Tuning Radio Resource in an Overlay Cognitive Radio Network for TCP: Greed Isn't Good

Teerawat Issariyakul, Laxminarayana S. Pillutla, and Vikram Krishnamurthy,
The University of British Columbia

This work was supported in part by the Canadian Natural Sciences and Engineering Research Council and by the Asia-Pacific Telecommunity.

ABSTRACT

In this article we illustrate the performance of Transmission Control Protocol in an overlay cognitive radio network under dynamic spectrum access. We show that the performance of TCP in overlay CR networks that implement DSA to be quite different from its performance in conventional networks, which do not allow DSA. The key difference is that secondary users in an overlay CR network have to cope with a new type of loss called service interruption loss, due to the existence of primary users. We demonstrate on an NS2 simulation testbed the surprising result: Excessive radio resource usage leads to a decrease in aggregate TCP throughput. This behavior is in contrast to the behavior of TCP in conventional networks, where throughput increases monotonically with the available radio resource.

INTRODUCTION

Cognitive radio (CR) is an emerging communications paradigm, wherein a wireless transceiver unit can sense the surrounding environment and adapt itself accordingly. CR technology, along with dynamic spectrum access (DSA), has the potential to alleviate the shortage of radio resource. For many years, it was believed that the spectrum shortage was due to an increasing number of wireless applications and their substantial bandwidth usage. However, spectrum measurements in Washington, DC, New Orleans, San Diego, Atlanta, Chicago, and other metropolitan areas show that vast portions of licensed spectrum are not in use (i.e., they are underutilized) [1]. This finding suggests that an improvement in spectrum utilization can alleviate the spectrum shortage problem. A transceiver equipped with CR capability can access and release spectrum with more agility. Such agility can allow secondary users (who are unlicensed) to access the spectrum when primary users (i.e., spectrum licensees) are inactive. Referred to in the literature as DSA, this approach clearly reduces spectrum underutilization, hence alleviating the spectrum shortage problem.

Most of the CR research focused on the lower layers (i.e., physical and medium access control [MAC] layers) of the network protocol stack. Despite sporadic works that study routing issues in CR networks, no work to date has considered transport layer issues in CR networks [2]. In this article we study the performance of Transmission Control Protocol (TCP) in a CR network.

TCP has become a ubiquitous transport layer protocol for reliable non-real-time data transmission over the Internet. TCP was originally designed to control network congestion in wired networks, and was later enhanced to cope with channel errors in wireless networks [3]. The advent of DSA introduces a new type of packet loss called *service interruption loss* (see the definition given later), in addition to the existing packet losses resulting from network congestion and channel errors. Consequently, there is a strong motivation to study the performance of TCP under service interruption loss in CR networks that employ DSA.

This article provides a detailed survey of CR networks. We present their salient features, access techniques, and architectures. We also discuss a surprising result regarding TCP performance in overlay CR networks (see the definition given in the next section): the TCP throughput of a secondary user does not always increase with the available radio resource. Due to service interruption from the primary user, TCP performance of a secondary user could degrade when it tries to acquire too much radio resource. This finding poses the need to optimally control how secondary users obtain the radio resource in order to fully utilize the available radio resource.

The rest of this article is organized as follows. The next section discusses CR evolution and provides a survey of various regulatory issues, access techniques, and architectures proposed in the literature. We then describe the basic operation of TCP in conventional networks, and emphasize the need to study TCP performance over a CR network. The following section shows our simulation study of TCP performance in CR networks allowing DSA. Conclusions are given in the final section.

 0163-6804/09/$25.00

A general centralized network architecture consists of two main entities. One is a base station, which schedules the data transmission of users in the network. The other entity is the spectrum broker, which is responsible for allocating the radio resource to users.

CR: Evolution, Access Techniques, and Architectures

Evolution and Salient Features of CR

CR is a key enabling technology to achieve DSA of licensed spectrum bands. The current interest in CR technologies is due to a report published by the Federal Communications Commission (FCC) pointing out that vast portions of spectrum are underutilized [1]. This finding suggested that granting access to unlicensed users could lead to significant improvement in spectrum utilization.

The term CR was coined by Mitola in [4], although its genesis can be traced to the concept of software defined radio (SDR). According to the FCC, "an SDR is one in which operating parameters such as frequency range, modulation type, or output power can be altered by software without making any changes to the hardware components." Taking the definition of SDR a step further, Mitola defined CR as "an SDR that senses its environment, tracks changes and reacts accordingly" [4].

