Routing Metrics of Cognitive Radio Networks: A Survey

Moustafa Youssef, Mohamed Ibrahim, Mohamed Abdelatif, Lin Chen, and Athanasios V. Vasilakos

Abstract—The majority of work in cognitive radio networks have focused on single-hop networks with mainly challenges at the physical and MAC layers. Recently, multi-hop secondary networks have gained attention as a promising design to leverage the full potential of cognitive radio networks. One of the main features of routing protocols in multi-hop networks is the routing metric used to select the best route for forwarding packets. In this paper, we survey the state-of-the-art routing metrics for cognitive radio networks. We start by listing the challenges that have to be addressed in designing a good routing metric for cognitive radio networks. We then provide a taxonomy of the different metrics and a survey of the way they have been used in different routing protocols. Then we present a case study to compare different classes of metrics. After that, we discuss how to combine individual routing metrics to obtain a global one. We end the paper with a discussion of the open issues in the design of future metrics for cognitive radio networks.

I. INTRODUCTION

The wide proliferation of the wireless technology and mobile devices becoming more ubiquitous have led to a new era where bandwidth is expected to be available everywhere in abundance. Currently, spectrum assignment is based on an auctioning mechanism by governmental agencies and licensees are granted the rights for the use of the frequency bands on a long term basis over vast geographical regions. However, as the demand for more wireless bandwidth increases, regulatory authorities have started to reassign bands used by old legacy systems, or by technologies that have transitioned to digital communications (e.g. Digital Dividend [1]). However, even with these reassignments, the spectrum shortage persists due to uneven spectrum usage. Recent studies by the spectrum regulatory authorities (e.g. the Federal Communications Commission (FCC)) highlight that many spectrum bands allocated through static assignment policies are used only in bounded geographical areas or over limited periods of time, and that the average utilization of such bands varies between 15-85% [2].

On the other hand, the use of wireless technologies operating in unlicensed bands, especially in the ISM band, has been prolific with a wide range of applications developed in different fields (e.g. WLANs, mesh networks, personal area networks, body area networks, sensor networks, etc.), which caused overcrowding in this band.

This highlights two main problems with wireless networks: exhaustion of the scarce wireless spectrum and under-utilization of the licensed spectrum in some areas. Cognitive Radio Networks (CRNs) emerged as a paradigm to address these problems. In CRNs, wireless nodes change their parameters to communicate efficiently, avoiding interference with licensed (primary users (PUs)) or unlicensed users (secondary users (SUs)). This alteration of parameters is based on monitoring the radio environment, such as the radio frequency spectrum, user behavior, and network state. CRNs are composed of cognitive, spectrum-agile devices capable of changing their configurations on the fly based on the spectral environment. This capability opens up the possibility of designing flexible and dynamic spectrum access strategies with the purpose of opportunistically reusing portions of the spectrum temporarily vacated by licensed PUs. On the other hand, the flexibility in the spectrum access phase comes with an increased complexity in the design of communication protocols at different layers.

Most of the work in CRNs has focused on the lower layers of the protocol stack, mainly at the physical and MAC layers [3], [4] with single-hop forwarding. Their goal is to address the channel scarcity problem and achieve efficient wireless communication. It allows CRNs to discover spectrum holes, and utilize them, which decreases contention on channels, minimizes interference between communicating nodes and improves the average channel efficiency.

Routing in multi-hop CRNs, however, is an important problem that affects the performance of the entire network. Different from traditional routing protocols in adhoc networks, routing in CRNs has to deal with a number of challenges, including adapting to the dynamic changes of spectrum availability due to the stochastic behavior of the primary and secondary users, the heterogeneity of resources such as the availability of different channels and radios on the same node, and synchronization between nodes on different channels. Therefore, deploying adhoc networks’ traditional routing protocols directly in CRNs will result in poor performance in terms of throughput, delay, and probability of packet loss. For example, using hop count without taking into account PU behavior may make the protocol to favor an extremely unstable route. Even route instability definition itself is different in adhoc networks from CRNs. In adhoc networks, instability means one or more nodes in the route becomes unreachable from its neighbors. However, in CRNs all nodes in the route are reachable by their neighbors according to adhoc networks definition, but one or
more of them are not reachable because of PUs behavior. One of the main components of a routing protocol is its routing metric that determines the quality of the different routes. In this paper, we present a survey and taxonomy of the different routing metrics for CRNs and mention other notable aspects of such metrics and the routing protocols that employ them. We highlight the challenges of designing a routing metric for CRNs, both inherited from traditional wireless networks and those unique to CRNs. We then provide a taxonomy for routing metrics in CRNs based on two main categories: single-path routing and multi-path routing. An important direction for research in routing metrics for CRNs is hybrid metrics that combine a number of atomic metrics. Therefore, we present different techniques for combining individual routing metrics to obtain a global one. We conclude the paper with a discussion of open issues in the design of new routing metrics.

