Chapter 10
Delay in Cognitive Radio Networks

Yaling Yang, Chuan Han, and Bo Gao

Abstract This chapter presents analysis for delays for both multihop cognitive radio networks and single-hop cognitive radio networks. For multihop cognitive radio networks, we analyze the amount of time that a packet spends to travel over the intermittent relaying links over multiple relaying hops and characterize it with the metric called *information propagation speed*. Optimal relaying node placement strategies are derived to maximize information propagation speed. For single-hop cognitive radio networks, we will analyze how delay is affected by multiple cognitive radio design options, including the number of channels to be aggregated, the duration of transmission, the channel separation constraint on channel aggregation, and the time needed for spectrum sensing and protocol handshake. How these different options may affect the delay under different secondary and primary user traffic loads is revealed. Methods for computing optimal cognitive radio design and operation strategy are derived.

10.1 Introduction

Cognitive radio networks (CRN) have a lot of potentials to improve wireless spectrum efficiency. Understanding the fundamental performance characteristics of this new type of networks is important for the optimal planning of CRN and the design of CRN applications. Hence, in this chapter, one important aspect of CRN performance, the delay, is analyzed.

For a multihop CRN, information propagation speed (IPS) is used as a metric for understanding the multihop end-to-end packet delays. Models of IPS and methods to maximize IPS in two cases are introduced. The first case, named the maximum network IPS, maximizes IPS across a network topology over an infinite plane. The second case, named the maximum flow IPS, maximize the IPS between a given pair of source and destination nodes separated by a fixed distance. The analysis shows that both maximum IPS are determined by the activity level of primary users (PU) and the placement of secondary user (SU) nodes. Optimal relay placement strategies will be identified to maximize these two IPS under different primary users' activity levels.

Y. Yang (✉)
Virginia Polytechnic Institute and State University, Blacksburg, VA, USA
e-mail: yyang8@vt.edu

F.R. Yu (ed.), *Cognitive Radio Mobile Ad Hoc Networks*,
DOI 10.1007/978-1-4419-6172-3_10,

For a single-hop CRN, we will analyze how delay is affected by multiple SU design options, including the number of channels to be aggregated, the duration of transmission, the channel separation constraint on channel aggregation, and the time needed for spectrum sensing and protocol handshake. How these different options may affect the delay under different SU and PU traffic loads is revealed. Methods for computing optimal CRN design and operation strategy are derived.

10.2 Optimal Information Propagation Speed Analysis in Multihop Cognitive Radio Networks

Similar to the delay in any networks, delay in CRN is the combination of two components [1]: the information propagation delay and the queuing delay. The information propagation delay is the amount of time that a packet spends to travel over the intermittent relaying links in a CRN and is determined by the underlying communication capabilities of the network. The queuing delay is the amount of time that a packet spends in waiting for other packets to finish their transmission. Queuing delay is determined by the specific traffic load, traffic pattern, and the scheduling algorithms adopted at all hops of CRN.

In this section, we will study the information propagation delay and understand how to plan node placement to minimize this important delay component in multihop CRN. The analysis of information propagation delay in CRN is different from it in other types of networks due to the unique two-tier architecture of CRN, where the information propagation delay in a CRN is related to not only the settings of the CRN network itself but also the traffic activities of primary users. Hence, existing works about multihop delay analysis for other types of networks [1–4] cannot be applied to CRN.

As a means to interpret the information propagation delay independent of propagation distance, we use *Information Propagation Speed (IPS)* as its measurement metric. Information propagation speed is defined as the speed that a piece of information (e.g., a packet of very small size) can be transmitted over a multi-hop CRN. In the remainder of this section, we will establish a model of IPS in CRN and categorize IPS maximization problem into two cases. The first case, named the maximum network IPS problem, maximizes IPS across a network topology over an infinite plane. The second case, named the maximum flow IPS problem, maximizes the IPS between a given pair of source and destination nodes separated by a fixed distance. We will reveal that both maximum network and flow IPS are determined by the primary user (PU) activity level and the placement of SU relay nodes. We will introduce numerical methods to compute the two maximum IPS and built optimal relay placement strategies to realize these two maximum IPS under different PU activity levels.

The results of the IPS study can be used as a benchmark for network design. It can be used to check whether a certain delay-sensitive traffic is supportable in a

particular network setting. The optimal node placement settings that can achieve the maximum IPS can also be used as useful design guidelines in CRN planning.

The rest of this section is organized as follows. The network model and IPS model are presented in Sections 10.2.1 and 10.2.2. The analyses for the network and flow IPS are in Sections 10.2.3 and 10.2.4, respectively. The analytical results are validated by simulations and numerical results in Section 10.2.5.

10.2.1 Network Model

A cognitive radio network is modeled as a network formed by secondary users (SU) and an overlaid primary user (PU) network in an infinite two-dimensional region. The location distribution of PU nodes is assumed to be a two-dimensional Poisson point processes. Assume that there are K channels and an active PU or SU only uses one channel. Based on the measurement study of realistic wireless PU activities in cellular networks [5], the PU traffic can be accurately modeled as a Poisson arrival process with the mean arrival rate λ_P per unit area, while the service time of PU is usually not Poisson.

We assume duplex communications between the SU transmitter and the receiver (e.g., the SU receiver sends ACK back for received data). Hence, both SU transmitter and SU receiver need to avoid interfering with PU receivers. Denote d_s as the sensing radius of SU for PU receivers[1] and let $U(d)$ be the union of the sensing regions of a pair of SU transmitter and receiver that are d distance apart. The shape of $U(d)$ is shown by the gray region in Fig. 10.1 (II). When there are no active PU receivers within $U(d)$, the SU transmitter and receiver can communicate and we call that the SU link between the SU transmitter and the receiver is feasible. Note that here we implicitly assume that SUs can cancel the interference from PU transmitters to the SU receiver through interference cancelation schemes [6].

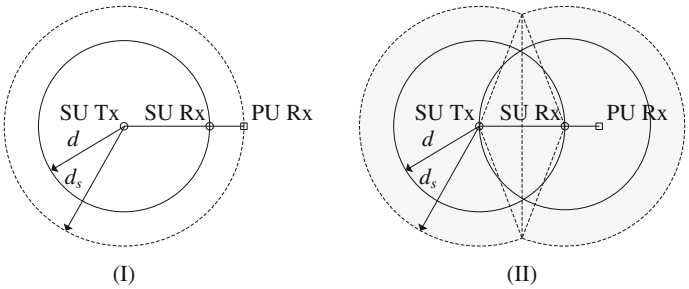

Fig. 10.1 One hop sensing region

[1] PU receiver detection can be realized by exploiting the feedback mechanisms in two-way PU communications as shown by [7, 8].

It is important to note that given cognitive radio's capability to adapt transmission power and hence its potential interference to PUs, the sensing radius of SU should not be treated as a fixed value. In our model, we assume that a SU controls its transmit power so that its communication range equals its distance d to its receiver. In this way, the interference to PUs is minimized. Under this optimal power control policy, we get:

$$\frac{P_{\mathrm{t}}}{d^{\alpha}} = T_{\mathrm{r}} \tag{10.1}$$

where $\alpha \geq 2$ is the attenuation exponent, P_{t} is the transmit power of the SU transmitter and T_{r} is the receiver's sensitivity level. Assume that when the signal of a SU exceeds the interference threshold T_{s} at a PU receiver, the PU receiver is interfered by the SU transmission. Therefore, SU's sensing radius d_{s} can be expressed as:

$$\frac{P_{\mathrm{t}}}{d_{\mathrm{s}}^{\alpha}} = T_{\mathrm{s}} \tag{10.2}$$

Combining (10.1) and (10.2), we have

$$d_{\mathrm{s}} = \left(\frac{T_{\mathrm{r}}}{T_{\mathrm{s}}}\right)^{\frac{1}{\alpha}} d = C_1 d \tag{10.3}$$

where $C_1 = \left(\frac{T_{\mathrm{r}}}{T_{\mathrm{s}}}\right)^{\frac{1}{\alpha}}$. The above equation shows that the sensing radius d_{s} is proportional to the SU communication range that equals the distance d between the SU transmitter and SU receiver under the optimal power control policy.

10.2.2 Problem Formulation

Based on the assumptions in the previous section, we can model IPS, denoted as w, in a multihop CRN as:

$$w = \frac{D}{\tau(D)} = \frac{\sum_i P(d_i)}{\sum_i \tau(d_i)} \tag{10.4}$$

where D is the distance between a pair of SU source node and destination node, and $\tau(D)$ is the expected information propagation delay over this distance D. d_i is the transmission distance of the ith hop, and $P(d_i)$ is the projection of d_i on the straight line from the source to the destination as shown in Fig. 10.2. $\tau(d_i)$ is the expected information propagation delay over the ith hop.

By observing (10.4), we could see a trade-off exists in the setting of d_i and w. While a large d_i increases the numerator in (10.4), it also increase $\tau(d_i)$ since it has

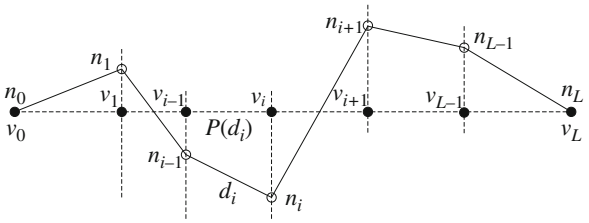

Fig. 10.2 One-hop progress distance

a large $U(d_i)$ and a SU link cannot be used if there is any active PU receiver in the area of $U(d_i)$. The objective of our research, hence, can be formulated as

$$\max_{\mathbf{d}} \frac{\sum_i P(d_i)}{\sum_i \tau(d_i)} \tag{10.5}$$

$$\text{subject to } d_i \leq r_c, \forall d_i \in \mathbf{d} \tag{10.6}$$

where r_c is the SU communication range constrained by the SU radio's maximum transmission power and $\mathbf{d}$ is the set of all hops.

In the following two sections, we study two different cases for solving the above optimization problem by optimally setting $\mathbf{d}$. In Section 10.2.3, we maximize IPS while assuming the network is over an infinite plane and all the SU nodes can be flexibly placed. We call it maximum network IPS. In Section 10.2.4, we maximize IPS for a specific flow with fixed source and destination locations. We call it maximum flow IPS.

10.2.3 Network IPS

Based on (10.6), when a CRN is over an infinite plane and all nodes' locations are flexible, the maximum network IPS, denoted as W_u, is as follows:

$$W_u = \max_{\mathbf{d}} \frac{\sum_i P(d_i)}{\sum_i \tau(d_i)} = \frac{P(d^*)}{\tau(d^*)} \tag{10.7}$$

where d^* is the optimal one-hop distance that results in the maximum IPS in that hop. Equation (10.7) means that the maximum network IPS is achieved when every hop has the same optimal one-hop distance d^*. Therefore, the problem for deriving the IPS upper bound in a multi-hop cognitive radio network has been transformed to the problem of finding the maximum one-hop IPS. In the reminder of this section, we derive the expected one-hop delay function, and then study its monotonicity and convexity, based on which we derive the optimal one-hop distance d^* that achieves the maximum network IPS.

10.2.3.1 One-Hop Delay Function

The expected one-hop delay over a transmission distance d can be expressed as

$$\tau(d) = E\{T\} + \tau_0 \tag{10.8}$$

where $E\{T\}$ is the mean one-hop delay induced by waiting for PU traffic to vacant a channel, τ_0 is the sum of other constant delay components, such as channel sensing delay for determining the channel availability, transmission delay determined by the channel capacity, and packet processing delay determined by the hardware processing capability.

While τ_0 is fixed given the cognitive radio design, the value of $E\{T\}$ is equivalent to the time that it takes for a channel to be vacanted by PU when a SU is ready to transmit a packet. Note that in Section 10.2.1, we have pointed out that measurement results [5] show that realistic PU traffic follows a Poisson arrival process and has complex service time distribution. Hence, we can treat PU as a high priority flow, SU as a low priority flow, and the K channels as K servers in a two-priority preemptive M/G/K queuing model. Under this model, $E\{T\}$ is equivalent to the queueing delay of the low priority flow when the packet service time of the low priority flow approaches 0.

The anlaytical work of two-priority preemptive M/G/K queue in [9] shows that the queuing delay of the low priority flow can be approximated as follows:

$$\text{delay}_1 = \begin{cases} \dfrac{1}{2K\rho_1}\left[\dfrac{s_1\rho_1 + s_2\rho_2}{1-\rho_1-\rho_2}\dfrac{(\rho_1+\rho_2)^K + (\rho_1+\rho_2)}{2} - \dfrac{s_2\rho_2}{1-\rho_2}\dfrac{\rho_2^K + \rho_2}{2}\right] \\ \rho_2 > 0.7,\ \rho_1 + \rho_2 > 0.7; \\ \dfrac{1}{2K\rho_1}\left[\dfrac{s_1\rho_1 + s_2\rho_2}{1-\rho_1-\rho_2}\dfrac{(\rho_1+\rho_2)^K + (\rho_1+\rho_2)}{2} - \dfrac{s_2\rho_2}{1-\rho_2}\rho_2^{\frac{K+1}{2}}\right] \\ \rho_2 < 0.7,\ \rho_1 + \rho_2 > 0.7; \\ \dfrac{1}{2K\rho_1}\left[\dfrac{s_1\rho_1 + s_2\rho_2}{1-\rho_1-\rho_2}(\rho_1+\rho_2)^{\frac{K+1}{2}} - \frac{s_2\rho_2}{1-\rho_2}\rho_2^{\frac{K+1}{2}}\right] \\ \rho_2 < 0.7,\ \rho_1 + \rho_2 < 0.7. \end{cases} \tag{10.9}$$

where subscript 1 is for low priority flow and 2 is for high priority flow and $s_i := \frac{1+C_{B_i}^2}{\mu_i}$, $C_{B_i}^2 := \frac{\sigma_i^2}{\bar{X}_i^2}$, $\rho_i := \lambda_i \bar{X}_i / K$ Here, X_i is the service time for priority i traffic, σ_i^2 is the variance of X_i, and $\bar{X}_i$ is the mean value of X_i.