CR is characterized by two main features:

- *Cognitive capability*: This refers to the ability of the radio to identify spectrum opportunities and adapt to the situation accordingly. The typical steps involved in this adaptive operation are *spectrum sensing*, *spectrum analysis*, and *spectrum decision*. Spectrum sensing involves detection of white spaces (i.e., the spectrum portion not in use). Spectrum analysis involves analyzing the characteristics of the detected white space (e.g., characterizing the white space based on its time-varying radio characteristics and the primary user activity). The spectrum decision specifies the action (e.g., adjusting transmission rate, power levels, and/or bandwidth) that should be taken after spectrum sensing and analysis [2, 5].
- *Reconfigurability*: This refers to the ability of the radio to be configured dynamically according to the environment. The various parameters that can be reconfigured are, for example, operating frequency, modulation type, and transmission power [5].

A general CR network consists of two subnetworks: a *primary network* and a *secondary network*. The primary network consists of primary users and the primary base station, whose function is to control and regulate the users of the network, similar to a cellular system. The secondary network does not own a license and tries to utilize the spectrum in an opportunistic manner. A secondary network may or may not have a base station, depending on the architecture.

We note that CRs can be conceived for even the unlicensed bands of the spectrum, as noted in [2]. However, in this article we focus our attention on CR networks for access to the licensed bands only.

Spectrum Access Policy: Own It or Share It?

The inception of CR has raised spectrum regulatory issues. Two schools of thought — *property rights* and *spectrum commons* — are being debated extensively [6]. Advocates of property rights argue that spectrum owners should have an absolute right over their spectrum. Proponents of spectrum commons, on the other hand, believe that sharing spectrum would be more efficient and greatly alleviate the problem of spectrum underutilization. Hence, there are constant regulatory and legal policy discussions on these two conflicting ideas, the result of which is still inconclusive.

Access Techniques for CR Networks

Access techniques for CR networks can be classified into two types:

- *Overlay approach (or interference-free approach)*: In this approach the secondary users access the portion of the spectrum that is not used by primary users. As a result, there is virtually no interference to the primary users. We refer to CR networks that employ overlay access techniques as *overlay CR networks*.
- *Underlay approach (or interference-tolerant approach)*: In this approach the secondary users access the network by spreading their signals over a wide frequency band. The underlay approach imposes severe constraints on the transmission power of secondary users. Operating below the noise floor of primary users, secondary users are allowed to interfere with primary users up to a certain tolerable level. We refer toCR networks that employ underlay access techniques as *underlay CR networks*.

The current spectrum management policy adopted by the FCC, based on the property rights model, does not allow secondary user operation in the licensed spectrum, irrespective of the access techniques. Hence, there may be regulatory and legal discussions on how secondary users should operate in the licensed spectrum.

Architectures for CR Networks

The architecture for CR networks can be either *centralized* or *distributed*. Table 1 summarizes various CR architectures proposed in the literature.

Centralized CR Network Architectures — A general centralized network architecture consists of two main entities. One is a base station, which schedules the data transmission of users in the network. The other entity is the spectrum broker, which is responsible for allocating the radio resource to users (primary, secondary, or both). The spectrum broker can be a primary or secondary base station [2], or a dedicated entity dealing with spectrum allocation [7, 9]. The sensing functionality in a centralized CR network can be performed by secondary users or the spectrum broker. The sensed spectrum information is used by the spectrum broker to create a spectrum allocation map for radio resource allocation.

The following are examples of centralized CR architectures proposed in the literature. A more comprehensive survey of CR architectures is given in [2].

	Centralized	Distributed
Sensing entity	Secondary users/spectrum broker	Secondary users
Spectrum broker	Separate entity/primary base station/secondary base station	None
Access granted to	Primary and secondary users	Secondary users
Examples	Spectrum pooling [7] DIMSUMnet [9]	CORVUS [8] Nautilus [10, 11]

Table 1. *Summary of cognitive radio architectures.*

In general, the suggested solutions to improve performance of TCP in wireless networks are classified into three basic groups based on their fundamental philosophies: end-to-end solutions, split-connection solutions, and link-layer solutions.

Spectrum Pooling — In this architecture spectrum from different owners is put together in a common pool [7]. The architecture is based on the well-known orthogonal frequency-division multiplexing (OFDM). The advantage of an architecture based on OFDM is that interference to primary users can be minimized by allocating zero power to the subcarriers occupied by them.

Dynamic Intelligent Management of Spectrum for a Ubiquitous Mobile Network — In dynamic intelligent management of spectrum for a ubiquitous mobile network (DIMSUMnet) [9], a spectrum broker owns a contiguous chunk of spectrum called the coordinated access band (CAB).[1] The CAB is then leased according to requests from different users for a specified time.

Distributed CR Network Architectures — This type of architecture has no centralized agent like a spectrum broker or base station to coordinate spectrum access of secondary users. It is similar to that used in wireless ad hoc networks. The main difference from wireless ad hoc networks is the presence of primary users and service interruption loss, which will be defined in the next section.

Distributed CR networks can be classified into *cooperative* or *non-cooperative* networks. In a cooperative network the users share the interference information and determine spectrum allocation based on this shared information. In a non-cooperative network the users access the spectrum based on local policies, since there is no communication regarding interference information among users of the network. In general, the cooperative approach achieves better throughput than the non-cooperative approach. However, a cooperative approach may incur extra overhead due to the communication among users.

