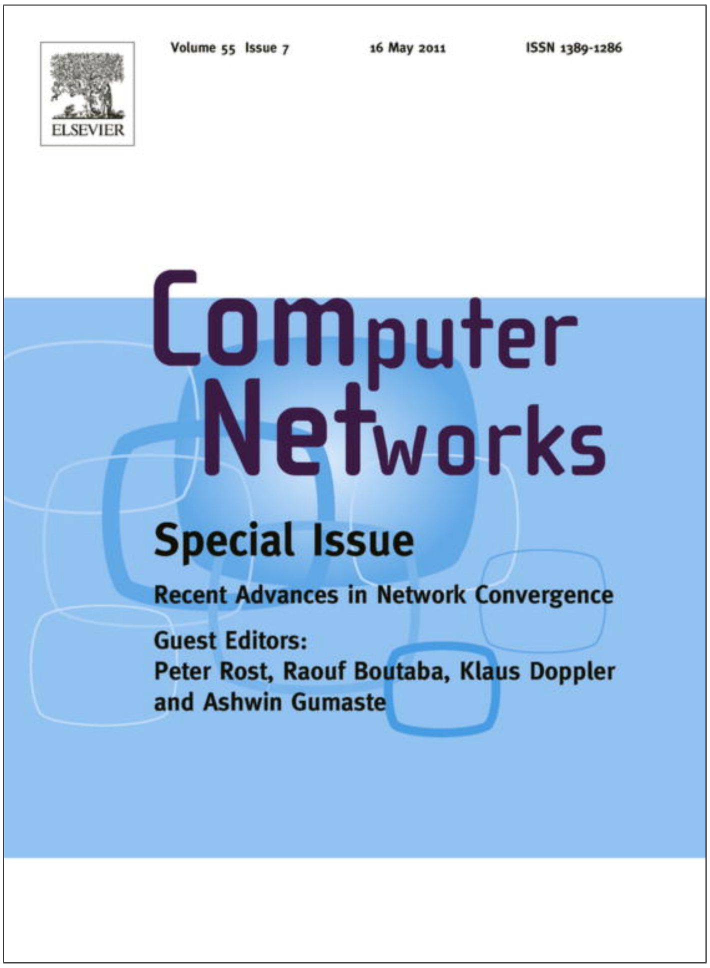
ELSEVIER
Volume 55 Issue 7
16 May 2011
ISSN 1389-1286
Computer Networks
Special Issue
Recent Advances in Network Convergence
Guest Editors:
Peter Rost, Raouf Boutaba, Klaus Doppler
and Ashwin Gumaste

Computer Networks 55 (2011) 1592–1613

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

QoS capacity of virtual wireless networks

B. Ishibashi [a], N. Bouabdallah [a,b], R. Boutaba [a,c,*]

[a] Cheriton School of Computer Science, University of Waterloo, Waterloo, Canada
[b] INRIA, Campus universitaire de Beaulieu, 35042 Rennes, Cedex, France
[c] Division of IT Convergence Engineering, POSTECH, Republic of Korea

ARTICLE INFO

Article history:
Available online 25 February 2011

Keywords:
Cognitive radio
Wireless networks
Virtual networks
Performance analysis
Service reliability

ABSTRACT

The allocation of wireless resources to many different services and service providers has resulted in a heavily fractured spectrum and inefficient use. However, new dynamic spectrum capabilities allow the recapture of this wastage. This work describes the creation of a virtual wireless network (VWN). Taking an approach based on cognitive radio techniques, the VWN allows new network services to be deployed utilizing only residual resources, while preserving the service to primary users. A Markov chain-based analysis is performed, and is used along with discrete event simulations to demonstrate the VWN as a management tool for supporting new services within the existing system.

1. Introduction

Recent growth in wireless services has created an interesting scenario. A wide range of services has been deployed, provided by numerous operators. Supporting all of these services is the basic underlying resource, radio spectrum. Despite being universal and self-replenishing, demand is quickly overtaking supply, making spectrum a scarce and tremendously valuable resource.

Despite this scarcity, an interesting scenario has emerged. Recent studies have demonstrated that overall utilization is actually quite low [1]. Due to the fixed allocation of resources through the assignment of frequency channels to specific purposes, some channels may be heavily congested, while others remain essentially idle. Other users are prevented from utilizing these spectrum holes by the regulations of the licensing system. As a result, the system operates inefficiently. Deploying new services into this scenario is also problematic, as obtaining new spectrum is difficult, slow, and expensive.

However, regulatory bodies have recognized the opportunity to make smarter use of all resources. They have acknowledged that there is a need for a new system for spectrum management – one that is more flexible, dynamic, and efficient [2]. This impetus, combined with advances in radio technology has created the vision secondary spectrum approaches such as cognitive radio (CR). With the ability to detect and exploit residual bandwidth left underutilized by licensed users, these approaches can provide a seemingly new supply of resources, without negatively affecting existing use.

Secondary spectrum use presents an opportunity to exploit existing wireless resources in new ways, in order to deploy new networks and services. Virtual wireless networks (VWNs) can be created, operating without obtaining any dedicated spectrum resources of their own. By effectively managing residual resources from all primary users, the VWN allows new services to fit into the existing environment. Although inherently dependent on the primary networks for their resources, they can operate essentially independently, free to utilize their resources in whatever manner they require.

As these resources are inherently opportunistic, their nature is not clear. If a network utilizes solely residual bandwidth, can it reliably offer services, with an acceptable quality? When primary networks are idle, it is clear that a

* Corresponding author at: Cheriton School of Computer Science, University of Waterloo, Waterloo, Canada.

E-mail addresses: bkishiba@uwaterloo.ca (B. Ishibashi), nizar.bouabdallah@inria.fr (N. Bouabdallah), rboutaba@uwaterloo.ca, rboutaba@bbcr.uwaterloo.ca (R. Boutaba).

1389-1286/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.comnet.2010.12.027

secondary network can succeed. However, what happens when primary networks are active, and resources are harder to come by. Can a virtual wireless network deliver anything more than best effort service? Can it satisfy quality-of-service (QoS) guarantees?

This work aims to show that service providers' unused bandwidth is still a valuable resource, and that a viable virtual operator can be created. This virtual operator can support its own services with the necessary service quality, despite not having any dedicated resources of its own. Therefore, the virtual wireless network created in this work is an effective tool for aggregating resources into a functional secondary service.

1.1. Related work

Cognitive Radio represents a long-term goal for making wireless communication more intelligent and efficient. The term was initially applied to extending software radios with a self-awareness about its characteristics and requirements, in order to determine, along with other devices, an appropriate radio etiquette to be used [3]. This self-awareness includes an awareness of its environment, through the monitoring, modelling, and prediction of the radio spectrum [4]. For an overview of cognitive radio, a survey is presented in [5].

While cognitive radio encapsulates many different concepts into a single vision, early work has latched onto one aspect of CR – the idea of dynamic management of the wireless spectrum. In particular, work has addressed the operation of new CR users within existing licensed spectrum. For example, the IEEE 802.22 working group on Wireless Regional Area Networks is working towards communications that can co-exist with TV broadcast service [6]. The IEEE Standards Coordinating Committee 41 on Dynamic Spectrum Access Networks (DySPAN) has also been established, arising out of the IEEE 1900 Standards Committee on Next Generation Radio and Spectrum Management [7], as an effort to specify terminology, study co-existence and conformance issues, and create an architecture for making radio resource decisions.

Previous works have addressed the capacity of cognitive radio systems. In [10,11], the capacity has been considered from an information theory perspective, at the physical level. They focus primarily on the capacity of a single transmission in the presence of primary users. However, QoS must also be considered at the higher level of a network model. Therefore, the capability of the CR system to support QoS traffic is evaluated, in terms of the primary network traffic. Thus, this work demonstrates the capacity gains of the networks and systems as a whole, rather than for an individual.

The secondary use of spectrum by a network has been considered in [12]. The spectrum pooling approach involves the collaboration of service providers, in servicing each otherś users. In [13], an analysis of a spectrum pooling system is performed, in order to evaluate the capacity gain that can be achieved. It is however limited to the support of additional users by a single license owner, opportunistically supporting more of its own users. [14] also applies continuous time Markov chains while considering a single primary user and additional unlicensed users. This work initially uses a similar approach for evaluating the capacity gain that can be achieved on multiple networks. It then focuses on the use of these resources to create new services, distinct from the original networks.

The term "virtual unlicensed spectrum" was used in [15]. It described the use of cognitive radio to re-use allocated but unused spectrum as if it were unlicensed. Although this unlicensed spectrum could be used to create a virtual network, their focus was solely on best-effort traffic.

1.2. Outline

The demonstration of the virtual wireless network is presented as follows. The next section presents the primary methodology. First, a classic system model is described, onto which the capabilities of cognitive radio are added. The virtual wireless network concept can be considered using this extended model. The classic and cognitive models are then analyzed, using a Markov model to derive expressions for several metrics.

These models are verified through simulations, and a variety of system characteristics are investigated. The key results of these investigations are presented in Sections 3 and 4. The former focuses on connection-based traffic and QoS, while the latter integrates the system performance of data traffic into the model.

In Section 5, a discussion of the results, limitations, and other issues is presented. Finally, Section 5.5 summarizes the conclusions of this work.

2. Materials and methods

2.1. System model

This work will consider the capacity gains achieved by utilizing cognitive radio in the presence of one or more existing primary networks. Primarily, the CR system will overlay the primary networks, opportunistically utilizing unused channels. Secondarily, adding CR abilities to the users of existing networks is also considered. In either case, the first step to quantifying the performance improvements of cognitive radio is to describe the classic network. The CR systems will then be created by extending the classic network with the additional requirements and capabilities of cognitive radio.

2.1.1. Classic networks

Consider the classic network of service provider i. Service provider i has wireless resources dedicated to its network, consisting of K_i channels. Initially, connection-based traffic will be considered, with each connection requiring the continuous service of an entire channel throughout its time in the system.

A connection or flow c_i arrives at its home network i and requests service. If the network has sufficient resources (the network has a channel available), then the flow can be accepted. Otherwise, the flow is blocked – service is refused and the flow ends unsuccessfully. A flow, once

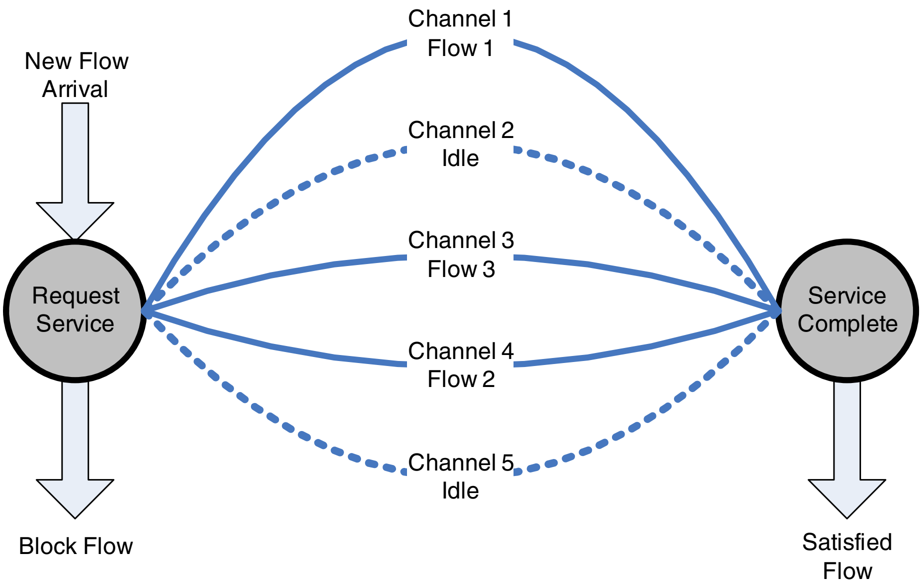

Fig. 1. A network with 5 channels.

accepted, must have continuous service until the completion of its service requirements. Any interruption to its service would be considered a failure in the service quality experienced by that user. After the complete service time has elapsed without any interruption, the call is completed successfully. At this point, the call is terminated and the channel can be released. A single network with 5 channels is depicted in Fig. 1.

The classical system consists of N classical networks. Each network has its own set of resources, with the total capacity of the system being $M = \sum_{i=1}^{N} K_i$. In the classical system, users are tied to their service provider's network by contract and/or technology. This means that c_i can only be serviced by a channel belonging to network i. It can be serviced by any channel on network i, although as all channels are identical, there is no motivation for a classical flow to change channels. Therefore, a classical flow, once accepted, will remain in the same channel until its service is completed.

In this scenario, traffic only occupies channels on its home network. Therefore, it is possible that two networks may experience different utilization levels at different times. A network with high utilization may be forced to block incoming flows, despite the fact that the overall system (the complete set of all networks) has a relatively low utilization.

2.1.2. Cognitive radio model

The above model for a system of classical networks represents the existing primary users of spectrum. The requirements and capabilities of cognitive radio devices are now considered.

