Improving the Reliability of Wireless Networks Using Cognitive Radios

Arash Azarfar, Jean-François Frigon, and Brunilde Sansò

Abstract—To ensure widespread deployment and popularity, next generation wireless services will require a Quality of Service (QoS), and particularly a reliability, that is independent of the radio transmission medium. However, because of the failure-prone nature of wireless networks, providing a reliable communication link and guaranteeing a consistent QoS to users become key issues. In this tutorial, we describe the most common source of failures in wireless networks and provide a systematic failure classification procedure. Drawing from the vast literature on reliability in wireline networks, we then explain how cognitive radios can use their inherent capabilities to implement efficient prevention and recovery mechanisms to combat failures and thereby provide reliable communications and consistent QoS under all circumstances.

Index Terms—Wireless Communications, Failure, Reliability, Cognitive Radio, Prevention, Protection and Restoration.

I. INTRODUCTION

WIRELESS services have recently enjoyed tremendous success because users increasingly appreciate the ability to access or share information anywhere and anytime. In return for these conveniences, users have accepted that wireless links are unreliable with inconsistent Quality of Service (QoS) in which problems (dropped or hung connections, variable data rates, delays, etc.) are frequent occurrences. Although users consider these problems as inherent characteristics of wireless networks, as wireless services become more pervasive and replace applications currently provided only in wireline networks, the question of reliability becomes more critical. For example, if a wireless multimedia distribution service is used to replace cable or DSL, users will not accept frequent disruptions during their favorite show. The requirement of reliability and seamless QoS calls for a paradigm change in the design of wireless networks in order for them to offer a reliability comparable to that produced by the fives nines design approach in wireline networks (i.e., 99.999% service availability). The problem is that, because of their nature, wireless networks are inherently more unreliable and error-prone. In addition to the basic hazards, including natural disasters, power failures and hardware failures, that threaten all communication networks, several other factors, such as the random nature of the communication channel and the presence of interferers, affect the reliability in wireless networks. The study of wireless networks reliability is thus a critical aspect for designing network architectures suitable for next-generation wireless services.

A mobile-user connection usually consists of a concatenation of fixed and mobile networks. Any consideration of reliability must consider the entire end-to-end connection. These notions have long been important areas of research in wireline networks [1]–[3] and in wireless network infrastructure [4], [5]. However, the end-to-end reliability is limited by its weakest components. Traditionally, the wireless access link has been perceived as this weakest component and numerous techniques, such as channel coding and diversity, have been proposed at the physical layer to improve the radio-link quality [6]. The design of dependable wireless access networks using a mesh topology has also been a topic of growing interest [7]. However, providing dependable wireless access requires a systematic approach that accounts for the different failure modes of wireless links, the available prevention and recovery mechanisms to combat these failures and the relevant performance metrics.

The objective of this tutorial is to discuss the reliability issues associated with wireless networks and explain how Cognitive Radios (CRs) can be used to improve the reliability of wireless networks. In this tutorial, we use the term reliability as a qualitative concept which includes all related parameters such as availability and performability and is used interchangeably with robustness or dependability. For the exact definition of these parameters, interested readers can refer to [3], [8]. We first present a different perspective on wireless networks based on the notion of reliability, which draws from the vast body of work on the reliability, availability, performability and survivability of fixed wireline networks.

We then discuss how CR technology can be used to implement one of those approaches to build dependable wireless networks. Cognitive Radio is a new paradigm that was introduced in 2000 by J. Mitola to solve the problem of spectrum scarcity and usage inefficiency [9]–[12], and a new wireless standard based on this technology is currently under development [13], [14]. Most of the research on CR networks has focused on its spectrum agility features to exploit available spectrum not used by licensed users. However, CR nodes also possess the necessary attributes to make considerable progress in the robustness and dependability of wireless networks [10], which has been less explored. Considering the failure classification and failure management approaches that we discuss, the main contribution of this tutorial is in discussing how, for various wireless network failure causes, the different cognitive capabilities of CR nodes can be used to prevent such failures, decrease their occurrence rate and severity or...
handle the failure after occurrence in a more efficient way. This tutorial provides to the reader the necessary insight to expand this view of cognitive radios in several new directions.

This tutorial is organized as follows. In Section II, we examine the concept of reliability in wireless networks by classifying failures and studying the most common causes of failure. In Section III, we survey the traditional concepts of protection, recovery mechanisms and related performance metrics in wireline networks to study how failures can be prevented or managed after occurrence. The concept of cognitive radio is introduced in Section IV and, in Section V, based on the wireless failure classification and the different failure management approaches discussed previously, we explain how CR features can be used to provide more efficient prevention and recovery mechanisms and build dependable wireless access. Section VI discusses the challenges associated with the development of cognitive radio networks and reviews some of the limiting factors that may slow down the deployment of those networks. Finally, the tutorial is concluded in Section VII.

II. FAILURES IN WIRELESS NETWORKS

The challenge in improving the reliability of wireless networks is that the causes and consequences of failures are extremely diverse. For instance, in addition to major failures (such as a base station or mobile malfunction), fading, interference, and battery power, just to name a few, can cause failures. Furthermore, a binary failure model is not adequate. For example, a channel fade might result in a new modulation and coding scheme with a lower data rate. Although the link is still functional, the data rate might be too low to sustain the traffic demand and thus a failure occurs. Therefore, prior to studying prevention and recovery mechanisms for wireless networks, it is important to correctly classify the different failures that can afflict them.

A. Failure Classification

As illustrated in Fig. 1, a classification based on the component type, severity, rate, duration, dimension and scope axes can be used to encompass the most important characteristics of failures in wireless networks. Similar classifications with slightly different failure parameters have been previously proposed [1], [3] and our classification includes all aspects discussed in the literature. The detection or estimation of these failure parameters can help to devise better prevention methods and aid the recovery mechanisms to select the most appropriate approach, as will be discussed in Section V.

The definition of each failure classification parameter is as follows.

1) Component type: This parameter indicates the component under failure. In wireless networks, two components can suffer from a failure: the nodes (fix or mobile mobile nodes, base station or spectrum server) or the transmission links.

2) Severity: Two levels of failure severity can be identified: hard and soft. A hard failure occurs when the communication flow is totally halted. In contrast, a soft failure refers to a situation where the communication flow is not stopped, but the service that can be offered (bandwidth, QoS, etc.) is degraded. This parameter is also known in the litterature as the Failure Degree [1].

3) Failure Rate (Frequency): The failure rate describes the number of times that a failure happens in a specified period. For example, a node failure due to power loss may happen once a month while a failure due to hardware defects happens once every two years.

4) Duration (Outage Time [1]): A wireless network failure can be either permanent or transient. For example, if a user is moving away from a base station, the link failure with this base station will be permanent, while a channel fade causes a transient failure whose duration is determined by the mobile speed.

5) Dimension (Failure Cardinality): The failure dimension indicates whether an event results in single or multiple failures. A single failure dimension implies that, in a short period of time, it is unlikely that multiple failures will occur, whereas a multiple failures dimension indicates that, if a failure occurs, there is a high probability that other failures will also appear somewhere else in the network. For example, a channel fade has a single failure dimension, whereas the appearance of an interferer has a multiple failures dimension.

6) Scope (Failure Propagation): The failure scope is related to the failure propagation concept. That is, a single failure might not only affect the component under failure but also influence the behavior of other components in the surrounding area. The failure scope indicates the area (number of links and nodes) affected by a failure. For example, a link bandwidth degradation in a wireless mesh network might affect the performance of other links in the neighborhood due to the congestion created by the re-routed traffic. However, in a single-hop network, a link degradation only has a local effect on the link.