Therefore, assuming that the traffic load of PU is reasonable (a.k.a. $\rho_2 < 0.7$), we can get $E\{T\}$ as

$$E\{T\} = \lim_{\rho_1 \to 0, s_1 \to 0} \text{delay}_1 \tag{10.10}$$

$$= \frac{\frac{K+1}{2} s_2 \rho_2^{\frac{K+1}{2}}}{2K(1-\rho_2)} + \frac{s_2 \rho_2^{\frac{K+3}{2}}}{2K(1-\rho_2)^2} \tag{10.11}$$

The reason that we assume $\rho_2 < 0.7$ is because CRN is usually used in scenarios where PUs have low utilization of licensed spectrum bands. Hence, the assumption that PU traffic load is light to medium is valid in our application scenario. The ρ_2 in (10.11) can be computed as follows. Note that within an unit area, the PU traffic is a Poisson arrival process with parameter λ_P. Hence, the aggregate PU traffic arrival rate within region $U(d)$ is Poisson with parameter $\lambda_a = \lambda_P A(d)$, where $A(d)$ is the area of region $U(d)$ as shown in Fig. 10.1. By (10.3) and Fig. 10.1 (II), it can be shown that $A(d) = Cd^2$, where $C = 2\pi C_1^2 - 2C_1^2 \cos^{-1} \frac{1}{2C_1} + \sqrt{C_1^2 - \frac{1}{4}}$ and $C_1 = \left(\frac{T_r}{T_s}\right)^{\frac{1}{\alpha}}$. Therefore, we have

$$\rho_2 = Cd^2 \lambda_P \bar{X_p} / K = Cd^2 \rho \tag{10.12}$$

where $\rho := \lambda_P \bar{X_p} / K$ and $\bar{X_p}$ is the mean active duration of a primary user.

Hence, combining (10.11) and (10.8), we get:

$$\tau(d) = \frac{(K+1) s_2 \rho_2^{\frac{K+1}{2}}}{4K(1-\rho_2)} + \frac{s_2 \rho_2^{\frac{K+3}{2}}}{2K(1-\rho_2)^2} + \tau_0 \tag{10.13}$$

10.2.3.2 Properties of One-Hop Delay Function

Next, we study two important properties of $\tau(d)$: monotonicity and convexity. We show that the one-hop delay function $\tau(d)$ is monotonically increasing and strictly convex. To simplify the mathematical derivation, we denote

$$h_1(\rho_2) = \frac{(K+1) s_2 \rho_2^{\frac{K+1}{2}}}{4K(1-\rho_2)} \tag{10.14}$$

and

$$h_2(\rho_2) = \frac{s_2 \rho_2^{\frac{K+3}{2}}}{2K(1-\rho_2)^2} \tag{10.15}$$

We can prove the following lemmas.

Lemma 1 *For* (10.14) *and* (10.15)*, it follows*

$$h_1'(\rho_2) > 0, h_2'(\rho_2) > 0 \tag{10.16}$$

and

$$h_1''(\rho_2) > 0, h_2''(\rho_2) > 0 \tag{10.17}$$

Proof See our technical report [10].

Lemma 2 *The function* $\tau(d), 0 < d \leq r_c$ *is monotonically increasing.*

Proof See our technical report [10].

Lemma 3 *The function* $\tau(d), 0 \leq d \leq r_c$ *is strictly convex.*

Proof See our technical report [10].

10.2.3.3 Speed Upper Bound Analysis

Clearly, the IPS can only be maximized when SU nodes are aligned on the straight line between the source and the destination. When this happens, the one-hop progress distance along the straight line from the source to the destination equals the one hop distance, i.e., $P(d) = d$. Therefore, the optimal one-hop distance d^* in (10.7) can be computed as

$$d^* = \arg \min_{0<d\leq r_c} \left\{ \frac{\tau(d)}{d} \right\} \tag{10.18}$$

Since the physical meaning of $\frac{\tau(d)}{d}$ is the slope of the line passing through the origin and a point on the function $\tau(d)$ curve, it follows that d^* is the d value that minimizes the line slope.

- *Optimal One-Hop Distance Analysis*
  By Lemma 2 and 3, $\tau(d), 0 < d \leq r_c$ is monotonically increasing and strictly convex. As shown in Fig. 10.3, there are two possibilities when determining d^*. Consider the tangent line of the curve $\tau(d)$ that passes the origin and touches $\tau(d)$ at a point $(d_0, \tau(d_0))$. When there is such a tangent line as shown in Fig. 10.3 (II), we have $d^* = d_0$. When there is no such a tangent line as shown in Fig. 10.3 (I), we have $d^* = r_c$. Mathematically, we have

$$d^* = \begin{cases} d_0, \text{ if } \exists 0 < d_0 \leq r_c, \ s.t. \ \tau'(d_0) = \frac{\tau(d_0)}{d_0} \\ r_c, \text{ if } \nexists 0 < d_0 \leq r_c, \ s.t. \ \tau'(d_0) = \frac{\tau(d_0)}{d_0} \end{cases} \tag{10.19}$$

  To determine d^*, we need to determine if there is a real root to equation $\tau'(d) = \frac{\tau(d)}{d}, 0 < d \leq r_c$. To solve this root existence problem, we define

$$\begin{aligned} f(\rho, d) &= \tau'(d) - \frac{\tau(d)}{d} \\ &= 2Cd\rho h_1'(Cd^2\rho) + 2Cd\rho h_2'(Cd^2\rho) \\ &\quad - \frac{h_1(Cd^2\rho) + h_2(Cd^2\rho)}{d} - \frac{\tau_0}{d} \end{aligned} \tag{10.20}$$

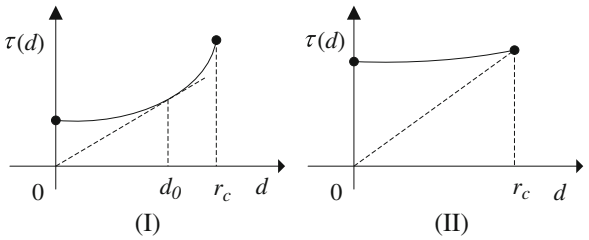

Fig. 10.3 Two examples of $\tau(d)$: (I) $\exists 0 < d_0 \leq r_c,\ s.t.\ \tau'(d_0) = \frac{\tau(d_0)}{d_0}$; (II) $\nexists 0 < d_0 \leq r_c,\ s.t.\ \tau'(d_0) = \frac{\tau(d_0)}{d_0}$

and study the root existence problem for the following equation

$$f(\rho, d) = 0, 0 < d \leq r_c \tag{10.21}$$

- *Threshold Property of d^**
  Next, let us solve the root existence problem in (10.21) and determine d^*. The analytical results are summarized in the following proposition.

Proposition 1 *There exists a threshold* $0 < \rho_u(r_c) < \frac{1}{Cr_c^2}$, $f(\rho_u(r_c), r_c) = 0$ *such that* $d^* = r_c$ *when* $\rho < \rho_u(r_c)$ *and* $d^* < r_c$ *when* $\rho > \rho_u(r_c)$.

Proof By Lemma 2 and 3, $\tau(d), 0 < d \leq r_c$ is monotonically increasing and strictly convex. By (10.20), it follows that $f(\rho, d)$ is an increasing function of d. Since $\lim_{d \to 0^+} f(\rho, d) = -\infty$, it follows that $f(\rho, d)$ increases from $-\infty$ to $f(\rho, r_c)$ when d increases from 0 to r_c. When $f(\rho, r_c) > 0$, there must exist a real root to the problem in (10.21). By (10.19), we have $d^* < r_c$. When $f(\rho, r_c) < 0$, we have $f(\rho, d) < 0, \forall 0 < d \leq r_c$, i.e., there does not exist a real root to the problem in (10.21). By (10.19), we have $d^* = r_c$.

Next, we study the positivity of $f(\rho, r_c)$. Since r_c is fixed, the positivity of $f(\rho, r_c)$ depends on ρ. We study the positivity of $f(\rho, r_c)$ when ρ changes. By (10.20), we have

$$\begin{aligned} f(\rho, r_c) = {} & 2Cr_c\rho h_1'\left(Cr_c^2\rho\right) + 2Cr_c\rho h_2'\left(Cr_c^2\rho\right) \\ & - \frac{h_1\left(Cr_c^2\rho\right) + h_2\left(Cr_c^2\rho\right)}{r_c} - \frac{\tau_0}{r_c} \end{aligned} \tag{10.22}$$

It can be shown that $\lim_{\rho \to 0^+} f(\rho, r_c) < 0$ and $\lim_{\rho \to \left(\frac{1}{Cr_c^2}\right)^-} f(\rho, r_c) = \infty$. Therefore, there is at least one real root to equation

$$f(\rho, r_c) = 0, 0 < \rho < \frac{1}{Cr_c^2} \tag{10.23}$$

In the following, we prove that there is only one real root. By (10.22), we have

$$\frac{\partial f(\rho, r_c)}{\partial \rho} = Cr_c h_1'\left(Cr_c^2\rho\right) + Cr_c h_2'\left(Cr_c^2\rho\right) + 2\rho C^2 r_c^3 h_1''\left(Cr_c^2\rho\right) + 2\rho C^2 r_c^3 h_2''\left(Cr_c^2\rho\right) \tag{10.24}$$

By (10.16) and (10.17) in Lemma 1, we have $\frac{\partial f(\rho, r_c)}{\partial \rho} > 0$, i.e., $f(\rho, r_c)$ is an increasing function with respect to ρ. Therefore, there exists only one real root to equation $f(\rho, r_c) = 0, 0 < \rho < \frac{1}{Cr_c^2}$. Denote the root as $\rho_u(r_c)$. Recall that $\lim_{\rho \to 0^+} f(\rho, r_c) < 0$ and $\lim_{\rho \to \left(\frac{1}{Cr_c^2}\right)^-} f(\rho, r_c) = \infty$. It follows that $f(\rho, r_c) > 0$ when $\rho > \rho_u(r_c)$ and $f(\rho, r_c) < 0$ when $\rho < \rho_u(r_c)$. Therefore, when $\rho > \rho_u(r_c)$ we have $d^* < r_c$, and when $\rho < \rho_u(r_c)$ we have $d^* = r_c$.

Although it is difficult to derive a closed form formula for the threshold $\rho_u(r_c)$ and d^*, we can numerically derive it from (10.23). The physical intuition behind Proposition 1 is as follows. When the actual ρ is small, the delay of SU traffic caused by yielding to PU transmissions in the region $U(d)$ is negligible. The SU's IPS is only constrained by the maximum transmission power. Therefore, the optimal one-hop distance $d^* = r_c$. When ρ is large, the delay caused by PU transmissions in the region $U(d)$ dominates other delay components. Hence, a shorter one-hop distance d incurs a smaller $U(d)$ size, resulting a smaller delay. Therefore, the optimal one-hop distance $d^* < r_c$.

10.2.4 Flow IPS

Beyond the network IPS upper bound for all possible flows, we are also interested in the IPS upper bound for a particular given flow, called the flow IPS upper bound. Since in this case the source–destination distance is fixed, the problem of maximizing the IPS is equivalent to minimizing the total propagation delay from the source to the destination. Therefore, for the flow IPS case, by (10.8), the IPS upper bound W_f can be modeled as

$$W_f = \sup\left\{\frac{D}{\sum_i \tau(d_i)}\right\} = \frac{D}{\inf\left\{\sum_i \tau(d_i)\right\}} \tag{10.25}$$

where $\tau(\cdot)$ is the expected one-hop propagation delay. It is clear that the total delay is minimized when all the SU nodes are placed on the straight line between the source and the destination. Mathematically, the problem of optimal node placement is transformed to

$$\min \sum_i \tau(d_i), s.t. \sum_i d_i = D. \tag{10.26}$$

We decompose the above minimization problem to two subproblems: how to place SU nodes given a fixed number of relay nodes and how many of them should be added between the source and the destination. We show that the IPS is maximized when an optimal number of relay nodes are evenly spaced along a straight line between the source and the destination.

10.2.4.1 Optimal Node Placement

We first study the problem of how to place relay nodes to minimize the total delay when the total number of relay nodes is fixed. We decompose the problem of multi-hop path delay to a series of two-hop path problems. Our analysis shows that the total delay is minimized when the inter-node distances are equal.