Rather than requiring service from a specific network, CR users require the availability of a channel – any channel, from any network. For this work, no assumptions are made on how this borrowed channel is actually used. It is only important that the cognitive system will capitalize on this available channel for its own purposes. For this same reason, it is difficult to characterize the actual CR communication, and therefore the capacity. Different modes of communication and transmission parameters could result in different capabilities. Therefore, this work characterizes the CR abilities in terms of the ability to support additional traffic identical to the traffic existing on the classical networks. This allows the capacity gain to be expressed relative to the original networks.

2.1.3. CR-enabled users

One way to consider this is if existing users are given the additional ability to obtain service from other networks. Although such a system would not necessarily require CR, CR would be used to implement this type of system. A CR-enabled node would request service from their home network, but if resources were not available, could obtain additional resources from another network.

Classical and cognitive flows are distinguished as $c_{CL}(i)$ and $c_{CR}(i)$ respectively. Flows from CR users have the same requirements as those of classic users. Each $c_{CR}(i)$ requires the use of a channel throughout its time in the system. However, while a flow $c_{CL}(i)$ can operate only on its home network i, $c_{CR}(i)$ can operate anywhere. They may move between different channels as required.

In a real system such a handoff might incur some overhead and possible disruption in service quality. In order to switch channels, the conflict must be detected, a new channel must be determined, and the communicating nodes must switch to the destination channel. However an ideal handoff is assumed here. For QoS traffic, this assumption simply requires that the switching delay incurred is less than a tolerable interval. For data traffic, the delay results in a small decrease in capacity, however if we expect that channel residency times are long relative to this delay, then the effect is negligible. An alternative option also exists: rather than forcing calls to switch channels, a spectrum trading scheme could be adopted [16]. Under this system, although channels switch between networks, a call only loses its channel if there is no new channel available (and it will be dropped). This approach might add to the delay that occurs to set up a new call, however would avoid disruption by eliminating channel handoffs.

The traffic in the system can be described in terms of two counts for each network: $n_{CL}(i)$ and $n_{CR}(i)$ represent the classic and cognitive flows belonging to network i that are currently occupying channels in the system. The total amount of traffic belonging to each network is $n(i)$, the sum of these counts.

An incoming classic flow is still blocked if the home network is fully occupied by its own flows (whether or not these flows have cognitive capabilities). However, a cognitive flow is only blocked if the above condition is true, and if all resources in the entire system – the home network and all other networks – are utilized.

However, this creates a new possibility. Consider when a network i is servicing a foreign flow belonging to network j ($j \neq i$) and there are no other channels available. If a new flow c_i arrives for network i (either classic or cognitive), the cognitive flow $c_{CR}(j)$ must terminate its use of the channel. This is the condition of cognitive radio – a cognitive flow must not interfere with the service of a primary user. Depending on the traffic that network j is currently servicing, it may be able to boot another flow (which in turn will be returned to its home network), otherwise it will be forced to drop $c_{CR}(j)$. In the end, one (and only one) flow must be dropped.

An occupancy threshold parameter Th_i is also included in the network model. This threshold applies to incoming classic flows, allowing the balance between blocking and dropping of flows to be adjusted. If the total number of network i flows is greater than or equal to the threshold ($n(i) \geqslant Th_i$) then arriving classic users are blocked. Blocking of classic users when the threshold is exceeded means that if a flow $c_{CR}(i)$ loses its resources while operating on another network's channel, it is more likely that it will be able to return to the home network, thus avoiding being dropped.

2.1.4. Virtual wireless networks

For the purpose of considering purely CR users, virtual wireless networks are created, possessing no spectrum resources of their own (that is, $K_i = 0$). Any spectrum resources must be borrowed (leased) from one of the primary networks.

Note that this means that VWNs never have capacity to service their own flows. If the system is full, any incoming calls must be blocked. If the system is not full, a channel is borrowed/leased from another network where all channels are not used. In addition, any VWN traffic forced off of another network must be dropped.

2.2. Analysis

In order to analyze the behaviour of the system, the network is represented as $X(t)$, defined by the tuple $((n_{CL}(i), n_{CR}(i)), i = 1,\ldots,N))$ at time t.

Flows arrive according to a set of Poisson processes. For each network, arrival process parameters $\lambda_{CL}(i)$ and $\lambda_{CR}(i)$ control the arrival rates for classic and cognitive flows. Similarly, $\mu_{CL}(i)$ and $\mu_{CR}(i)$ represent the exponentially distributed service rates. Therefore, $\{X(t), t \geqslant 0\}$ is a Markov process with continuous time and finite state space $\mathcal{S} = \{(n_{CL}(i),\, n_{CR}(i)), i = 1,\ldots,N | n_{CL}(i) \leqslant Th_i, \sum_{i=1}^{N} n(i) \leqslant M\}$.

In the following, the transition probability matrix Q is derived for the process X when leaving a generic state $s \in \mathcal{S}$, with $s = ((n_{CL}(i), n_{CR}(i)), i = 1,\ldots,N)$. Let e denote the next occurring event and $s' \in \mathcal{S}$ the associated new network state. e can be either a new flow arrival or a flow departure. The transition probabilities regarding each case are derived as follows.

2.2.1. Arrival of a new flow

Let f denote the new arriving connection, belonging to home network i. The flow can be either a classic or a cognitive flow, and each case will be handled separately.

- f is a Classic connection ($c_{CL}(i)$) f will be blocked unless the total number of network i's flows in service is less than the threshold (i.e., $n(i) < Th_i$). If it can be accepted, then f will be served on the only possible network, i, its home network. If there is one or more free channels on the network, then f will occupy one of those channels. This occurs when $\sum_{k=1}^{N} n(k) < M$. When this occurs, the network transits from state s to s' as follows:
$$Q(s, s') = \lambda_{CL}(i), \tag{1}$$
where:
$$s' = ((n_{CL}(0), n_{CR}(0)), \ldots, (n_{CL}(i) + 1, n_{CR}(i)), \ldots, (n_{CL}(N), n_{CR}(N))).$$
If $\sum_{k=1}^{N} n(k) = M$, then the network must interrupt one of the foreign cognitive connections, in order to free a channel for f. However, the interrupted flow may belong to any of the other networks whose usage currently exceeds its physical number of channels. Therefore, different transitions may occur, depending on what flow is preempted. For each $j \in E_j$, with $E_j = \{j \in \{1,\ldots,N\}, j \neq i | n(j) > K_j\}$, the network state transits from state s to the state s'_j as follows:
$$Q\left(s, s'_j\right) = \lambda_{CL}(i) \times \frac{(n(j) - K_j)}{h(s)}, \tag{2}$$
where $h(s) = \sum_{k=1}^{N} 1_{|n(k)>K_k} \times (n(k) - K_k)$ denotes the current number of connections served outside of their home networks. $1_{|\mathcal{A}}$, the indicator function of condition $\mathcal{A}$, is equal to 1 if $\mathcal{A}$ is true, and 0 otherwise. The resulting network state is:
$$s'_j = ((n_{CL}(0), n_{CR}(0)), \ldots, (n_{CL}(i) + 1, n_{CR}(i)), \ldots, (n_{CL}(j), n_{CR}(j) - 1), \ldots, (n_{CL}(N), n_{CR}(N))).$$
- f is a Cognitive Flow ($c_{CR}(i)$) The service of f depends on whether or not all of the N networks' resources are fully used. When $\sum_{k=1}^{N} n(k) < M$, the system can carry f on an available channel, and transits to state s' according to:
$$Q(s, s') = \lambda_{CR}(i), \tag{3}$$
where:
$$s' = ((n_{CL}(0), n_{CR}(0)), \ldots, (n_{CL}(i), n_{CR}(i) + 1), \ldots, (n_{CL}(N), n_{CR}(N))).$$

However, if f arrives and all the network resources are in use, then f will be either blocked or served by interrupting another cognitive flow being served on f's home network i. Which event occurs depends on the current state of network i's active flows. If $n(i) \geqslant K_i$, then f will be blocked. Otherwise, f will be served by interrupting a flow whose home network j is currently using more resources than it has channels (that is, $n(j) > K_j$). Recall that this implies that $j \in E_j$. Therefore, for each $j \in E_j$, the network state transits from state s to state s'_j according to:

$$Q\left(s, s'_j\right) = \lambda_{CR}(i) \times \frac{(n(j) - K_j)}{h(s)} \tag{4}$$

with the resulting state being:

$$s'_j = ((n_{CL}(0), n_{CR}(0)), \ldots, (n_{CL}(i), n_{CR}(i) + 1), \ldots, (n_{CL}(j), n_{CR}(j) - 1), \ldots, (n_{CL}(N), n_{CR}(N))).$$

2.2.2. Departure of a flow

The second possibility for e is that it is a flow departure. Flow departure transitions occur only when a flow completes its service.

If f is a classic flow ($c_{CL}(i)$):

$$Q(s, s') = n_{CL}(i) \times \mu_{CL}(i), \tag{5}$$

where:

$$s' = ((n_{CL}(0), n_{CR}(0)), \ldots, (n_{CL}(i) - 1, n_{CR}(i)), \ldots, (n_{CL}(N), n_{CR}(N))).$$

If f is a cognitive connection ($c_{CR}(i)$):

$$Q(s, s') = n_{CR}(i) \times \mu_{CR}(i), \tag{6}$$

where:

$$s' = ((n_{CL}(0), n_{CR}(0)), \ldots, (n_{CL}(i), n_{CR}(i) - 1), \ldots, \times (n_{CL}(N), n_{CR}(N))).$$

To summarize, the transition probabilities of the process X when leaving a generic state $s \in \mathcal{S}$, with $s = ((n_{CL}(i), n_{CR}(i)), i = 1, \ldots, N)$ to a state $s' \in \mathcal{S}$ i:

$\forall s \in \mathcal{S}$ with $s = ((n_{CL}(i), n_{CR}(i)), i = 1, \ldots, N)$
For $i = 1, \ldots, N$
% A new $c_{CL}(i)$ classic connection arrival
- if $(n(i) < Th_i) \& \left(\sum_{k=1}^{N} n(k) < M\right)$ then derive $Q(s,s')$ as given by Eq. (1), along with the respective s'.
- if $(n(i) < Th_i) \& \left(\sum_{k=1}^{N} n(k) = M\right)$ then $\forall j \in E_j$, the system transits from state s to s'_j with a rate $Q\left(s, s'_j\right)$ given by Eq. (2).

% A new $c_{CR}(i)$ cognitive connection arrival
- if $\left(\sum_{k=1}^{N} n(k) < M\right)$ then use Eq. (3).
- if $\left(\sum_{k=1}^{N} n(k) = M\right) \& (n(i) < K_i)$ then $\forall j \in E_j$, derive $Q\left(s, s'_j\right)$ using Eq. (4).

% The departure of a classic connection $c_{CL}(i)$
- if $(n_{CL}(i) > 0)$ then use Eq. (5).

% A departure of cognitive connection $c_{CR}(i)$
- if $(n_{CR}(i) > 0)$ then use Eq. (6).

% End

With all of the transition possibilities considered, the transition probability matrix $Q = [q_{ss'}]$ can now be derived. The steady state probabilities $\pi = [\pi_s]$ of the process X is then obtained by resolving the following system:

$$\pi Q = \pi \quad \text{and} \sum_{s \in \mathcal{S}} \pi_s = 1. \tag{7}$$

2.2.3. Performance metrics

Building on the above results regarding the matrix Q and the vector π, the performance of the different classes of flows can now be evaluated. Four classical QoS metrics are considered: blocking, dropping, and failure probabilities, as well as network resource utilization.