By the proposed parameters, we try to cover different aspects of failure in wireless networks. The proposed classification is not completely orthogonal and the correlation between the parameters depends on the other factors such as network topology, type of redundancy and application. For instance, in general, permanent failures are hard failures; however, this is not always true and depends on other parameters. For example, a channel failure due to interference which forces a radio to

---

**Fig. 1. Failure classification chart.**
change its operating frequency is assumed permanent. But, the severity depends on the availability of other channels. If the node quickly finds a new channel, the failure can be assumed soft, otherwise it is a hard failure.

It is also important to consider that a failure can be classified differently depending on the perspective. For instance, in a mesh network, a permanent node failure can be interpreted as a soft failure for other nodes as they are able to change their route. However, for the failed node (user) this failure is hard because it causes the user to get disconnected from the network. Next important point is the perspective and terminology of the failure. As another example, when a protected link is disconnected and the backup link is used, from the link perspective, this event is a soft failure which decreases the overall resource availability in the network. However, this link failure is masked and tolerated at the network operational level. Moreover, the classification of failures also depends on other parameters such as the applications and the specified QoS thresholds like acceptable delay and packet loss ratio.

### B. Failure Causes

Failures in wireless networks occur for various reasons. In this section, we discuss and classify the most common causes of failures and Table I presents their classification according to the previously proposed criteria.

1) **Node Failure**: There are several possible sources of node failure. For example, a power outage, hardware defects and severe software faults are sources of hard node failures because the connectivity is completely lost. In a single-hop wireless network, there is no recovery from such a failure unless a spare node is employed. In a mesh network, the node failure will affect multiple links, but the traffic going through this node may be re-routed. Note that, in both cases, all traffic originating from or targeting the failed node will be lost.

Several types of backup resources may be used (if no backup resource is available, the node failure is permanent). For example, multiple antennas or transceivers can be used. If one component fails (partial failure), the communication link can still use the other antennas or transceivers. However, this technique might result in a lower data rate or link reliability after the recovery. A backup power resource (for example, a battery) can also be used to cope with a main power outage. To preserve its energy, the failure recovery algorithm might elect to reduce the transmit power such that transmission is now only possible with closer neighbors or at a lower data rate. A similar situation can also occur for mobile nodes when the battery level goes below a threshold.

A node failure is permanent and, depending on the availability of backup resources, it can be either a hard or a soft failure. Possible redundancies and the quality of hardware components are such that the rate of node failure is normally low. A node failure will affect a variable number of surrounding links and nodes depending on the network topology. However, it is unlikely that multiple node failures will occur simultaneously.

2) **Link Failure**: Path loss, shadowing, multipath fading and interference are the major wireless channel impairments that can cause link failures. A wireless link completely fails when the performance metrics (bit error rate, signal-to-noise ratio, throughput, etc.) are not acceptable. However, in most cases, the signal can still be received with degraded metrics. Bit Error Rate (BER) is the most widespread performance metric and link quality indicator in wireless communication. In general, the BER is inversely proportional to the Signal-to-Noise Ratio (SNR) at the receiver but the exact relation depends on the exact modulation scheme and diversity techniques that are used [15]. In a high SNR regime we further have that:

\[
BER \propto SNR^{-L} \quad (L > 0)
\]

where \( L \) represents the diversity order of the communication system [6].

Path Loss: In a wireless network, when the distance between the source and the destination of a transmission link increases due to the users’ mobility, the received signal power decreases thereby increasing the BER and packet loss and degrading the link quality. Let \( d \) be the distance between the transmitter and the receiver (assuming a constant noise and interference power), we then have:

\[
SNR \propto \frac{1}{d^n}
\]

where \( n \) represents the path loss exponent which depends on the characteristics of the environment. In urban areas, \( n \) is generally between three and four [15]. Because the distance varies gradually, the failure caused by distance is a soft failure, but it can become a hard failure as the nodes become farther apart. The failure due to distance is considered permanent because it cannot be assumed that the nodes will come closer in the future.

Environment Effects (Shadowing and Fading): Stochastic signal variations, such as shadowing and multipath fading, usually cause transient soft failures. For example, signal degradation due to a building shadow will disappear when the user moves away and small-scale fading causes large signal variation with a displacement on the order of the wavelength. Estimating the duration of those failures can help in implementing efficient recovery mechanisms. These variations decrease the power of the received signal which in turn increase the BER [15].

Interference: In a wireless environment, several users can simultaneously transmit on the same channel, which can create interferences. The SNR at the receiver is proportional to the inverse of the interference:

\[
SNR = \frac{P_r}{N+I}
\]

where \( P_r \) is the power of the received signal, \( N \) represents the power of the noise and \( I \) stands for the total interference. Higher interference thus directly increases the BER of the link. Some technologies, such as spread-spectrum communications, are more immune to interference than others (such as narrowband systems). Therefore, depending on the communication technique, the impact of an interferer can vary from a soft failure to a total link failure [6], [15]. In addition, the failure duration depends on the nature of the interferer and can be transient or permanent. For example, a cordless phone will create interference on a wireless network during the time of a conversation but, if a neighbor sets up his wireless network on the same frequency channel, the failure will be permanent.
Furthermore, due to the broadcast nature of wireless media, an interferer will usually simultaneously trigger failures on several links.

**Congestion:** In wireline networks, a high volume of traffic can generate packet loss and delays that can cause severe failures in higher layer communication protocols. In a wireless network, similar phenomena can occur. However, because the wireless channel is shared among several users, a source with a large volume of traffic will degrade not only the performance of his link but also that of the other surrounding users. For example, in random-access protocols such as in IEEE 802.11, a node with a large amount of traffic will increase the contention delay (collision probability) of all users in the network [16]. Therefore, traffic increase in one node can cause failures somewhere else in the network.

Special care should also be taken when classifying the cause of a failure. For example, when a node has several operational transceivers and one of them experiences a hardware failure (partial node failure as explained earlier), one of the operating links fails and the radio handles this failure by switching to other transceivers. This implies that we can model these types of node failure as a link failure and consider hardware problems as a new cause of link failures for multi-transceiver nodes. However, a failure in a spare transceiver which is not operational represents a degradation of hardware redundancy and reliability and can not be modeled as such as a link failure.

### III. Traditional Reliability in Communication Networks

Prior to studying the use of cognitive radios to improve the wireless network reliability in Section V, we will review in this section the traditional reliability concept developed over the years for wireline communication networks. Network robustness has been a major driving factor in the design of wireline networks (such as public switched telephone networks (PSTN) and asynchronous transfer mode (ATM) networks) partly due to regulatory requirements and customer expectations. Network robustness implies network reliability, which generally in a communication network is related to the ability to [5]:

1) Prevent the occurrence of failures;
2) Solve and recover from failures.

#### A. Prevention Mechanisms

Networks use prevention mechanisms to decrease the occurrence or the severity of failures. Most of these approaches are based on the use of dependable hardware and software for the transmission links and nodes. Other solutions such as selecting less-hazardous environments and equipping communication cables with protective covers are also classified as prevention methods.

The objective of a prevention mechanism is to postpone the occurrence of failures. The most appropriate performance metrics to evaluate a prevention mechanism are thus the number of failure occurrences in a period, the probability of a failure occurrence and the duration between two consecutive failures known, respectively, as the **Failure Rate**, the **Reliability** and the **Mean Time Between Failures (MTBF)** of the system.

The **Failure Rate** is the frequency at which a failure happens and can be obtained mathematically from the failure probability density function (PDF) or cumulative distribution function (CDF).