Lemma 4 *Consider a K-hop ($K \geq 2$) SU path between a pair of given source and destination nodes. The total expected delay from the source to the destination is minimized when all the $K-1$ relay nodes on the path are evenly placed on the straight line from the source to the destination.*

Proof Consider a two-hop SU path, whose source node and destination node is $0 < y < 2r_c$ distance apart. Then, the total delay of the two-hop path is

$$\tau_2(x) = \tau(x) + \tau(y-x) \tag{10.27}$$

where $y - r_c < x \leq r_c, r_c \leq y < 2r_c$ or $0 < x < r_c, 0 < y < r_c$. By Lemma 2 and 3, we have $\tau_2''(x) = \tau''(x) + \tau''(y-x) > 0$, and $\tau_2(x)$ is strictly convex. Therefore, $\tau_2(x)$ is minimized when $\tau_2'(x) = \tau'(x) - \tau'(y-x) = 0$, i.e., $x = \frac{y}{2}$. Physically, the total delay of a two-hop path is minimized when the relay node is placed in the middle point between the source node and the destination node.

Next, we prove the lemma by contradiction. Given a fixed number of relay nodes, suppose that the minimum total expected delay from the source to the destination is achieved when nodes are not evenly spaced along the straight line between the source and the destination. Denote such a path as P_u. It follows that there exists a two-hop subpath on the path P_u such that the middle SU node of the two-hop subpath is not in the middle position between the source and the destination of the subpath. By placing the middle SU node to the middle position, the total expected delay of path P_u decreases. This contradicts that path P_u minimizes the total expected delay from the source to the destination. Therefore, the lemma follows.

10.2.4.2 Optimal Number of Relay Nodes

Next, we determine the optimal number of relay nodes to minimize the total delay between the source and the destination. Note that to guarantee connectivity between the source and the destination, there are at least $\left\lfloor \frac{D}{r_c} \right\rfloor$ SU nodes placed on the straight line between the source and the destination. Denote n as the number of SU nodes to add, and $m = n + 1$ as the number of hops between the source and the destination.

It follows that $n \geq \left\lfloor \frac{D}{r_c} \right\rfloor$ and $m \geq \left\lfloor \frac{D}{r_c} \right\rfloor + 1$. By (10.26) and Lemma 4, given a m hop SU path ($n = m - 1$ relay nodes), the minimum total delay is

$$t(m) = m\tau\left(\frac{D}{m}\right) \tag{10.28}$$

Therefore, the optimization problem in (10.26) can be transformed to

$$\min t(m) = m\tau\left(\frac{D}{m}\right), s.t.\ m \geq \left\lfloor \frac{D}{r_c} \right\rfloor + 1, m \in Z^+ \tag{10.29}$$

Consider its continuous counterpart problem

$$\min t(x) = x\tau\left(\frac{D}{x}\right), s.t.\ x \geq \left\lfloor \frac{D}{r_c} \right\rfloor + 1, x \in R^+ \tag{10.30}$$

It can be shown that

$$t'(x) = \tau\left(\frac{D}{x}\right) - \frac{D}{x}\tau'\left(\frac{D}{x}\right) \tag{10.31}$$

and

$$t''(x) = \frac{D^2}{x^3}\tau''\left(\frac{D}{x}\right) > 0 \tag{10.32}$$

By (10.32), $t(x)$ is a strictly convex function over $x \geq \left\lfloor \frac{D}{r_c} \right\rfloor + 1$. There are two possibilities when solving the problem in (10.30). When $t'\left(\left\lfloor \frac{D}{r_c} \right\rfloor + 1\right) > 0$, the optimal solution $x^* = \left\lfloor \frac{D}{r_c} \right\rfloor + 1$. When $t'\left(\left\lfloor \frac{D}{r_c} \right\rfloor + 1\right) \leq 0$, the optimal solution $x^* = x_0$, where $t'(x_0) = 0, x_0 \geq \left\lfloor \frac{D}{r_c} \right\rfloor + 1$. Therefore, the optimal solution to the problem in (10.29) is as follows:

$$m^* = \begin{cases} \left\lfloor \frac{D}{r_c} \right\rfloor + 1, \text{ if } t'\left(\left\lfloor \frac{D}{r_c} \right\rfloor + 1\right) > 0, \\ \arg \min\limits_{m \in \{m_1, m_2\}} \{t(m)\}, \text{ if } t'\left(\left\lfloor \frac{D}{r_c} \right\rfloor + 1\right) \leq 0 \end{cases} \tag{10.33}$$

where $m_1 = \lfloor x^* \rfloor$, $m_2 = \lceil x^* \rceil$, and $t'(x^*) = 0$.

10.2.4.3 An Iterative Method of Calculating m^*

Note that directly computing m^* from (10.33) involves solving the equation $t'(x_0) = 0, x_0 \geq \left\lfloor \frac{D}{r_c} \right\rfloor + 1$, which may be computationally intensive. This motivates

us to find alternative methods to determine m^*. Since we have proved that $t(x)$, $x \geq \left\lfloor \frac{D}{r_c} \right\rfloor + 1$ is convex, $t(m)$, $m \geq \left\lfloor \frac{D}{r_c} \right\rfloor + 1$ can be either monotonically increasing or first monotonically decreasing and then monotonically increasing. Therefore, m^* is the smallest m such that $t(m+1) > t(m)$. Mathematically,

$$m^* = \min\{m | (m+1)\tau\left(\frac{D}{m+1}\right) > m\tau\left(\frac{D}{m}\right) \\ m \geq \left\lfloor \frac{D}{r_c} \right\rfloor + 1, m \in Z^+\} \tag{10.34}$$

Based on (10.34), it is straight forward to develop an iterative algorithm to compute m^*.

10.2.4.4 A Table Look-Up Method Based on Threshold Property of m^*

While it is possible to determine m^* by (10.34), the iterative algorithm may incur heavy computation overheads. It may take many steps before finding m^*, when m^* is much larger than $\left\lfloor \frac{D}{r_c} \right\rfloor + 1$. This motivates us to find another easy method of determining m^*.

Our basic idea is to determine m^* by considering whether adding a relay node decreases the total delay. Our analysis shows that there exists a threshold PU activity level when deciding whether to add a relay SU node. By (10.34), adding a relay decreases the total delay when $(m+1)\tau\left(\frac{D}{m+1}\right) < m\tau\left(\frac{D}{m}\right)$. To determine m^*, we define

$$\begin{aligned} g(\rho, m) &= (m+1)\tau\left(\frac{D}{m+1}\right) - m\tau\left(\frac{D}{m}\right) \\ &= (m+1)\left[h_1\left(\frac{\rho CD^2}{(m+1)^2}\right) + h_2\left(\frac{\rho CD^2}{(m+1)^2}\right)\right] \\ &\quad - m\left[h_1\left(\frac{\rho CD^2}{m^2}\right) + h_2\left(\frac{\rho CD^2}{m^2}\right)\right] + \tau_0 \end{aligned} \tag{10.35}$$

When $g(\rho, m) < 0$, adding a relay node decreases the total relay. When $g(\rho, m) > 0$, adding a relay node increases the total relay. Given m, the positivity of $g(\rho, m)$ depends on ρ. Next, we study the positivity of $g(\rho, m)$ when ρ changes.

Lemma 5 *Consider a m-hop SU path, whose source–destination distance is D. There exists $0 < \rho_f(m) < \frac{m^2}{CD^2}$, $g(\rho_f(m), m) = 0$ such that the following properties hold.*

- *When $\rho > \rho_f(m)$, adding a relay node and evenly spacing all relay nodes decreases the total delay.*

- *When $\rho < \rho_f(m)$, adding extra relay nodes increases the total delay.*
- *The function $\rho_f(m)$ is monotonically increasing.*

Proof Note that $\lim_{\rho\to 0^+} g(\rho, m) = \tau_0 > 0$, and $\lim_{\rho\to\left(\frac{m^2}{CD^2}\right)^-} g(\rho, m) = -\infty < 0$. Therefore, there is at least one real root to equation $g(\rho, m) = 0, 0 < \rho < \frac{m^2}{CD^2}$. Next, we show that there is only one real root. We prove this by showing that $g(\rho, m)$ is monotonically decreasing with respect to ρ. By (10.35), we have

$$\frac{\partial g(\rho, m)}{\partial \rho} = \frac{CD^2}{m+1}\left[h_1'\left(\frac{\rho CD^2}{(m+1)^2}\right) + h_2'\left(\frac{\rho CD^2}{(m+1)^2}\right)\right] - \frac{CD^2}{m}\left[h_1'\left(\frac{\rho CD^2}{m^2}\right) + h_2'\left(\frac{\rho CD^2}{m^2}\right)\right] \tag{10.36}$$

By (10.16), we have $h_1'\left(\frac{\rho CD^2}{(m+1)^2}\right) + h_2'\left(\frac{\rho CD^2}{(m+1)^2}\right) > 0$, and

$$\frac{\partial g(\rho, m)}{\partial \rho} < \frac{CD^2}{m}\left[h_1'\left(\frac{\rho CD^2}{(m+1)^2}\right) + h_2'\left(\frac{\rho CD^2}{(m+1)^2}\right)\right] - \frac{CD^2}{m}\left[h_1'\left(\frac{\rho CD^2}{m^2}\right) + h_2'\left(\frac{\rho CD^2}{m^2}\right)\right]. \tag{10.37}$$

By (10.17), it follows that $h_1'(\rho_2)$ and $h_2'(\rho_2)$ are monotonically increasing. Therefore, by (10.37) we have $\frac{\partial g(\rho,m)}{\partial \rho} < 0$. Hence, there is only one real root to equation $g(\rho, m) = 0, 0 < \rho < \frac{m^2}{CD^2}$. Denote the root as $\rho_f(m)$. Recall that $\lim_{\rho\to 0^+} g(\rho, m) > 0$, and $\lim_{\rho\to\left(\frac{m^2}{CD^2}\right)^-} g(\rho, m) < 0$. We have the following conclusions.

There exists a threshold value $0 < \rho_f(m) < \frac{m^2}{CD^2}$ such that $g(\rho_f(m), m) = 0$. When $\rho < \rho_f(m)$, it follows that $g(\rho, m) > 0$. By (10.35), adding a relay node increases the total delay. When $\rho > \rho_f(m)$, it follows that $g(\rho, y) < 0$. By (10.35), adding a relay node and placing all the nodes equal distance apart decreases the total delay.

Next, we prove that $\rho_f(m)$ is a monotonically increasing function of m, i.e., $\rho_f(m+1) > \rho_f(m)$. Recall that $\lim_{\rho\to 0^+} g(\rho, m) = \tau_0 > 0$, which is not dependent on m, and $\frac{\partial g(\rho,m)}{\partial \rho} < 0$. To show $\rho_f(m+1) > \rho_f(m)$, it is equivalent to show that $g(\rho_f(m), m+1) > 0$. By (10.35) and the definition of $\rho_f(m)$, we have $g(\rho_f(m), m) = (m+1)\tau\left(\frac{D}{m+1}\right) - m\tau\left(\frac{D}{m}\right) = 0$. Since we have proved that $t(x) = x\tau\left(\frac{D}{x}\right)$ is strictly convex, we have $t(m+2) - t(m+1) > t(m+1) - t(m)$, i.e., $(m+2)\tau\left(\frac{D}{m+2}\right) - (m+1)\tau\left(\frac{D}{m+1}\right) > (m+1)\tau\left(\frac{D}{m+1}\right) - m\tau\left(\frac{D}{m}\right)$. Therefore, given $\rho = \rho_f(m)$, we have $g(\rho_f(m), m+1) = (m+2)\tau\left(\frac{D}{m+2}\right) -$

$(m+1)\tau\left(\frac{D}{m+1}\right) > (m+1)\tau\left(\frac{D}{m+1}\right) - m\tau\left(\frac{D}{m}\right) = g(\rho_f(m), m) = 0$. Hence, $\rho_f(m)$ is a monotonically increasing function of m.

The significance of Lemma 5 is that it can be used to determine the value interval of ρ corresponding to a given optimal m^* value. Given $m \geq \left\lfloor \frac{D}{r_c} \right\rfloor + 1$, by equation $g(\rho, m) = 0$, we can numerically derive the threshold $\rho_f(m)$. The optimal hop count m^* can be determined as follows.

Proposition 2 *Given an actual ρ, we have $m^* = m$, when*

- $\rho \in (0, \rho_f(m)], m = \left\lfloor \frac{D}{r_c} \right\rfloor + 1$;
- *or* $\rho \in (\rho_f(m-1), \rho_f(m)], m > \left\lfloor \frac{D}{r_c} \right\rfloor + 1$.

Proof When $\rho > \rho_f(m)$, adding a relay node decreases the total delay and increments m, which in turn increases the threshold value $\rho_f(m)$ by Lemma 5. Keep incrementing m, until ρ is less than the new threshold value $\rho_f(m)$. At this stage, adding extra relay nodes increases the total delay. Therefore, the optimal hop count $m^* = m$, for $\rho \in (0, \rho_f(m)], m = \left\lfloor \frac{D}{r_c} \right\rfloor + 1$ or $\rho \in (\rho_f(m-1), \rho_f(m)], m > \left\lfloor \frac{D}{r_c} \right\rfloor + 1$.

Note that each interval of ρ corresponds to an optimal hop count m^*. Since $\rho_f(m)$ can be numerically computed, they can be computed off-line and stored in a table. When there are different ρ values, the optimal hop count m^* can be derived simply by looking up the table. This saves a lot of online computation overhead. With m^* computed, the optimal number of relay nodes $n^* = m^* - 1$ can be easily determined. Therefore, we conclude the following proposition.

