>AP</td>
<td>UE1</td>
<td>A node outside CR</td>
</tr>
<tr>
<td>6</td>
<td>Downlink (Data slot)</td>
<td>AP</td>
<td>UE1</td>
<td>A node outside CR</td>
<td>UE1</td>
</tr>
</tbody>
</table>
TABLE III

<table>
<thead>
<tr>
<th>Call Type</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type-(in, in, c₁, cₖ, k) call, or type-(1, c₁, cₖ, k)</td>
<td>This type represents the calls among UEs in the same CCRN, and the call needs at least c₁ and at most cₖ slots for transmissions. When there are Nₜ antennas working simultaneously, k may be larger than Nₜ. For example, if c₁=3, cₖ=4, Nₜ=3 and there are enough slots for the channel allocation, then 2 slots per channel should be allocated for the call. The number of slots allocated to the sender/receiver are k = 2Nₜ=6, and the total number of slots allocated to the communication stream is 12.</td>
</tr>
<tr>
<td>Type-(in, out, c₁, cₖ, k) call, or type-(2,c₁, cₖ, k)</td>
<td>This type represents a call between a UE and the node outside the CCRN. In this case, AP should relay the message to the nodes outside the CCRN.</td>
</tr>
<tr>
<td>Type-(out, in, c₁, cₖ, k) call, or type-(3, c₁, cₖ, k)</td>
<td>The source is a node outside of CRRN, and the destination is UE, so AP should relay the message to the nodes outside the CCRN though wire connections.</td>
</tr>
</tbody>
</table>

can be sent to AP successfully. There is an interrupt from physical layer of UE every 10 ms. The major symbols we used in the paper are listed in Table I.

II. BASIC DEFINITION AND ASSUMPTIONS

In the proposed infrastructure based Cognitive radio networks, we assume that there are N PCs (primary channels) that are used by the PUs (Primary Users). In many scenarios, Nₚₚ PCs are used by PUs while Nₑₑ (Nₑₑ = N - Nₚₚ) PCs can be used by the SUs. The maximum number of channels the SUs can use is M. M is the number of antennas each UE is equipped. When 0 < Nₑₑ < M is satisfied, only Nₑₑ channels can be used. If Nₑₑ = 0, then all the antennas should be turned off and the communication of the CRRN will be interrupted. Each channel will be divided into T sub channels and T₁ of them should be used as signalling such as time synchronization and time slot allocation. The other T₂ (T₂ = T - T₁) time slots are only used for data transmissions. If the number of channels the CRRN used is Nₜ, then Nₜ antennas are working and (M - Nₜ) antennas should be turned off. If there are total D types of multimedia traffic in CRRN, the least number of time slots should be guaranteed for each traffic in order to assure the QoS. For example, some multimedia traffic require at least c₁ sub slots for transmissions, then the system should allocate k (k ≥ c₁) time slots. On the other hand, any type of traffic require at most cₖ sub slots for data transmissions. For each node equipped with multi-antennas, number of k slots should be allocated to the user, and k may be larger than cₖ. However, it must follow cₖ ≤ k < (cₖ + 1) * Nₜ. This is because the working antennas will be in receiving/sending status at a same time due to the limitation of hardware.

A. Time Slot Definition

Multiple slot allocation patterns can applied for different communication modes. For the communication among users in the CRRN, the number of time slots for the uplink allocated to the sender is equal to the number of time slots for the downlink allocated to the receiver. Because the communication between a user within CRRN and a node outside of CRRN involves both wireless and wired connections, we should consider the difference for the time slot allocation specifically. In addition, both AP and UE need to exchange signalling information for the uplink and downlink. The detail of multiple time slot definition is shown in Table II.