There are different accepted definitions for **Reliability**. Qualitatively, the reliability of a system can be associated to the ability of the system to perform its tasks under some performance and timing constraints [4], [8]. Mathematically, the reliability is given by the probability that no failure occurs during a certain period of time [17]. Let \( f(t) \) be the failure probability density function, the reliability is then defined as

\[
R(t) = 1 - \int_0^t f(x) \, dx \tag{4}
\]

and represents the probability that no failure occurs between the time zero and \( t \).

The term **Mean Time To Failure (MTTF)** represents the average time between the return of the system to its normal state and the next failure (i.e., average time to failure). Mathematically, MTTF can be defined based on the failure probability density function as follows:

\[
MTTF = \int_0^\infty t f(t) \, dt = \int_0^\infty R(t) \, dt \tag{5}
\]

To represent the time between two consecutive failures, the term **Mean Time Between Failures (MTBF)** is often used instead of MTTF and is given by:

\[
MTBF = MTTF + MTTR, \tag{6}
\]

where **Mean Time To Repair (MTTR)** stands for the average repair time after a failure occurred. For further definitions, interested readers can refer to [17, Chap. 9]. Advanced readers can also find more mathematical details in [18].

These different metrics revolve around the same concept and may be used interchangeably in different networks. In this tutorial, we will show in Section V how the cognitive radio features can be employed to decrease the probability of failure occurrence or equivalently to increase the MTTF of a wireless network.
B. Recovery Mechanisms

Recovery mechanisms are divided in Protection and Restoration methods [19], [20]. Protection mechanisms are network design and capacity allocation techniques [5] which assign backup resources in advance, whereas restoration methods attempt to find a solution after the occurrence of a failure. Usually, recovery mechanisms are hybrid and use a mixture of the two approaches. The ability of a network to recover from failures is also studied in conjunction with other concepts such as Survivability, Fault-Tolerance and Healing. For example, survivability is a qualitative concept and is defined as the capability of the system to continue performing its specified tasks when a failure happens [4]. In the other words, survivability discusses how a system handles failures by using recovery mechanisms.

1) Protection Methods: In general, protection methods specify some reserved (spare, backup) resources that will be used when a failure happens, and these backup resources are substituted for the failed ones. The resource substitution can be done automatically by the network or manually by a network administrator. Redundant resources may also be employed actively or passively [21], [22]. In active protection, before the failure, the backup resource performs the same tasks or plays the role of a load-balancer and shares the specified tasks with the main resource. Aggregated links are an example of active protection [23]. In passive protection, the redundant backup resource only monitors the status of the system when the main resource is functional. Blocked (backup) links in a Spanning Tree Protocols (STP) without aggregation are an example of passive protection [24].

One of the most popular protection mechanisms is Automatic Protection Switching (APS). In APS, a predefined backup resource is substituted automatically when the main resource fails. The most basic approaches are the 1+1 and 1:1 APS schemes, where each resource has a separate backup [25]. Obviously, the cost of this mechanism is high, but it is simple and provides very good and fast recovery in instances of single failures. M:N is another, often employed, APS scheme where one of the M backup resources is substituted when one of the N resources fails. In this case, signaling is required to perform this switching, which can incur a small delay. Obviously, this scheme cannot handle more than M simultaneous resource failures. APS measures can be deployed at different layers of communication protocol and for various resources. For example, at the physical layer, protection could be applied to optic fibers, time slots, subcarriers or wavelengths. In the IEEE 802.22 standard, backup frequencies can be specified for each link that can be used in case of licensed users’ appearance or quality concerns [13]. However, a physical layer APS measure is sensitive to hazards, such as a backbone accident, that affect all the links routed together.

In the network layer, and jointly with the physical resources, the APS mechanisms are usually based on the network’s connectivity and are more robust to single failures. Ring-based, mesh-based and p-cycles can be classified in these survivability methods. A ring-based protection measure is a simple survivable topology that is based on the ability to transmit information in both ring directions. Mesh-based protection is based on more advanced graph connectivity than ring-based protection and can be used to provide link or path protection [25]. In link protection (local protection), each link has backup links or paths (paths involve multiple links). If the backup resources are dedicated to the protection of a link, we have disjoint (dedicated) link protection (similar to 1:1), and if the resources are used to simultaneously protect several links, we have shared link protection [25] (similar to M:N). Path protection (global protection) provides end-to-end protection in the network. It is thus more efficient but more complex than link protection. We can again have either dedicated or shared path protection. P-cycle protection [26] uses the idea of mixing rings and mesh protection mechanisms. P-cycles provide good protection and low recovery delay similar to ring mechanisms, but their required redundancy is similar to mesh protection [26]. These concepts can be applied to ad-hoc or mesh wireless networks, especially when wireless nodes are equipped with more than one transceivers.

Multiprotocol Label Switching (MPLS) Fast Reroute (FRR) is another local protection approach which is applied to Label Switched Paths (LSPs) [27]. A LSP (1:1) or a group of LSPs (1:N) passing through a node is protected by a backup LSP. When this node detects the failure in the main LSP, it activates and uses the backup one. For more details about different protection mechanisms, readers can refer to [19], [26].

2) Restoration Methods: In restoration methods, when an active resource fails, there is no pre-assigned backup resource and the substitute resources should be found dynamically in reaction to the failure occurrence. Depending on the resource type and operation layer, different schemes can be used. For example, at the routing layer in mesh-based or ad-hoc networks, restoration can be applied at the link or path level [25]:

- Link Restoration: when a link fails, the nodes at both link ends dynamically find a new path to locally re-route the information around the failed link;
- Path Restoration: when a path fails (due to one or more link or node failures), both ends of the path dynamically find a new end-to-end route around the failures.

In some cases, the recovery methods are hybrid and include both protection and restoration approaches. For instance, in a M:N path protection method, if more than M failures happen (protection fails), a restoration mechanism like re-routing is used to find a new path. Most of these hybrid approaches are cross-layer (multi-layer) healing mechanisms where protection is used for some resources in one layer (usually lower layers) and restoration is employed in the upper layers when the protection fails. For example, a backup frequency may be assigned to a wireless link. When the main frequency fails due to high fading or interference, the user switches to the backup channel. However, if the user finds the backup channel also occupied or unusable, the node requests the base station a vacant channel (infra-structured topology) or routing algorithms are activated to find a new path to the destination (mesh or ad-hoc topologies). An example of such an implementation of cross-layer hybrid approaches in wireless networks can be found in [28] where the authors propose protection channels in the physical layer and restoration at the transport layer. For more information about multi-layer survivability, interested readers are referred to [29].
3) **Performance Metrics:** Whereas the recovery mechanisms directly affect the network reliability [30], appropriate performance metrics are required to evaluate and select the most suitable recovery mechanisms for a communication network. Depending on the network topology, applications, performance goals and traffic type, different scenarios might have different requirements and constraints. Various evaluation metrics have thus been defined [4, 20, 26, 28, 30, 31] and the most common and important parameters are listed in Table II.

The **Recovery Delay**, which is also known as the MTTR or outage duration, is the average time between the occurrence of a failure and the return to the normal network state after recovery. This parameter is a function of the failure detection time and the speed of the recovery algorithm.

The **Time to Next Failure** is related to the recovery method intelligence and accuracy. It characterizes how the recovery method is able to prevent or postpone the occurrence of failures in the future. This means that an intelligent recovery method can also act like a prevention approach. This parameter has a strong relation with other performance metrics depending on the technology and topology of the wireless network. In a long operational duration, similar to prevention mechanisms, this parameter is evaluated by the MTTF that stands for the average time between the return of the system to the normal state and the next failure.