Proposition 3 *In the flow IPS case, the IPS upper bound is achieved when $n^* = m^* - 1$ relay nodes are evenly spaced along the straight line between the source and the destination, where m^* is given by* (10.33)*,* (10.34)*, or the table lookup method.*

10.2.5 Simulation and Numerical Validation

In this section, we validate the correctness of our upper bounds by simulations and show the correctness of the analytical results in Proposition 1 and 2 by numerical experiments.

10.2.5.1 Validation of the Theoretical Upper Bound

- *Network IPS Case*
  To validate the correctness of our theoretical results, we next compare our theoretical IPS upper bound with the actual IPS computed from simulations. The

simulation region is a square with edge length 10,000 m. The PU transmitters are uniformly distributed within the simulation region. We simulate one-hop, two-hop, and three-hop SU paths. For each path length, we generate 50 paths and for each path, we generate 50 PU transmitter distribution. For each of the setting, we measure the delays for 20,000 packet deliveries between the source and the destination. In the simulation, $K = 20$, $r_c = 110\,\text{m}$, $\tau_0 = 0.1\,\text{ms}$, $\frac{T_r}{T_s} = 2$, and $\mu_P^{-1} = 1\,\text{ms}$. Three possible PU service time distributions are simulated: exponential distribution, uniform distribution, and constant. Their simulation results are shown in Fig. 10.4a–c, respectively.

We perform two sets of simulations for each distribution. In the first set of simulations, we randomly position SU nodes. The maximum IPSs are shown in Fig. 10.4a (I), b (I), and c (I). The mean IPSs and the standard derivations are shown in Fig. 10.4a (II), b (II), and c (II). The maximum IPSs from the simulation are below the theoretical IPS upper bound, validating the correctness of the IPS upper bound. When the path hop count increases, the simulated IPS decreases. This is because a longer SU path has a higher probability that SU nodes may not be aligned on the straight line between the source and the destination, causing an excessive delay. When the ρ value increases, the simulated IPS decreases. This is because that the PU traffic becomes heavier when ρ increases, causing a larger delay. Also, we observe that our theoretical upper bound is tight compared with the maximum IPS, validating the correctness of our approximation.

In the second set of simulations, SU nodes are evenly spaced along the straight line between the source and the destination. We focus on examining the delay of a 3-hop SU path. The one-hop distance d is set to d^*, $0.8d^*$, and $1.2d^*$. When the one-hop distance $d > r_c$, it is rounded to r_c. The simulated mean IPSs and their standard deviations are shown in Fig. 10.4a (III), b (III), and c (III). When $d = d^*$, the simulated mean IPS curves almost match the theoretical IPS upper bound curve. When $d = 0.8d^*$, $1.2d^*$, the simulated mean IPS curves are below the theoretical upper bound curves. This proves that our IPS upper bound can be achieved when SU nodes are optimally deployed.

- *Flow IPS Case*
  Next, we compare our theoretical IPS upper bound with the actual IPS computed from simulations in the flow IPS case. The simulation region is a square with edge length 10, 000 m. The PU transmitters are uniformly distributed within the simulation region. The source–destination distance is 500 m. We simulate m^*-hop, $m^* + 1$-hop, and $m^* + 2$-hop SU paths, where m^* is the optimal number of relay nodes. For each path length, we generate 50 paths and for each path, we generate 50 PU transmitter distribution. For each of the setting, we measure the delays for 20, 000 packet deliveries between the source and the destination. In the simulation, $K = 20$, $r_c = 110\,\text{m}$, $\tau_0 = 0.1\,\text{ms}$, $\frac{T_r}{T_s} = 2$, and $\mu_P^{-1} = 1\,\text{ms}$. Three possible PU service time distributions are simulated: exponential distribution, uniform distribution, and constant. Their simulation results are shown in Fig. 10.5(a–c), respectively.

We perform two sets of simulations for each distribution. In the first set of simulations, we randomly position SU nodes as long as they maintain connectivity between

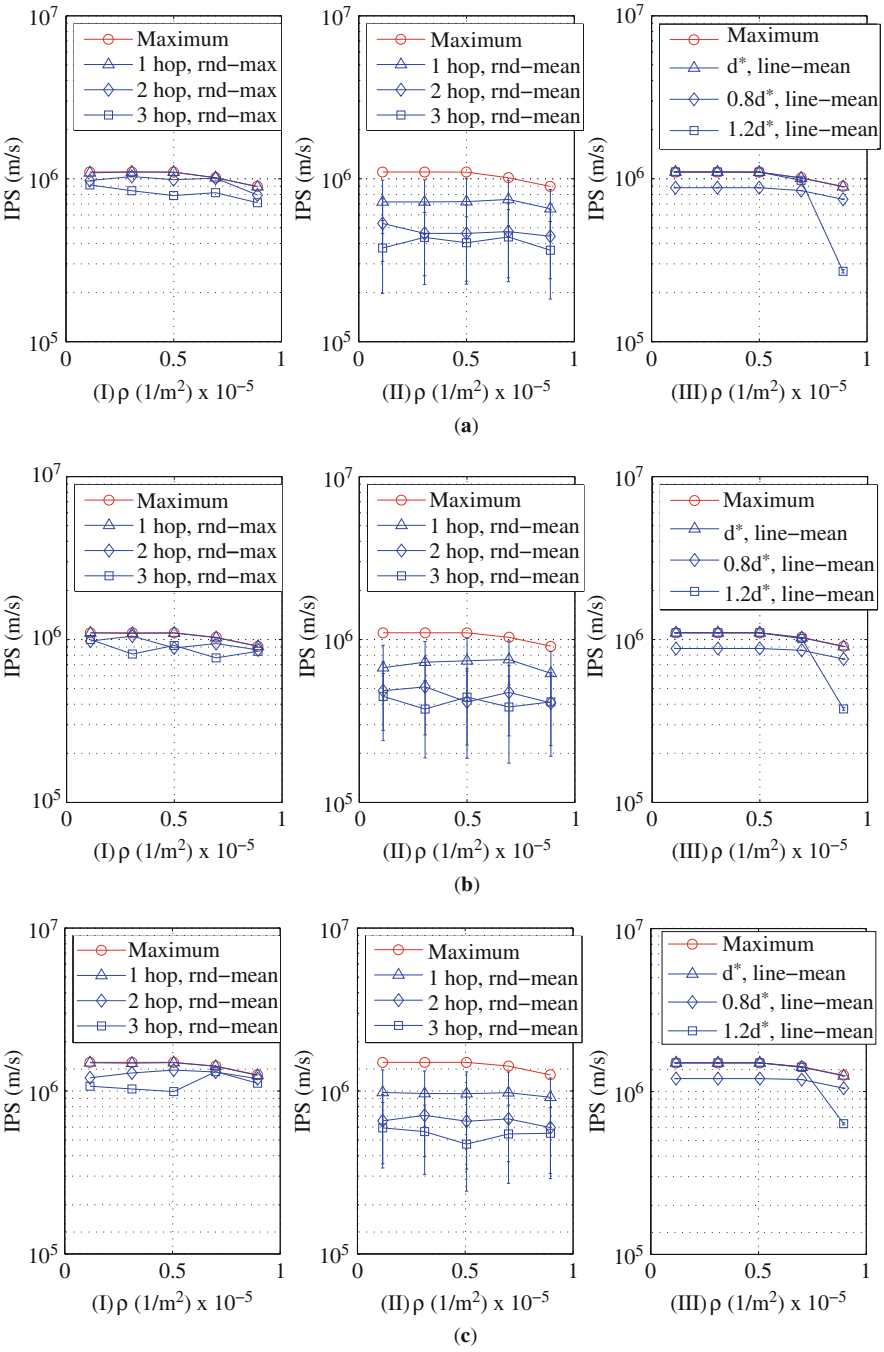

Fig. 10.4 Theoretical IPS upper bound and the simulated IPS for the network IPS case. **(a)** Exponential distribution service time; **(b)** Uniform distribution service time; **(c)** Constant service time

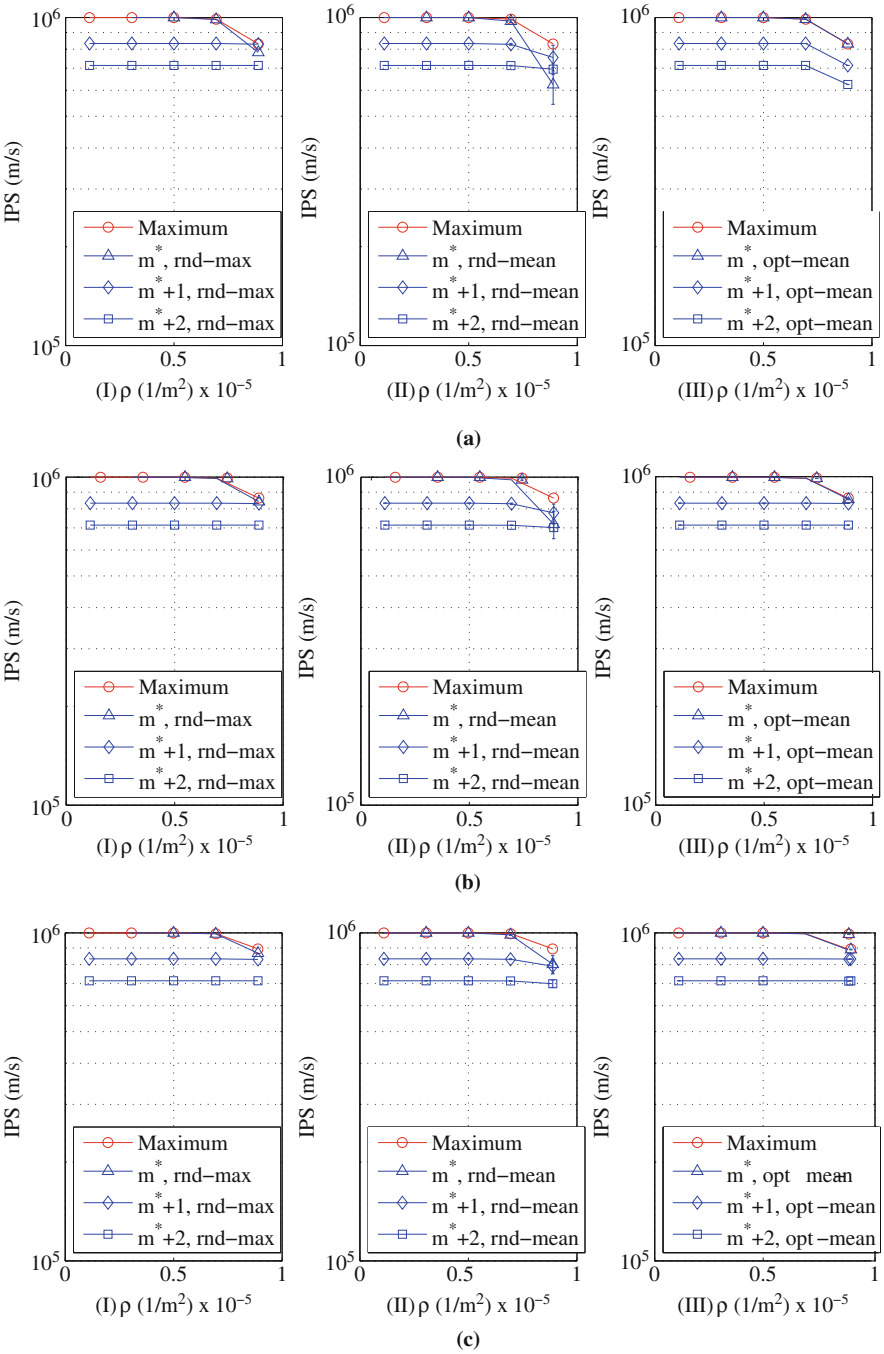

Fig. 10.5 Theoretical IPS upper bound and the simulated IPS for the flow IPS case. **(a)** Exponential distribution service time; **(b)** Uniform distribution service time; **(c)** Constant service time

the source and the destination. The maximum IPSs are shown in Fig. 10.5a (I), b (I), and c (I). The mean IPSs and the standard derivations are shown in Fig. 10.5a (II), b (II), and c (II). The maximum IPSs from the simulation are below the theoretical IPS upper bound, validating the correctness of the IPS upper bound. Also, we observe that our theoretical upper bound is tight compared with the maximum IPS, validating the correctness of our approximation.

In the second set of simulations, SU nodes are evenly spaced along the straight line between the source and the destination. The simulated mean IPSs and their standard deviations are shown in Fig. 10.5a (III), b (III), and c (III). When $m = m^*$, the simulated mean IPS curves match the theoretical IPS upper bound curve. When $m = m^* + 1, m^* + 2$, the simulated mean IPS curves are below the theoretical upper bound curves. This proves the correctness of Lemma 4 and Proposition 3.

10.2.5.2 Optimal One-Hop Distance, Optimal Number of SU Relay Nodes, and Theoretical Upper Bounds

- *Network IPS Case*
  To demonstrate the correctness of Proposition 1, we numerically compute the optimal one-hop distance d^* and the corresponding IPS upper bound for different network settings, e.g., communication range r_c, and PU service time distribution. The d^* is derived from (10.19) and (10.21). We perform numerical experiments for multiple cases with different r_c values and PU service time distributions. In the experiments, we have $K = 20$, $\frac{T_r}{T_s} = 2$, $\alpha = 3$, $\tau_0 = 0.1$ ms, and $\mu_P^{-1} = 1$ ms. We consider three PU service time distributions: exponential distribution, uniform distribution and constant as shown in Fig. 10.6(a–c), respectively.