B. Call Type Classification

According to the different communication modes in Table II, we can classify the calls into three types as shown in Table III. Among the call types described in Table III, for a given type-(in, in, cl, ch, 0), type-(in, out, cl, ch, 0) or type-(out, in, cl, ch, 0) call, the call will be allocated k (k ≥ cl) slots for transmission if it is accepted by the system. Then the call on communication is denoted as type-(in, in, cl, ch, k), type-(in, out, cl, ch, k), or type-(out, in, cl, ch, k). If some channels CRRN used are reoccupied by the arriving PUs, the AP will try to re-allocate the slots to the ongoing users to avoid the call termination. In this scenario, some calls may be dropped due to the unavailability of time slots. If only a single direction link (uplink or downlink) slots allocation is considered and there are enough number of slots to be allocated, total c slots on a single direction link will be allocated for the SU. Because the number of working antenna is Nₜ, the allocated total slots are φ(c, Nₜ), φ(c, Nₜ) ≥ c. The function φ can be defined as follows:

\[
φ(c, Nₜ) = \begin{cases} 
  c, & \text{if } \text{mod}(c, Nₜ) = 0 \\
  \left\lfloor \frac{c}{Nₜ} + 1 \right\rfloor Nₜ, & \text{if } \text{mod}(c, Nₜ) \neq 0 
\end{cases}. 
\]

In (1), \(φ(c, Nₜ)\) represents the total time slots allocated to a SU. The time slots allocated to the SU per channel (or per antenna) is \(Ω(c, Nₜ)\), which is expressed in the following:

\[
Ω(c, Nₜ) = \frac{φ(c, Nₜ)}{Nₜ}. 
\]

C. States Definition and Basic Formulas

Given a fixed Nₜ, the state of the SUs can be modeled as follows:

\[
E_{Nₜ} = \left\{ \begin{array}{l} J₁^1, \phi(1, Nₜ), \\ J₁^2, \phi(2, c₁, cₖ, k), \\ J₁^3, \phi(3, c₁, cₖ, k), \\ J₁^{11}, \phi(1, d₁, d₂, d₄₁), Nₜ, + (d₂ - d₄₁)Nₜ, \\ J₁^2, \phi(2, c₁, cₖ, k), \phi(3, c₁, cₖ, k), \\ J₁^{11}, \phi(1, d₁, d₂, d₄₁), Nₜ, + (d₂ - d₄₁)Nₜ, \\ J₁^{12}, \phi(2, c₁, cₖ, k), \phi(3, c₁, cₖ, k), \\ J₁^{21}, \phi(1, d₁, d₂, d₄₁), Nₜ, + (d₂ - d₄₁)Nₜ, \\ J₁^{31}, \phi(2, c₁, cₖ, k), \phi(3, c₁, cₖ, k), \\ J₁^{32}, \phi(1, d₁, d₂, d₄₁), Nₜ, + (d₂ - d₄₁)Nₜ, \end{array} \right\} . 
\]

(2)

Where \(J_{n,b,c} \) is the number of calls of type \((k, a, b, c)\) as described above. The elements in \(E \) are indexed by four integer \(a, b, c, \) and \( k \). We reordered the index of
and only use one integer to index the elements. The total number of elements for a fixed \( N_w \) is \( N_{state} = 3d_1(d_2 - d_1 + 1)(d_2, 1 - \varphi(d_1, 1) + 1) \) and the number of total states is \( N_{total} = \sum_{k=0}^{N_w} n_{state,k} \). For a certain \( N_w \), the element indexed by \((a, b, c, k)\) will be indexed by \( n_{index} \), \( 1 \leq n_{index} \leq N_{state,N_w} \). After we map the four elements index \([a, b, c, k]\) to \( n_{index} \), \( E_{N_w} \) can be described in (3):

\[
E_{N_w} = \{ (e_1, d_1, N_w), \ldots, e_c, d_1, d_2, N_w + \varphi(c, d_1, 1)N_w, e_3, d_1, d_2, N_w + \varphi(c, d_1, 1)N_w, e_4, d_1, d_2, N_w + \varphi(c, d_1, 1)N_w \}
\]

or

\[
E_{N_w} = \{ (e_1, e_2, c_1, 1, \ldots, e_3, e_4, c_1, 1, \ldots) \}
\]