Without considering other impacts of a failure such as degraded performance or partial failures as described in the resource availability parameter (see below for its definition), the network either performs recovery and is in the recovery/repair state with an average duration equal to MTTR, or is functional with an average duration equal to MTTF. The **availability** metric measures the network’s ability to perform its designated function at any given time and indicates the percentage of the time that the system is functional. Since it is a function of the MTTF and the MTTR as follows:

$$\text{Availability} = \frac{\text{MTTF}}{\text{MTTF} + \text{MTTR}}, \quad (7)$$

we did not include it as a separate metric in Table II.

The **Loss** metric indicates how many packets may be dropped or lost during the recovery period, (e.g., due to the absence of a operational transmission link while a suitable backup resource is established). The number of packets lost depends on several factors such as the packet-dropping policy and buffer management used during the recovery interval, the recovery delay and the traffic characteristics.

The **Complexity** metric quantifies the complexity of the recovery algorithm implementation. For example, a re-routing algorithm that searches the entire network graph is more computationally intensive than a recovery algorithm that acts locally. It can be interpreted as timing or processing cost.

The **Costs** incurred by the implementation of a recovery mechanism are also an important metric for network operators. Resources that are not used actively in the normal network state and are reserved for recovery are a major contributor to the cost of a recovery mechanism. For example, recovery schemes that reserve a backup resource for each active resource (i.e., 1+1 or 1:1) are considered the most expensive methods.

The **Resource Availability** metric measures the relationship between the available resources before the failure occurrence and after the recovery. It is a very general but important metric. Recovery policies affect this parameter considerably. For example, whereas an exhaustive search might yield the same resources as before the failure, a recovery mechanism might decide to select a backup transmission link or a path with less resources in order to reduce the delay, loss, cost and complexity. The difference between throughput, bandwidth, quality of service and the number of users that can be served can all be included in this parameter. In telephony networks, usually the blocking rate (number of users that can not be serviced) is compared before and after the failure.

The **Performability** parameter, which takes into account the performance of the system after a failure [32], is often used to quantify the resource availability metric. Performability discusses what percentage of the designated tasks for a system can be performed after a failure. This parameter is directly related to the resource availability parameter as discussed above but no unique definition can be proposed because it depends on several factors such as the topology and network model, application and traffic model and the technology employed in a communication network [2], [32]. Performability measures can be used in the classification of failures as soft or hard, as discussed in Section II-A.

The **Failure Masking** metric indicates the ability of the recovery mechanism to mask the failure and prevents its propagation. This term can be defined by the number of services or users affected by the failure.

Considering these performance metrics, we will show in Section V how the efficiency of the recovery mechanisms in a wireless network can be improved (lower MTTR, higher MTTF, less failure propagation and etc.) utilizing the cognitive radio features.

### IV. Cognitive Radio Networks

In this section we will review the different features of cognitive radio, and particularly the ones that will be used in Section V to improve the reliability of wireless networks. A Cognitive Radio node is an intelligent radio that can observe and learn from the environment and adapts its communication parameters based on this knowledge [9]. This technology has been proposed as an extension to Software Defined Radio (SDR) [33]–[35] where this software architecture provides re-configurability for CR to adjust its communication parameters based on its knowledge and observations. A collection of CR nodes that organize themselves using these cognitive features is called a Cognitive Radio Network (CRN). CRN is a general concept and includes any wireless network from a point-to-point link between two stationary cognitive nodes to a complex cellular or mobile ad-hoc network.

Nowadays, the most known objective of Cognitive Radio (CR) is to solve the problem of spectrum scarcity and usage inefficiency. From 2000 to 2004, reports by the Federal Communications Commission (FCC) stated that on one side, due to extension of the wireless applications and mobile telephony networks, some parts of the spectrum are over-utilized and new requests for spectrum have dramatically
increased (spectrum scarcity), while on the other side, several assigned frequency bands are under-utilized and partially used (spectrum usage inefficiency). To solve these problems, as a sub-method for dynamic spectrum access [36], the idea of using spectrum holes or spectrum opportunities to deploy new networks was proposed. A spectrum hole is a portion of the spectrum which is licensed to a primary network but is vacant in a specific time and geographic area [10] and can be temporarily used by a secondary network. This secondary manner of communication implies that a secondary user is able to use a vacant channel when the primary (licensed) user is not present. In wideband technologies, the “vacant” term means that the interference level experienced by primary users will be acceptable (lower than the required threshold) if a secondary network starts operating in this channel. Cognitive Radio, based on its spectrum-awareness, was selected as the best candidate for implementing this concept. A secondary CRN consists of wireless nodes that are spectrum-aware. Using their spectrum-sensing capabilities, they detect chunks of unused spectrum licensed to primary users (e.g., television channels with no broadcaster in the geographic area) and deploy a secondary CRN in the available spectrum. The IEEE 802.22 standard, which is based on this concept, is currently under development [13], [14]. Note that this mode of operation adds a new cause of link failure when a primary user appears in a CRN operating channel.

The main requirement for a CR secondary user is to respect the priority of primary users. If the physical layer technology allows the co-existence of primary and secondary users in the same channel (e.g., wideband technologies), the CR node should make sure that it does not exceed a given interference level. Otherwise, the CRN has to vacate the channel immediately, which results in a link failure unless a multi-channel (aggregated) communication technology is used.

However, for the purpose of improving wireless network reliability, we consider in this tutorial the general definition of a CR node [9] which possesses the following cognitive features:

- Spectrum-awareness;
- Location-awareness;
- Learning and History-awareness;
- Adaptability and Reconfigurability;
- Reasoning and Decision-making.

Using these features, a CR node operates in a cognitive cycle illustrated in Fig. 2.

A. Spectrum-awareness

As mentioned earlier, spectrum-awareness implies that a CR node is able to sense the spectrum to find the spectrum holes and estimate their quality considering the interference and environmental effects. Based on the spectrum sensing results, the transmitter and receiver select a common spectrum hole as their operating channel. Sensing may be done periodically or occasionally to verify if the channel is still vacant of the primary users’ activity and/or to verify if the channel quality is acceptable. If one of these conditions is violated, the CR node decides either to change its configuration (e.g., transmission power, modulation and coding scheme) to decrease the interference level and compensate the channel effects or to switch to a new vacant channel.

In the sensing process, the CR node attempts to detect the signal of other users. If the CR node has some information about the interference signal characteristics, sensing can be done more accurately and faster using coherent detection mechanisms (e.g., Matched Filter Detection [37]). Otherwise, the CR node can use Energy Detection mechanisms [37] which is fast but not as accurate since different signal sources in the channel can not be distinguished and the possibility of incorrect decisions increases. A more accurate approach is cyclostationary feature detection where the CR node analyzes the spectrum for a long enough period and detects the other users’ modulated signal based on their periodicity [37]. Noise signals are not cyclostationary so the CR is able to more easily
and accurately differentiate them from modulated signal. The cost that is paid for this accuracy is a longer sensing time [37].

When a CR node senses the spectrum, obstacles, distance and fading conditions may affect its sensing accuracy. In a rich fading or high interference environment, it is possible that a CR node detects an unused channel occupied (false-alarm) or a busy channel vacant (miss-detection). When a false-alarm occurs, the CR node does not use the channel for communication although it is vacant. For a miss-detection, the CR node will either interfere in the communication of other users or suffer from high interference level. When a CR node incorrectly detects the status of one channel, one of its neighbors may sense this channel correctly due to its different position and environment. Therefore, the exchange of sensing information among CR nodes in a neighborhood decreases the possibility of false-alarms or miss-detections. Moreover, depending on the available bandwidth, the sensing of the whole spectrum for a CR node could be time-consuming. Again, by cooperation of CR users, the task of each CR node can be simplified and the sensing task can be distributed among the nodes, each node senses a narrower band of the spectrum. The distribution of this information among all the nodes makes all of them aware of the spectrum. These methods are called Cooperative or Collaborative Sensing [38], [39].