The rest of this paper is organized as follows: in Section II we present the challenges of designing routing metrics for CRNs. Section III presents a taxonomy of the different routing metrics of CRNs. In Section IV we provide a case study that compares different classes of metrics. Then we present the different ways of combining atomic metrics into a global one in Section V. Finally, we discuss future and open issues for routing metrics in CRNs in Section VI.

II. ROUTING METRICS DESIGN CHALLENGES

In this section, we present an overview of the different challenges that face CRNs and their effect on designing a routing metric for multi-hop CRNs. We start by the challenges inherited from traditional networks followed by challenges unique to CRNs. Figure 1 summarizes the different challenges.

A. Inherited from Traditional Networks

Traditional routing metrics (designed for both wired and wireless environments) for link state or distance vector paradigms are not well suited to be applied to CRNs. The main reason is that there are frequent dynamic changes in the CRN that may trigger a large number of updates and lead to rapidly changing routing tables. These dynamic changes inherit the wireless mobile network characteristics including: nodes mobility, nodes limited power and the network life time, the wireless medium properties (higher error rates, higher interference, signal fading), channel scarcity, and conflict with other ISM-based devices; and has the added constraints imposed by the primary users. These challenges serve as the basic set of challenges inherited from traditional networks that a routing metric for CRNs should address. As we discuss later, a number of metrics extend the classical routing metrics, such as delay, to fit the characteristics of CRNs.

B. Spectrum Availability

One of the main tasks of the cognitive radio (CR) is to determine whether a spectrum band is available or not. This includes analyzing the spectrum in its vicinity and analyzing the transmission to know the type (PU or SU) of the transmitter. In addition, this analysis has to be done carefully in order not to produce (or produce minimum) interference to the transmitter, especially if it is from the PU.

A good routing metric for CRNs has to assign different weights to different spectrum ranges (channels) based on their availability and the probability that a transmission will be interrupted due to the PU’s activity and/or other SUs. It also can estimate the future activity of the PU to minimize the route interruption time and maintainable cost.

C. Interruption Time

One of the main assumptions of the CRNs model is that the PU is the owner of the spectrum and has higher priority in using it for transmission over SUs. Therefore, as soon as a SU detects the transmission of the PU on the channel that it currently uses, it has to seize transmission immediately and switch the channel.

Therefore, a routing metric for CRNs has to take into account the channel switching time that needs to be paid when a channel switching decision is taken. This channel switching time may involve the time required to communicate the channel switching decision to the next hop neighbor and the time to tune the radio to the new channel. This tuning time is usually a function of the difference between the original and the new channels [5], [6]. This time has to be minimized to avoid stalling the data forwarding session. In addition, interruption can also affect transport layer protocols, such as TCP, that may timeout and trigger congestion avoidance measures.

D. Signaling and Deafness Problem

A notable problem when dealing with multi-channel communication is that a node tuned on one channel band (whether receiving on this channel band or not) cannot sense signals (and thus receive transmission) transmitted on a different channel band. This is referred to as the deafness problem. The traditional solution to this problem is to use a common control channel (CCC) shared between all nodes [7], [8], [9], [10] to disseminate the route initialization and maintenance data. However, this solution makes the CCC a bottleneck for communication. Other solutions involve sending the same data on all channels or the use of channel synchronization schemes [11], [12]. These solutions usually incur higher delays and power consumption (due to frequent channel switching operations). The solution of the deafness problem affects the choice of the routing metric and the performance of the routing protocol in general.

E. Cross Layering

Based on the above challenges, a cross-layering approach for routing in CRNs is a must, not an option. Decisions at the network layer have to reflect the channel status (PU activity, error rate, etc) which is collected at the physical and MAC layers. Therefore, routing metrics for CRNs have to take into account features from different layers.


III. TAXONOMY

In this section, we present a taxonomy of the different routing metrics that have been used in multi-hop CRNs (Figure 2). We categorize them into two main groups: metrics for single-path routing algorithms and metrics for multi-path routing algorithms. For each category, we start with the traditional routing metrics borrowed from wireless adhoc networks and how they have been adapted to reflect the nature of CRNs. We then describe the new metrics that have been designed from the beginning with CRNs in mind. For each metric, we describe its merit and the different ways to use it within the different routing protocols. The next section discusses open issues and future directions.