The state of \( E \) includes a series of \( N_w \) sub state blocks. Space \( E \) is represented as \( \{ i, E_1, E_2, \ldots, E_{N_w}, \ldots, E_W \} \), where \( i \) is the number of PUs. The total state space can be shown as \( h^1, h^2, \ldots, h^{N_{state}} \). All elements in \( E \) but not belonging to \( E_{N_w} \) are zero. We defined two functions. One function is to translate four elements index to one integer index as shown in the following:

\[
n_{index} = \begin{cases} (a, b, c, k) \\
\end{cases}
\]

The other is to convert one integer index to four elements index:

\[
(a, b, c, k) = \begin{cases} n_{index} - 1 \\
\end{cases}
\]

The sub index of these four elements can be expressed in (6):

\[
a = n_1(n_{index}), b = n_2(n_{index}), c = n_3(n_{index}), k = n_4(n_{index}).
\]

Therefore, the elements in the state \( E \) can be noted as \( e_1, e_2, \ldots, e_4(k) \), or \( e_1(k), e_2(k), e_3(k), e_4(k) \) separately, where \( k > 1 \). If \( k = 0 \), then \( e_0 \) is \( i \). The elements for state \( h^i \) can be noted as \( e_1^i, e_2^i, \ldots, e_4^i(k) \) or \( e_1(k), e_2(k), e_3(k), e_4(k) \). For the communication within CCRN, the time slots must be allocated to the sender and the receiver simultaneously so that the reduced number of available slots is the twice of that allocated to the sender. Therefore, \( N_f \) and \( N_o \) can be expressed in (7)–(8):

\[
N_f = N_o = 2 \sum_{c=1}^{d_2} \sum_{a=1}^{d_1} \sum_{c \neq c_{a,b,c,1}} e_{a,b,c,1} \frac{\varphi(b, c_{a})}{b-1} a-1
\]

\[
N_u = 2 \sum_{c=1}^{d_2} \sum_{a=1}^{d_1} \sum_{c \neq c_{a,b,c,1}} e_{a,b,c,1} \frac{\varphi(b, c_{a})}{b-1} a-1
\]

III. CHANNEL ALLOCATION MODELING FOR MULTIMEDIA TRAFFIC

For multimedia traffic over cognitive radio networks, there are eight possible events which lead to the state change in \( E \) space, which are listed as follows.

1. Accepting a type-\(\{in, in, cu, c0\}\) call.
2. Leaving of a type-\(\{in, in, cu, c0\}\) call.
3. Accepting a type-\(\{in, out, cu, c0\}\) call.
4. Leaving of a type-\(\{in, out, cu, c0\}\) call.
5. Accepting a type-\(\{out, in, cu, ch\}\) call.
7. Accepting a PU.
8. Leaving of a PU.

These events above can change the state from \( E \) to \( E' \). For any events, if the destination state and the corresponding transition rate can be determined, multiple Markov chain can be solved.

A. Accepting a Type-\(\{in, in, cu, c0\}\) or Type-\(\{in, cu, ch\}\) Call

In this case, \( c_{lw} \) and \( c_{uh} \) represent the low and high numbers of slots allocated to the call, respectively. Then \( c_{lw} = 2\varphi(cl, N_w) \) and \( c_{uh} = 2\varphi(ch, N_w) \). In addition, the transition rate equals to \( \lambda_{c,\alpha} \) (here \( \alpha = 1 \) if the destination states exist. The destination states can be analyzed in the following:

1. If \( N_f \) \( N_u < c_{uh} \)

The call will be allocated \( \alpha = \varphi(ch, N_w) \) slots to the sender and the receiver per channel (or per antenna), and becomes an ongoing call of type-\(\{in, in, cl, ch, \alpha N_w\}\), where \( \alpha N_w \) is the number of slots allocated to the sender and receiver simultaneously. The number of time slots on these \( N_w \) channels allocated to a pair of UEs within CCRN is \( 2\alpha N_w \). For the new state \( E' \), \( \frac{\alpha N_w}{j} \) is \( \frac{\alpha N_w}{j} + 1 \) in state \( E \) as same as that of \( E' \).

2. If \( c_{lw} \leq N_f \) \( N_u < c_{uh} \)

The call will be allocated \( \alpha = N_f \) slots per channel, and become an ongoing call of type-\(\{in, in, cl, ch, \alpha N_w\}\). Then for \( E' \), \( \frac{\alpha N_w}{j} \) is \( \frac{\alpha N_w}{j} + 1 \) in state \( E \) as same as that of \( E' \).