  For each case, the optimal one-hop distances d^* and the corresponding theoretical IPS upper bounds with respect to ρ values are shown. As shown in these figures, for all different r_c values there is a threshold ρ value. Below this threshold, the optimal one-hop distance $d^* = r_c$. Above this threshold, the optimal one-hop distance $d^* < r_c$. We also observe that the optimal one-hop distance d^* decreases when the ρ value increases. This is because a higher ρ value indicates heavier PU traffic, which causes more excessive delay. Therefore, for a higher ρ value, the optimal one-hop distance d^* is shorter. These results validate the correctness of Proposition 1. For all different r_c values, the bound decreases when the ρ value increases. This is because a higher ρ value indicates heavier PU traffic, which slows down the IPS.
- *Flow IPS Case*
  To demonstrate the correctness of Proposition 2, we numerically compute the optimal number of relay nodes n^* and the corresponding theoretical IPS upper bound for different network settings, e.g., communication range r_c and PU service time. The n^* is iteratively derived from (10.34) and the fact that $n^* = m^* - 1$. We perform numerical experiments for multiple cases with different r_c values and PU service time distributions. In the experiments, we have $K = 20$, $\frac{T_r}{T_s} = 2$,

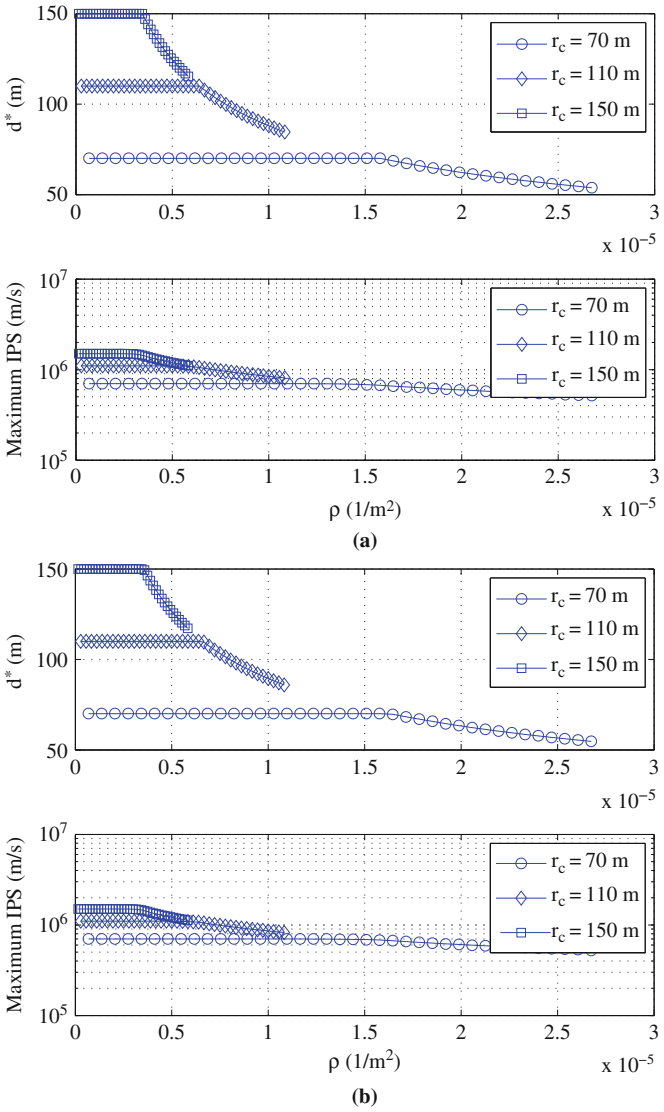

Fig. 10.6 Optimal one-hop distance d^* and IPS upper bound for the network IPS case. **(a)** Exponential distribution service time; **(b)** Uniform distribution service time; **(c)** Constant service time

$\alpha = 3$, $\tau_0 = 0.1$ ms, and $\mu_P = 1$ ms. We consider three PU service time distributions: exponential distribution, uniform distribution, and constant as shown in Fig. 10.7(a–c), respectively.

For each case, the optimal number of relay nodes n^* and the corresponding theoretical IPS upper bounds with respect to ρ values are shown. As shown in these figures, for each r_c value there are threshold ρ values. Above these thresholds,

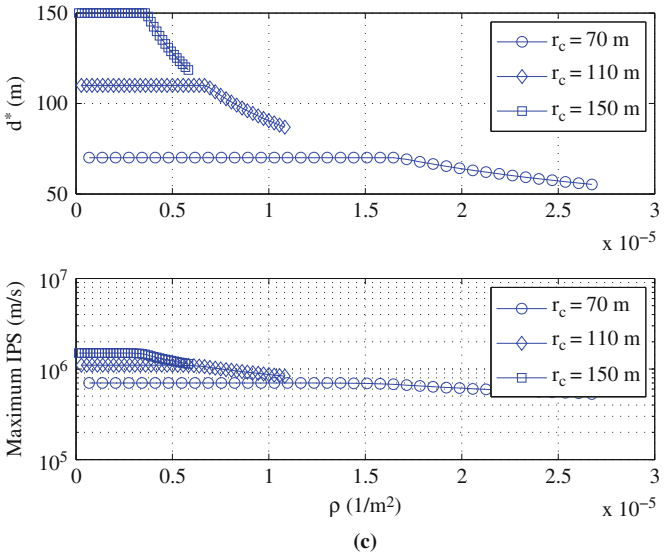

Fig. 10.6 (continued)

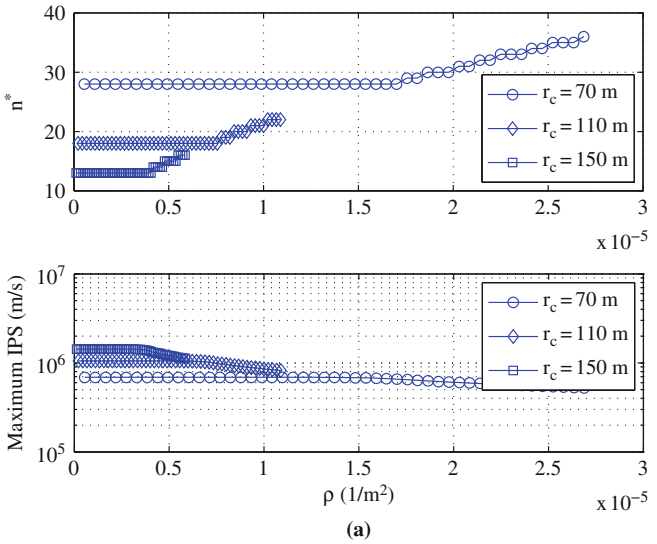

Fig. 10.7 Optimal number of relay SU nodes n^* and IPS upper bound for the flow IPS case. **(a)** Exponential distribution service time; **(b)** Uniform distribution service time; **(c)** Constant service time

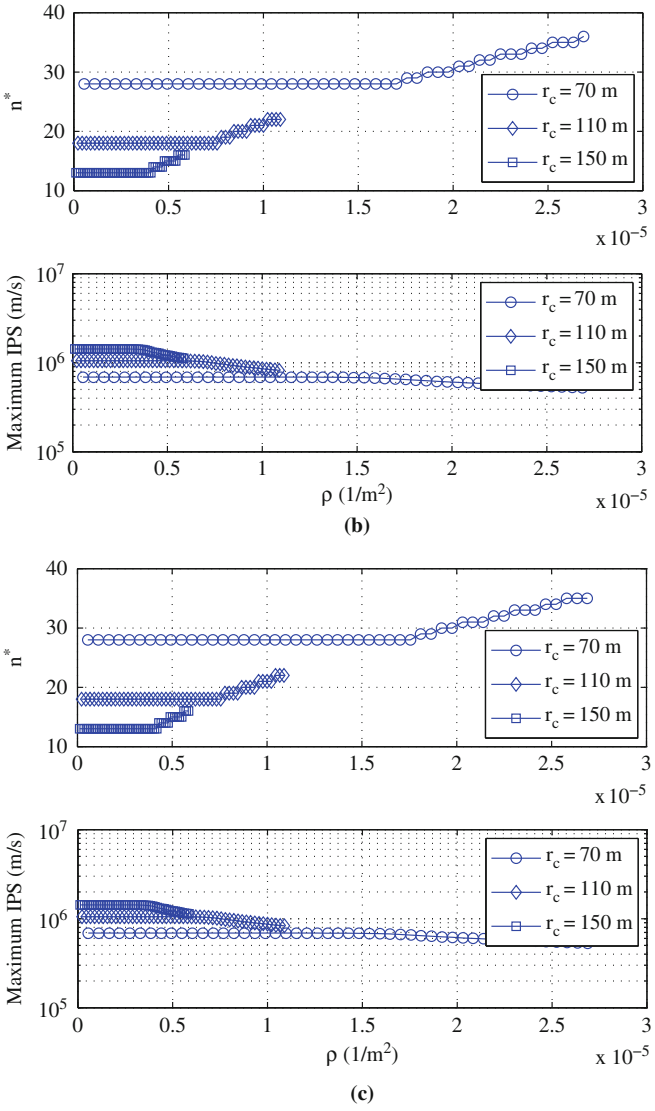

Fig. 10.7 (continued)

n^* increments. We also observe that n^* is a non-decreasing function of ρ. This is because when ρ is large, the IPS is mainly constrained by the interference from PU traffic. A larger number of relay nodes results in a shorter one-hop distance and a shorter sensing range, rendering less interference from PU traffic. Therefore, n^* is a non-decreasing function of ρ. These results validate the correctness of Proposition 2. The general trends and underlying rationales are the same as that of the network IPS case. The bounds are slightly below the bounds of the network IPS case. This

is because the source–destination distance is not necessarily a multiple of d^* in the network IPS case, which decreases the IPS.

10.3 Delay Analysis in Single-Hop Cognitive Radios Networks

As discussed in Section 10.2, the total delay of a flow is a combination of both information propagation delay and queueing delay. While Section 10.2 gives an analysis for information propagation delay in multihop CRN, accurate analysis of queueing delays in such networks is still an open problem due to the complex correlations between packet loss, queue length, scheduling algorithms, and interference among all the hops of a flow. However, in a single-hop CRN network, it is possible to provide accurate analysis of the total delay that includes both information propagation delay and queueing delay with detailed consideration of many CRN design characteristics. In this section, we will provide such an analysis.

One design characteristic that we consider in CRN is *channel aggregation*. Usually, licensed spectrum is divided into a number of discrete channels. As in the Shannon's theorem, channel capacity is proportional to channel width (bandwidth). Hence, efficient utilization of *white spaces* can be achieved by properly enabling each SU to access multiple channels at a time [11]. This assembling of noncontiguous channels for communication is called *channel aggregation* as defined in IEEE 802.22 draft [12]. Technically, channel aggregation can be implemented based on orthogonal frequency division multiplexing (OFDM) [13–15] or multiple radios [16]. Other design characteristics that are considered in our analysis include the duration of each transmission attempt, the delay in channel sensing and channel switching, and the handshake delay for channel negotiation among communication peers.

This section is organized as follows. First, we propose a new channel usage model to investigate the impact of both PU and SU behaviors on the availability of white spaces for channel aggregation. Unlike the ON-OFF process, this general model can capture a wider range of user behaviors. Next, we derive the delay costs for performing channel aggregation under this model. User demands in both frequency and time domains are considered to evaluate the costs for making negotiation and renewing transmission. Further, an optimal channel aggregation strategy is defined in order to minimize the cumulative delay for transmitting data. Finally, numerical analysis and discrete-event simulation are used to illustrate and validate our model and the optimal channel aggregation strategy.

10.3.1 System Model

10.3.1.1 Basic Assumptions

A SU is assumed to be equipped with a dedicated radio for operating on data channels in vacant licensed spectrum bands and another *scanner radio* for

sending/receiving control messages on a dedicated control channel in unlicensed spectrum band [17, 18]. In addition, the scanner radio is responsible for sensing the licensed spectrum bands to discover spectrum white spaces, denoted as $\mathcal{W}_n$, in its sensing region V_n. Denoting $\mathcal{F} = \{f_1, \cdots, f_K\}$ as the set of K channels in the licensed spectrum bands, we have $\mathcal{W}_n \subseteq \mathcal{F}$. For these data channels, each SU with a b-channel *bandwidth demand* can assemble b channels at a time so as to form an *aggregated channel* $\mathcal{A}^{(b)} = \{f_{1'}, \cdots, f_{b'}\} \subset \mathcal{F}$ for data communication. Due to limitations on radio design complexity, there exists a limit B on b such that $b \leq B$. Usually, B is a small positive number. For example, in the IEEE 802.22 standard, $B = 3$. In addition, if any two channels, say f_l and $f_{l+\delta}$, are too far apart in $\mathcal{F}$, they cannot be aggregated. This constraint for the channel separation is denoted as Δ (a.k.a. $\delta \leq \Delta$). In other words, an $\mathcal{A}^{(b)}$ can only be selected from $\mathcal{C}_{l,\Delta} = \{f_l, \cdots, f_{l+\Delta}\} \subset \mathcal{F}$, which is a set of *candidate channels* satisfying the Δ constraint.