In centralized CRNs, instead of exchanging the information with the neighbors in a distributed manner, gathered spectrum information by each user can be sent to the cognitive base station or a spectrum server which analyzes the spectrum, assigns a channel as the next operating channel to each CR node and sends a list of channels to each of them to be sensed in the next sensing period. Although collaborative sensing seems very useful and interesting, several problems may arise. First, synchronization among users, especially in an ad-hoc network, is hard and increases the complexity. Second, the dissemination of the sensing information introduces additional overhead and interference, and needs efficient signaling and routing algorithms. For more details about the different mechanisms of spectrum sensing and related challenges readers can refer to [37], [38], [40], [41].

B. Location-awareness

Location-awareness can greatly help a CR node to have an accurate view of the network. The propagation delay estimation, analysis of the position of the base stations and routing in an ad-hoc network are some of the tasks that can be simplified or improved when a CR node has location-aware mechanisms. CR nodes may employ a Geographical Positioning System (GPS) tool to accomplish this task [11], [42]. When the use of GPS tool is not practical, positioning information may be sent to users by a central entity. In ad-hoc networks, some users can be equipped and send their position to the neighbors. Using directional antennas and distance estimation, other users will be able to find their approximate positions. In [11, Chap. 8], more information about the usefulness of location-awareness in CRNs and its implementation details is provided. For location-aware routing mechanisms, readers can refer to [43], [44].

C. Learning and History-awareness

In the definition of the CR by Mitola, learning capability implies that a CR node possesses a database which can save the observation results and experienced events in order to use them to take history-aware decision in the future. This capability is implemented as a Knowledge Base in some proposed CR architectures and prototypes [9], [42], [45]. The learning phase is the outcome of observation, planning and decision-making phases. Mitola explains that learning can be very time-consuming and computationally intensive. So, the CR node may have special intervals for learning and in those intervals it does not function as a wireless node. That is, the learning intervals are similar to a sleep period [9].

D. Adaptability and Reconfigurability

As an extension to SDR, reconfigurability is one of the main features of a CR which makes it enable to adjust different communication parameters based on the current system state. For example, at the physical layer, the frequency, operating bandwidth, modulation and coding scheme, number and configuration of antennas and the transmission power are some of the parameters that can be adjusted. Correspondingly, the sensing process parameters, e.g. the sensing duration or signal power threshold levels, can be adapted.

E. Reasoning and Decision-making

The huge amount of information that a CR node should process to make a decision and the variety of alternative configurations motivate the need for a reasoning unit in a CR architecture. Just as an example, when a link fails due to the increase of the distance between two nodes, several options are available to them:

- Changing the location if one of the node is mobile;
- Recovering the link by changing the modulation and coding scheme;
- Recovering the link by switching to a lower frequency with lower path loss;
- Finding a new route.

The reasoning unit should analyze all of these possibilities in a very short time to make the best decision. Some performance bounds should be defined and the decision will be made based on these bounds. Spectrum occupancy, SNR, BER, delay, packet loss or a mixture of these parameters may be used as performance bounds.

F. Cognitive Cycle

All the previously described capabilities are operating inside the framework of the CR cognitive cycle (Fig. 2). The cognitive cycle consists of five main stages completed by the learning stage. In the observation stage, the radio senses and identifies the environment to obtain a variety of facts about it. Spectrum awareness and location-awareness methods are part of this stage. During the orientation stage, the CR node adapts its architecture according to the priority and importance of the observed events. Based on the available resources and environmental parameters, the CR creates different plans, decides which plan will be selected and applies the decision
by changing the required parameters in various layers. Finally, the CR node can learn from its observations and decisions for future uses.

V. COGNITIVE RADIO NETWORKS AND WIRELESS RELIABILITY

Our main objective is to design a wireless system architecture that can counter wireless failures and improve wireless network reliability using approaches similar to those currently in place in wireline networks. As will be discussed in this section, considering its cognitive features and intelligence, the Cognitive Radio has the necessary attributes to achieve this objective. The modified CR cognitive cycle presented in Fig. 3 illustrates the inherent capability of CRNs to prevent or recover from failures to improve wireless network reliability. In stage 3, after the environment observation phase and the monitoring of the performance and QoS parameters (stages 1 and 2), the cognitive radio detects whether any new event has occurred or may be occurring in the near future. To make the most appropriate decision, the CR node classifies the new event as a Warning or Failure in stage 4. In the former case, the CR deploys failure prevention measures. For example, if a CR mobile station detects that its distance from the base station is increasing, it can switch to a lower modulation and coding to prevent path loss failure. In the later case, the CR node characterizes the failure according to the failure classification chart (Fig. 1) and uses the appropriate protection and restoration techniques (stages 5 and 6). The CR node can also learn from the current experiences and observations to help it in the development of more efficient plans in the future (stage 7).

In the following sections, considering the failure causes studied in Section II-B, we discuss how cognitive radios are able to prevent the occurrence of such failures or recover from them using the cognitive features of CRs. Table III summarizes some of the prevention and recovery methods implemented using cognitive features that can be employed for different failure causes.

Note that the goal of this tutorial is to provide a perspective to the reader on the different possible approaches to use the cognitive radio features to implement prevention and recovery mechanism in wireless networks. Unfortunately, due to the wide array of possibilities, we do not go in details on any of those approaches. However, the material in this section along with the appropriate references provided in this tutorial are sufficient to allow the interested reader to dive further and research or implement the specific CR prevention or recovery mechanism which best suits its needs.

A. Prevention Methods

1) Reliable Transmission Techniques: Transmission techniques with higher level of reliability can be employed in CRNs to reduce the probability of link failure (failure rate) or its severity. For example, wideband transmission techniques, such as spread spectrum, frequency hopping and Orthogonal Frequency Division Multiple Access (OFDMA) can be used to increase the wireless network reliability in environments with high levels of interference. Using the reconfigurability and reasoning of CR nodes, when a CRN detects an environment with high level of interference or a primary network using widebands technologies, after the required coordination, it can reconfigure the physical layer to a more appropriate wideband technology. Similarly, transmission parameters such as the channel coding type and rate, the signaling rate and the modulation can be adjusted to increase the reliability of distant users operating with a higher noise level, or to mitigate the impact of interference. In severe fading environments, time, frequency and spatial diversity techniques can be used to increase the system reliability and prevent the occurrence of failures [6]. It is also important to note that, generally, there is a tradeoff between reliable transmission techniques that facilitate failure prevention and costs, complexity and throughput.

2) Sensing Mechanisms: An important feature of cognitive radios is their ability to perform spectrum sensing. A reliable operating frequency channel can therefore be selected based on its interference level (from primary or secondary users) and its attenuation, shadowing and fading characteristics. Increasing the accuracy of the sensing algorithms is thus a primary factor in providing an accurate channel characterization and in improving network reliability. Better sensing algorithms and longer sensing periods can be used to improve the accuracy. In addition, when obstacles, distance, hidden nodes and other special situations limit the sensing abilities of a node, collaborative sensing could be a key in improving the reliability provided by this prevention method.