3. If \( N_f \) \( N_u < c_{lw} \) and \( N_u < c_{lw} \), there will be not enough idle time slots that can be utilized. A call will be allocated \( \alpha = \varphi(cl, N_w) \) slots per channel and become an ongoing call of type-\(\{in, out, cl, ch, \alpha N_w\}\). Therefore, the system should try to release time slots from other type-\(\{in, out, in, cl, cl, ch, \alpha N_w\}\) calls (\( \beta > c_{lw} \)) and allocate them to new calls. The time slot allocation has two steps: 1) releasing time slots; 2) allocating released slots and other idle slots to the new call. After the releasing, \( E \) becomes a temporary state \( E^0 \). Let function \( \varphi(n, N_w) \) return the number of slots, which can be re-allocated per channel from the calls indexed from 1 to \( n \) in the \( E_n \) set block.

\[
\varphi(n, N_w) = \sum_{i=1}^{n} \left[ \left( 1 + \varphi(1, i-1) \right) \frac{\varphi(i, \alpha-1)}{N_w} e_{c} \right]
\]
1, the call is a type-\(\text{-}(in, in, m, h)\) call. The uplink and the downlink should release equal number of time slots. There must exist \(n_{\xi} \) that satisfies \(\varphi_{1}(n_{\xi} - 1, N_{w}) < \alpha - N_{f}\) and \(\varphi_{1}(n_{\xi}, N_{w}) \geq \alpha - N_{f}\) Also, there must exist non-negative integers \(p\) and \(q\) in the following,

\[
p = \frac{\varphi_{1}(n_{\xi} - 1, N_{w})}{(1 + \sum_{n_{\xi} = 1}^{\infty}) \varphi_{2}(n_{\xi}, N_{w})}
\]

\[
q = n_{\xi} - p
\]

where \(p\) is the number of de-allocation candidates of type-\(\text{-}(n_{\xi}N_{w}, N_{w})\), and \(q\) is the number of calls of type-\(\text{-}(n_{\xi}N_{w}, N_{w})\) which need not be de-allocated. Therefore, the accepted call index position of \(E^{0}\) can be located according to the element of \(F^{i}\).

**B. Leaving of a Type-\(\text{-}(in, in, c_{i}, c_{h}, \alpha N_{w})\), or Type-\(\text{-}(1, c_{i}, c_{h}, \alpha N_{w})\) Call**

The transition rate is

\[
e_{c_{i}, c_{h}, \alpha N_{w}, 1} = \frac{N_{w}}{u_{s}}
\]

though the call using total \(2\alpha N_{w}\) slots, but the uplink and the down link use \(\alpha N_{w}\) slots respectively.

The transition state: if type-\(\text{-}(1, c_{i}, c_{h}, \alpha N_{w})\) call leaves, then \(2\alpha N_{w}\) slots will be released. AP should re-allocate those \(2\alpha\) slots per channel to the other ongoing calls as possible as it can. Firstly, AP should find who should be upgraded. Secondly, AP reallocates slots to them. After the call leaves, the system came into a temporary state \(E^{0}\).

**C. Accepting of a Type-\(\text{-}(in, out, c_{i}, c_{h}, \alpha N_{w})\) or Type-\(\text{-}(2, c_{i}, c_{h}, \alpha N_{w})\) Call in the Up Link of UE**

The transition rate is \(\lambda_{2}^{c_{i}, c_{h}}\).

Destination state: for only the uplink slots should be allocated, the time slot allocation strategy is as the same as \(1\) and \(c_{i_{\text{new}}} = c_{i}, c_{\text{high}} = c_{h}\).

**D. Leaving of a Type-\(\text{-}(in, out, c_{i}, c_{h}, \alpha N_{w})\) or Type-\(\text{-}(2, c_{i}, c_{h}, \alpha N_{w})\) Call**

Transition rate is \(e_{c_{i}, c_{h}, \alpha N_{w}, 2} = \alpha N_{w}\) \(\times u_{s}\).