Whenever a sender n tries to send data packets to its receiver n', a *negotiation* between n and n' via control channel is necessary for an agreement on the forming of the aggregated channel $\mathcal{A}^{(b)}$. It is assumed that each SU does not have full knowledge of spectrum usage in its vicinity. Thus, multiple negotiation attempts between n and n' may be needed to finalize $\mathcal{A}^{(b)}$ that satisfy the Δ constraint.

Specifically, n should first sense a set of channels to find common white spaces in both V_n and $V_{n'}$. As shown in the example in Fig. 10.8, first, n picks a spectrum range $\mathcal{C}_{l,\Delta}$ (e.g., $\mathcal{C}_{7,4} = \{f_7, f_8, f_9, f_{10}\}$ in the 1st attempt) that satisfies the Δ constraint to sense. This leads to the discovery of the white space $\mathcal{T}_{l,\Delta} = \mathcal{C}_{l,\Delta} \cap \mathcal{W}_n$ (e.g., $\mathcal{T}_{7,4} = \{9\}$ in 1st attempt). Then, n checks if $|\mathcal{T}_{l,\Delta}| \geq b = 2$. If not the case (e.g., 1st attempt), we call it a blocking incident at the sender and n will pick another spectrum range $\mathcal{C}_{l,\Delta}$ to sense for white spaces until finally it finds a white space $\mathcal{T}_{l,\Delta}$ that is larger than b. Then, n initiates a *handshake* with n' to see if enough channels in $\mathcal{T}_{l,\Delta}$ is also available in n' such that $|\mathcal{P}_{l,\Delta}| = |\mathcal{T}_{l,\Delta} \cap \mathcal{W}_{n'}| \geq b$. If not true (e.g., attempt 2nd), a blocking incident at the receiver happens and n is informed to go

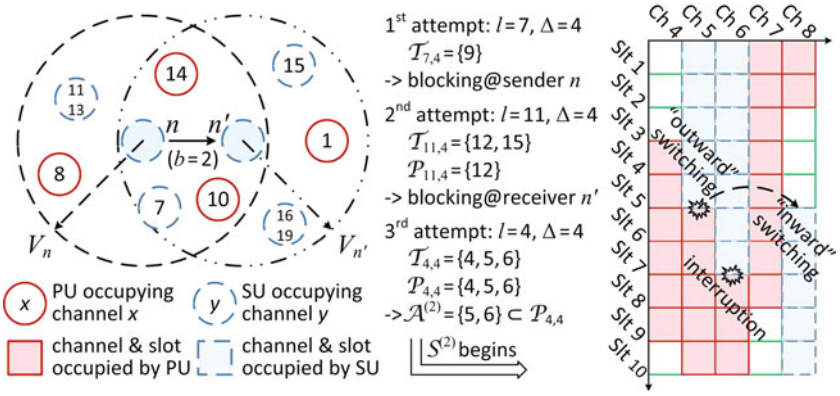

Fig. 10.8 An example of negotiation and transmission between n and n'

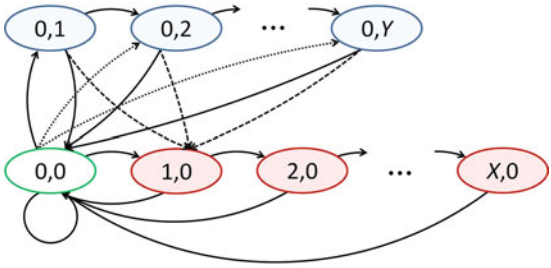

Fig. 10.9 Channel usage model

over all the spectrum sensing and handshake steps again until finally a viable $\mathcal{P}_{l,\Delta}$ is found (e.g., attempt 3rd). Then, n' selects an $\mathcal{A}^{(b)} \subseteq \mathcal{P}_{l,\Delta}$ and replies to n. Then, a *transmission* $\mathcal{S}^{(b)} = \{s_{1'}, \cdots, s_{b'}\}$ is initiated, which includes b parallel subflows with a d-slot *duration demand*.

However, a successful negotiation does not mean a reliable transmission, attributed to the low priority of SU service. To overcome this, *spectrum switching* is employed. Specifically, whenever a PU arrival to any $f_k \in \mathcal{A}^{(b)}$ is detected, the pair of n and n' needs to vacate the preempted f_k immediately and then tries to renew the corresponding s_k on a *backup channel* $f_{k'} \in \mathcal{B}_{l,\Delta}$, where $\mathcal{B}_{l,\Delta} = \mathcal{P}_{l,\Delta} \backslash \mathcal{A}^{(b)}$. For ease of presentation, such spectrum switching is divided into two steps: "outward" switching from f_k and "inward" switching to $f_{k'}$. If $|\mathcal{B}_{l,\Delta}| = 0$, an *interruption* incident occurs to the expelled s_k. But all the other ongoing ones in $\mathcal{S}^{(b)}$ may not be affected as long as the independence of these parallel subflows is guaranteed [19].

10.3.1.2 Channel Usage Model

In a certain SU n's vicinity, any channel $f_c \in \mathcal{F}$ may be occupied by an active PU or SU service for a period of time. As in Fig. 10.9, the average channel occupancy on such f_c is modeled as a Markov chain, in which channel state transits on a slot basis with a τ-second slot duration. Three groups of channel states are defined as follows: (i) *idle state* (0,0), in which f_c is a white space; (ii) *PU service states* (x,0)'s, $x \in \{1, \cdots, X\}$, in which f_c has been occupied by a PU service for x slots; (iii) *SU service states* (0,y)'s, $y \in \{1, \ldots, Y\}$, in which f_c has been occupied by a SU subflow for y slots. Both X and Y are large enough such that $\Pr[x > X]$ and $\Pr[y > Y]$ are negligible. If there are multiple PU or SU services sharing f_c, the statistical data of the service with maximum duration can be applied.

The availability of white spaces is characterized by the steady-state probabilities of channel states, denoted by $\pi_{(\text{current state})}$'s, especially $\pi_{(0,0)}$ for idle state. To derive them, the transition probabilities, denoted by $\omega^{(\text{next state})}_{(\text{current state})}$'s, are obtained as follows.

(1) First, state transitions from (0,0) and (0,y)'s to (1,0) represent a PU arrival to f_c. Each transition from (0,y) to (1,0) also indicates an "outward" switching of the SU from f_c. The transition probabilities are actually equal to the PU arrival

probability, denoted by λ_α, which further depends on the PU arrival process with average arrival rate α_n learnt in V_n. Namely, we have

$$\omega_{(0,0)}^{(1,0)} = \omega_{(0,y)}^{(1,0)} = \lambda_\alpha,\ y \in \{1, \ldots, Y\} \tag{10.38}$$

(2) Next, state transitions from $(x,0)$'s to (0,0) and that among $(x,0)$'s are defined by the distribution of PU service duration on f_c. Note that any closed-form distribution function is not necessarily required here. Instead, one can directly input the statistical distribution of service duration collected from a real network to determine the following transition probabilities

$$\begin{cases} \omega_{(x,0)}^{(0,0)} = \Pr\left[(x-1)\tau < S_{\text{pu}} \le x\tau\right],\ x \in \{1, \cdots, X\}; \\ \omega_{(x,0)}^{(x+1,0)} = 1-\omega_{(x,0)}^{(0,0)},\ x \in \{1, \cdots, X-1\} \end{cases} \tag{10.39}$$

where S_{pu} denotes the random variable of PU service duration.

(3) In a similar way, state transitions from $(0,y)$'s to (0,0) and that among $(0,y)$'s are defined by the distribution of SU service duration on f_c, which can also be general. Hence, we have

$$\begin{cases} \omega_{(0,y)}^{(0,0)} = \left(1-\omega_{(0,y)}^{(1,0)}\right) \Pr\left[(y-1)\tau < S_{\text{su}} \le y\tau\right],\ y \in \{1, \cdots, Y\}; \\ \omega_{(0,y)}^{(0,y+1)} = 1-\omega_{(0,y)}^{(1,0)}-\omega_{(0,y)}^{(0,0)},\ y \in \{1, \cdots, Y-1\} \end{cases} \tag{10.40}$$

where S_{su} denotes the random variable of SU service duration.

(4) State transition from (0,0) to (0,1) is triggered by a SU arrival. The SU arrival probability, denoted by λ_β, is determined by the SU arrival process with average arrival rate β_n learnt in V_n.

State transition from (0,0) to $(0,y)$ where $y > 1$ is triggered by an "inward" switching to f_c. Each transition from (0,0) to $(0,y)$ indicates the case that a subflow $s_{c'}$ switches to f_c when it has last y-1 slots on another channel $f_{c'}$ and has been forced to leave $f_{c'}$ due to the arrival of PU activities on $f_{c'}$. The subflow is renewed on f_c starting from the yth slot.

To derive the state transition probability in the above two cases, we first derive the probability that a certain subflow has successfully switched into f_c, denoted by γ. Due to Δ constraints, only the $2{\cdot}\Delta$ channels excluding f_c in $\mathcal{C}_{c-\Delta,2\Delta}$ can perform an "inward" switching to f_c. If there are u preempted channels and v idle channels out of such $2{\cdot}\Delta$ channels, γ is equivalent to the probability that one of the u expelled subflows successfully chooses f_c out of the total v+1 idle channels for an "inward" switching. Here the worst case is analyzed, in which any incoming subflow neglects the idle channels that are not included in $\mathcal{C}_{c-\Delta,2\Delta}$. Then, we have

$$\gamma = \sum_{u=1}^{2\Delta} \sum_{v=0}^{2\Delta-u} \frac{(2\Delta)!}{u!v!(2\Delta-u-v)!} \left[\lambda_\alpha \left(\sum_{y=1}^{Y} \pi_{(0,y)}\right)\right]^u \left(\pi_{(0,0)}\right)^v \cdot \left[1-\lambda_\alpha \left(\sum_{y=1}^{Y} \pi_{(0,y)}\right) - \pi_{(0,0)}\right]^{2\Delta-u-v} \min\left\{\frac{u}{v+1}, 1\right\} \tag{10.41}$$

Given that a subflow, say $s_{c'}$, has switched into f_c, we further derive $\xi^{(0,y)}$, the probability that $s_{c'}$ has finished $(y-1)$ slots on $f_{c'}$. Assuming that the primary user arrival is independent of the secondary user activities, we have

$$\xi^{(0,y)} = \sum_{z=y}^{Y} \Pr\left[(z-1)\tau < S_{su} \le z\tau\right], \ y \in \{2, \ldots, Y\} \tag{10.42}$$

Using (10.4) and (10.5), we have the following transition probabilities

$$\begin{cases} \omega_{(0,0)}^{(0,1)} = \left(1-\omega_{(0,0)}^{(1,0)}\right)\lambda_\beta; \\ \omega_{(0,0)}^{(0,y)} = \left(1-\omega_{(0,0)}^{(1,0)}\right)\gamma\xi^{(0,y)}, \ y \in \{2, \ldots, Y\} \end{cases} \tag{10.43}$$

At last, we have the transition probability from (0,0) to itself

$$\omega_{(0,0)}^{(0,0)} = 1-\omega_{(0,0)}^{(1,0)} - \sum_{y=1}^{Y} \omega_{(0,0)}^{(0,y)} \tag{10.44}$$

10.3.2 Delay Analysis Under Channel Aggregation

Both the negotiation and the transmission between a sender n and its receiver n' can involve service failures and delay costs. In this section, we investigate the corresponding service failure probabilities and model the delay costs for performing channel aggregation under the influence of PU activity.

10.3.2.1 Delays in Negotiation Process

The efficiency of negotiation between n and n' is restricted by the availability of white spaces in both V_n and $V_{n'}$. In general, the delays for making a successful negotiation include: (i) *sensing delay* $T_{\text{ss}}^{(b)}$, which is the time required for sensing channels at both n and n'; (ii) *handshake delay* $T_{\text{hs}}^{(b)}$, which is the time required for accessing control channel and making handshakes on it back and forth. In the following, these delays in negotiation processes are analyzed.

Note that after each blocking incident caused by $|\mathcal{P}_{l,\Delta}| < b$, a new round of handshake needs to be started until the maximum limit of blocking incidents, denoted as $\hat{N}_{\text{bl}}$, is reached. Hence, negotiation delays are related to the number of blocking incidents during a negotiation process.

To analyze the number of blocking incidents, note that the blocking probability, denoted as $\theta^{(b)}$, can be computed as:

$$\theta^{(b)} = \sum_{v=0}^{b-1} \binom{\Delta+1}{v} \left(\pi_{(0,0)}\right)^{v} \left(1-\pi_{(0,0)}\right)^{(\Delta+1)-v} \tag{10.45}$$

where $\pi_{0,0}$ is the probability that a channel is idle in the joint sensing range of a pair of SUs n and n' (a.k.a. $V_n \cup V_{n'}$). The PU and SU's arrival rates in $V_n \cup V_{n'}$ can be derived as $\alpha_{n,n'} = c_{n,n'} \cdot \alpha_n$ and $\beta_{n,n'} = c_{n,n'} \cdot \beta_n$ respectively, where $c_{n,n'}$ denotes the correlation factor between the arrival rate in V_n and the arrival rate in $V_n \cup V_{n'}$. Replacing α_n and β_n with $\alpha_{n,n'}$ and $\beta_{n,n'}$ in the channel model in Section 10.3.1.2, $\pi_{(0,0)}$ in (10.45) can be computed.