We now briefly summarize the existing distributed CR network architectures.

Cognitive Radio Approach for Usage of Virtual Unlicensed Spectrum — In the CR approach for usage of virtual unlicensed spectrum (CORVUS) [8], secondary users form a communication group called a secondary user group (SUG). The secondary users in an SUG help each other in sensing the channel for primary user activity. If primary users are inactive on a channel, the SUG takes that channel for communication within the group.

We note that CORVUS can be used even in the centralized mode. In each SUG a secondary user could be selected to control spectrum access of other secondary users.

Nautilus — This project devises distributed, efficient, and scalable algorithms for spectrum sharing in CR ad hoc networks [10, 11]. The project proposes spectrum access algorithms based on two approaches: one based on graph coloring [10] and another based on local bargaining [11]. In the former approach the spectrum allocation is transformed into a graph-coloring problem, where a secondary user corresponds to a vertex, while a radio channel corresponds to a color. Since a graph-coloring problem is NP-hard, the authors propose heuristic approximation algorithms to solve the problem [10]. In the latter approach the secondary users affected by topological changes form local bargaining groups and adapt their spectrum assignment to a new optimal conflict free assignment [11].

TCP in Conventional vs. CR Networks

TCP in Wired and Conventional Wireless Networks

TCP was originally conceived for wired networks as a means to avoid and control network congestion, and to provide reliable end-to-end delivery of user data. Whenever the sender identifies loss of data (either in the form of several duplicated acknowledgments or an absence of acknowledgment for a timeout interval), TCP reacts to the loss by reducing its transmission window size before retransmitting lost packets. This reduction in window size eases the load on intermediate links, hence controlling congestion in the network. The subsequent increase in window size depends on whether TCP is in *slow start phase* or *congestion avoidance phase*. In the slow start phase the window size grows linearly in response to every acknowledged packet, while the window size grows sublinearly in the congestion avoidance phase. In general, the essential functionality of TCP remains the same, despite the modifications to improve its performance. For example, TCP variants (TCP-Tahoe, TCP-Reno, and TCP-Vegas) implement the same window-based adaptation mechanism, but have minor differences.

[1] *CABs can be designated frequency bands in cellular, PCS, or TV broadcasting systems.*

TCP performance under the congestion losses, channel error losses, and/or data collision losses has been studied extensively. However, due to the service interruption loss, TCP performance in an overlay CR network could be significantly different from that in a conventional network.

TCP was originally developed for wired networks and later enhanced for wireless networks. Unlike in a wired network where packet loss is mainly caused by congestion, the majority of packet loss in a wireless network is due to channel errors (caused by fading, interference, and shadowing) or packet collision (i.e., when more than one mobile accesses the channel simultaneously). While two consecutive packet losses due to network congestion are highly correlated, those due to channel errors or collision may not be related at all.

In general, the suggested solutions to improve performance of TCP in wireless networks are classified into three basic groups based on their fundamental philosophies: *end-to-end solutions*, *split-connection solutions*, and *link-layer solutions* [3]. Among these three solution approaches, link-layer solutions are the most popular, since they do not require any modification of the network architecture or TCP operation.

SERVICE INTERRUPTION LOSS IN OVERLAY CR NETWORKS

The concept of DSA introduces a new type of loss called service interruption loss for the secondary users of an overlay CR network. *Service interruption loss refers to the loss experienced by secondary users due to the intervention of primary users while transmitting data*. Service interruption loss is absent in underlay CR networks, since in an underlay approach the secondary users operate far below the power level of the primary users, hence ruling out the possibility of collisions with the primary users of the spectrum.

The amount of service interruption loss for the secondary users of the overlay CR network is determined by the amount of primary user activity in the network. To capture the amount of primary user activity, we define load factor (A) as the amount of traffic *all* the primary users generate on a *single* channel. Mathematically, it is stated as

$$A = \frac{T_{ON}}{T_{ON} + T_{OFF}} \times \frac{N_p}{L}, \quad (1)$$

where T_{ON} and T_{OFF} denote the mean ON time and OFF time of the primary users, N_p denotes the number of primary users, and L denotes the total number of channels in the spectrum.

The rationale behind Eq. 1 is as follows. First, the average traffic generated by each primary user is $T_{ON}/(T_{ON} + T_{OFF})$. Hence, the total traffic generated by all the primary users is $N_p \cdot T_{ON}/(T_{ON} + T_{OFF})$. Since this traffic is generated over L radio channels, the total traffic (or load) generated by the primary users on a single channel is given by (A) defined as in Eq. 1. In other words, if the traffic generated by all the primary users is distributed equally among all the radio channels, A in Eq. 1 denotes the amount of primary user traffic on one radio channel. For example, $A = 0.5$ implies that every radio channel is used by all primary users on average 50 percent of the time. In general, a higher value of the load factor (A) implies lower aggregate TCP throughput of the secondary users.