Destination state: as the situation of \(B\), if a type-\(\text{-}(2, c_{i}, c_{h}, \alpha N_{w})\) call leaves, then \(\alpha N_{w}\) slots will be released. AP should re-allocate those time slots to the other ongoing calls as possible as it can. The situation is as similar as \(2\), only the release slots number is \(N_{w}\) instead of \(2N_{w}\).

**E. Accepting a Type-\(\text{-}(out, in, c_{i}, c_{h}, 0)\) or Type-\(\text{-}(3, c_{i}, c_{h}, 0)\) Call in the Down Link of UE**

The transition rate is \(\lambda_{3}^{c_{i}, c_{h}}\).

**F. Leaving of a Type-\(\text{-}(out, in, c_{i}, c_{h}, \alpha N_{w})\) or Type-\(\text{-}(3, c_{i}, c_{h}, \alpha N_{w})\) Call**

Transition rate is \(e_{c_{i}, c_{h}, \alpha N_{w}, 3} = \alpha N_{w}/u_{s}\).

Destination state: The situation is the same as \(D\).

**G. Accepting a PU**

Translation rate is \(\lambda_{p}\)

**H. Leaving of a PU**

The transition rate is \(1/u_{p}\). Destination state is determined below: Once a PU leaves the system, the scan antenna of the AP can find the event, then AP will decide whether occupies the released SC or not. If \(N_{w} = M\), this means all \(M\) antennas are working, AP will not use the released SC all element of \(E^{i}\) and that of \(E^{0}\) are the same. If \(N_{w} > M\), AP will start a idle antenna with the channel of the released SC, \(N_{w} = N_{w} - 1\), at that time, additional number of \(N_{w}\) slots can be re-allocated. In this scenario, AP starts an idle antenna and sends a control command to all the UEs in the CCRN. The new slot allocation table created by AP also will be send to the UEs through downlink header slot. The AP will try to allocate those \(N_{w}\) slots to the UEs for full channel utilization. Once all the translation states \(E^{i}(1 \leq i \leq N_{\text{total}})\) and the corresponding transition rates are obtained, each state probability can be obtained through queueing theory.

**IV. NETWORK PERFORMANCE METRICS**

In our study, we evaluated the proposed instructress based cognitive networks performance using three metrics: blocking probability, dropping probability and throughput. The blocking of PUs only occurs when an arrival SU find that there are no available channels in the system. A simple BCC (blocking call cleared) queuing model can be used to describe it. An arriving SU is blocked only if the AP can not find the enough
available slots to satisfy the basic slot requirement of the SU type \((i, a, b, c, d, e, f, g, h, k)\). The blocking probability of SUs of type\((k, m, n, 0)\) is expressed in the following:

\[
P_{b,k}^n = \sum_{N_a(\{E^i\} \subset \{1, 2, \ldots, |M(m, N_a)\})} \rho_{E^i}.
\]  

Where \(N_a(\{E^i\})\) is the available slots of state \(E^i\) according formula (8).

The blocking probability of PUs is expressed as follows:

\[
P_{b,k}^{pu} = \sum_{E^i} P_{E^i}.
\]

The forced termination probability is equal to the ratio of the number of terminated SU connections to the sum of terminated and complete SU connections. For a state \(E^i\), the number of working antennas is \(N_w\) and the total ongoing calls is \(\sum_{j=1}^{N_{state}(E^i)} e_j^i\). A channel must be released when a PU arrives. For a state \(E^i\) where there are \(N_w^\prime\) working antennas, the new number of working antennas is \(N_w^\prime = N_w - 1\). Let integer number \(n_{\xi}\) follow \(\sum_{j=1}^{N_{state}(E^i)} (1 + 1) \Omega(\{N_i\}, N_w^\prime) e_j^i \leq N_w^\prime\). \(n_{\xi}\) represents the number of ongoing calls that will be reserved. The total number of dropped calls is \(\sum_{j=1}^{N_{state}(E^i)} e_j^i - n_{\xi}\).

If the number of working antenna number decreases, only the calls indexed from 1 to \(n_{\xi}\) are preserved and the calls indexed by \(n_{\xi} + 1\) to \(N_{state}(N_w)\) will be dropped in the new state \(E^\prime\).