- Blocking probability – $P_B(i, CL)$ is the blocking probability for arriving classic flows belonging to network i. A classic flow f is blocked if it arrives while its home network is already using more resources than its service threshold (i.e., $n(i) \geqslant Th_i$). Therefore, the probability of blocking is:

$$P_B(i, CL) = \sum_{s \in \mathcal{S}} 1_{|n(i) \geqslant Th_i} \times \pi_s. \tag{8}$$

Similarly, $P_B(i, CR)$ is the blocking probability for cognitive flows. However, the conditions for blocking are different for these flows – in order to be blocked, two conditions must be true when the flow arrives: (i) all network resources must be occupied $\left(\sum_{j=1}^{N} n(j) = M\right)$; and, (ii) network i already uses all of its own resources ($n(i) \geqslant K_i$). Therefore:

$$P_B(i, CR) = \sum_{s \in \mathcal{S}} 1_{|\sum_{j=1}^{N} n(j) = M} \times 1_{|n(i) \geqslant K_i} \times \pi_s. \tag{9}$$

- Dropping probability – it is easy to see that the dropping probability of an in-service classic connection is null. Therefore, only $P_D(i, CR)$, the dropping probability of network i's cognitive flows, must be considered. Several conditions must occur in order for a cognitive flow to be dropped upon the arrival of a new flow f belonging to network j. These conditions differ slightly, depending on whether the incoming flow is a classic ($c_{CL}(j)$) or cognitive ($c_{CR}(j)$) connection. In both cases, these two conditions must occur:
  – All M channels in the system are occupied

$$\left(\sum_{k=1}^{N} n(k) = M\right). \tag{i}$$

  – Network i uses more resources than it possesses

$$(n(i) > K_i). \tag{ii}$$

If f is an arriving classic flow ($c_{CL}(j)$), then (iii) must also be true:
  – Network j uses less resources than its threshold

$$(n(j) < Th_j). \tag{iii}$$

If f is an arriving cognitive flow ($c_{CR}(j)$), condition (iv) must occur:

– Network j uses less resources than its capacity

$$(n(j) < K_j). \tag{iv}$$

Therefore, for an arriving classic flow f belonging to network j, a flow belonging to network i is interrupted with a probability:

$$\frac{Q(s, g_{CL}(s,i,j))}{\lambda_{CR}(i)},$$

where $s \in \mathcal{S}$ is the current network state that verifies conditions (i), (ii), and (iii). $g_{CL}(s,i,j)$ is a function that returns the new network state, expressed as follows:

$$g_{CL}(s,i,j) = ((n_{CL}(0), n_{CR}(0)), \ldots, (n_{CL}(i), n_{CR}(i) - 1), \ldots, (n_{CL}(j) + 1, n_{CR}(j)), \ldots, (n_{CL}(N), n_{CR}(N))).$$

For arriving cognitive flows belonging to network j, a cognitive flow from network i is dropped with a probability equal to:

$$\frac{Q(s, g_{CR}(s,i,j))}{\lambda_{CR}(i)}.$$

In this case, $g_{CR}(s,i,j)$ can be expressed as:

$$g_{CR}(s,i,j) = ((n_{CL}(0), n_{CR}(0)), \ldots, (n_{CL}(i), n_{CR}(i) - 1), \ldots, (n_{CL}(j), n_{CR}(j) + 1), \ldots, (n_{CL}(N), n_{CR}(N))).$$

Therefore, the dropping probabilities for network i are:

$$P_D(i, CL) = 0, \tag{10}$$

$$P_D(i, CR) = \left[\sum_{1 \leqslant j \leqslant N} \sum_{j \neq i} \sum_{s \in \mathcal{S}} 1_{\left|\sum_{k=1}^{N} n(k) = M\right.} \times 1_{|n(i) > K_i} \times 1_{|n(j) < Th_j} \times 1_{|\lambda_{CL}(j) > 0} \times \pi_s \times \frac{Q(s, g_{CL}(s,i,j))}{\lambda_{CR}(i)} + \sum_{\substack{1 \leqslant j \leqslant N \\ j \neq i}} \sum_{s \in \mathcal{S}} 1_{\left|\sum_{k=1}^{N} n(k) = M\right.} \times 1_{|n(i) > K_i} \times 1_{|n(j) < K_j} \times 1_{|\lambda_{CR}(j) > 0} \times \pi_s \times \frac{Q(s, g_{CR}(s,i,j))}{\lambda_{CR}(i)} \right] \times \frac{1}{1 - P_B(i, CR)}. \tag{11}$$

- Failure probability – cognitive traffic experiences a reduction in blocking, but an increase in dropping compared to classic traffic (where $P_D(i,CL) = 0$). Therefore, P_F is used to compare the overall performance of cognitive and classic traffic, where P_F is the probability that an arriving flow will not receive the service it requires. P_F is calculated using P_B and P_D.

$$P_F(i) = P_B(i) + (1 - P_B(i)) \times P_D(i).$$

For classic traffic, $P_D(i,CR) = 0$, therefore $P_F(i,CL)$ is:

$$P_F(i, CL) = P_B(i, CL). \tag{12}$$

For cognitive traffic:

$$P_F(i, CR) = P_B(i, CR) + (1 - P_B(i, CR)) \times P_D(i, CR). \tag{13}$$

- Utilization – U_i is the average occupancy of all channels on network i. Therefore it is the average number of channels that are in use, divided by the total number of channels.

$$U_i = \sum_{s \in \mathcal{S}} \pi_s \times \frac{1}{K_i} \times \left(\max(K_i - n(i), 0) \frac{h(s)}{t(s)} + \min(n(i), K_i) \right), \tag{14}$$

where $\max(K_i - n(i), 0)\frac{h(s)}{t(s)}$ represents the number of foreign flows in service on network i, while $\min(n(i),K_i)$ is the number of network i's own flows (in service on i). $h(s)$ denotes the current number of connections served outside their home networks, and $t(s)$ denotes the number of channels that can accommodate cognitive flows. $h(s)$ and $t(s)$ are:

$$h(s) = \sum_{k=1}^{N} 1_{|n(k) > K_k} \times (n(k) - K_k),$$

$$t(s) = \sum_{k=1}^{N} 1_{|n(k) \leqslant K_k} \times (K_k - n(k)).$$

3. Results

A simulation model has been developed in order to validate the analytic results. The system of networks was implemented as a discrete event simulation. Numerous evaluations were performed in order to confirm that the analytic results obtained in Eqs. 8, 9, 11 and 14. In all cases, the results matched very closely.

For the simulations, the result obtained is dependent on the length of the simulation (the number of arrivals). As an example, Fig. 2, shows the mean blocking, dropping, and failure probabilities for a system of 10 cognitive-enabled networks, each with 5 channels. The traffic arrival rates on each network are $\lambda_{CL}(i) = 0$ and $\lambda_{CR}(i) = 10$, $\forall i = 1,\ldots,10$. The service rate for all traffic is maintained as a normalized $\mu = 1$.

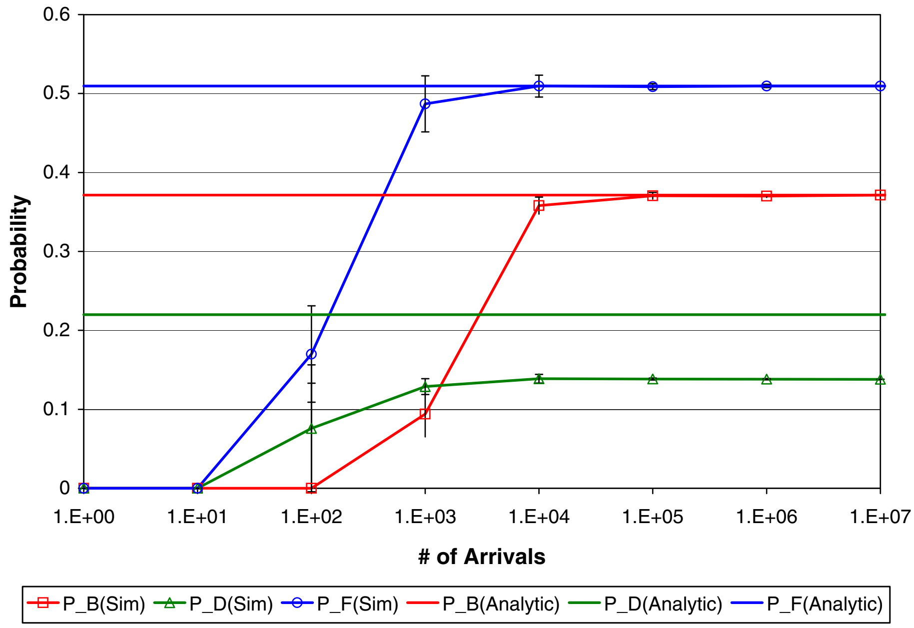

Fig. 2. Simulations approaching analytic results – length of simulation.

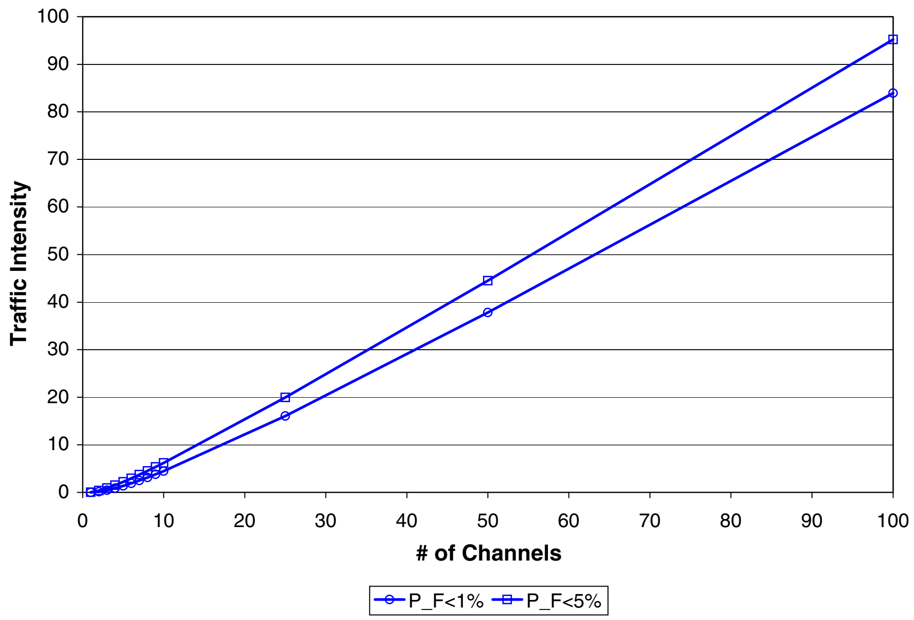

Fig. 3. Maximum traffic intensities at $P_F \leqslant 1\%$, 5% – single network.

Due to the presence of an initial transient effect caused by all networks starting in the 0-load state, simulations run for a short period of time have considerable error relative to the analytical result. However, as the number of arrivals increases, the result converges, reducing the error to a very low level (less than 0.1% for 10^6 arrivals). Confidence intervals for the simulation results are also shown. Even with a relatively low sample size of only 5 runs, the 99% confidence intervals are +/− less than 0.3% for run lengths longer than 10^5 arrivals.

For the remainder of the results, it has been confirmed that there is a consistently good fit between the simulation and analytical results. Given these results, for presentation purposes, all remaining figures show only a single curve, as the error is indistinguishable within the scales presented.

Additionally, this section discusses only systems where all networks have the same number of channels (with the exception of virtual networks), however it has been confirmed through numerous evaluations that the analytical model holds for other non-homogeneous system parameters.

3.1. Classic network baseline

The case of a single classic network was used to develop a baseline to which CR improvements could be compared.

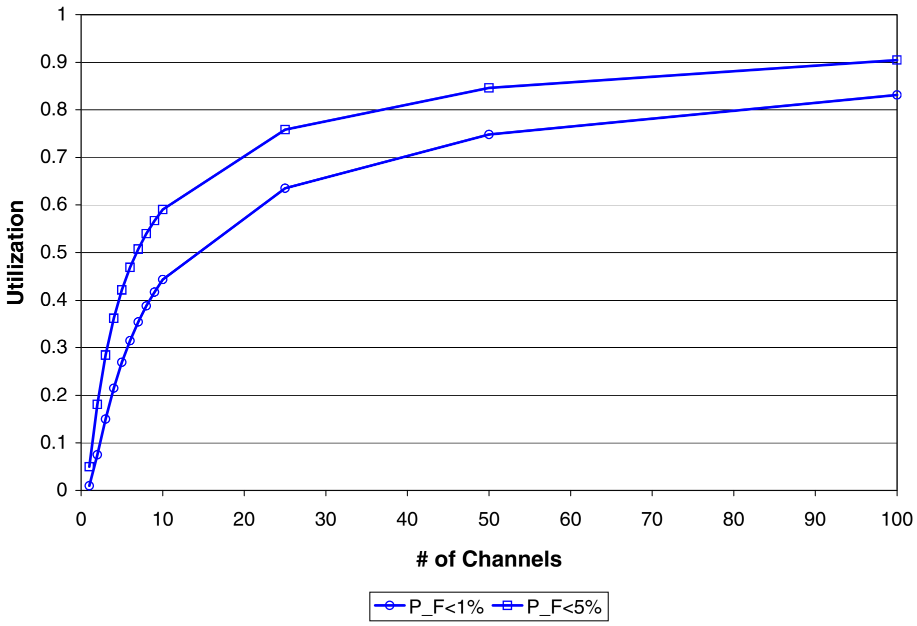

Fig. 4. Utilization at $P_F \leqslant 1\%$, 5% – single network.

Results from the single network case can easily be extended to multiple networks as the networks are, by definition, independent of each other. Obviously, for a single network, only classic traffic must be considered.

It is assumed that the network service provider must deliver a certain level of QoS, based on the P_F (which for the classical networks is equal to P_B). It is then important to know the traffic intensity ($\rho = \lambda/\mu$) that can be supported by the network while achieving this QoS. Fig. 3 shows the maximum traffic intensities supported while limiting blocking to either 1% or 5%, QoS levels commonly seen in experimental literature. Results for network sizes ranging between 1 and 100 channels are shown.

Fig. 4 shows the utilization levels of the network when the traffic intensity is at the maximum level for 1% and 5%. At small numbers of channels the utilization is very low, meaning that a large number of channels are unoccupied much of the time. Utilization increases as the number of channels increases, as a greater number of simultaneous flows are required to push the network into a blocking state - a single flow represents a decreasing share of the total network capacity. However, utilization must remain below 100%, in order to maintain the QoS requirements. It is also important to note that for large networks, although utilization is high (approx. 80% for 100-channel networks), this still means that a large number of channels are, on average, unused.