A. Category 1: Single-path Routing

A number of routing metrics in CRNs are designed for single path routing. On one hand, these metrics depend on classical routing metrics including delay, hop count, and location-aided metrics, and extend them to adapt to the specifics of the CRN paradigm. On the other hand, new metrics that target the characteristics of CRNs have also been developed including spectrum availability and route stability. Usually, a single routing protocol may combine a number of these metrics. Figure 2 shows a taxonomy of the proposed techniques.

1) Delay: This classical routing metric captures all factors that contribute to the end-to-end delay. These include the channel switching time, MAC backoff time, queuing delay, transmission delay, among others. It is usually used alone or combined with other metrics.

In [13], the Effective Transmission Time (ETT) metric is proposed. The metric captures the transmission delays on a link as \( \frac{L}{T(1-p)} \), where \( L \) is the average packet length, \( T \) is the transmission rate, and \( p \) is the packet error rate. This effectively captures the transmission time taking the expected number of retransmissions into account.

In the STOD-RP [14] protocol, the used routing metric combines route stability and the channel switching delay, which is assumed to be constant by a channel switch, i.e., independent of the particular start and end channels.

The work in [5], [6] uses a routing metric that combines two aspects of delay: channel switching time and multi-flow interference. The first component reflects the time required to change the channel, which is taken to be proportional to the difference between the initial and final channels. The proportionality constant is taken to be 10ms/10MHz in the spectrum range 20MHz-3GHz. The second component reflects the backoff delay caused by contention between the different nodes and is equal to: \( \frac{1}{(1-p_c)(1-\frac{W_0}{\sum_{n=1}^{\text{Num}} 1})} \), where Num represents the number of contending nodes, \( p_c \) is the probability that a contending node experiences collision, and \( W_0 \) represents the minimum contention window size.

The DORP protocol [15], [16] extends the previous metric to include queuing delays. Queuing delays for a flow \( n_0 \) is caused by the transmission of packets from other flows. Based on a round robin scheduler between flows and a fair share of wireless capacity between flows, the queuing delay term is added as \( \sum_{n=1,n \neq n_0} P/B \), where Num represents the number of contending nodes, \( P \) is the packet length and \( B \) is the bandwidth. The combined objective function captures different delay effects. Assigning a new active frequency band for the flow results in increased channel switching time. On the other hand, letting the flow use an existing active frequency band increases the number of contending nodes thus increasing the backoff and queuing delays.

In SEARCH [17], the destination, receiving a path on each available channel based on a greedy location-based metric (see Sect. III-A4), runs an algorithm to combine all these paths to select the optimal path that can span multiple channels. The selected path is the one that minimizes the end-to-end delay which includes the cost of the channel switching time between paths on different channels along with the path delay on each channel.

2) Hop count: Hop count as a metric is usually used as a reflection of other metrics, based on a lower-is-better principle, such as an indication of “faster” transmission routes (delay) or routes that consume less networking resources (as they pass through a lower number of nodes). It has been used as the main routing metric or as a filter to select among the candidate paths. For example, SAMER [18] takes a two-tier routing approach that balances between long-term optimality (based on the hop count) and a short-term opportunistic gain (in terms of higher spectrum availability). SAMER builds a forwarding mesh which is centered around the long-term shortest path and opportunistically expands or shrinks periodically to exploit spectrum availability.
CAODV [19] is a modified version of the AODV protocol [20] that avoids active primary users’ regions during both route formation and packet forwarding without requiring a dedicated common control channel. During the route formation, channels that are currently used by PUs are excluded from the route discovery process. During forwarding, when a PU becomes active on a certain channel, all the neighbouring SUs invalidate the routes that use this channel. The hop count is used as a filter to select between the candidate routes that have the same sequence number.

In SEARCH [17], the hop count is used as a filtering metric. This is done by comparing the hop count used in the original route formation to the number of hops used in the current path, which differs from the original route due to route maintenance based on PUs’ activity, periodically. If the difference is above a threshold, it signals the need of a new route formation. This limits the sub optimality of the path, specially in high mobility scenarios.