Assume that an entire negotiation process is considered failed when the number of blocking incident reaches the threshold $\hat{N}_{\text{bl}}$ with no success. Then, the negotiation failure probability, denoted as $\varepsilon^{(b)}$, can be expressed as:

$$\varepsilon^{(b)} = 1 - \sum_{r=0}^{\hat{N}_{\text{bl}}} \left(\theta^{(b)}\right)^{r} \left(1-\theta^{(b)}\right) \tag{10.46}$$

With (10.45) and (10.46), the expected number of blocking incidents in a successful negotiation process becomes:

$$N_{\text{bl}}^{(b)} = \frac{1}{1-\varepsilon^{(b)}} \sum_{r=0}^{\hat{N}_{\text{bl}}} \left(\theta^{(b)}\right)^{r} \left(1-\theta^{(b)}\right) r \tag{10.47}$$

Note that not every blocking incident costs the delay of a handshake since n initiates a handshake only when $|\mathcal{T}_{l,\Delta}| \geq b$. Therefore, we also need to obtain the blocking probability due to $|\mathcal{T}_{l,\Delta}| < b$, which is denoted as $\tilde{\theta}^{(b)}$. The same formula in (10.45) can be used to compute $\tilde{\theta}^{(b)}$ in a similar way as $\theta^{(b)}$. The only difference is that the $\pi_{(0,0)}$ in the expression of $\tilde{\theta}^{(b)}$ is computed using α_n and β_n, which are the PU and SU arrival rates in V_n. With $\tilde{\theta}^{(b)}$ computed, the expected number of handshake attempts for a successful negotiation can be computed as:

$$N_{\text{hs}}^{(b)} = N_{\text{bl}}^{(b)} \left(1 - \frac{\tilde{\theta}^{(b)}}{\theta^{(b)}}\right) + 1 \tag{10.48}$$

where $1 - \tilde{\theta}^{(b)}/\theta^{(b)}$ denotes the probability that $|\mathcal{T}_{l,\Delta}| \geq b$ but $|\mathcal{P}_{l,\Delta}| < b$.

Assuming that sequential sensing is used [20, 21], based on (10.48) and (10.47), we can compute the sensing delay $T_{\text{ss}}^{(b)}$ as follows. In sequential sensing, whenever a new $\mathcal{C}_{l,\Delta}$ is chosen, n needs to sense the channels in it to keep $\mathcal{T}_{l,\Delta}$ fresh, while n' senses the channels in $\mathcal{T}_{l,\Delta}$ to complete the entire negotiation. Hence, we can get:

$$T_{\text{ss}}^{(b)} = N_{\text{hs}}^{(b)} \left[\sum_{v=b}^{\Delta+1} \binom{\Delta+1}{v} \left(\pi_{(0,0)}\right)^{v} \left(1-\pi_{(0,0)}\right)^{(\Delta+1)-v} v\right] \tau_{\text{ss}} + \left(N_{\text{bl}}^{(b)}+1\right) (\Delta+1)\, \tau_{\text{ss}}, \tag{10.49}$$

where τ_{ss} denotes the average time for sensing one channel, and $\pi_{(0,0)}$ is computed using α_n and β_n. We can also compute the handshake delay as:

$$T_{\text{hs}}^{(b)} = N_{\text{hs}}^{(b)} \left(\tau_{\text{ma}} + \tau_{\text{rt}}\right), \tag{10.50}$$

where τ_{ma} denotes the average time for accessing control channel, which would be given by classic analytical models [22]; and τ_{rt} denotes the round-trip time for one handshake.

10.3.2.2 Delays in Transmission

The success of transmission cannot be guaranteed due to the occurrence of interruption incidents. In general, the delay costs for completing a successful transmission include: (i) *switching delay* $T_{\text{sw}}^{(b,d)}$, which is the time required for vacating the preempted channels and renewing the corresponding subflows; (ii) *transmission delay* T_{tx}, which is the amount of time it takes to transmit a given size of data. Typically, such transmission costs depend not only on b but also on d, i.e., the number of slots demanded for service duration after each successful negotiation. Sometimes, dividing a large size of data into smaller segments and transmitting them separately in several shorter periods can be a better choice due to reduced switching needs in each transmission periods. In the following, the switching delay $T_{\text{sw}}^{(b,d)}$ and transmission delay T_{tx} will be analyzed.

Transmission delay T_{tx} depends on the data transmission rate R_{tx}. To get the data transmission rate, note that when a PU arrival to a subchannel in an aggregated channel $\mathcal{A}^{(b)}$ is detected, the switching subflow in $\mathcal{A}^{(b)}$ is interrupted if the SU finds no available white space in the spectrum range that it is sensing (a.k.a. $|\mathcal{B}_{l,\Delta}| = 0$). Then, the transmission on that subflow fails and the overall capacity of the aggregated channel reduces. Hence, to study such bandwidth reduction of $\mathcal{A}^{(b)}$ in each slot, we define a one-step interruption probability matrix

$$\boldsymbol{\Phi} = \begin{pmatrix} \varphi_{0,0} & & & & & \\ \varphi_{1,0} & \varphi_{1,1} & & & & \\ \varphi_{2,0} & \varphi_{2,1} & \varphi_{2,2} & & & \\ \varphi_{3,0} & \varphi_{3,1} & \varphi_{3,2} & \varphi_{3,3} & & \\ \vdots & \vdots & \vdots & & \ddots & \\ \varphi_{B,0} & \varphi_{B,1} & \varphi_{B,2} & \cdots & \cdots & \varphi_{B,B} \end{pmatrix} \tag{10.51}$$

in which each $\varphi_{i,j}$ denotes the probability that $(i-j)$ subflows in $\mathcal{A}^{(i)}$ are interrupted in one slot. Note that here we assume that more than one subflow in $\mathcal{A}^{(b)}$ can be interrupted in the same slot. $\varphi_{i,j}$ can be expressed as

$$\begin{cases} \varphi_{i,j} = \sum_{u=i-j}^{i} \binom{i}{u} (\lambda_\alpha)^u (1-\lambda_\alpha)^{i-u} \binom{(\Delta+1)-i}{u-i+j} \left(\pi_{(0,0)}\right)^{u-i+j} \\ \qquad\qquad \cdot \left(1-\pi_{(0,0)}\right)^{(\Delta+1)-u-j}, \ i \in \{1, \cdots, B\}, \ j \in \{0, \cdots, i-1\}; \\ \varphi_{i,i} = 1 - \sum_{j=0}^{i-1} \varphi_{i,j}, \ i \in \{1, \cdots, B\} \end{cases} \tag{10.52}$$

Further for a d-slot transmission, the d-step interruption probability matrix $\boldsymbol{\Phi}^d = \boldsymbol{\Phi}^{d-1}\boldsymbol{\Phi}$ is used instead, in which each $\varphi_{i,j}^{(d)}$ defines the corresponding

bandwidth reduction of $\mathcal{A}^{(i)}$ within d slots. Note that $\varphi_{b,0}^{(d)}$ is the complete transmission failure probability, where all the subflows of $\mathcal{A}^{(b)}$ are interrupted. With $\boldsymbol{\Phi}^d$, the average transmission rate for a transmission attempt can be expressed as

$$R_{tx} = \left(\sum_{j=0}^{b} \varphi_{b,j}^{(d)} R j d \tau \right) \tag{10.53}$$

where R denotes the bit rate on one channel and τ is the duration of a time slot.

Next, the switching delay $T_{\text{sw}}^{(b,d)}$ for a successful d-slot subflow is analyzed. Given that there is no interruption, let $\chi^{(b)}$ be the probability that a switching operation succeeds in one slot. We have

$$\chi^{(b)} = \frac{\lambda_\alpha \left[1 - \left(1 - \pi_{(0,0)} \right)^{(\Delta+1)-b} \right]}{1 - \lambda_\alpha \left(1 - \pi_{(0,0)} \right)^{(\Delta+1)-b}} \tag{10.54}$$

in which $\pi_{(0,0)}$ is computed using $\alpha_{n,n'}$ and $\beta_{n,n'}$. Within d slots, the expected number of switching operations is

$$N_{\text{sw}}^{(b,d)} = \sum_{z=0}^{d} \binom{d}{z} \left(\chi^{(b)} \right)^z \left(1 - \chi^{(b)} \right)^{d-z} z \tag{10.55}$$

Accordingly, for a successful d-slot subflow, we have

$$T_{\text{sw}}^{(b,d)} = N_{\text{sw}}^{(b,d)} \tau_{\text{sw}} \tag{10.56}$$

where τ_{sw} denotes the time required for one switching operation. The sensing time for locating backup channels can be negligible due to the simultaneous operations of the cognitive radio and scanning radio.

10.3.3 Optimal Bandwidth Duration Decision

Based on the derived negotiation and transmission costs for performing channel aggregation, in this section, we further define an optimal channel aggregation strategy that minimizes the cumulative delay costs.

10.3.3.1 Cumulative Delay

As shown above, the delay costs for channel aggregation are closely related to the values of b and d, i.e., the user demands on aggregated bandwidth and service duration. On one hand, the choice of b should consider the trade-off between channel capacity and blocking (interruption) probability during a negotiation (transmission). More white spaces are needed to meet a higher requirement of b. On the other hand, the choice of d should consider the trade-off between negotiation overhead and interruption probability during a transmission. With certain data to transmit, one can

choose to divide the entire data into d-slot segments. A larger value of d results in fewer data segments and thus fewer negotiation operations. However, there may be more spectrum resources wasted due to higher interruption probability. Therefore, there exists an optimal combination of b and d to achieve optimal efficiency of channel aggregation.

The optimal efficiency is represented by a metric named *cumulative delay*, which is the total amount of time needed for transmitting a M-bit size data. Note that the average amount of data that can be successfully transmitted by each attempt is

$$\tilde{M}^{(b,d)} = \left(1-\varepsilon^{(b)}\right) R_{\text{tx}} \tag{10.57}$$

In addition, an attempt can meet three cases: (i) fails at negotiation stage, (ii) succeeds in negotiation but fails at transmission; and (iii) succeeds in both negotiation and transmission. The average cumulative delay, denoted as $T_{\text{cm}}^{(b,d)}$, must account for all the cases. Hence,

$$\begin{aligned} T_{\text{cm}}^{(b,d)} = \frac{M}{\tilde{M}^{(b,d)}} \Bigg\{ & \varepsilon^{(b)} \left(\hat{T}_{\text{ss}}^{(b)} + \hat{T}_{\text{hs}}^{(b)}\right) + \left(1-\varepsilon^{(b)}\right) \Big[T_{\text{ss}}^{(b)} + T_{\text{hs}}^{(b)} \\ & + \varphi_{b,0}^{(d)} \left(T_{\text{sw}}^{\left(b,\frac{d}{2}\right)} + \frac{d}{2}\tau \right) + \left(1-\varphi_{b,0}^{(d)}\right) \left(T_{\text{sw}}^{(b,d)} + d\tau \right) \Big] \Bigg\} \end{aligned} \tag{10.58}$$

in which $\hat{T}_{\text{ss}}^{(b)}$ and $\hat{T}_{\text{hs}}^{(b)}$ are computed by replacing the expected negotiation times $N_{\text{bl}}^{(b)}$ with negotiation failure threshold $\hat{N}_{\text{bl}}$ in (10.49) and (10.50), respectively, and the expected duration of failed service related to $\varphi_{b,0}^{(d)}$ is assumed to be $d/2$.

10.3.3.2 Optimal Channel Aggregation Strategy

A *channel aggregation strategy* (b, d) is defined as the combination of both bandwidth and duration demands. The objective is to find the optimal (b^*, d^*) that minimizes $T_{\text{cm}}^{(b,d)}$:

$$(b^*, d^*) = \arg \min_{(b,d) \in \mathcal{G}} T_{\text{cm}}^{(b,d)} \tag{10.59}$$

It is not hard to find (b^*, d^*) by searching the finite set of all possible (b, d)'s. Note that in CR networks, both PU and SU behaviors that affect the availability of white spaces are stochastic. Hence, the optimal channel aggregation strategy defined in (10.59) is actually optimal in the sense of average performance.

10.3.4 Numerical Analysis and Simulation Results

To illustrate and validate the analytical results, figures that are derived from numerical analysis and discrete-event simulation of a few of our analytical results are shown in this section, including the negotiation failure probability $\varepsilon^{(b)}$ in (10.46), the transmission failure probability $\varphi_{b,0}^{(d)}$ defined in Section 10.3.2.2, and the cumulative delay $T_{\text{cm}}^{(b,d)}$ in (10.58). These figures will show the impact of PU activity and channel aggregation strategy on the efficiency of negotiation and transmission between secondary users.

In the numerical analysis, a pair of PU sender and receiver, n and n', in a distributed CR network is considered. Poisson PU arrival process is assumed in both V_n and $V_{n'}$ areas. The distribution of PU service duration is set according to the statistical distribution of call duration collected from a real cellular network [5]. As for SU behavior, Poisson SU arrival process and random SU service duration are assumed. Note that the choice of service duration for the pair of n and n' is a part of their optimal decision, but we fix the patterns of SU activity in the background. The constant parameters are set as follows: $K = 50$; $B = 3$; $\beta_n = 0.02$ user/s; $c_{n,n'} = 1.5$; $E[Y] = 3$ s; $\tau = 10$ ms; $\tau_{\text{ss}} = 10$ ms [18, 21]; $\tau_{rt} = 200$ ms; $\tau_{\text{sw}} = 600$ ms [20]; $\hat{N}_{\text{bl}} = 5$; $M = 50$ Mb; $R = 5$ Mb/s. The others are viewed as variables.