Any division of the spectrum to allow multiple service providers and networks results in inefficiencies. As an example, consider a single 100-channel network versus ten 10-channel networks. The traffic intensity is set to the maximum that maintains the acceptable QoS ($P_F \leqslant$ 1%) for each network (see Fig. 3). Although the same bandwidth is provided to each system, the single network system operates at 82% utilization, while the multiple networks are at only 44%. As a result, the single network solution is more efficient and supports a larger number of users. Frequently however, other circumstances necessitate that the available spectrum be divided.

3.2. Cognitive radio network scenarios

With the basic classic network case established, cognitive radio networks can now be explored. This exploration begins with the simplest cases involving only two networks, then continues with larger systems.

3.2.1. One classic, one virtual wireless network

The simplest scenario involving a cognitive radio network is the system consisting of one classic network and one VWN. If the classic network traffic is already at the maximum traffic intensity for a particular QoS level (e.g., traffic intensity for $\leqslant$1% blocking), then obviously the virtual network cannot make the same QoS guarantee. One possibility is to simply support additional traffic, without any QoS guarantees – this is often the approach suggested in CR.

Fig. 5 shows the failure probabilities if the CR network carries the same amount of traffic as the classic network. As the network (and its traffic intensity) grows, while the classic traffic performance remains unaffected, the QoS performance of the additional virtual network traffic deteriorates rapidly. However, as shown in Fig. 6, this traffic does increase the network utilization.

However, the QoS of this additional traffic is very low. Although a smaller intensity of traffic could be used on the virtual network, it will always have a lower QoS than the classic traffic. The more interesting scenarios occur as more networks are introduced.

3.2.2. Multiple classic networks, one VWN

Consider the scenario where there are n independent classic networks, each supporting the maximum classic traffic for $P_F \leqslant 1\%$ blocking. While any additional classic

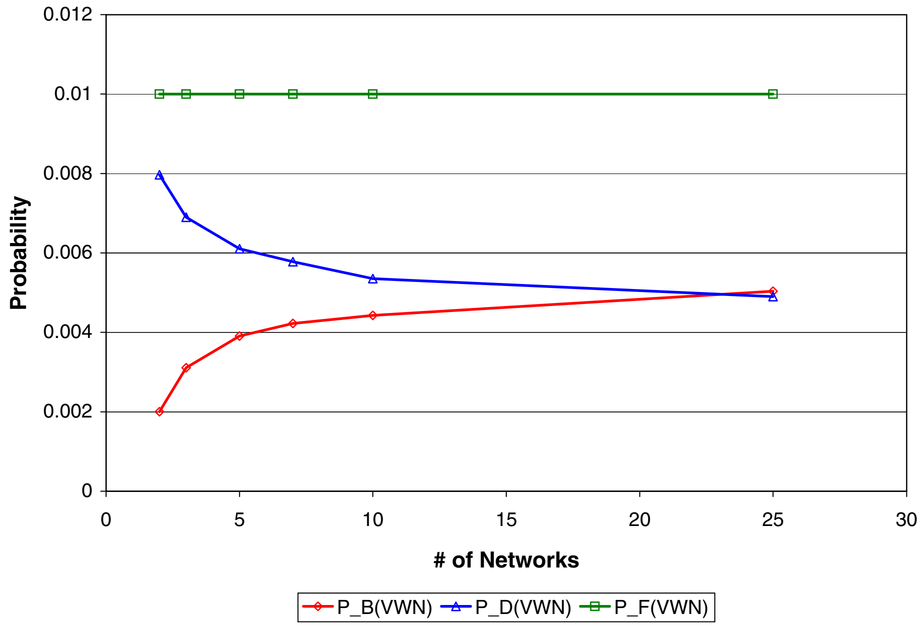

Fig. 5. Provided QoS levels – VWN with 1 CL network.

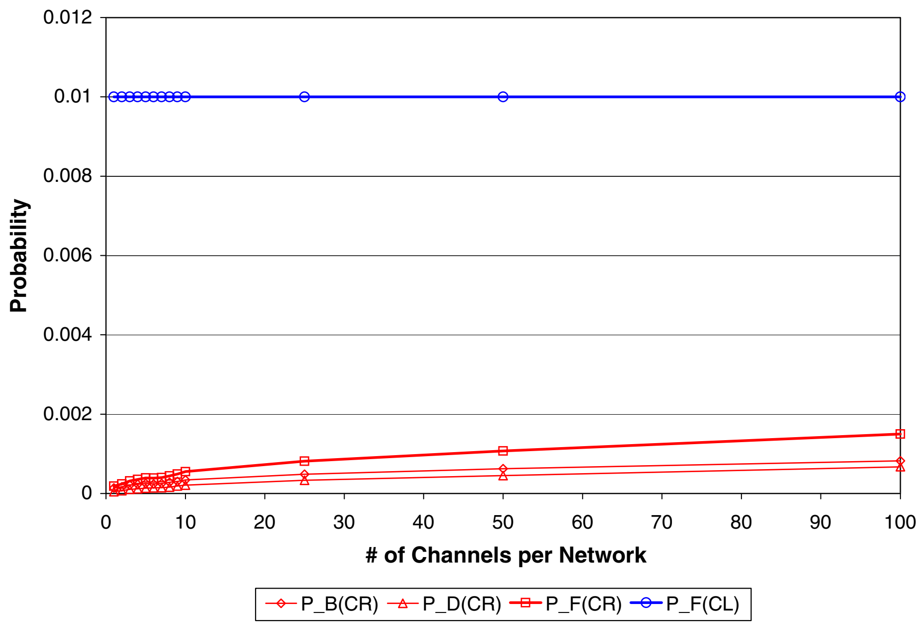

Fig. 6. Utilization – CL with VWN vs. baseline.

traffic would increase the P_F above the 1% constraint, if an additional virtual network is added using CR-techniques, it can make use of unused capacity on any of the existing networks, without affecting the P_F constraint of the classic networks. Fig. 7 shows the resulting QoS for a VWN supporting the same baseline intensity as the classic networks.

As can be seen in the figure, the resulting performance depends on both the number of networks and the number of channels per network. Recall that the baseline traffic intensity increases as the number of channels K increases. Therefore, Fig. 7 shows two important properties. First, as the number of classic networks increases, $P_F(VWN)$ decreases, as there is an increased probability that one of the networks has available resources. Second, increases in network size (number of channels) result in an increase in the $P_F(VWN)$. This is due to the fact that larger classic networks (greater K) have a higher baseline utilization level than smaller networks, leaving a lower probability of having resources available for cognitive use.

Fig. 8 shows the maximum traffic intensity that can be supported on the virtual network while maintaining a QoS level (the usual 1% failure rate). Again, all classic networks continue to maintain the baseline traffic intensity and QoS performance. In this case, each classic network possesses

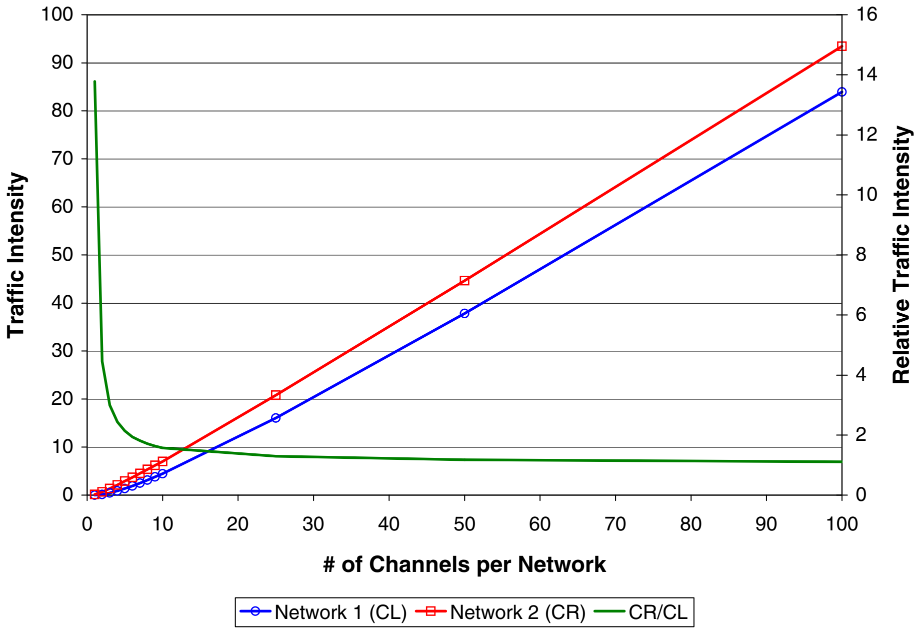

Fig. 7. QoS Levels provided by VWN – varying number of CL networks.

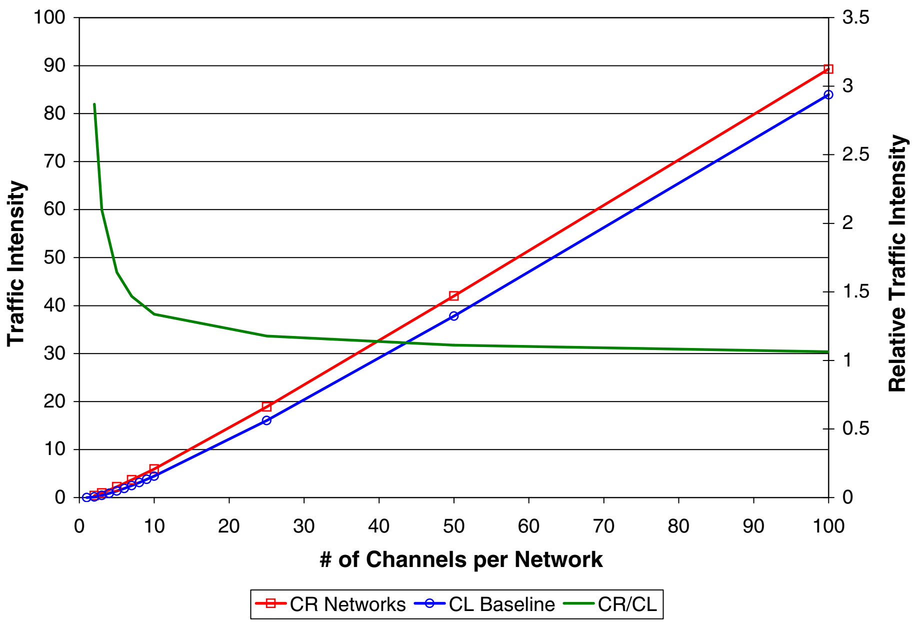

Fig. 8. Max VWN traffic intensity for 1% failures.

$K = 10$ channels. Considering that the baseline intensity for $K = 10$ is $\rho = 4.48$ (from Fig. 3), for the 25-network system, the VWN actually carries a traffic intensity of $\rho = 112$. This is the equivalent traffic of 25 additional classic networks. Although this is not quite as efficient as the equivalent single network (all channels and all traffic accessing one network – i.e. $N = 1$, $K = 250$), it is a tremendous improvement over multiple classical networks.

The QoS behaviour of these networks is interesting. If the P_B, P_D, and P_F are plotted for this virtual network, a shift can be seen. Fig. 9 shows that if the VWN is carrying the maximum allowable traffic, the cause of failures shifts from mostly dropping to an even split of blocking and dropping, as the number of networks increases. In networks where drops are considered more problematic (or costly) than blocks, this may be an important characteristic to consider.

3.3. Primary users with CR capabilities

The case where primary users are given CR capabilities will now be considered. The results in this scenario provide motivation for the adoption of CR by existing licensees, as it is shown that the resulting system can support improved QoS or additional users.

3.3.1. One classic, one network with CR-Enhanced users

In the first scenario, consider two nearly identical networks – both networks have the same number of channels, and the same traffic intensities. However, while network 1 is a classic network, network 2's clients are equipped with cognitive radios, enabling them to operate on either network.

Setting both flows to the $P_F \leqslant 1\%$ traffic intensity obtained in the baseline case, Fig. 10 shows the P_F for each of the two networks, as well as the P_B and P_D for the CR network. While the classic network's P_F has remained the same as expected (i.e. 1%), network 2's P_F has been dramatically reduced.