The service interruption loss is different from the following three major losses that occur in a conventional network. First, losses due to network congestion are usually caused by buffer overflow and are assumed to be correlated. Second, losses due to channel errors depend on channel characteristics. Depending on whether the user is moving slowly or rapidly, the losses due to channel errors can be assumed correlated or independent. The third and final type of losses are due to collision between various users of the network that depend on the underlying medium access control (MAC) mechanism. Service interruption loss due to DSA, on the other hand, does not depend on network conditions, channel characteristics, or the underlying MAC mechanism. It depends on the amount of primary user activity in the spectrum, which in turn depends on extraneous factors such as geographical location and time of the day. The report in [1] notes that in a given geographical area, for the most part the spectrum is devoid of primary user activity. In such cases the duration of service interruption loss may be small, since it is possible for the secondary user to find an alternate channel for transmission almost immediately. However, if a given geographical region has heavy primary user activity, it may not be possible for a secondary user to find an alternate channel immediately, hence leading to a longer duration of service interruption loss.

TCP performance under congestion losses, channel error losses, and/or data collision losses has been studied extensively. However, due to service interruption loss, TCP performance in an overlay CR network could be significantly different from that in a conventional network. This poses a clear need to study TCP performance in an overlay CR network. As we shall see from our simulation study presented in the next section, a definite number of channels must be carefully chosen in order to optimize TCP performance in an overlay CR network.

SIMULATION STUDY OF TCP PERFORMANCE IN AN OVERLAY CR NETWORK

SYSTEM MODEL OF AN OVERLAY CR NETWORK

In this article we consider an overlay CR network with N_p primary users and N_s secondary users (Fig. 1). Primary users send their data to a primary network via its base station. On the other hand, each secondary user transfers a file with infinite size to the server connecting to its base station. Both primary and secondary users share L radio channels to communicate with their base stations. We assume that the data rate and propagation delay of each radio channel are b_r b/s and d_r s, respectively. Also, the link that connects the base station and server has a data rate of b_l b/s and a propagation delay of d_l s. Each secondary user is allowed to access the free channels only. While using a channel, a secondary user may be asked (by a primary user) to give up its ongoing transmission and return the channel to the primary user. In this case the packet transmitted by the secondary user will be lost due to service interruption.

We model the primary user activity according to a continuous-time Markov chain with two states, ON and OFF. This modeling assumption implies each primary user generates a data packet that holds the channel for an exponentially distributed time with mean T_{ON}. After transmitting a packet, a primary user turns off for an exponentially distributed time with mean T_{OFF}. When switching to an ON state, a primary user first looks for a free channel to transmit data. If there is no free channel, a primary user will force one of the secondary users (if any) to give up transmission on the radio channel. If all the channels are occupied by primary users, the new primary user will not be able to transmit its data.

Each secondary user transfers a file using TCP as its transport layer protocol. Again, based on a perceived packet loss pattern, TCP feeds and stops feeding packets to the underlying link layer. A secondary user attempts to transmit each link layer packet according to the following spectrum access mechanism. At any time, a secondary user can access a maximum of L_s radio channels, which implies there are always $(L - L_s)$ channels reserved for the primary users. Among the L_s channels, a secondary user chooses to transmit (or) not to transmit on a free channel with probability p and $1 - p$, respectively.[2] If more than one secondary user chooses to transmit on the same channel simultaneously, all the transmitted packets will be lost. The above access mechanism is invoked repeatedly until all the link layer packets are transmitted or the number of occupied channels is greater than L_s.