So we have the dropping probabilities of the three types of SUs:

\[
P_{s, k}^s = \frac{\sum_{1 \leq N_w \leq W} \left[ \sum_{j=1}^{N_{state}(E^i)} e_j^i \right] \rho_{E^i} \lambda^p}{\sum_{n=1}^{2} \sum_{m=1}^{N_w^\prime} (1 - P_{b,k}^{nu}) \lambda^m}
\]

where \(k = 1, 2\) or 3 means the dropping probability of three types of SUs. \(N_{state}(E^i)\) is the number of states in the sub state space \(E^i_{N_w}\) of the state \(E^i\).

\(T\) represents the throughput that is defined as the average number of service completions for SUs per second, which can be written in the following:

\[
T = \sum_{\psi \in \psi} \left( u^s \sum_{i=1}^{N_{state}(E^i)} (1 + 1) cN_{state}(E^i) e_i^i \right)
\]

where \(N_{state}(E^i)\) is the number of working antennas in state \(E^i\). \(N_{state}(E^i)\) is the number of states of \(E^i\) and \(\psi\) is the total states set of \(E^i\).

In this paper, we will conduct simulations to verify our analytical model described before. The simulation is independent of the theoretical analysis. The behavior of type-\((i, a, b, c, d, e, f, g, h, k)\) calls is as similar as that of type-\((i, a, b, c, d, e, f, g, h, k)\) calls because they only use single direction slots. Therefore, we only consider one of the two types. In case study, the parameters are set as follows: \(N = 2, M = 2, N_c = 2\); There are two types of calls in the system: 1) type-\((i, a, b, c, d, e, f, g, h, k)\) call, or type-\((1, c_1, c_2, k), c_1 = 1, c_2 = 2\); 2) type-\((i, a, b, c, d, e, f, g, h, k)\) call or type-\((2, c_1, c_2, k), c_1 = 1, c_2 = 2\); \(\lambda^p = 1\) to 6 calls/s. The arrival rate of PUs \(\lambda^p\) ranges from 1 to 6 calls/s. The arrival rate \(\lambda^p\) of the two types of SUs is 2 calls/s and \(\lambda^p\) of SUs is 1. The flow directions of SUs are explained in Fig. 4.

In this paper, we will conduct simulations to verify our analytical model described before. The simulation is independent of the theoretical analysis. The behavior of type-\((i, a, b, c, d, e, f, g, h, k)\) calls is as similar as that of type-\((i, a, b, c, d, e, f, g, h, k)\) calls because they only use single direction slots. Therefore, we only consider one of the two types. In case study, the parameters are set as follows: \(N = 2, M = 2, N_c = 2\); There are two types of calls in the system: 1) type-\((i, a, b, c, d, e, f, g, h, k)\) call, or type-\((1, c_1, c_2, k), c_1 = 1, c_2 = 2\); 2) type-\((i, a, b, c, d, e, f, g, h, k)\) call or type-\((2, c_1, c_2, k), c_1 = 1, c_2 = 2\); \(\lambda^p = 1\) to 6 calls/s. The arrival rate \(\lambda^p\) of PUs is 2 calls/second and the service rate \(\lambda^p\) of SUs is 1. The flow directions of SUs are explained in Fig. 4.

The simulation time is 10 000 seconds. The simulation and theory results are shown from Figs. 5–7. From these figures, we can see that the simulation result perfectly matches the theoretical results.

Table IV shows the number of calls handled by the simulation system when \(\lambda^p\) is set to 6 calls/s. For example, for the first
and second rows, there are total 59 907 arrived PUs during the first 10 000 s, and 28 038 of them are accepted while the rest of 31 869 are blocked due to lack of PCs. When the simulation time is expired, there are still 2 PUs in communications. The number of successful serviced PUs is 28 036. For the third and fourth rows, there are total 20 021 SUs for arrived multimedia stream 1, in which 4450 of them are accepted and the other 15 571 SUs are blocked. There are no remaining SUs of stream 1 when the simulation time exceeds 10 000 s. Therefore, the number successful serviced SUs are 1073. The simulation scenarios of multimedia stream 2 are similar with stream 1.