3) Power Consumption: A major issue when dealing with protocols for mobile devices is being energy-efficient to conserve the limited battery resource. This applies in CRNs as well as in traditional ad-hoc networks. However, in CRNs each SU has the extra overhead of continuously sensing the presence of PUs. Therefore, it is even more important to have power consumption as a metric for CRNs.

A power-aware routing protocol [21] was proposed using five different metrics based on battery power consumption at nodes. These five metrics are the following:

1) Minimize energy consumed/packet: where energy consumed per packet is

\[ e_j = \sum_{i=1}^{k-1} T(n_i, n_{i+1}) \]

Where \( T(a, b) \) is the energy consumed in transmitting (and receiving) one packet over one hop from a to b.

2) Maximize time to network partition: using the max-flow-min-cut theorem, a minimal set of nodes (the cut-set) the removal of which will cause the network to partition can be found. This routing procedure balance the load over this set of nodes to maximize the life of the network.

3) Minimize Variance in node power levels: send packets to the neighbors with the least amount of data waiting to be transmitted. This metric is based on that all nodes in the network are equally important and try to remain up all the nodes in the network.

4) Minimize cost/packet: where the node cost is the total energy consumed by this node so far.

5) Minimize maximum node cost.

In the Minimum Weight Routing Protocol (MWRP) [7], the link weights are defined as the transmission power required to reach the receiver over a certain interface based on a free-space propagation model as \( P_{Rx} = P_{Tx} \left( \frac{1}{4\pi d^2} \right) \). Here, \( P_{Tx} \) is the transmitted power, \( P_{Rx} \) is the received power, \( \lambda \) is the wave length, and \( d \) is the distance between the transmitter and receiver. Therefore, the transmission power is proportional to the square of distance. Each neighbor selects the next hop neighbor (and interface) that minimizes this metric locally.

4) Location-based: Many of today’s wireless devices are location-enabled, e.g. through the GPS system or network-based localization, and this is expected to become more ubiquitous in the future. In addition, location information of CRN nodes can be obtained via FCC Geolocation-Databases [22] or estimated via measurements [23], [24], [25], [26], [27], [28], [29], [30] accurately. This motivates the work on location-based metrics for CRNs that do not require global knowledge. For example, SEARCH [17] uses a greedy location-based approach similar to the GPSR [31] protocol in classical adhoc networks. In its greedy forwarding phase that works on each available channel, SEARCH selects the next hop neighbor as
the neighbor closest to the destination from the current node within a focus region. A focus region is a sector around the line joining the current node and the destination used to limit the deviation from the straight line path (Figure 3). As shown in Figure 3 and according to the definition of focus region, the closest node to the destination is not because is out of the focus region.

MP-JSRCA [32] uses a greedy location-based approach similar to SEARCH [17] that selects candidates from the next hop from within a controlled sector region towards the destination. However, at each hop the packets are forwarded to the neighbor with the lowest data transmission cost (DTC) within the controlled region, where DTC is a weighted sum of the mobility cost and channel interference cost to PUs and other SUs.

IPSAG [33] is an IP-based location-based spectrum aware routing protocol. It uses the IP notion of piggybacking control information as headers in the data packets, instead of sending them separately, to avoid the control overhead. The selected next hop is the closest neighbor to the destination that has at least one common available channels and satisfies a certain SNR threshold.

The work in [34] proposes a routing protocol for vehicular cognitive adhoc networks that exploits channel and geo-location information. Each node forms a forwarding set consisting of the geographically close neighbors to the destination that minimizes the Expected Transmission Time (ETT). This is based on the PU activity and the distance between the sender and receiver. Similar work was proposed in [35] that takes the links reliability into account. Each node selects as the next hop the nearest two-hop neighbor to the destination from the nodes that have an expected transmission count (ETX, which reflects the quality of the link) greater than a certain threshold. Then a set of relay nodes are used to deliver the packets to the intermediate destination node where a priority-based opportunistic routing is used; The relay nodes are virtually ordered based on their ETX values and the highest priority node will start sending the packets earlier. Leveraging the broadcast nature of the wireless channel, the remaining nodes can transmit in order, sending the packets which were not acknowledged by the higher priority relay nodes. This concept of anypath routing increases the reliability of the routes by providing backup nodes in the forwarding relay set to transmit the packets in case of failure of the higher priority nodes.

In [36], a greedy routing protocol that uses two modes of operation is proposed. When the PU is not active, the closet neighbor to the destination is selected as the next hop. When a PU is detected, the second mode is activated, where the nearest neighbor to the current hop is selected as the next hop and the transmission power is adjusted to avoid interference with the PU transmission.