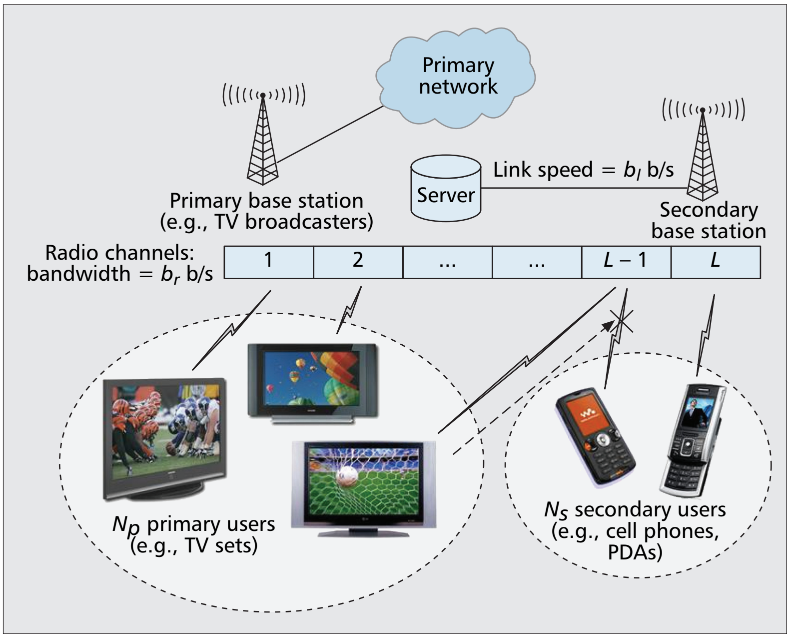


Figure 1. *System model of an overlay CR network:* N_p *primary users and* N_s *secondary users access their base stations via* L *shared radio channels. Each radio channel has a bandwidth of* b_r *b/s. The primary base station connects to the primary network, while the secondary base station connects to a server via a wired link with a bandwidth of* b_l *b/s. When needed, a primary user can force a secondary user to give up transmission so that it can use the released channel.*

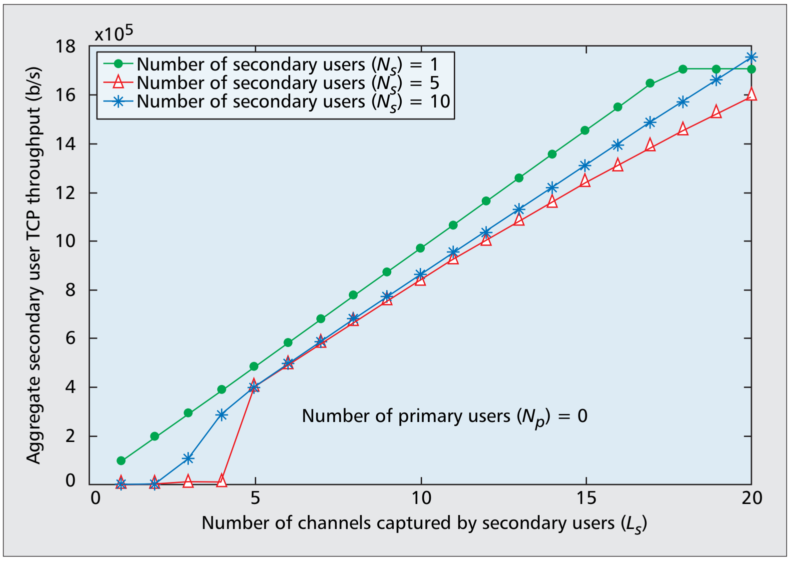


Figure 2. *Aggregate secondary user TCP throughput vs. the maximum channels for secondary users* (L_s)*when* $N_p = 0$*. Due to the absence of service interruption from the primary users, this case is similar to an ad hoc network with random access. The figure shows that the aggregate TCP throughput of secondary users increases monotonically with* L_s.

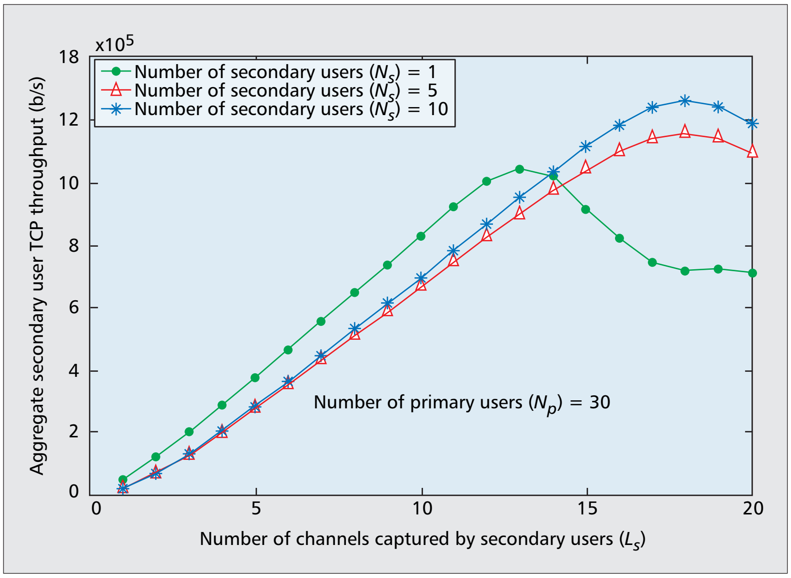
Simulation Parameters

We run the simulation using an open source network simulation tool, NS2. Unless specified otherwise, we assume $N_p = 30$, $L = 20$, $T_{OFF} = 0.1$ s, $b_r = 100$ kb/s, $d_r = 2$ ms, $b_l = 100$ Mb/s, and $d_l = 2$ ms. The value of T_{ON} is determined by the choice of the load factor value (A). The value of b_r chosen above is typical for wireless data applications. For example, the code-division multiple access (CDMA) 2000 1xEV-DO standard (which is one of the six radio interfaces under IMT-2000) supports a maximum data rate of 154 kb/s [12]. We run the simulation for a duration of 500 s, which we found to be sufficient for TCP to reach its steady state. Each secondary user employs TCP New Reno[3] to upload the file, and chooses transmission probability value (p) equal to $1/N_s$, where N_s denotes the number of secondary users.[4] Finally, we assume there are no losses due to channel errors (or) buffer overflow on all the TCP flows of secondary users.