If instead the traffic on the CR-enhanced network is increased, failure probability rises, but can still be limited to the previous QoS level (or another chosen threshold). Without changing network 1's traffic, the traffic intensity of network 2 was increased in order to find the maximum traffic intensity that can be supported while maintaining a $P_F \leqslant 1\%$. Fig. 11 plots this traffic intensity as the network size changes, as well as the intensity relative to the classic baseline intensity. It is important to note that the QoS of network *1*'s traffic continues to be unaffected. Clearly, the cognitive abilities provide an advantage over the classical network, either by improving QoS through lowering failure probability (see Fig. 10), or by supporting additional users,

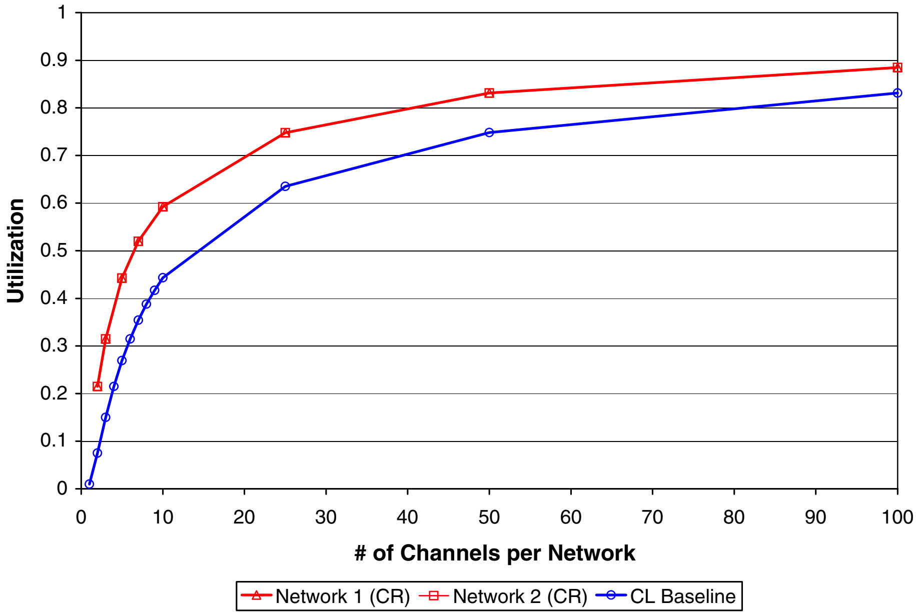

Fig. 9. Behaviour of the VWN at max traffic intensity, $P_F \leqslant 1\%$.

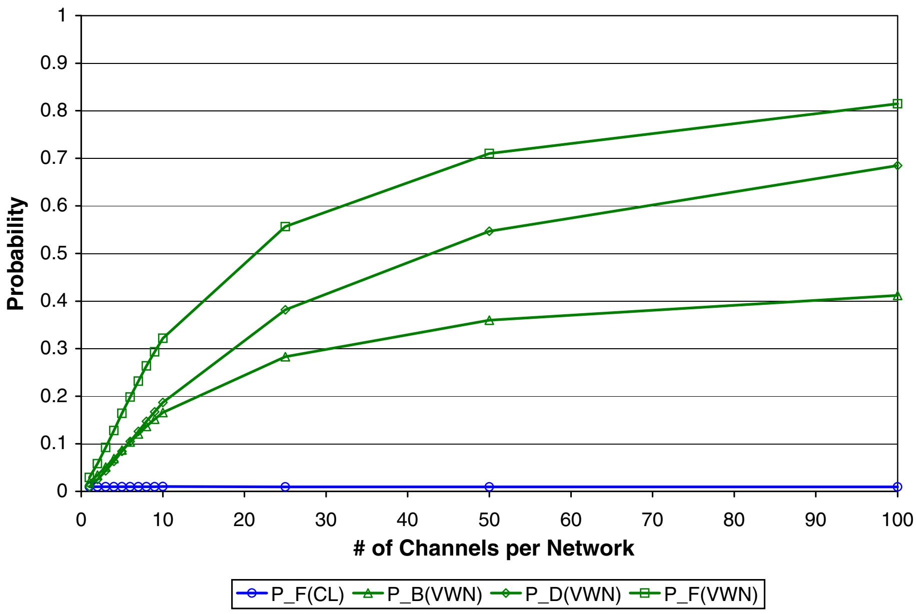

Fig. 10. Provided QoS levels – 1 CL + 1 CR network.

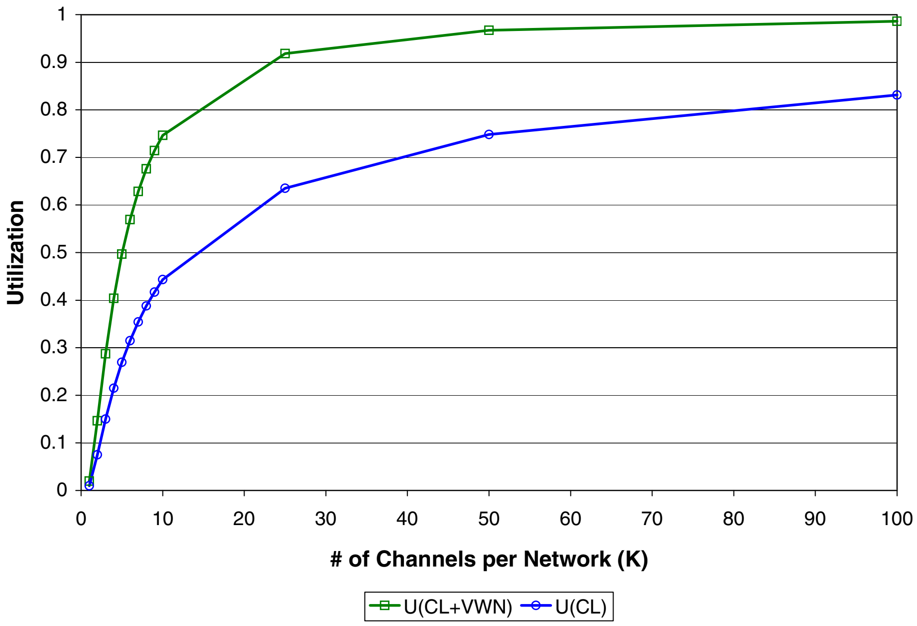

Fig. 11. Maximum and relative traffic intensity – 1 CL + 1 CR network, $P_F \leqslant 1\%$.

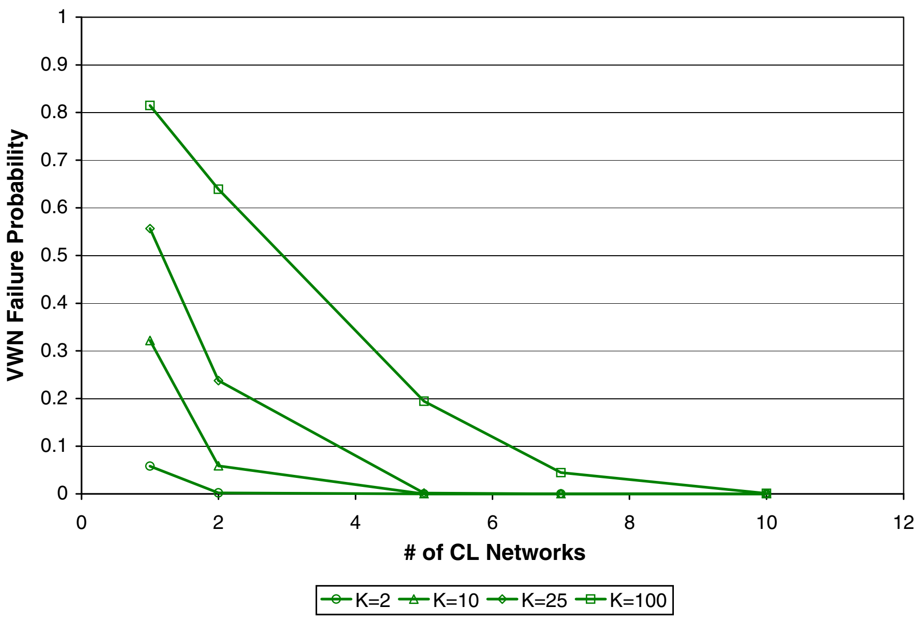

Fig. 12. Relative traffic intensity – 2 CR vs. Baseline CL.

without impacting clients on network 1 (see Fig. 11). For reference, the Fig. 11 also shows the relative increase in traffic intensity of the cognitive network compared to the classic network $\left(\frac{\rho_{CR}}{\rho_{CL}}\right)$. Although the magnitude of the increase regarding supported traffic intensity increases as the size of the network grows, the relative improvement decreases.

3.3.2. Two networks with CR-enhanced users

The same evaluation has been performed for the two network scenario, with both networks supporting CR-enhanced clients.

If the networks carry the baseline traffic intensity, the resulting performance is very similar to the cognitive network in the previous 2-network scenario. As even the classic failure rate was very low (1%), the change to cognitive radio users has very little effect on the overall utilization.

As in the previous scenario, if traffic is increased subject to the 1% constraints on P_F, then the maximum supportable traffic can be found for a certain QoS level. In this scenario, traffic can now be adjusted on both CR networks. As shown in Fig. 12, there is a significant improvement over the classic baseline performance, although the relative gains are most pronounced in small networks.

This additional traffic has some effect on utilization. As seen in Fig. 13, the utilization increased more than in the previous scenario. However, there is still additional spectrum that cannot be utilized without reducing the QoS requirements.

Clearly, the cognitive systems are a major improvement over the classic networks. The primary drawback is the resulting dropping of flows. This is a direct consequence of the cognitive radio requirement that cognitive users cannot interfere with primary users, so if a primary flow arrives, it has priority over a cognitive flow. However, despite the drops, the overall P_F is far better when CR is enabled.

3.3.3. Mixed user types and QoS

With the same traffic intensities, CR-enhancement greatly reduces P_F relative to a classic network. In addition, cognitive abilities can also differentiate QoS for different user types on the same network.

Consider the following scenario. Each network in the system has 10 channels. Traffic arrives at each network with a total arrival rate of 10 and a service rate of 1. However, three different cases are considered for the traffic: (*i*) all traffic is classic traffic (i.e., $\lambda_{CL}(i) = 10, \lambda_{CR}(i) = 0$); (*ii*) all traffic is cognitive ($\lambda_{CL}(i) = 0, \lambda_{CR}(i) = 10$); and (*iii*) half the traffic is classic, half is cognitive ($\lambda_{CL}(i) = 5, \lambda_{CR}(i) = 5$).

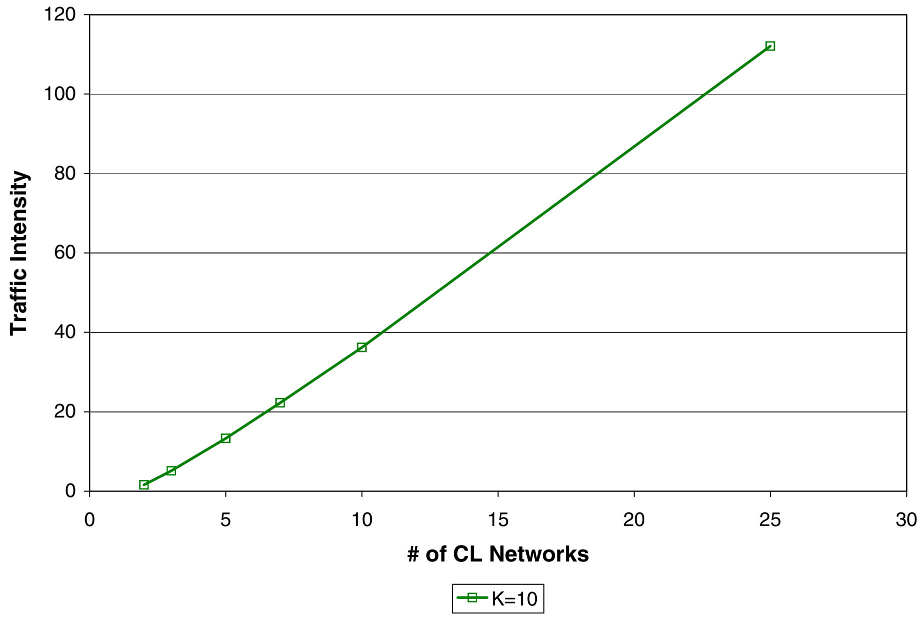

Fig. 13. Utilization increase – 2 CR vs. CL Networks.

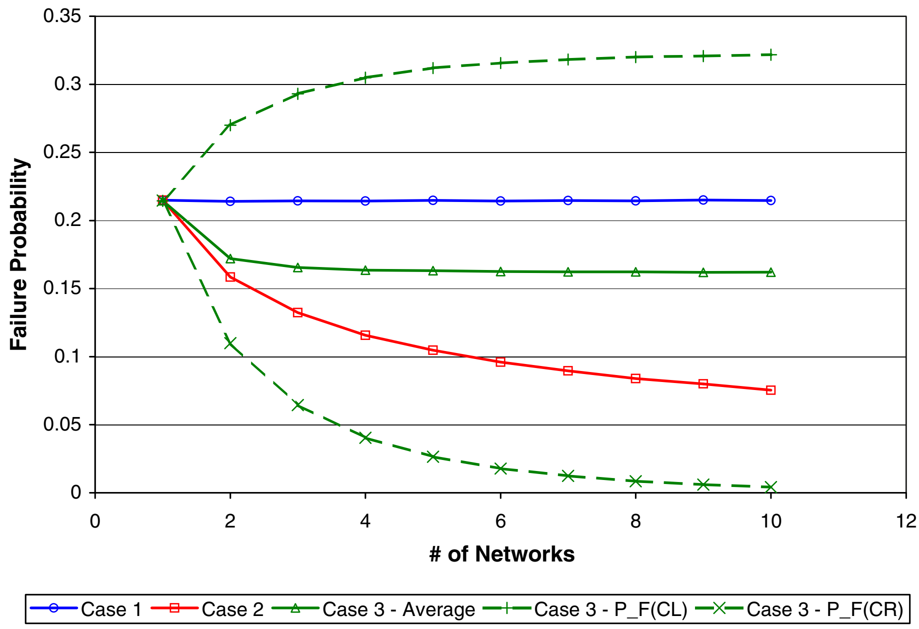

Fig. 14. Failure probabilities for different traffic scenarios: Case 1: all classic users; Case 2: all cognitive; and Case 3: half classic and half cognitive.