Work in [37] proposed a location-based routing protocol where the path is determined at the receiver. The sender broadcasts a request-to-forward (RTF) message using the maximum transmission power through the control channel including the needed transmission rate, the set of available channels, and the location of the destination and itself. Each neighbor then delays the reply a period of time based on its proximity to the destination and ability to satisfy the demanded rate. Therefore, there is no need in this beaconless approach for the sender to continually monitor the locations of his neighbors since the neighbor with the best metric will reply first.

LAUNCH [38] presents a location-based PU-aware routing protocol for CRNs that combines different metrics including PU activity, switching delay, and location information. The next hop neighbour is selected in a greedy manner based on the combined metric. A channel-locking mechanism is used if multiple interfaces are available, where each node is locked in a particular channel for transmitting and another for reception to minimize the channel switching delay.

5) Spectrum availability: The spectrum availability metric between two nodes refers to the bandwidth available for the two nodes for communication taking both the PUs’ activity and other SUs’ activity into account. In SAMER [18], the spectrum availability is used to select the forwarding path. The spectrum availability between two nodes is defined as: \[ S_{ab} = T \times B \times (1 - P_{loss}) \]

Where \( T \) is the fraction of time during which the node is free to transmit and/or receive packets during a spectrum opportunity block which can be calculated using MAC layer information, \( B \) is the available bandwidth, and \( P_{loss} \) is the loss probability of the spectrum block which can be calculated by measuring the loss rate of broadcast packets between pairs of neighboring nodes. Loss rate depends both on each frequency band’s properties and the interference it perceives from both primary and secondary users.

6) Route Stability: A routing solution that produces stable routes in CRNs is highly desirable as it is one of the main challenges of CRNs due to the PUs’ activities. Unstable routes will lead to frequently firing new re-routing events which consumes the network resources and degrades its performance. Route stability can be captured explicitly in the routing metric or implicitly as in [17].

The STOD-RP [14] protocol uses a routing metric that combines route stability and end-to-end delay. For the stability part of the routing metric, the metric reflects the stability in terms of link’s available time: \[ \frac{[O_{ca} + O_{pt} + P_{et}]}{T - O_{ca}} \]

Where \( O_{ca} \) is the channel access overhead in \( \mu s \), \( O_{pt} \) is the protocol overhead in \( \mu s \), and \( P_{et} \) is the size of a packet, all of which are taken as constants for a specific access technology. \( r \) and \( e_{pt} \) are the link rate in Mbps and the packet error rate, respectively. \( T \) is the timeduration during which a spectrum band is available to the link which can be predicted from the statistical history of PUs’ activities. The division by \( T \) is equivalent to the integration of the link stability. The path stability cost is the sum of the individual costs of the links constituting the path.

Moreover, in the Coolest Path protocol [39], other operators, i.e. other than the sum operator, are used for calculating the total path cost. These include the maximum stability cost over the path links or a mixed cost between the maximum and the accumulated cost along the path. In this work, the link stability cost is the probability that the common channel over the link is not available. The work showed that, in case of frequent PU activities, the accumulated cost achieves better performance in terms of path switching ratio and path longevity. When the PU
arrival rate is low or the PU channel occupancy time is long, the maximum cost performs better. The mixed metric achieves the best performance when he PU activity is neither low nor high.

In SEARCH [17], stability of the routes is implicit as the process of constructing the routes avoids routing through nodes in the vicinity of active PUs and circumvent such areas by information gathering in periodic beacons.

The work in [40] defines the link cost metric as a function of both the link holding time \( h \) and communication capacity \( b \). The first metric reflects the usage pattern of the primary users, while the second metric is a measure of communication conditions which include the available bandwidth on that link. They assume that the two parameters are independent. Therefore, the link cost metric is defined as: \( \frac{1}{\lambda h} \), where \( \lambda \) is a tunable parameter which determines the relative weight of the two parameters. The link cost is maintained and updated by the routing protocol periodically.

The work in [41] presents a new route stability metric that reflects the cost that will be paid if a route needs to be changed. The route maintenance cost reflects the node and link switching operations required as a PU becomes active. The link cost is defined as \( \frac{C_{\text{sw}}}{h^2} \), where \( C_{\text{sw}} \) is the cost of switching from the current link to another link in terms of number of hops, \( C_{\text{rep}} \) is the expected cost to repair the link in the future in terms of link failure probability and expected switching time, and \( \alpha \) allows gauging of different cost contributions. \( E[\text{TTS}] \) represents the average time the link remains stable before switching. Exact expressions for the different components are given under the assumption of a random ergodic ON/OFF PU’s activity and knowing the first order statistics of the PUs’ activity.