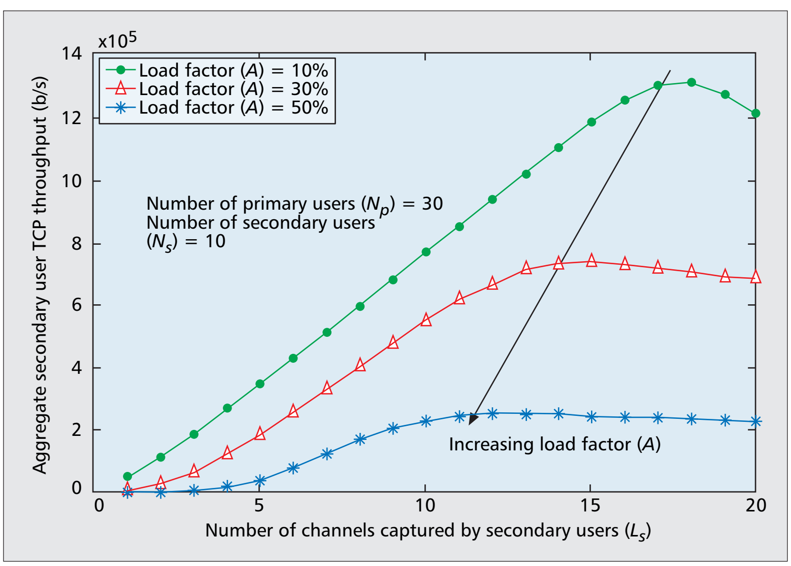
Simulation Results

Figure 2 shows the plot of aggregate TCP throughput of secondary users vs. the maximum channels for secondary users (L_s) for the case when there are no primary users (i.e., $N_p = 0$). With no primary users, this case corresponds to a conventional random access ad hoc network. As can be seen from Fig. 2, without service interruption from primary users, the aggregate TCP throughput of secondary users increases monotonically as L_s increases. Using random access, the secondary user experiences no collision for $N_s = 1$. Therefore, this case has higher aggregate TCP throughput than the multi-user case (i.e., $N_s = 5$ and $N_s = 10$). Note that the aggregate TCP throughput for $N_s = 1$ saturates at $L_s = 18$ since the TCP window size has reached its maximum value. However, for the multi-user case (i.e., $N_s = 5$ and 10), the aggregate TCP throughput increases continuously since in these

[2] *This type of random access is widely used for ad hoc wireless networks, where there is no base station to control and coordinate the transmission of secondary users.*



■ **Figure 3.** *Aggregate secondary user TCP throughput vs. the maximum channels for secondary users* (L_s) *in the presence of service interruption due to 30 primary users. The figure shows that the throughput is no longer monotone with respect to* L_s*. There exists a well defined optimal value of* L_s *that maximizes the aggregate TCP throughput.*



■ **Figure 4.** *Aggregate secondary user TCP throughput vs. the maximum channels for secondary users* (L_s) *for various values of load factor* (A) *with service interruption due to 30 primary users. The figure shows that the aggregate TCP throughput of secondary users decreases as the load factor increases.*

cases the TCP window size has not reached its maximum.

Figure 3 shows the impact of service interruption on the aggregate TCP throughput of secondary users. Here, we set the number of primary users (N_p) equal to 30. In this case the throughput gradually increases as the maximum channels for secondary users (L_s) increase up to a certain point beyond which the throughput decreases. Hence, there exists a well defined optimal value of L_s (denoted L_s^*) that maximizes the aggregate TCP throughput.

This non-monotonic behavior of TCP in overlay CR networks that implement DSA can be explained as follows. Intuitively, as the maximum channels for secondary users (L_s) increase, we expect an increase in aggregate TCP throughput. However, an increase in L_s also increases the probability of service interruption from a primary user, leading to increased packet loss. This packet loss (detected via timeout or duplicate acknowledgment) leads to a decrease in TCP window size, and hence significant degradation of aggregate TCP throughput of secondary users. When adopting DSA in overlay CR networks, secondary users need to be judicious in choosing the number of channels for data transmission.

We also observe from Fig. 3 that the optimal value of L_s (L_s^*) also increases as the number of secondary users increases. Since every secondary user accesses the channel with the same probability $p = 1/N_s$, on average, all the free channels are distributed equally among the secondary users. For example, suppose the number of free channels are 15. In a single-user case all 15 free channels are used by only one user. When N_s = 5, each secondary user employs, on average, 15/5 = 3 channels to transmit TCP packets. With a small number of channels per user, a success in single transmission could gain a significant increase in TCP throughput compared to the packet loss due to service interruption. Therefore, the peaks (i.e., L_s^*) in Fig. 3 shift to the right as the number of secondary users increases.