Fig. 14 shows the failure probabilities, as the number of networks in the system is varied between 1 and 10. When there is only one network, there is obviously no difference in performance between the different types of traffic. For networks with all classic traffic, failure probability remains constant, as each of the additional networks operates independently of the others. As previously illustrated, networks carrying all cognitive traffic see a decrease in failure probability as the number of networks increases.

The system of networks with mixed traffic shows that the QoS delivered to classic and cognitive users is strongly differentiated as the number of networks increases. The classic users see an increase in their $P_F(i, CL)$. However, the $P_F(i, CR)$ of the cognitive users rapidly decreases, below even the probability of the all cognitive user system. Most importantly, the overall probability for the entire mixed system also improves relative to the all classic case.

4. Data traffic

Up to this point, only connection-based traffic has been considered. In this section, data traffic is included in the system, and several performance characteristics are investigated.

4.1. Model

Unlike connection-based traffic, data traffic does not require continuous service. Service can start and stop, without greatly affecting the service quality. For a network making opportunistic use of available resources, servicing data traffic can allow a much higher level of resource utilization than can be achieved for calls due to QoS requirements.

To study data traffic, a dataflow traffic type is added to the existing network and system model. The data flow is considered to be a bulk transfer consisting of an exponentially determined data volume, where successful service requires that the entire volume must be successfully delivered.

Each network maintains a queue of pending dataflows. When a channel is available, a packet is taken from the first available flow in the queue. Packets are essentially very short connections, which occupy a channel for the duration of their service. However, unlike previous connections, any interruption causes the packet to be dropped (a packet cannot switch channels), and any dropped packet must be re-transmitted in its entirety when another channel becomes available. The maximum length of a packet is P_{Length}. All packets are maximum length, with the possible exception of the final packet in the dataflow. Service time of each packet was determined based on a channel rate parameter (C_{Rate}).

A flow can transmit multiple packets simultaneously, however it is limited to a maximum number of channels by the parameter P_{Limit}. Multiple dataflows operate simultaneously, depending on the number of channels available.

In this model, data is considered to have lower priority than connections. Therefore, if a network has the opportunity to accept an incoming connection (or requires a channel to maintain service to a flow that has been booted from another network), then packets will be dropped. Also, due to the short nature of individual packets, they are never dropped for other packets, regardless of the home network or channel ownership. However, a network will only loan a channel for data traffic from other networks (including a virtual wireless network) if it does not have any data of its own waiting to be sent.

The nature of the preemptive priority between connection and data traffic, combined with the non-preemptive priority among data traffic (i.e., between primary and foreign packets), makes the system very difficult to analyze with the methods used in the connection-based analysis.

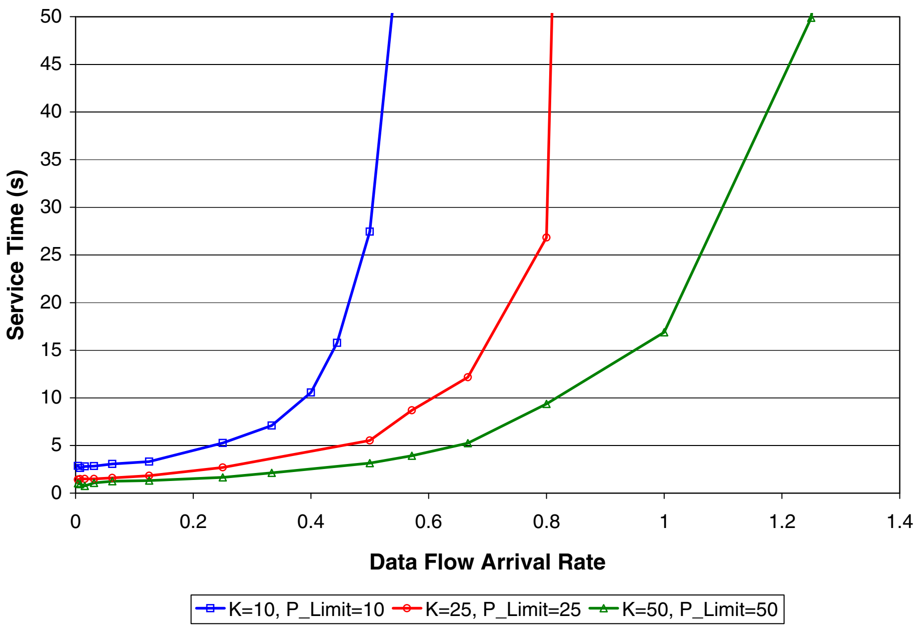

Fig. 15. Single network service time for networks of 10, 25, and 50 channels.

Therefore, the system characteristics were studied using simulations of this model.

Two metrics were considered. Service time is the time required for the system to complete service of a data flow. This is the interval from the time of the flow's arrival, to the time when all of its packets have been transmitted successfully, including all time waiting for service (queuing). Utilization is also studied, and is defined as before. Due to the nature of the model, throughput has not been directly shown, although it can generally be inferred from the utilization.

4.2. Performance evaluation

4.2.1. Single network baseline

As was done for connection-based traffic, the single network baseline is first considered. A single network is considered, carrying existing connection-based traffic. This traffic arrives at a baseline rate giving a $P_B \leqslant 1\%$. Data traffic is then added to the network. Data traffic intensity is adjusted by changing the arrival rate of data flows, while keeping the average flow length constant at 100 packets.

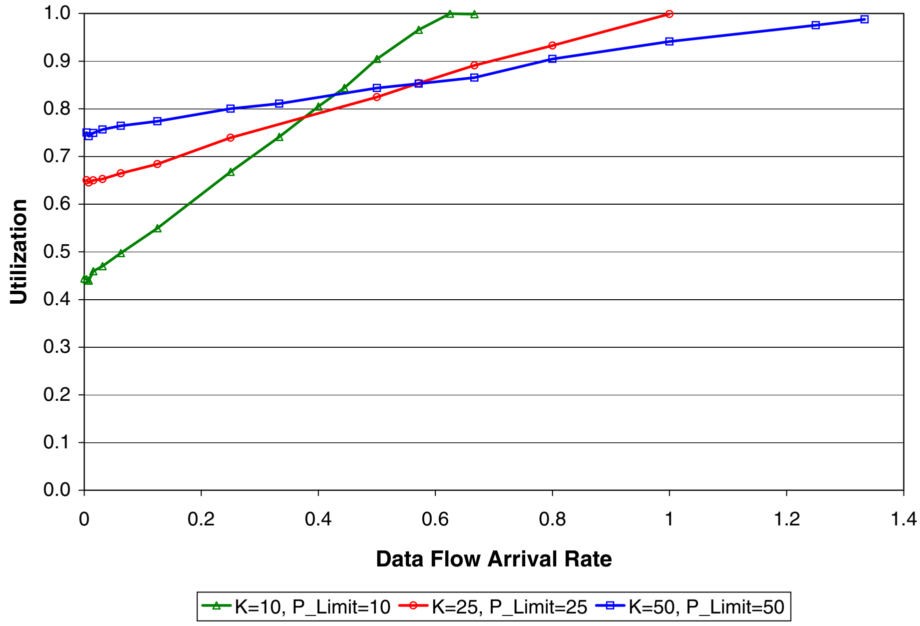

Fig. 16. Single network utilization for networks of 10, 25, and 50 channels.

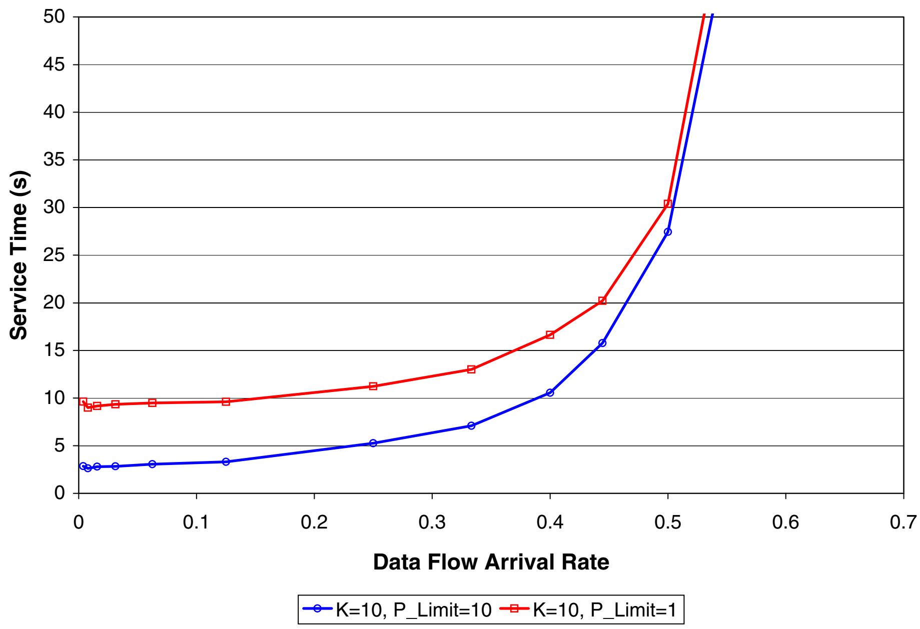

Fig. 17. The Effect of P_{Limit} on service time in a single network.

Fig. 15 shows the service time experienced for networks of K equal to 10, 25, and 50 channels. In each case, P_{Limit} is set to equal K – that is, any flow can fully use all available resources. The service times increase slowly at low arrival rates, with a minimum based on the time required to transmit packets over the channel. As traffic increases past a threshold, transfer times start to increase rapidly as data flows become backlogged and are forced to wait longer for service in the queue. Above a certain arrival rate, the network cannot handle all the incoming flows, and the queues grow. Simulations are terminated if data flow queues grow beyond a certain threshold.

Fig. 16 shows the utilization for the network. With no data traffic, each network starts with a utilization level based on the baseline call-based traffic. The utilization increases linearly as data traffic increases, until the network reaches full capacity. For larger networks, baseline utilization is higher, but the network can also handle larger volumes of data traffic as more channels remain available. In all cases it should be noted that the sharp increase in service times experienced by data flows occurs at utilization levels well below full capacity.

The same scenario is repeated with $K = 10$ and P_{Limit} set to either 1 or 10. Figs. 17 and 18 show the service time and

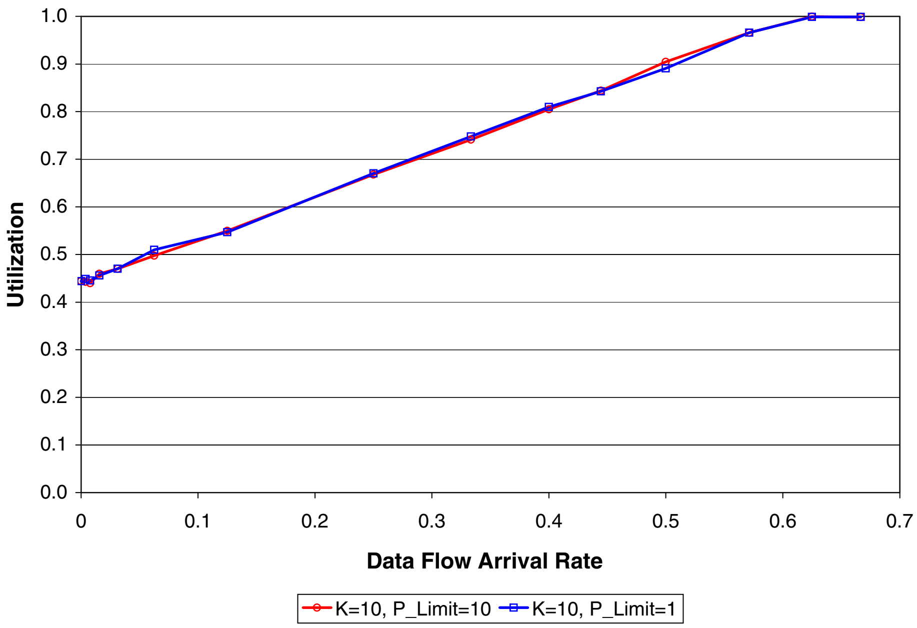

Fig. 18. The Effect of P_{Limit} on utilization in a single network.