3) Historical and Predefined Data: Location-awareness is one of the important features of a CR node [9]. Geographical and environmental information can be obtained through a GPS in the CR node, embedded information in packets exchanged between nodes or a central server that sends the most up-to-date global Radio Environment Map (REM) information [11, Chap. 11]. When a CR node knows its location, using learning capabilities, it can record several events (normally not transient) experienced in different locations and times. Fig. 4 illustrates this capability. The CR node uses the geographical coordinates of the previously explored areas to remember that, between point 1 and 2, there was a WiFi network. The road between points 2 and 3 is in an urban area and the possibility of interference, shadowing and multipath fading is very high. The area between points 3 and 4 is located near a highway, an airport and a train station. The Doppler spread is therefore higher in this location. After point 4, there was a hilly area and the multipath spread was much larger. When the CR node is approaching these geographical coordinates, a warning alarm is generated that notifies the CR to take adequate measures to prevent the occurrence of failures. For example, the CR node can select a more reliable frequency band or a more robust transmission technique to prevent failures. The CR node or base station may also analyzes alternative paths and decides that the path should be modified to reduce the failure probability and thus proposes a more appropriate alternative path. Otherwise, the CR node can activate protection and restoration methods to mitigate the impact of possible failures.

Another example is using previous observation information in the spectrum assignment. Considering the sensing results of
Fig. 3. Modified cognitive cycle for failure management.

TABLE III
PREVENTION AND RECOVERY METHODS FOR DIFFERENT FAILURE CAUSES.

<table>
<thead>
<tr>
<th>Failure Type</th>
<th>Prevention Methods</th>
<th>Recovery (Protection + Restoration) Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Node failure</td>
<td>Components reliability</td>
<td>Re-routing and mesh networking</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Planning based on HW avl.</td>
</tr>
<tr>
<td>Distance</td>
<td>Sensing mechanisms</td>
<td>Adaptive transmission techniques</td>
</tr>
<tr>
<td></td>
<td>Historical and predefined data</td>
<td>Channel switching</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Re-routing and mesh networking</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Motion estimation</td>
</tr>
<tr>
<td>Shadowing and Fading</td>
<td>Reliable transmission techniques</td>
<td>Adaptive transmission techniques</td>
</tr>
<tr>
<td></td>
<td>Historical and predefined data</td>
<td>Motion estimation</td>
</tr>
<tr>
<td></td>
<td>Sensing mechanisms</td>
<td>Backup channels</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Channel switching</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Re-routing and mesh networking</td>
</tr>
<tr>
<td>Interference</td>
<td>Sensing mechanisms</td>
<td>Backup channels</td>
</tr>
<tr>
<td></td>
<td>Reliable channel assignment</td>
<td>Channel switching</td>
</tr>
<tr>
<td></td>
<td>Reliable transmission techniques</td>
<td>Motion estimation</td>
</tr>
<tr>
<td></td>
<td>Historical and predefined data</td>
<td>Re-routing and mesh networking</td>
</tr>
<tr>
<td>Traffic congestion</td>
<td>Traffic monitoring</td>
<td>Channel aggregation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Backup channels</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Channel switching</td>
</tr>
</tbody>
</table>
the other channels or previous sensing periods, the CR node is able to prune its decision space when some dependencies between contiguous channels or consecutive time slots exist. For instance, in a network where primary users operate in bonded (aggregated) channels, the CR user may postpone the sensing of channels $N+1$ and $N+2$ when channel $N$ is sensed busy (current slot time) or postpone the sensing of channel $N$ itself when this channel was sensed busy in the previous slot and its probability of return to the available state is low according to its recorded background.

4) **Reliable Channel Assignment:** The objective of this prevention method is to select channels that minimize the probability of future failures. For instance, in a multi-channel and aggregated transmission technology, using distributed and non-contiguous channels decreases the probability of simultaneous interference or appearance of primary users in all of the channels [46]. Sensing and observation results of a long period (statistical information through history-awareness) besides the information that a CRN may obtain through central entities are analyzed to make the best decision for the next operating channel. Some of the proposed CR MAC protocols use statistical data of the channel occupancy to decide the next operating channel more accurately [47], [48]. Therefore, even when there is no a central entity or negotiation between the neighbors, a CR node does not sense all channels one by one in a random or sequential order. Instead, it uses memorized background information about the channels to create a more efficient sensing order or channel access which decreases the possibility of false-alarms and miss-detections.

In [10], the author discusses that cognitive features improve the channel identification tasks such as estimation of channel-state information (CSI) and prediction of channel capacity. This information can be used during the channel selection to decrease the probability of failure in future.

CRNs should also be aware of other neighboring networks. To prevent interference among CR nodes, channel assignment can be done centrally by a server or base station. In an ad-hoc topology where central spectrum management is not possible, collaborative methods can be employed. In this case, spectrum occupancy and available channel information in each node is sent to the other nodes periodically. When a CR node wants to select a new channel, it will consider both its local and its neighbors’ information.

In a cellular scenario, the absence of collaboration among neighboring base stations increases the probability of interference noticeably. In this situation, CR base stations may coordinate with each other to establish an efficient static or dynamic channel assignment. The location-awareness features of CR nodes can also be used for a location-aware channel assignment. For example, base stations will not assign the same channels to CR nodes located near the border of neighbor cells.

5) **Traffic Monitoring:** A traffic monitoring unit can be employed to periodically measure the status of buffers (level, waiting time, packet loss ratio, etc.), the number of backoffs for random access technologies or other QoS thresholds [28]. When a node detects that there is a traffic congestion, it can apply congestion-avoidance algorithms e.g., Random Early Detection (RED) to prevent the occurrence of failures caused by congestion. It can also change the frequency band if possible or increase the backoff time to decrease the probability of a deadlock.

6) **Components Reliability:** More reliable hardware and software can be used to improve wireless node reliability. Note that this approach is not specific to CR nodes and is applicable to any wireless system. However, component reliability does not have any impact on link reliability unless for the case that we model a partial node failure (i.e., if one of the multiple operational transceivers fails) as a link failure.

### B. Protection and Restoration Methods

1) **Backup Channels:** During the CRN setup phase, the network chooses an operating channel but can also select, in a way similar to APS in wireline networks, a backup channel for protection against failures in the main channel [31]. This idea is implemented in IEEE 802.22 standard [13]. These backup channels can be shared between links if M:N protection is adopted. However, because these channels are devoted to improving the network availability, they are not available for data transmission and the network total throughput is therefore decreased. An important issue is thus to maximize
the reliability while minimizing the spectral efficiency loss due to the use of redundant resources.

2) Channel (Frequency) Switching: If the channel conditions reach a certain threshold where a transmission link failure occurs, the CRN should determine if the failure duration exceeds the acceptable level. In this case, it searches for a new channel and several selection criteria can be used. First, the CRN should switch to a channel with a lower interference level and no primary users. It can also, as illustrated in Fig. 5a, use a frequency band with better shadowing and fading characteristics. In the cases where the failure is related to traffic congestion, when the channel bandwidths are not the same, the CRN can search for a wider channel that can accommodate more traffic. This approach is shown in Fig 5b. A channel with better transmission characteristics (e.g., a higher signal-to-noise ratio (SNR)) can also help alleviate the congestion by enabling a higher spectral efficiency.

Some wireless technologies, such as multi-band OFDM, also let wireless nodes communicate using simultaneously several contiguous or non-contiguous channels [49]. For instance, IEEE 802.22 provides channel bonding by merging two or three TV channels [13]. In case of congestion, the CRN can search for unoccupied subchannels and bond them to its current channels, as shown in Fig. 5c where two subchannels are used to increase the operational bandwidth.

Note that several similarities exist between the channel-switching restoration mechanism and the link-restoration mechanism used in wireline networks.

3) Adaptive Transmission Techniques: A software defined structure allows a CR to adapt its transmission strategies as a function of the current environment to ensure protection against link failure.