Figure 4 shows the variation of aggregate TCP throughput with respect to the maximum number of channels L_s for different values of the load factor (A). In this case we set N_p = 30 and N_s = 10, and the load factor A is chosen to be 10, 30, and 50 percent, respectively. A higher value of the load factor implies an increase in the probability of service interruption from the primary user and also a decrease in transmission opportunities for a secondary user. Hence, the aggregate TCP throughput of the secondary users decreases as the load factor increases.

Figure 5 plots the optimal value of L_s (L_s^*) vs. the load factor (A). We set the number of primary and secondary users to be 30 and 1, respectively. The non-monotone behavior of L_s^* with respect to the load factor A in Fig. 5 can be explained as follows. First, a small load factor implies that primary users rarely need the channel; service interruption is rare in this case. With small service interruption probability, a secondary user should take a chance and transmit on as many channels as possible to increase TCP throughput. Second, as the load factor increases, the service interruption probability increases. In this case we should decrease L_s^* to avoid service interruption. The decrease of L_s^* (with respect to A) stops when the traffic load is extremely large. In this case primary users tend to occupy all the channels for most of the time, and secondary users may not be able to acquire a channel. Therefore, secondary users should take a chance and transmit packets on as many channels as possible, even though they are prone to increased service interruption from the primary users. Due to the above non-monotone behavior, a secondary user needs to be judicious in setting the maximum number of channels for the secondary users viz. L_s.

[3] *TCP New Reno is an improved version of Reno that avoids multiple reductions of congestion window size when several TCP packets from the same window of data are lost [13].*

[4] *This choice of* $p = 1/N_s$ *minimizes the collision probability among* N_s *secondary users.*

Figure 5 also shows the effect of radio channel bandwidth on L_s^*. An increase in radio channel bandwidth decreases packet transmission time. In this case packet transmission could be complete before a primary user service interruption. Increasing radio channel bandwidth therefore decreases the service interruption probability. In a high bandwidth case we can allow a secondary user to acquire more radio channels without increasing the risk of service interruption. Hence, we observe in Fig. 5 that increasing the radio channel bandwidth can lead to a higher optimal value of L_s (i.e., L_s^*).

CONCLUSIONS

In this article we review the evolution of CR and the various architectures envisaged for its realization. The DSA technique introduces a new type of loss called service interruption loss for the secondary users in an overlay CR network. We show via simulations that *TCP performance under service interruption loss in an overlay CR network is significantly different to its performance in a conventional network*. In particular, we observe that in an overlay CR network that adopts DSA, there is an optimal number of channels the secondary users need to capture to maximize their aggregate TCP throughput. After all, a secondary user cannot blindly utilize all the available channels and expect the best result.

REFERENCES

[1] FCC, "Spectrum Policy Task Force Report," ET Docket no. 02-135, Nov. 15, 2002.

[2] F. Akyildiz *et al.*, "NeXt Generation/Dynamic Spectrum Access/Cognitive Radio Wireless Networks: A Survey," *Elsevier Comp. Net.*, vol. 15, no. 13, Sept. 2006, pp. 2127–59.

[3] H. Balakrishnan *et al.*, "A Comparison of Mechanisms for Improving TCP Performance over Wireless Links," *IEEE/ACM Trans. Net.*, vol. 5, no. 6, Dec. 1997, pp. 756–69.

[4] J. Mitola III, *Cognitive Radio: An Integrated Agent Architecture for Software Defined Radio*, Ph.D. thesis, KTH Royal Inst. Technology, Sweden, 2000.

[5] S. Haykin, "Cognitive Radio: Brain-Empowered Wireless Communications," *IEEE JSAC*, vol. 23, no. 2, Feb. 2005, pp. 201–20.

[6] J. M. Peha, "Approaches to Spectrum Sharing," *IEEE Commun. Mag.*, vol. 43, no. 2, Feb. 2005, pp. 10–12.

[7] T. Weiss and F. K. Jondral, "Spectrum Pooling: An Innovative Strategy for the Enhancement of Spectrum Efficiency," *IEEE Commun. Mag.*, vol. 42, no. 3, Mar. 2004, pp. 8–14.

[8] R. W. Brodersen *et al.*, "CORVUS: A Cognitive Radio Approach for Usage of Virtual Unlicensed Spectrum," White Paper, Berkeley Wireless Research Center, 2004, pp. 1–21.

[9] M. M. Buddhikot *et al.*, "DIMSUMNet: Spectrum Management in Coordinated Dynamic Spectrum Access Based Cellular Networks," *Proc. IEEE DySPAN '05*, Nov. 2005, pp. 299–307.