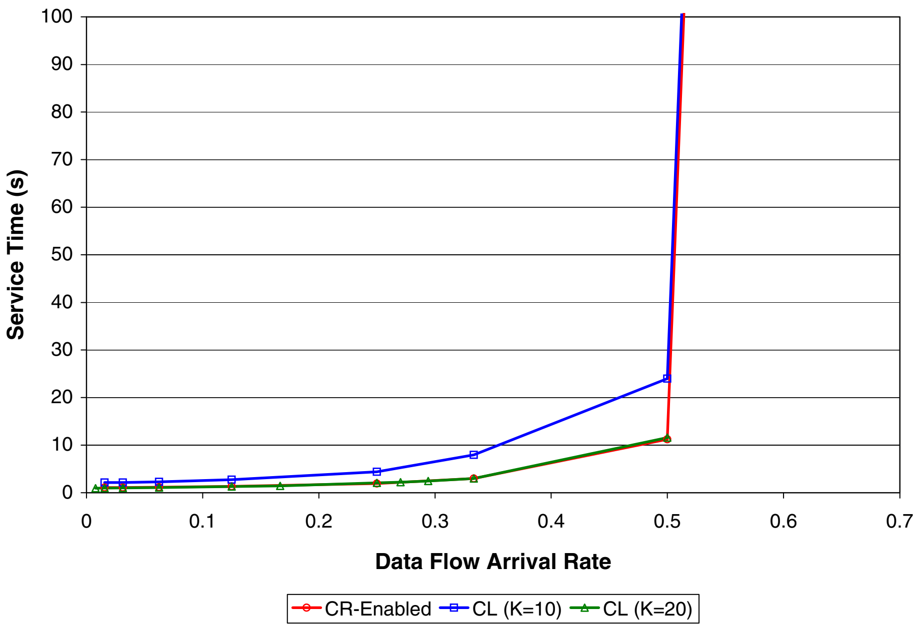

Fig. 19. Service time in two networks with CR-enabled data traffic.

utilization curves for these cases. Setting P_{Limit} to 1 increases the service time experienced by data flows, particularly at low traffic intensities. At these low intensities, service time is dominated by the time required to transmit the packets – very little time is spent waiting for available channels. There is very little difference in utilization as nearly the same amount of data is sent in either case.

4.2.2. Two networks

With a baseline for single network data traffic established, the support of CR-based data traffic can be considered. Two networks, each with 10 channels were created. Each network carrys the baseline level of classic call traffic. Both networks are enabled so that data traffic (but not call traffic) can utilize CR techniques. These results are then compared to both the scenario where data traffic is handled classically on both networks (i.e. two independent networks) and the case where resources are combined into a single network carrying identical loads. For all remaining simulations, P_{Limit} is set to K so that spectrum is utilized to maximum extent possible.

Figs. 19 and 20 show the service times and utilization respectively for this scenario. Dividing the system resources again results in a lower efficiency for the system, as the smaller classic network displays higher service

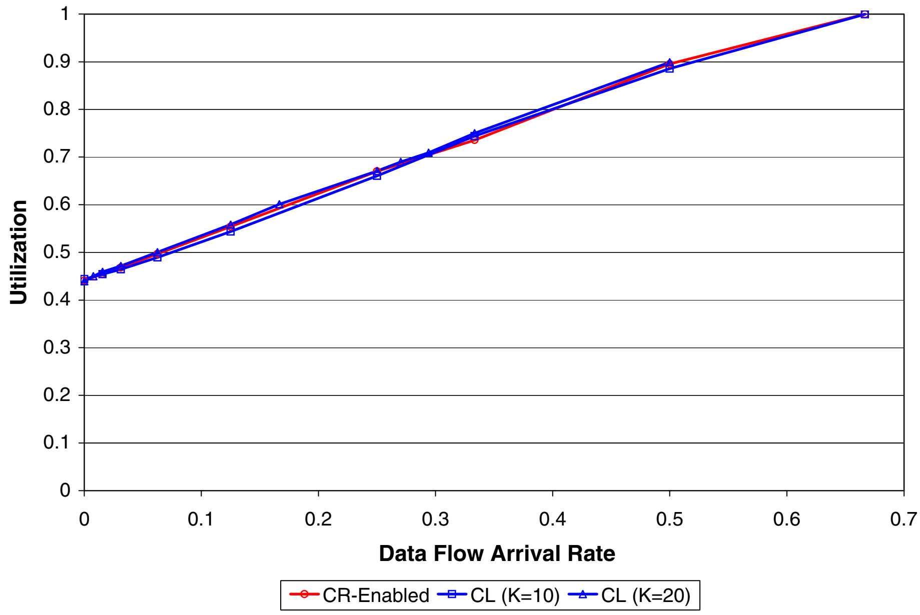

Fig. 20. Utilization in two networks with CR-enabled Data Traffic.

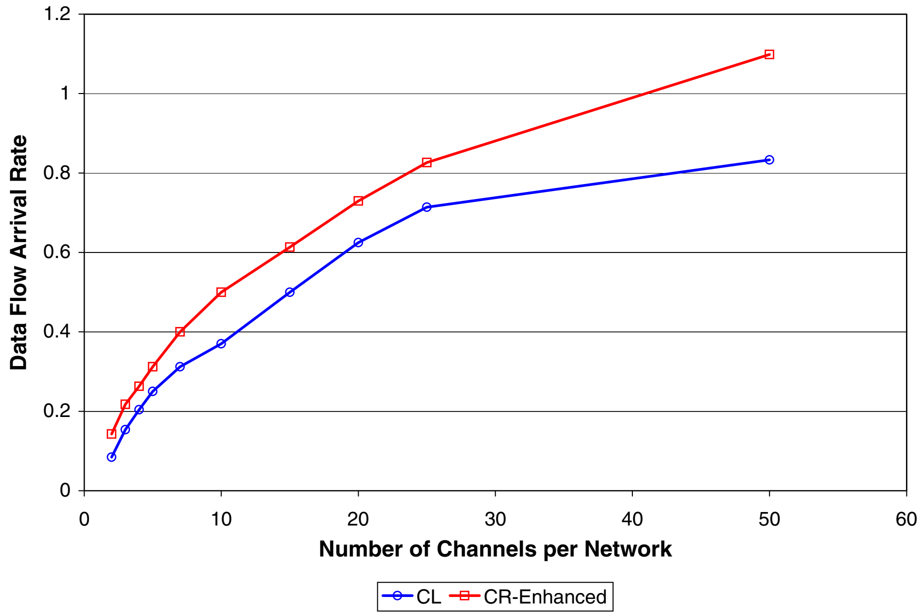

Fig. 21. Maximum data traffic rate to support a service time ⩽10 s.

times due to fewer channels available for data. The addition of cognitive abilities makes the system performance nearly identical to the larger $K = 20$ network for data traffic. The resources are now used effectively, reducing the time required for each flow. The utilization data shows that the overall difference in utilization is very small.

The capacity of the system to support data traffic can also be viewed in terms of QoS requirements. Consider if the system requires that the average service time for all data flows to be less than a certain length of time. Using this requirement, the maximum data capacity of the network can be determined. For the experiments presented, a period of 10 s was used to represent a reasonable service requirement.

Fig. 21 shows the supported arrival rate for data flows. Different sizes of networks are considered, with each network also supporting classic connection-based traffic at a $P_B \leqslant 1\%$ intensity for that network size. From the figure, it is evident that although data capacity increases as the network size increases, the gains are diminishing for large networks. This is due to the increasing intensity (and utilization) of the background connection-based traffic that has priority over the data traffic. However, CR-enhancement clearly delivers a considerable improvement

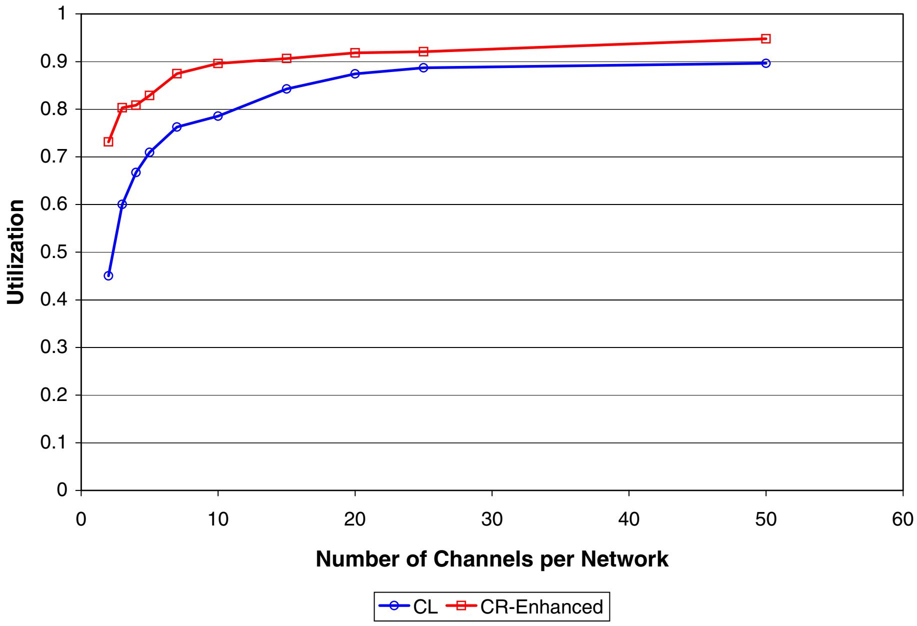

Fig. 22. Utilization while supporting service time $\leqslant$10 s.

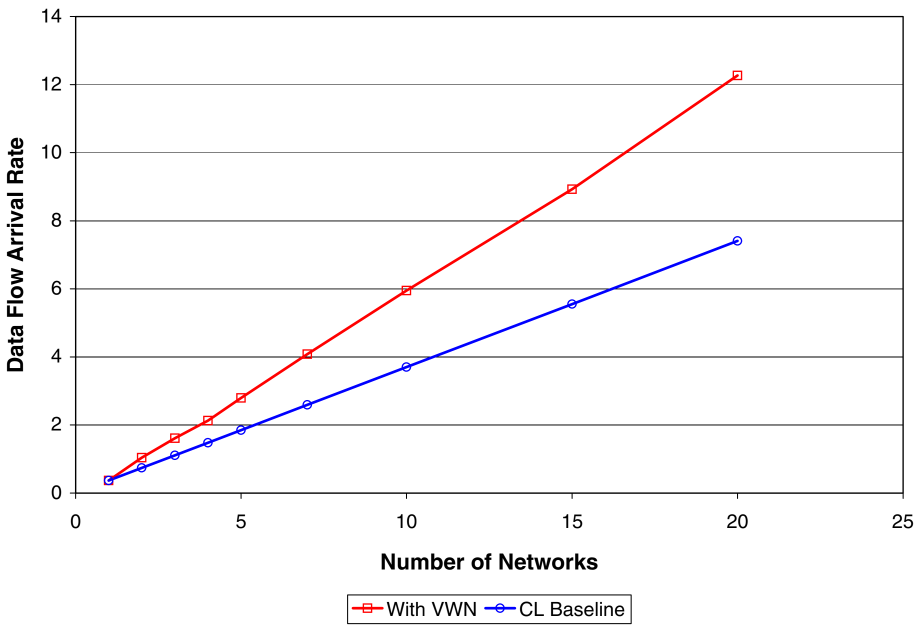

Fig. 23. Maximum data traffic rate supported by a VWN.

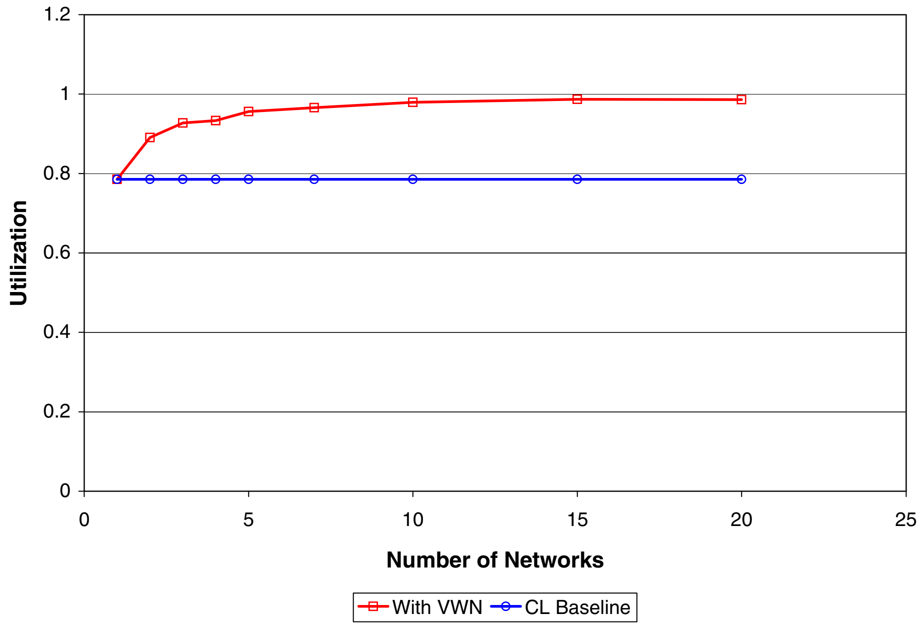

Fig. 24. Utilization of system resources by data traffic from a VWN.

over the classic network case, an improvement that grows as the network size increases. Fig. 22 shows the utilization for the same experiment. Obviously, the increase in supported data traffic directly increases the utilization of system resources.

4.2.3. Data in virtual wireless networks

Finally, the data traffic performance of VWNs is considered. In this scenario, the data traffic capacity of the VWN is studied, while increasing the number of classic networks. Similar to the previous scenario, the capacity is considered under the condition that the average service time for data flows be less than or equal to 10 s.