7) Probabilistic: When the exact status of the spectrum is not available, or is difficult to reconstruct in a distributed way, routing decisions should be based on probabilistic metrics.

In [10], the authors propose a probabilistic metric that captures the probability of PU interference at a given SU over a given channel. The metric is used to determine the most probable path to satisfy a given bandwidth demand \( D \) (rate demand \( D \) in bits/second) in a scenario with \( N \) nodes that operate on a maximum of \( M \) orthogonal frequency bands. The probability that a channel with capacity \( C \) can support the demand \( D \) is calculated as: \( Pr[C \geq D] = Pr[P \leq \frac{D}{1-h} - N_0] \), where \( P \) is the received power from the PU, \( W \) is the channel bandwidth in Hz, \( N_0 \) is the power of the white Gaussian noise. The value of \( P \) is taken such that a certain outage probability, at the MAC layer, can be guaranteed for PUs. The paper also assumes a log-normal distribution for \( P \), which is the power of the received signal.

Once this probability is calculated, the routing metric is taken as: \(-\log Pr[C \geq D + U]\), where \( U \) is a system memory that accounts for the interference in the vicinity of the nodes constituting the link and \( C \) is the maximum channel capacity given by Shannon’s Theorem.

B. Category 2: Multi-path Routing

With the challenge of meeting the stochastic activity of the PUs in CRNs, multi-path routing, along with its associated metrics, appeared as a way to provide redundant paths for transmission between the source and destination to reduce the effect of disruption by the PU. Traditionally, multi-path routing refers to topology-wise disjoint paths, i.e. paths that share no common node except the source and the destination. This concept is extended for CRNs to cover spectrum-wise disjoint paths, which are the paths that may share a common node but different bands/channels are assigned for the links around the common node. Another important aspect of multi-path routing is the criteria used to select the secondary routes. We note that many of the multi-path metrics are also shared with the single-path routing. However, we repeat them here for better presentation with the routing protocols they were introduced in.

1) Delay: As in single-path routing, delay as a routing metric captures different aspects including channel switching time, end-to-end delay, transmission delay, among others. It can be used as the main metric or as a filtering metric.

In [42], the Opportunistic Link Transmission (OLT) metric is proposed. This metric captures three aspects of delay: \( \text{OLT} = d_{tx} + d_q + d_{\text{access}} \), where \( d_{tx} \) denotes the link transmission delay, \( d_q \) denotes the packet queueing delay in a node, and \( d_{\text{access}} \) denotes the link access delay.

In Urban-X [43], the path cost is taken to reflect a function of the packet transmission delay of a flow via the node to a destination.

The work in [44] provides a statistical model for the end-to-end delay in single path routing and then extends it to duplication-based and coding-aided multi-path routing schemes. The paper considers the impact of interference and dynamic spectrum access to derive the end-to-end delay, including medium access and retransmission delay.

In [45], [46], [47], delay is used implicitly to select the main route, which is the route whose RREQ packet arrives first. It is also used as a tie breaker for secondary routes.

2) Hop count: This classical routing metric is used to filter out routes or as a tie breaker.

For example, the Multi-path Routing and Spectrum Access (MRSA) framework [48] uses the standard DSR route discovery mechanism. A RREQ packet received with an old ID will be forwarded only if its hop count is smaller than the previously received one. Final route selection is performed at the destination based on the available route bandwidth capacity metric (see Sect. III-B4). Other paths can be selected in the same way iteratively. Ties are broken preferring paths that share the minimum nodes with previously selected routes and then based on the minimum number of hops. SPEAR [49] also uses the hop count metric as a tie breaker.

In Urban-X [43], in order to limit the forwarding structure and to avoid routing loops, the routes are limited to have a maximum number of hops.

The hop count metric has also been used in [45], [46], [47] as a filter for selecting secondary routes.