[10] H. Zheng and L. Cao, "Device-Centric Spectrum Management, " *Proc. IEEE DySPAN '05*, Nov. 2005, pp. 56–65.

[11] L. Cao and H. Zheng, "Distributed Spectrum Allocation via Local Bargaining," *Proc. IEEE SECON '05*, Sept. 2005, pp. 475–86.

[12] 3GPP2, " CDMA 2000 High Rate Packet Data Air Interface Specification," tech. rep. no. C.S20024 v. 2.0, Oct. 2000.

[13] J. F. Kurose and K. W. Ross, *Computer Networking: A Top-Down Approach*, Addison-Wesley, 2008.

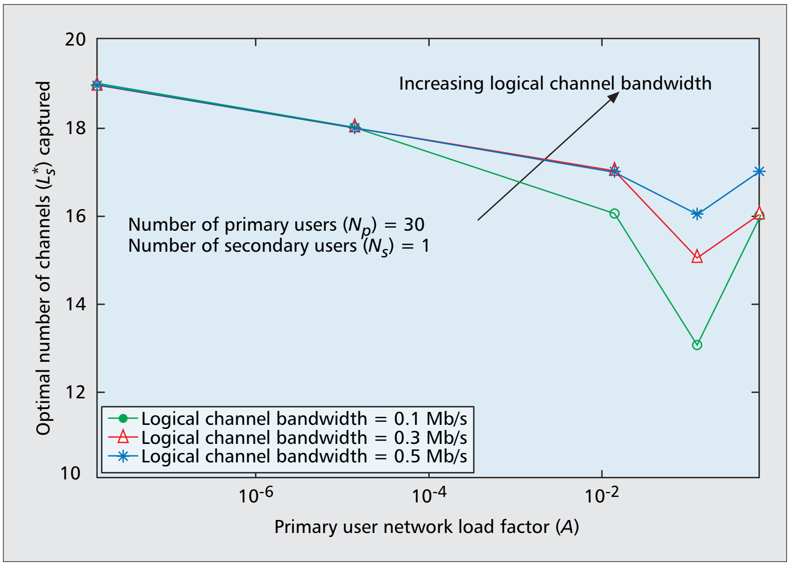


Figure 5. *Optimum number of channels (L_s^*) a secondary user captures to maximize its TCP throughput vs. load factor (A) for different values of bandwidth. The figure shows that for a given bandwidth, the optimal number of channels (L_s^*) captured is high for small and large load factor values, and low for moderate load factor values.*

BIOGRAPHIES

TEERAWAT ISSARIYAKUL [M] (teerawat@ece.ubc.ca) received his B.Eng and M. Eng. from Thammasat University and the Asian Institute of Technology (AIT), Thailand, in 1997 and 1999, respectively. In 2005 he completed his Ph.D. at the Department of Electrical and Computer Engineering, University of Manitoba. In 2001–2005 he was also affiliated with Telecommunication Research Laboratory (TRLabs). In 2005-2006 he worked as a postdoctoral fellow at the Department of Electrical and Computer Engineering, University of British Columbia. Currently, he is a system engineer at TOT Public Company Limited, Thailand. He also holds an adjunct faculty position at the Asian Institute of Technology. His current research area is modeling and optimization of cognitive radio networks and intelligent transportation networks.

LAXMINARAYANA S. PILLUTLA [S'04] (laxp@ece.ubc.ca) obtained his B.Eng. from the University of Madras, India, and his M.S. from Wichita State University, Kansas, in electrical engineering. He received a Ph.D. from the University of British Columbia in electrical engineering under the supervision of Dr. Krishnamurthy in 2008. His research interests are in general communication theory (CDMA, OFDM, and MIMO systems), cross-layer design of wireless networks, wireless sensor networks, cognitive radio, and vehicular telematics.

VIKRAM KRISHNAMURTHY [S'90, M'91, SM'99, F'05] (vikramk@ece.ubc.ca) received a B.E. degree from the University of Auckland, New Zealand, in 1988 and a Ph.D. degree from the Australian National University, Canberra, in 1992. Since 2002 he has been a professor and Canada Research Chair of the Department of Electrical Engineering, University of British Columbia, Vancouver, Canada. Prior to 2002 he was a chaired professor with the Department of Electrical and Electronic Engineering, University of Melbourne, Australia. His current research interests include stochastic control and computational game theory, stochastic optimization and scheduling, and statistical signal processing. He has served as Associate Editor for several journals including *IEEE Transactions on Signal Processing*, *IEEE Transactions on Aerospace and Electronic Systems*, *IEEE Transactions on Automatic Control*, *IEEE Transactions on Nanobioscience*, and *Systems and Control Letters*. From 1009 to 2010 he is serving as a Distinguished Lecturer for the IEEE Signal Processing Society.