Each classic network again consisted of 10 channels, and each carried baseline connection-based traffic. Therefore, regardless of the number of networks in the system, each network carried the same existing traffic. Similarly, for the classic case of supporting additional data traffic, each network had the same capabilities, and therefore the same capacity and utilization.

Fig. 23 shows the comparison of the data traffic capacity of the VWN versus the classic network case. Obviously, the classic system grows linearly as each additional network supports the same amount of new traffic. The VWN capacity also appears to grow linearly, but at a faster rate than in the classic networks. This indicates that the formation of a VWN can better utilize residual network capacity for data traffic than any of the networks can individually.

This is also shown in terms of utilization in Fig. 24. The classic network data traffic maintains a constant utilization regardless of the number of networks in the system. Each additional network carries the same intensity of connection and data traffic. However, this means that as the number of networks increases, the total volume of wasted resources also grows. However, the VWN data traffic makes very good use of these resources, and as the system grows, the VWN makes increasingly efficient use of the residual capacity. Although the presence of a service constraint prevents utilization from reaching 100%, it comes very close, even at relatively low numbers of networks.

5. Discussion and conclusions

The results of this work demonstrate that a cognitive radio system can support a considerable amount of traffic. It should be stressed that this traffic is strictly in addition to the traffic supported by the primary networks, and observes the fundamental requirement of non-interference with the primary traffic.

5.1. Traffic models

Within this work, Poisson processes have been used to control traffic. While necessary for the analytical method, this may not accurately reflect the true nature of traffic in such a system. However, consider the effect if a bursty model is used. In this case, an individual network is more likely to have periods with very high or very low utilization. When there is high utilization on a network, the resource sharing abilities would have a greater benefit. When there is low utilization, there are more resources available to be shared with other networks (including VWNs). Also, with such a model, each network would require a greater number of channels to maintain QoS guarantees.

Therefore, we believe other traffic patterns could increase, not decrease the benefits of this approach, compared to the exponential traffic. An obvious exception would be the case where traffic is highly correlated across the entire system. In this case, it is likely that all networks would be congested simultaneously. Although this reduces the benefit of resource sharing, it should be stressed that the system would never perform worse than an entirely classical system.

5.2. Traffic handling

Both connection-based and data flows have been considered. While best-effort traffic has typically been considered for CR due to the opportunistic nature of the system, the ability to support QoS guarantees is very important to current application demands. These results have demonstrated that with multiple primary networks present, QoS traffic can be supported even with all of the primary networks operating at maximum traffic intensities.

The residual capacity to be used by CR is dependent on the traffic on the primary networks. The results presented relied on failure probabilities of 1 and 5 percent. In fact, real world requirements could be even more strict. In addition, burstiness in traffic also requires providers to over-provision their networks in order to ensure acceptable QoS levels. As a result, the actual utilization of the primary networks could be lower than these results assume. On the other hand, real-world traffic patterns of different networks may not be independent. Heavy-use periods on different networks may coincide, reducing the available capacity of CR networks.

Support of data traffic shows very little difference between supporting the traffic on the primary networks, and support through CR. The opportunistic nature of the CR system places the CR traffic at a lower priority than primary network traffic. Data traffic effectively utilizes residual capacity of the primary networks. The very short nature of data packets means that packet drops are relatively rare and have little effect on the network performance.

5.3. Spectrum rights

Opening regulated spectrums has raised concerns from primary users and providers. As a resource, spectrum is extremely valuable, and many licensees have invested heavily in acquiring their spectrum rights [8,9]. They are therefore very hesitant to allow sharing of their frequency bands with secondary "cognitive radio" users. However, CR could offer considerable value to service providers.

Primary service providers can harness cognitive radio abilities to their network and user devices. By collaborating with other service providers, a significant improvement in capacity and/or QoS can be obtained. It has also been demonstrated that the addition of CR capabilities creates a differentiation within the network – CR-enabled nodes realize a significant QoS advantage over classic nodes. This can be used by a service provider to motivate users to upgrade equipment.

Alternatively, the ability of CR to support both QoS and best-effort traffic suggests that a network relying solely on resources obtained through CR can be viable. Although CR has typically been considered as a replacement or extension to unlicensed spectrum, this work suggests that a primary networks' residual bandwidth remains a valuable resource. The leasing of residual capacity from primary service providers is sufficient to create a viable virtual service provider. This potentially provides additional revenue/profit to the primary spectrum owner, and entry into the market for the virtual provider, without the expense of traditional spectrum licensing.

5.4. Limitations and issues

Obviously, this work has relied on a number of assumptions and limitations. The implications of these assumptions will be considered.

This model only considers overlay type traffic, where cognitive traffic occupies unutilized spectrum holes. Cognitive radio may also include an underlay of traffic, where transmissions are made below an interference threshold to avoid affecting primary users. There has also been some discussion of less strict constraints on cognitive radio users, where impact to primary user service is limited or minimal, rather than not at all.

The propagation characteristics of wireless communication makes cognitive radio applications very sensitive to position. Spectrum usage varies geographically, and the power requirements and interference characteristics at any particular location. Depending on the communication methods and parameters used, spectrum may be re-used multiple times within an area. However, this is difficult to characterize due to the wide range of variables involved. Therefore, while this work has taken a broader approach, there is clearly considerable room for improvements in this area. As the next step, the virtual wireless network concept is being applied to wireless mesh networks.

The system considered allows for perfect allocation of channels to traffic. Incoming flows are assigned with no delay, and switching of flows between networks requires no overhead. In particular, cognitive flows require no time to detect an available channel, or to vacate a channel if a primary user requires the resource. Obviously, a real system will require some overhead. However this approach should yield an upper bound for this system. The extent to which overhead deteriorates this performance is dependent on the approaches used.

For example, the time required to detect the availability of an available channel, and the granularity to which the spectrum can be shared, may depend on the level to which a primary network supports CR. A primary network that advertises opportunities might react much faster than CR relying on medium sensing. Similarly, a sensing-based system might have a more difficult time ensuring that the channel is released immediately when the primary network requires the resources again. In these simulations, all control runs directly through the primary networks, with CR traffic assigned to channels (or channels assigned to CR traffic) by the primary network control.

5.5. Summary of conclusions

The secondary use of spectrum will lead to a drastic change in the way spectrum is viewed and utilized. The reuse and and exploitation of spectrum that is under-utilized by primary users could be achieved relatively quickly, and can deliver considerable resources with which new services can be deployed.

Although the opportunistic nature of CR makes it natural to focus on supporting best-effort traffic, this work demonstrates that this may be an unnecessary limitation. While there can be a clear benefit to data traffic, which can yield a very efficient use of resources, classic network models and QoS metrics have been used to show that QoS can be supported. In fact, a significant volume of additional traffic can be supported, without adding resources. By considering QoS at the same level as existing traffic, and network loads at a threshold, it is conclusive proof that this additional traffic is directly attributable to the secondary spectrum capabilities.

The capabilities increase the efficiency of the system, by recapturing much of the system capacity that is wasted due to QoS guarantees. By combining residual capacity form multiple primary networks, additional flows can be supported, either as increased service quality or increased load on the primary networks, or as new traffic on a virtual wireless network. This is accomplished without affecting the primary users, who retain absolute priority over secondary users.

The VWN could be an important tool for managing spectrum in a crowded wireless environment. It not only increases the efficiency of the overall system, it allows new services to be deployed within the existing framework. Despite its inherent dependence on the primary networks, it can operate independently. Allthough for measurement purposes this work has assumed traffic with identical characteristics as the primary flows, the VWN is free to support traffic however the virtual operator decides, including different levels of service quality.

This work has demonstrated that the capacity wasted by the primary networks has considerable value. Primary spectrum licensees and service providers should therefore be advocates of developing cognitive radio technologies, and active facilitators of spectrum sharing. There is an opportunity for these service providers to yield additional revenue through CR, either by harnessing it themselves (through additional users and improved QoS) or by sub-leasing the bandwidth to other service providers, real or virtual.

Acknowledgement

This work was supported in part by the Natural Science and Engineering Council of Canada (NSERC) under its Discovery program, and in part by the WCU (World Class University) program through the Korea Science and Engineering Foundation funded by the Ministry of Education, Science and Technology (Project No. R31-2008-000-10100-0).

References

[1] M.A. McHenry, et. al., Chicago spectrum occupancy measurements & analysis and a long-term studies proposal, in: Proceedings of the ACM 1st International workshop on Technology and policy for accessing spectrum, 2006.
[2] FCC Spectrum Policy Task Force Report, Federal Communication Commission (USA), ET Docket No. 02-135, 2002.
[3] J. Mitola III, Cognitive Radio: An Integrated Agent Architecture for Software Defined Radio. Ph.D. Thesis, KTH Royal Institute of Technology, Stockholm, Sweden, 2000.
[4] S. Haykin, Cognitive radio: brain-empowered wireless communications, IEEE Journal on Selected Areas in Communications 23 (2) (2005) 201–220.
[5] I.F. Akyildiz et al., NeXt generation/dynamic spectrum access/cognitive radio wireless networks: a survey, Computer Networks 50 (13) (2006) 2127–2159.
[6] C. Cordeiro, et. al., IEEE 802.22: the first worldwide wireless standard based on cognitive radios, in: Proceedings of the Firstst IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, pp. 328–337, 2005.
[7] M. Much, et. al., IEEE P1900.B: Coexistence Support for Reconfigurable, Heterogeneous Air Interfaces. In proceedings of the 2nd IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, pp. 381–389, 2007.
[8] N. Jesuale, Spectrum policy issues for state and local government, International Journal of Network Management (Wiley) 16 (2) (2006) 89–101.
[9] J.A. Hoffmeyer, Regulatory and standardization aspects of DSA technologies – global requirements and perspective, in: Proceedings of Second IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, pp. 700–705, 2007.
[10] N. Devroye, P. Mitran, V. Tarokh, Achievable rates in cognitive radio channels, IEEE Transactions on Information Theory 52 (5) (2006) 1813–1827.
[11] S.A. Jafar, S. Srinivasa, Capacity limits of cognitive radio with distributed and dynamic spectral activity, IEEE Journal on Selected Areas in Communications 25 (3) (2007) 529–537.
[12] T. Weiss, F.K. Jondral, Spectrum pooling: an innovative strategy for the enhancement of spectrum efficiency, IEEE Communications Magazine 42 (3) (2004) S8–S14.
[13] F. Capar, et. al., Comparison of bandwidth utilization for controlled and uncontrolled channel assignment in a spectrum pooling system, in: Proceedings of the IEEE 55th Vehicular Technology Conference, vol. 3, pp. 1069–1073, 2002.
[14] B. Wang, Z. Ji, K.J.R. Liu, Primary-Prioritized Markov Approach for Dynamic Spectrum Access, in: Proceedings of the Second IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, pp. 507–515, 2007.
[15] D. Cabric, et. al., A Cognitive Radio Approach for Usage of Virtual Unlicensed Spectrum, in: proceedings of the Fourteenth IST Mobile Wireless Communications Summit, Dresden, Germany, 2005.
[16] C.E. Caicedo and M.B.H. Weiss, Spectrum Trading: An Analysis of Implementation Issues, in Proceeings of the Second IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, pp. 579–584, 2007.

Brent Ishibashi received the B.Sc. degree in Computer and Information Sciences from the University of Guelph, Canada, in 2000, and the M. Math degree in Computer Science from the University of Waterloo, Canada in 2004. He is presently working towards his Ph.D. degree at the University of Waterloo under the supervision of Prof. Boutaba. His research interests include wireless resource management, wireless ad hoc and mesh networks, and cognitive radio.

Nizar Bouabdallah received the B.S. degree in Telecommunications Engineering from Ecole Superieur des Communications (Sup'Com), Tunis, Tunisia, in 2001, and the M.S. and Ph.D. degrees in Computer Science from the University of Paris VI, Paris, France, in 2002 and 2004, respectively. He joined Alcatel Research Laboratories, Marcoussis, France, in 2002, while working on his Ph.D. degree. In 2005, he was with the North Carolina State University, Raleigh, NC, USA, as a Postdoctoral Fellow. He is currently a researcher at INRIA (Institut National de Recherche en Informatique et en Automatique). In 2007, he spent six months as a Visiting Researcher at the School of Computer Science, University of Waterloo, Canada. His research interests include

optical networking, wireless and sensor networks, performance evaluation, network planning and modeling, as well as control and management architectures.

Raouf Boutaba received the M.Sc. and Ph.D. Degrees in Computer Science from the University Pierre & Marie Curie, Paris, in 1990 and 1994 respectively. He is currently a Professor of Computer Science at the University of Waterloo. His research interests include network, resource and service management in multimedia wired and wireless networks. Dr. Boutaba is the founding Editor-in-Chief of the IEEE Transactions on Network and Service Management (2007–2010) and on the editorial boards of several other journals. He is currently a distinguished lecturer of the IEEE Communications Society, the chairman of the IEEE Technical Committees on autonomic communications. He has received several best paper awards and other recognitions such as the Premier's research excellence award.