In Fig. 5, we can see with the increasing of arrival rate of PUs, the blocking probability of PUs increase, it is clearly for the service resource is limited, and simulation and theoretical results perfectly match each other; they also match the classic queueing theory of BCC (blocked calls clear) model and Cooper’s result (see [8, Fig. A.1, p. 316], with $S = 2$), the curves in Fig. 5 is almost the same as [8].

Fig. 6 shows some interesting results. With the increasing rate of PUs, the blocking probability of stream 1 will decrease firstly and then increase later. This is reasonable because there are only two PCs and each PC is divided into two slots. AP should allocate the stream 1 to double direction slots, half for uplink (allocate to the source) and half for downlink. With the increasing of $\lambda^P$, fewer antennas are working and the blocking probability of stream 2 will increase. Because the arrival rate of steam 2 is the twice of that of steam 1, there are less stream 2 calls in the system and the stream 1 have larger probability to be accepted. Because $\lambda^P$ is large, the acceptance of all stream 1 and stream 2 will be affected.

Fig. 7 demonstrated that the dropping probability of two multimedia streams increase with the growth of the arrival rate of PU. Larger $\lambda^P$ represents that the UEs have higher probability to turn off its working antennas. Under large $\lambda^P$ some ongoing SUs must be dropped and therefore the throughput is reduced. In Fig. 7, we observed that the dropping probability of stream 1 is larger than that of stream 2. This is because that the 1st stream
TABLE V
TRAFFIC STREAMS IN THE SYSTEM

<table>
<thead>
<tr>
<th>Stream index</th>
<th>Traffic</th>
<th>Comm type</th>
<th>The least slots required</th>
<th>The maximum slots required</th>
<th>Arrival rate</th>
<th>Service rate/slot</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>(in, in)</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>(in, out)</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>(out, in)</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>(in, in)</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>(in, out)</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>(out, in)</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>(in, in)</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>3</td>
<td>(in, out)</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>3</td>
<td>(out, in)</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>2</td>
</tr>
</tbody>
</table>

requires two-way (uplink and the downlink) slot allocations between two nodes within the CCRN. The 2nd stream only needs one direction link (uplink) slots allocation. If a PU arrives and wants to recycle the channel used by SUs, stream 1 will have larger probability to be dropped. In Fig. 8, the throughput of stream 2 is larger than that of stream 1, for both the blocking probability and dropping probability of stream 2 are smaller (see Figs. 6 and 7) than those of stream 1. The calls of stream 2 have larger accepted probability and less dropped probability by the service system. Therefore, the radio of successful serviced calls of stream 2 is larger than that of stream 1 and the throughput of stream 2 is larger than that of stream 1.

From the simulation and analysis results described from Figs. 5–8, the correctness of the proposed modeling is verified. If there are many types of stream in the system, or the numbers of channels, slots and antennas are large, the multiple Markov chain is hard to be expressed due to page limit in this paper. Next, we studied a more complex scenario. We assume there are total nine types of traffic in terms of different communication range and time slots requirements in the system, which is shown in Table V.

Let $N = 4$, $W = 3$, $N_c = 5$, $\lambda^p = 1$ to 6 calls/s, $u^p = 2$ calls/s. All the arrival rates of nine types of SUs are equal to 2 calls/s, and the service rate of each SU is assumed to 1 call/s if the SU only uses one slot for transmission. In our simulation, the traffic intensity of PUs are from 0.5 to 3, and the traffic intensity of each group of SUs are two if only one slot used. The number of data antenna of each UE is set to three (not included the scanning antenna). The total number of PCs is four, and each channel can be divided into five slots for CCRN communication. The simulation time lasts 5000 s.

From Fig. 9, we can observe that the blocking probability of PUs will become larger with the increasing traffic intensity of PUs. Our simulation curve perfect matches the classic blocked call clear modal. Fig. 9 also shows that the behavior of PUs is not impacted by SUs.