3) Power Consumption: In the NDM_AODV protocol [47], secondary routes are selected based on the remaining energy at each node along the route. The protocol calculates the total remaining energy of all nodes in the path and then selects the path with the maximum total remaining energy.
4) Route bandwidth capacity: This metric takes into account the bandwidth of each link based on the number of nodes sharing it. For example, MRSA [48] uses the standard DSR route discovery mechanism to discover multiple candidate paths. Final route selection is performed at the destination based on the available route bandwidth capacity metric. Each candidate path is evaluated in terms of the minimum bandwidth capacity of the radios of all the nodes in the path, assuming all flows get a fair share of the bandwidth. For example, two flows sharing a radio each get half of its bandwidth. The path bandwidth capacity is the minimum of all the nodes constituting it. The path with the maximum bandwidth capacity is selected. A similar approach is also used in the SPEctrum-Aware Routing Protocol (SPEAR) [49] with hop count as a tie breaker.

The minimum number of common nodes is usually used as a tie breaker between secondary routes. For example, in [48], once the main route is selected based on the route bandwidth capacity metric, candidate routes are selected based on the same metric and ties are broken by preferring routes that share a lower number of nodes with previously selected routes.

5) Route closeness: The route closeness metric [9] selects routes based on how far away they are from each other. The intuition is that selecting non-close routes makes them less vulnerable to mobile PUs. In other words, if selected routes are far enough, a single active mobile PU would not be able to interrupt all of them at the same time.

Closeness of two routes is defined as the sum of the pairwise closeness of their links. The closeness of two links is the area of intersection between the PU’s effective region of the two links. The PUs effective region of a link is defined as the region around the link where a PU is able to interrupt that link, which is a function of the PU’s transmission range. Figure 4 shows an example of the link and route closeness metrics [9]. Figure 4.a shows an example of PUs Effective Region (PuER) of a link L between two SUs: S1 and S2. This region is used to quantify the closeness of two links as shown in Figure 4.b. And finally, the more the pairwise closeness of two routes’ links, the more the closeness of the two routes themselves (Figure 4.c).

6) Dead Zone Penetration: Current routing protocols reconfigure established routing paths to avoid active zones of PUs. The Dead Zone Penetration metric (DZP) [50] avoids reconfiguring the routing paths and tolerates PU activity using cooperative beamforming by forming cooperative links between neighbouring nodes that null out the transmission at the primary receiver. In other words, this work uses multipath routing in a new way to penetrate active PUs zones by making the relay selects one of its neighbours to cooperatively send the data to the next hop. Figure 5 shows a motivating scenario for DZP, in which Node 1 maintains the constructed route (1-2-4), even in the presence of a PU, by allowing nodes 2 and 3 to cooperatively send data packets to Node 4. This is better than using the alternate route that goes through Node 5.

7) Metrics that capture SU interference: It is very important in CRNs to pay attention to PU’s activity and handle its interference with ongoing SUs’ traffic sessions. Most of the routing metrics tend to select routes that are expected to be the least affected with future PU activities. Other metrics take into consideration the interference among SUs themselves where the decisions of channel assignments and next hop neighbors are based on SUs’ interference.

For example, in MRSA [48], channel assignment is based on avoiding using the same band that is already selected within two hops in order to reduce intra-path contention and interference.

The work in [51], [52], [53] define the metric Bandwidth Footprint Product (BFP) metric. This metric captures network resources in terms of both frequency usage (bandwidth) and
spatial occupancy (footprint). The footprint is an indication of the interference area of a node for a given transmission power. Since each node in the network uses a number of bands for transmission and each band has a certain footprint corresponding to its transmission power, decreasing the BFP for every possible node is a way to decrease the network’s BFP. Therefore, the work in [51], [52], [53] formulate an objective function whose goal is to minimize this sum. The objective function allows for flow splitting, i.e. multi-path routing, to achieve better performance.

A simplified version, assuming all footprints are equal, in which case the BFP reduces to bandwidth, is used in [54]. Similarly, another simplification, assuming each band has the same bandwidth, in which case the BFP reduces to the footprint, is given in [51].

8) Route stability: The route stability in CRNs is highly influenced by the behavior of the PUs which affects the connectivity of the network. In Gymkhana [55], the authors present a routing metric that captures the degree of connectivity of possible paths towards the destination by avoiding network zones that do not guarantee stable and high connectivity. This is achieved by modeling the problem as a graph and obtaining its Laplacian spectrum that can be used to compute the connectivity of the different network paths, which depends on PU behavior and mobility patterns [56].

The work in [57] selects the most stable route in terms of channel stability time and channel switching delay as the main route. Channel stability time is the time the channel is available to SUs. The route that has the highest divergence from the main route, in terms of hop count and route stability, is selected as a backup route.

A non-cooperative game theoretic approach for stable-aware traffic assignment among $n$ disjoint paths was proposed in [58], [59]. The game models each path as a player and the destination pays for every path/player according to the amount of received information through this path.