Following a received SNR variation, Adaptive Power Management can be employed to maintain the link throughput and BER [10], [42]. However, increasing the power should be done while taking into consideration the interference level in the network and the impact of an interference increase on other nodes.

Adaptive Modulation and Coding (AMC) is another mechanism available to cope with SNR variations. AMC adapts the constellation size, the coding scheme and coding rate to achieve a constant BER as a function of the received SNR [15]. For example, if there is a link failure due to a SNR decrease, the link can be restored by using a lower constellation size and a lower coding rate. However, the AMC technique affects the link throughput and traffic congestion might then occur.

Another approach to restore the transmission link under fading conditions is to employ Adaptive Diversity Techniques [6]. Several parameters of the time and frequency or spatial diversity techniques, e.g. the length of the repetition in a time diversity or the number of parallel channels in a frequency diversity, can be adjusted to the state of the network to provide more robustness following link failures. Adaptive diversity can therefore increase the Time to Next Failure after a link failure at the cost of additional complexity, larger bandwidth or lower throughput.

Other transmitter and receiver internal parameters can also be adapted to the current system state. For example, as a function of the current environment fading characteristics, the equalizer parameters of a narrowband system, the number of fingers in a direct sequence spread spectrum Rake receiver or the number of subcarriers in an OFDM system can be changed and adjusted [11, Chap. 4]. Those changes are more related to the system complexity but do not affect directly other system metrics such as the data throughput. Other adaptive parameters, such as the training and synchronization sequence duration, sensing period or the cyclic prefix length have an impact on the system spectral efficiency. Reconfigurability and adaptability are also present in the higher layers of the communication protocol in CRNs. For example, the spectrum access protocol can be changed to time-slotted following the detection of a cluster/cell with a cluster head or base station.

4) Re-routing and Mesh Networking: In a multihop network, if a node or a link failure occurs, the information can be transmitted via new routes in the network. In CRN, location-awareness and history of the cognitive nodes help the CR nodes to find a backup path quickly and more efficiently. Location-aware routing protocols have been a research topic in mobile ad-hoc networks (MANET) [43], [44] and can be implemented more easily and efficiently in CR wireless networks due to the adaptive structure of these networks. Thanks to adaptive transmission techniques in CRNs (discussed above), a broken link and consequently a route may be recovered faster and with less complexity by changing the transmission parameters such as frequency, power and modulation scheme without re-routing the broken link locally or changing the end-to-end path globally (link and path restoration mechanisms were discussed in Section III-B2).

Just as with wireline networks, the backup routes can be also predetermined or dynamically discovered. However, in the former case, due to the nodes mobility, the backup routes must be periodically updated.

It is also important to underscore the difference between the protection offered by backup frequency channels and backup routes. In the first case, if a link failure occurs, the involved nodes select a different operating frequency to reestablish a communication link between them. Additionally, backup channels do not offer protection against node failures. In the second case, the information is re-routed via a new path involving other nodes when a failure occurs. If a link fails between two nodes and other protection and restoration methods are unavailable or inefficient, the CRN can change from a single-hop communication link to a multihop path to maintain the connectivity between these two nodes. Route backups thus offer protection against both node and link failures. However, if the same frequency band is used, the performance after recovery from some causes of link failure, such as interference, is not as good as the one offered by channel protection [31]. A better approach involves a combination of route and channel backups. A CRN also supports dynamic protection and restoration methods. For example, different routing algorithms may be used for path restoration in a multihop scenario depending on the situation. Re-routing in ad-hoc CRNs and mesh networking are hot research topics in the continuing effort to improve the survivability of cognitive radio networks [50]–[52].

5) Planning Based on Hardware Availability: In CRNs, it is possible to include the current availability status of the
Fig. 5. (a) Recovery by switching to a frequency channel with lower path loss and shadowing, (b,c) congestion recovery approaches: (b) CR node switches to a wider available channel. (c) CR node occupies two subchannels to increase its available bandwidth.

Fig. 6. Considering the status of the hardware availability to make a new decision.

hardware parts with the other parameters used to make a new decision. Fig. 6 represents a simple case where the availability of backup power (BPWR) is involved in the decision-making. In the normal state, where both the main and redundant resources are available, the CR node has certain functionalities. If a failure occurs, the redundant resource is substituted and the system reliability decreases. The CR node can then decide to change its functionalities, such as its transmission range or relaying in multihop networks, to prevent the occurrence of failures after the recovery. This strategy can also be used by considering the state of the components, such as the battery charge level.

6) Motion Estimation: A CR node can use an estimate of its trajectory, direction and speed to evaluate a link-failure duration. The failure-duration estimation can also be improved if the CR node has location-awareness capabilities and information about the current environment is available in its knowledge base. Then, considering the performance requirements and available resources, the CR node can decide on the most appropriate protection and restoration method. For example, if the link-failure duration is estimated to be short, adaptive transmission techniques might be a sufficient protection measure. However, if the mobile speed is low and the failure might last longer, other protective techniques such as using backup channels may be more appropriate.

It should be noted that the cognitive recovery methods we described are not disjoint and may have weak or strong relations based on the topology and type of recovery. For instance, when a CR node misses both its main and backup channel, it uses spectrum sensing and frequency switching capabilities to find a new operating channel. In cognitive routing, the CR node can employ almost all of other recovery methods such as motion estimation, frequency switching and the state of hardware resources to select the best route or routing algorithm.

We can conclude that the cognitive capabilities enable CRNs to use specific recovery methods such as motion estimation and full adaptability that are not available in traditional wireless networks to improve the reliability of wireless networks. It can also provide improvements for existing methods like backup channels. In the latter case, the improvement that can be achieved is quantified and compared using the performance metrics reviewed in Section III-B3. For example, a cognitive assignment of backup channels considers the statistical information of the channels (learning and reasoning) and specifies/selects a channel with a low correlation with the main channel to minimize the possibility of concurrently missing both channels. Therefore, compared to the traditional assignment of backup channels, the MTTF (probability of failure) and availability of resources after failure increase, however the MTTR, cost and failure masking are almost the same. Generally, the complexity and hardware cost of the cognitive recovery methods are higher.

VI. CHALLENGES AND LIMITATIONS

Although in an ideal scenario a CR node is assumed to be a fully adaptive, reconfigurable and intelligent radio, in implementation, several limitations exist. These limitations have
caused that the current deployments and available standards based on the CR technology mainly focus on the spectrum awareness and sensing capabilities of CR. However, with the advance and development of related technological areas such as artificial intelligence, machine learning and digital signal processing, it is expected that other capabilities of CR technology will be developed and implemented in near future [42], [53]. To the best of our knowledge, the limitations to implement an ideal CRN can be categorized in three main areas:

- Decision-making time as an overhead for the main communication;
- Hardware complexity: cost, processing and power consumption limitations;
- Channel variations due to secondary manner of communication.

A. Decision-Making Time

For simplicity, let assume a time-slotted CRN where all activities in the network are synchronized to the boundaries of the fixed slot times with duration $T$. Then each CR node spends, as illustrated in Fig. 7.a, a portion of the slot time, called decision-making time, for observations, spectrum sensing, negotiation and message exchange and recovery (if necessary). The time spent for decision-making is an overhead for the CR node as it is generally not able to use this portion of the slot for transmission. So, a longer decision time implies shorter transmission time or lower throughput. A CR node cannot use this portion for data communication mostly because for sensing mechanisms based on energy detection, the CRN should be silent to be able to accurately detect the presence of the primary users or estimate the level of interference [38]. With feature detection sensing, the necessity of being silent is not as strict, but the sensing time noticeably increases which is not desirable for CR network. Also, when a failure occurs and the CR node decides to change the operating frequency, it should stop the communication in the current operating channel and spend some time to sense other channels to find a new one. This period of time is the recovery delay as discussed in Section III-B3 and normally can not be used for data communication.