10.3.4.1 Illustration of Negotiation Failure Probability

As in (10.46), $\varepsilon^{(b)}$ defined for negotiation failure which characterizes the repeated blocking incidents caused by lack of enough common white space at n and n' (a.k.a. $|\mathcal{P}_{l,\Delta}| < b$). The impact of α_n and Δ on $\varepsilon^{(b)}$ with fixed transmission duration of $d \cdot \tau = 3$ s is plotted in Fig. 10.10. Generally, the numerical results (marked as "ana") and simulation results (marked as "sim") match well with each other under the same settings. It can be seen that with a higher demand on b or a drop in the availability of white spaces, $\varepsilon^{(b)}$ increases. In addition, a relaxation of the hardware limitation on Δ offers more candidate channels and thus lowers $\varepsilon^{(b)}$.

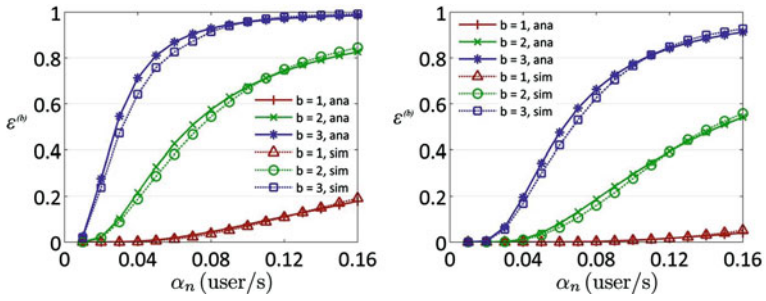

Fig. 10.10 Negotiation failure probability vs. PU arrival rate: (i) $\Delta = 10$ (*left*); (ii) $\Delta = 20$ (*right*)

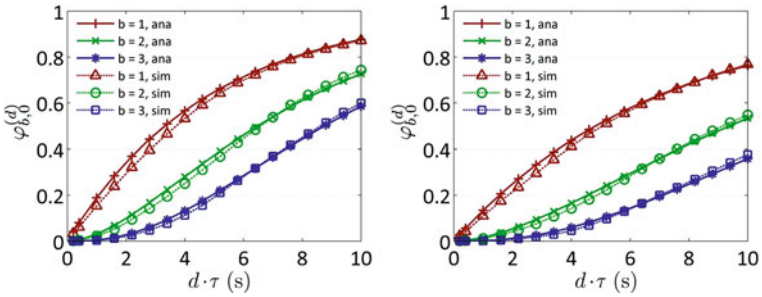

Fig. 10.11 Transmission failure probability vs. duration demand: (i) $\Delta = 10$ (*left*); (ii) $\Delta = 20$ (*right*)

10.3.4.2 Illustration of Transmission Failure Probability

The impact of transmission duration d on transmission failure probability $\varphi_{b,0}^{(d)}$ is shown at Fig. 10.11, where the PU traffic arrival rate is fixed at $\alpha_n = 0.1$ user/s. Defined in Section 10.3.2.2, $\varphi_{b,0}^{(d)}$ is the probability that all subflows of an aggregated channel are interrupted halfway during a transmission by PUs. Clearly, with longer transmission duration $d\tau$ or a smaller channel separation constraint Δ, the $\varphi_{b,0}^{(d)}$ increases. Interestingly, unlike negotiation failure probability $\varepsilon^{(b)}$, a larger transmission duration b actually reduces transmission failure probability $\varphi_{b,0}^{(d)}$. Intuitively, this is because the transmission consisting of more subflows would tolerant more interruption incidents. Hence, a trade-off obviously exists among the negotiation failure probability and transmission failure probability to achieve the overall optimal transmission strategy.

10.3.4.3 Illustration of Cumulative Delay

For the transmission of M-bit data, the related cumulative delay $T_{\text{cm}}^{(b,d)}$ has been chosen as our objective function as in (10.58). In Figs. 10.12 and 10.13, the impact of α_n and (b, d) on $T_{\text{cm}}^{(b,d)}$ is plotted, respectively. It can be seen that $T_{\text{cm}}^{(b,d)}$ rises rapidly with the increase of α_n due to the increase of $\varepsilon^{(b)}$ and $\varphi_{b,0}^{(d)}$ in such cases. A larger channel separation constraint Δ lowers $T_{\text{cm}}^{(b,d)}$. To achieve the lowest $T_{\text{cm}}^{(b,d)}$, the optimal channel aggregation strategy (b^*, d^*) is evaluated under different settings in Fig. 10.13. The marked point that represents the optimal decision varies significantly with the availability of white spaces. Obviously, when there are plenty of white spaces as in Fig. 10.13-i, larger b and d are the optimal solution to achieve the highest utilization of licensed spectrum. Note that the range of d differs for different values of b for transmitting the same size of data. However, if there are few white spaces as in Fig. 10.13-iii, both b and d should be low to avoid the huge costs for repeated negotiation and transmission attempt.

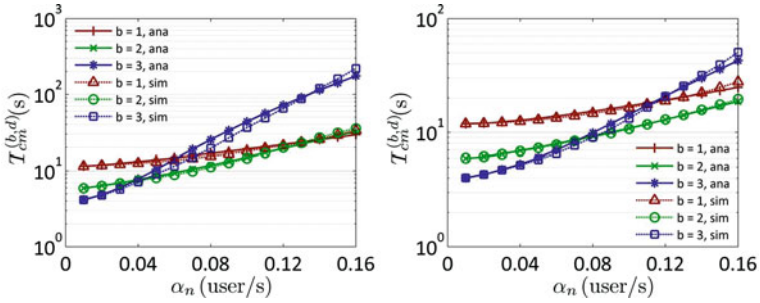

Fig. 10.12 Cumulative delay (in log scale) vs. PU arrival rate: (i) $\Delta = 10$ (*left*); (ii) $\Delta = 20$ (*right*)

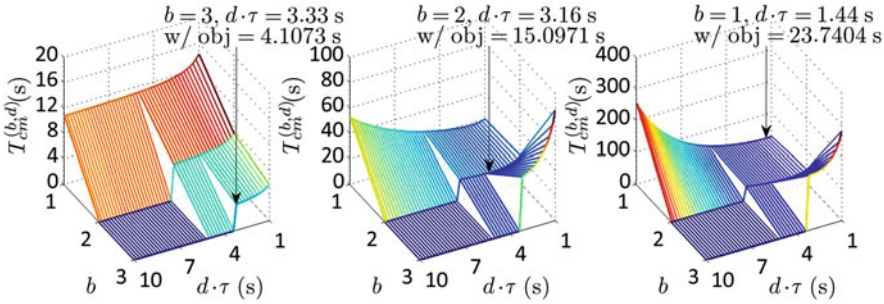

Fig. 10.13 Cumulative delay vs. channel aggregation strategy: (i) $\alpha_n = 0.01$ user/s (*left*); (ii) $\alpha_n = 0.1$ user/s (*middle*); (iii) $\alpha_n = 0.16$ user/s (*right*)

10.4 Summary

In this section, we analyzed the delay for both multihop and single-hop CRN.

For multihop CRN, we derive the IPS upper bound under two cases: the network IPS and the flow IPS. In the network IPS case, we discover that the network IPS upper bound is related to a threshold value of the PU activity level. Below the threshold, the IPS upper bound is achieved when the one-hop distance equals the communication range of cognitive radios. Above the threshold, the upper bound speed is achieved when using an optimal one-hop distance which is less than the communication range. We design efficient numerical methods to compute the optimal one-hop distance and the corresponding IPS upper bound. In the flow IPS case, we discover that the IPS upper bound is achieved when an optimal number of SU relay nodes are evenly spaced on the straight line between the source node and the destination node. The optimal number of relay nodes shows a stair-like incremental trend, when the PU activity level increases. We design multiple numerical methods to compute the optimal number of SU relay nodes. The simulation and numerical results prove the correctness of our analysis.

For single-hop CRN, we have studied the delay under considerations of various practical constraints and costs. A new channel usage model based on general assumptions is introduced to investigate the negotiation and transmission costs for

utilizing channel aggregation under the influence of PU activity. We have found that user demands on both aggregated bandwidth and service duration affect the delay performance. Hence, an optimal channel aggregation strategy has been defined and validated to achieve the lowest delay for data transmission.

Acknowledgments This work was supported in part by the US National Science Foundation under grant CNS-0831865 and the Institute for Critical Technology and Applied Science (ICTAS) of Virginia Tech.

References

1. Y. Xu and W. Wang, "The speed of information propagation in large wireless networks," in *Proc. of IEEE INFOCOM*, Phoenix, AZ, 2008.
2. R. Zheng, "Information dissemination in power-constrained wireless networks," in *Proc. of IEEE INFOCOM*, Barcelona, Catalunya, Spain, 2006.
3. P. Jacquet, B. Mans, P. Muhlethaler, and G. Rodolakis, "Opportunistic routing in wireless ad hoc networks: Upper bounds for the packet propagation speed," in *Proc. of IEEE MASS*, Atlanta, Georgia, USA, 2008.
4. P. Jacquet, B. Mans, and G. Rodolakis, "Information propagation speed in mobile and delay tolerant networks," in *Proc. of IEEE INFOCOM*, Rio de Janeiro, Brazil, 2009.
5. D. Willkomm, S. Machiraju, J. Bolot, and A. Wolisz, "Primary user behavior in cellular networks and implications for dynamic spectrum access," *IEEE Communications Magazine*, vol. 47, pp. 88–95, 2009.
6. P. Popovski, H. Yomo, K. Nishimori, R. D. Taranto, and R. Prasad, "Opportunistic interference cancellation in cognitive radio systems," in *Proc. of IEEE DySPAN*, Dublin, Ireland, 2007.
7. R. Zhang and Y.-C. Liang, "Exploiting hidden power-feedback loops for cognitive radio," in *Proc. of IEEE DySPAN*, Chicago, Illinois, 2008.
8. G. Zhao, G. Y. Li, and C. Yang, "Proactive detection of spectrum opportunities in primary systems with power control," *IEEE Transactions on Wireless Communications*, vol. 8, pp. 4815–4823, 2009.
9. G. Bolch, S. Greiner, H. de Meer, K. S. Trivedi, H. de Meer, and K. S. Trivedi, *Queueing Networks and Markov Chains: Modeling and Performance Evaluation With Computer Science Applications*. Wiley-Interscience, Wiley, 2006.
10. C. Han and Y. Yang, "Information propagation speed in cognitive radio networks: Network and flow analysis," Virginia Tech, Tech. Rep., 2010.
11. R. Chandra, R. Mahajan, T. Moscibroda, R. Raghavendra, and P. Bahl, "A Case for Adapting Channel Width in Wireless Networks," in *Proc. ACM SIGCOMM'08*, Seattle, WA, Aug. 2008.
12. IEEE 802.22 WG, "IEEE P802.22/D0.1 Draft 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 TV Bands," *IEEE Standard*, May 2006.
13. S. Haykin, "Cognitive Radio: Brain-Empowered Wireless Communications," *IEEE Journal on Selected Areas in Communications*, Vol. 23, No. 2, pp. 201–220, Feb. 2005.
14. H. Kim, and K. G. Shin, "Efficient Discovery of Spectrum Opportunities with MAC-Layer Sensing in Cognitive Radio Networks," *IEEE Transactions on Mobile Computing*, Vol. 7, No. 5, pp. 533–545, May 2008.
15. F. Huang, W. Wang, H. Luo, G. Yu, and Z. Zhang, "Prediction-Based Spectrum Aggregation with Hardware Limitation in Cognitive Radio Networks," in *Proc. IEEE VTC'10-Spring*, Taipei, Taiwan, May 2010.

16. P. Bahl, A. Adya, J. Padhye, and A. Wolman, "Reconsidering Wireless Systems with Multiple Radios," *ACM SIGCOMM Comp. Comm. Rev.*, Vol. 34, No. 5, pp. 39–46, Oct. 2004.
17. Y. Yuan, P. Bahl, R. Chandra, P. Chou, J. Ferrell, T. Moscibroda, S. Narlanka, and Y. Wu, "KNOWS: Kognitiv Networking Over White Spaces," in *Proc. IEEE DySPAN'07*, Dublin, Ireland, Apr. 2007.
18. Y. Yuan, P. Bahl, R. Chandra, T. Moscibroda, and Y. Wu, "Allocating Dynamic Time-Spectrum Blocks in Cognitive Radio Networks," in *Proc. ACM MobiHoc'07*, Montreal, QC, Canada, Sep. 2007.
19. J. Lee, and J. So, "Analysis of Cognitive Radio Networks with Channel Aggregation," in *Proc. IEEE WCNC'10*, Sydney, Australia, Apr. 2010.
20. D. Xu, E. Jung, and X. Liu, "Optimal Bandwidth Selection in Multi-Channel Cognitive Radio Networks: How Much Is Too Much?," in *Proc. IEEE DySPAN'08*, Chicago, Illinois, Oct. 2008.
21. T. Shu, and M. Krunz, "Throughput-Efficient Sequential Channel Sensing and Probing in Cognitive Radio Networks under Sensing Errors," in *Proc. ACM MobiCom'09*, Beijing, China, Sep. 2009.
22. G. Bianchi, "Performance Analysis of IEEE 802.11 Distributed Coordination Function," *IEEE Journal on Selected Areas in Communications*, Vol. 18, No. 3, pp. 535–547, Mar. 2000.