Fig. 10 shows the performances of SUs with different types and slot requirements. In Fig. 10, we can observe that multimedia streams within the CCRN (first, fourth, and seventh stream in the figure) have higher blocking probability than the that of multimedia streams between the nodes in CCRN and outside.
CCRN. This is because that each communication with CCRN needs double time slots including both the uplink and the downlink and the other communications only need slot allocation in either uplink or downlink. Therefore, they have lower blocking probability (streams 2, 3, 5, 6, 8, and 9).

We can observe that the blocking probability for streams 1, 4, and 7 increases with the increasing the least required number of the time slots. However, for the communications between the nodes in CCRN and nodes outside CCRN, the blocking probability also decrease if the number of the least slots requirements decrease. Thus, stream 8 and 9 have higher blocking probability than that of streams 2, 3, 5, and 6.

In Fig. 11, the dropping probability for all SUs increases with the increase of arrival of PUs. For streams 8 and 9, they have less dropping rate because they have larger minimum /maximum slots requirements. Fig. 11 also shows that the dropping probability is sensitive to the calls ability of release slots. If the interval between its minimum and maximum slots requirement is larger, the call has less dropping rate. For example, streams 2, 3, 5, and 6 have larger dropping rate than streams 8 and 9. It indicated that the communications within CCRN have larger dropping rates. This is because those calls communication within CCRN should release double number of slots including uplink and downlink. For example, stream 1 (stream 4 or stream 7) has larger dropping rate than that of streams 2 and 3 (streams 5 and 6 or streams 8 and 9).

In Fig. 12, the throughput of each stream decreases with the increase of the arrival rate of PUs. Obviously, more PUs in the system, less channels can be used by CCRN. The number of working antennas of CCRN reduces, the slots in CCRN will decrease and less SUs can be successfully served, which reduces the throughput of each stream (and of all the streams of SUs). Streams 1, 4, and 7 have less throughput than that of other streams because the SUs of the three streams are hardly to be accepted in the CCRN due to double directional slot allocation.

V. CONCLUSIONS

In the paper, we proposed a novel resource allocation model for an infrastructure based cognitive radio networks. In the CCRN, each communication node is equipped with multiple antennas and these antennas are dynamically set to sending/receiving/idle status. The number of working antennas is dynamically changed according to the arrival of PUs. They are in sending/receiving state at the same time due to the hardware limits. Three types of calls with different slot requirements are studied. We proposed to use a multi-dimensional Markov chain to model the channel allocation for multimedia streaming over CCRN. The behavior of multimedia traffic over cognitive radio networks was studied by both theoretical analysis and simulation. The match between the simulation and theoretical analysis proved that our proposed model is accurate.

REFERENCES


Tigang Jiang received the B.Eng degree in material engineering from Xi’an Jiaotong University, China, in 1998, the M.Eng. degree in computer engineering from Southwest Jiaotong University, China, in 2001, and the Ph.D. degree in communication engineering from the Southwest Jiaotong University in 2005.

Currently, he is an Associate Professor in the School of Communication and Information Engineering, University of Electronic Science Technology of China (UESTC), Chengdu. He has published more than 30 journal and conference papers. His research interests include parallel and distributed computing, wireless networks, cognitive radio networks and pattern recognition.

Honggang Wang (M’06) received the Ph.D. degree in computer engineering from the University of Nebraska-Lincoln in 2009.

He is an Assistant Professor in the Electrical and Computer Engineering Department, University of Massachusetts-Dartmouth. His research interests include networking, wireless sensor networks, multimedia communication, network and information security. He serves as an Associate Editor of Wiley’s Security and Communication Networks (SCN).


Yan Zhang (SM’10) received the Ph.D. degree from Nanyang Technological University, Singapore.

Since August 2006, he is working with Simula Research Laboratory, Lysaker, Norway, where he is currently Senior Research Scientist. He is also an Associate Professor (part-time) at the University of Oslo, Norway. He is a regional editor, associate editor, on the editorial board, or guest editor of a number of international journals. He is currently serving the Book Series Editor for the book series on Wireless Networks and Mobile Communications (Auerbach Publications, CRC Press, Taylor & Francis Group). His research interests include resource, mobility, spectrum, energy, and data management in wireless communications and networking.