Although the decision-making time is an overhead, a longer decision period implies higher accuracy and more reliable decisions, which increase the reliability of future communications. Therefore, a considerable part of the research work on CRNs is focused on proposing methods that provide shorter decision-making time with an acceptable level of accuracy and reliability. Particularly, a noticeable part of the decision-making process is spent on spectrum sensing activities: either to monitor the current operating channel to verify the quality, appearance of licensed users and interference level (see Fig. 7.b) or when finding a new channel during the recovery period (see Fig. 7.c). The sensing time is thus a critical but challenging factor: on one side a longer sensing period increases the accuracy of the sensing and therefore decreases the possibility of false-alarms or miss-detections, which results in higher reliability and throughput (less interference). On the other side, a longer sensing time is equivalent to longer decision-making and recovery delay, which decreases the portion of the slot time available for transmission and therefore degrades the throughput. This implies that there is a trade-off between throughput and reliability and an optimum sensing time should be found which meets the required thresholds and requirements [40], [54], [55].

In addition to the sensing time, the recovery delay also depends on the algorithm that the CRN uses to select the operating channel for the whole network or each CR user known as spectrum decision, spectrum selection or spectrum searching scheme. Normally, the shortest decision time is desirable and the objective in this area of research is to propose faster mechanisms [55], [56]. When the quality of the channels are different, channels may have a quality index (channel characteristics) which is updated centrally by a spectrum server or base station, or in a distributed manner by each CR user itself based on the users’ perspective [37]. In this case, the best spectrum decision approach is the one that finds in the shortest time the first channel which satisfies the performance and reliability expectations. Also, using historical information of channel occupancy (see Section V-A), the CR node can reduce the decision space and decrease the time spent for finding the next channel.

In conclusion, the decision-making time is an overhead which reduces the resources available for data communications. However, a good decision-making increases the future network reliability and its ability to deliver higher throughput to the user. Also, it is expected that research advances will be able to provide faster decision-making algorithms with good performance.

B. Complexity and Power Consumption Limitations

Implementing cognitive features such as spectrum sensing, learning and reasoning capabilities considerably increases the hardware complexity, cost and power consumption of the node, which can be a limiting factor in deployment and wide-spread adoption of CRNs. However, recent advances in computer hardware and signal processing as can be seen in PDAs, mobile phones and notebooks demonstrate that the hardware complexity should not be considered as a limiting factor.

Concerning the cost, as explained by Mitola [53], it is not necessary for every CR node to possess all cognitive capabilities. A basic architecture with some primary capabilities
e.g., spectrum awareness and adaptability can be provided with lower price for the public markets while governmental entities and organizations would use more advanced CR nodes. Furthermore, the ever decreasing cost of technology is an encouraging factor.

For applications such as mobile and wireless sensor networks where power efficiency is critical, giving an important role to CR is more challenging since most of the CR capabilities considerably increase the complexity and processing time and consequently the node power consumption. However, recent advances in related technology fields and the capabilities of the CR networks to employ more power-efficient spectrum management and routing algorithms have provided the possibility of such implementations as has been proposed by Mitola in 2001 [57]. Some proposals for CR-based wireless sensor networks [58], [59] and CR mobile cellular networks [60] have also been proposed in the literature recently.

C. Channel Variation

Returning to the scenario illustrated in Fig. 7.c, a high channel variation implies that in almost all time slots, the CR user performs the recovery and switches to a new channel, which considerably decreases the useful time available for data communication. For example, when a secondary CRN is deployed, although it lessens the problem of spectrum scarcity and usage inefficiency by opportunistically using the unused spectrum licensed to primary users, it adds a new source of failure which is the appearance of primary users in the operating channel which obliges the CR to vacate this band and switch to a new one. The rate and duration of this failure type is dependent on the primary users’ activity and traffic type. As the dynamism of the primary users increases, the secondary CRN experiences more failures and spends more time performing link recovery. In [61], it is shown that a threshold value for channel variations can be found where using CR technology yields a lower performance than employing a traditional radio with static spectrum assignment and no capability of channel switching. This threshold depends on the rate of channel variations and also the decision-making time. For more examples of the impact of channel variations, readers can refer to [47], [62]–[66] where the authors discuss about the performance of cognitive users considering the behaviors of primary networks (channel variations) for different models of medium access.

In [52], [67], the authors discuss how the routing in multihop CRNs is affected by the channel variations and is consequently more challenging compared to traditional ad-hoc and mesh wireless networks. They show that when the channel variation increases, the efficiency of the existing ad-hoc routing algorithm decreases and the need to propose more specific routing methods for CRNs increases. These appropriate routing algorithms for a CR network, especially for environments with high spectrum variation, should be opportunistic and spectrum-aware [37], [52], [67]–[69]. A spectrum-aware routing algorithm stands for a cross-layer algorithm that finds the route and operating channel jointly. In Opportunistic Routing (OR), a wireless node transmits over any available spectrum opportunity until at least one proper node in the path to the destination receives the packet. OR which is also known as opportunistic forwarding has received considerable attention from researchers in the scope of multihop wireless networks. For further studies, interested readers are referred to [70]–[72].

The impact of channel variation explains why licensed TV bands are the most favorite channels for implementation of a CRN [13], [14]. These bands are usually underutilized and have regular and predictable usage patterns while a WiMAX or WiFi primary network is more dynamic and less predictable.

VII. CONCLUSION

We have presented a broad view on failure in wireless networks and network robustness and described a wireless network architecture based on cognitive radios to improve the reliability offered to next generation wireless services. This higher reliability is achieved thanks to the cognitive capabilities of cognitive radio networks that empower these networks to prevent failure occurrence, decrease their severity or recover from failures more efficiently. This approach builds on the abundant literature on reliability in wireline networks and adapts it to the particular context of wireless networks. This tutorial article opens the way for new designs and evaluation approaches for wireless networks with an aim of improving their reliability. However, a detailed investigation of some of the different methods described in this paper and an evaluation of their performance with regards to reliability related metrics is still needed.

VIII. ACKNOWLEDGMENTS

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada under Strategic Grant STPG365205. We would also like to thank the Editor-in-Chief and the anonymous reviewers for their instructive and very valuable comments.

REFERENCES


Arash Azarfar received the B.Sc. degree in 2003 and the M.Sc. degree in 2005 both in computer engineering from Sharif University of Technology, Tehran, Iran. He is currently a Ph.D. student at the Department of Electrical Engineering, École Polytechnique de Montréal. His research interests include quality of service, reliability and performance evaluation in cognitive radio and multimedia communication networks.

Jean-François Frigon received the B.Eng. degree from École Polytechnique de Montréal, Montréal, QC, Canada, in 1996, the M.A.Sc. degree from the University of British Columbia, Vancouver, BC, Canada, in 1998, and the Ph.D. degree from the University of California at Los Angeles (UCLA), in 2004. From 2001 to 2003, he worked as the director of wireless communications systems at Innovics Wireless where he oversaw the design of a diversity enabled WCDMA terminal. He joined the Electrical Engineering department at École Polytechnique de Montréal in 2004 where he is currently an Associate Professor. His research interests include wireless networks, MAC protocols and design of multiple antennas digital communication systems for high speed wireless communications.

Brunilde Sansó is a full professor of electrical engineering at École Polytechnique de Montréal and director of the LORLAB. Her interests are in performance, reliability, design, and optimization of wireless and wireline networks. She is a recipient of several awards, Associate Editor of Telecommunication Systems, and editor of two books on planning and performance.