C. Discussion

Tables I and II summarize the different metrics used for single-path and multi-path routing in CRNs respectively. The metrics have evolved from traditional ones used in adhoc networks, such as delay and number of hops, to more CRNs specific such as spectrum availability and route closeness. Different routing techniques combine more than one metric to achieve different goals or to break the ties when a number of routes are equal under the primary metric.

1) Single Path Routing: Table III compares the different CR protocols used for single-path routing. In this table, we provided four CR routing protocols characteristics in order to be able to compare these protocols qualitatively:

- Centralized/distributed: Centralized routing algorithms are performed at centralized nodes that collect the entire network topology information and the activity information of PUs and estimate the optimal route.
- Route maintenance support: Where the routing protocols are able to reconfigure the routing paths when a PU becomes active.
- Mobility support: Supporting SU mobility.
- Common control channel: Where the routing protocols require a pre-established common control channel that is fixed and known to all the SUs in the network as compared to forwarding the control packets on all the available channels, as in SEARCH [17].

A distributed CR routing protocol that supports route maintenance, mobility, and does not require common control channel is considered one of the most suitable routing protocols to CRNs. However, designing such protocol while ensuring minimum end-to-end delay is not an easy task. Therefore, most of the protocols in Table III lack some of these characteristics while trying to support the rest of them. However, in infrastructure based networks like cellular networks, centralized approaches are more suitable compared to the distributed approaches. Moreover, in static networks there is no need for the overhead of mobility support. Therefore, each of these protocols are valid for certain networks configurations.

Table IV shows the main characteristics of location-based routing protocols. The routing decision in location-based routing protocols can be taken locally, i.e. based on next or 2-hop neighbourhood information, or globally, i.e. based on the whole path information from the source to the destination. Local approaches sacrifice the route optimality for the ability of fast adaptation to the network dynamics and the low overhead for route establishment. These routing protocols use the location information to limit the forwarding decision to a subset of nodes that are closer to the destination. This subset can be those nodes that fall within an angular focus region, where an angular focus region is a sector with a certain angle around the line joining the current node and the final destination, or the neighbouring nodes that are closer to the destination than the current node.

The next hop is chosen from this subset according to some routing metric: a node can forward the data to the nearest neighbour to the destination, thus decreasing the total number of hops to the destination or the farthest neighbour to destination (nearest neighbour to the current node) to lower the transmission power and avoid interfering with any active PU. Alternatively, a node can select the next hop based on a certain metric.

Finally, to obtain the location of the neighbours, the protocols may use an explicit beaconing-based discovery approach or a beaconless approach as in [37].

2) Multi-path Routing:

a) Multi-routes Selection Techniques: The selection of different routes in multi-path routing (Table V) can be performed incrementally, that is one route at a time, or all routes can be selected concurrently at the initial route setup. Another approach (hop by hop) is to select the next hop on a per packet basis, as compared to selecting the routes at connection establishment, at each node. This means that each packet will traverse just one path, which may change dynamically at every hop. Optimization based techniques (optimization-centralized) perform route selection in a centralized node. This allows for more sophisticated optimization techniques.
Finally, the cooperative relaying/network coding techniques implicitly use multiple routes based on leveraging the overhearing of packets. Some metrics may be more suitable for a specific route selection technique than other metrics. For example, channel state metrics are more suitable for cooperative relaying techniques. All techniques of routing selection are distributed except the optimization based techniques.

b) Traffic Distribution Techniques: Once the routes are selected, there are different techniques for distributing the packets among them. One category of these techniques uses only the primary route and uses secondary routes as backup. Another technique distributes the packets in a round robin fashion over different routes. Another category uses only one route that is changed dynamically based on the current conditions. Optimization based techniques distribute packets over different routes implicitly as a result of the optimization process based on their quality using flow splitting. Cooperative relaying/network coding techniques is another technique that can be used where the broadcast nature of the wireless medium is used to increase the reliability by leveraging the overhearing of packets to retransmit through other routes. Such traffic splitting techniques affect the performance of the entire network. For example, using the round robin technique may lead to out of order delivery of packets, which may affect the performance of some transport layer protocols, such as TCP. Finally, cooperative beamforming can be used to penetrate
dead zones, i.e. areas where PUs are active, by cooperatively sending data to next hop and nulling out the transmission at the primary receiver.

IV. CASE STUDY

In this section, we compare three of the listed classes of metrics in the previous section under basic scenarios via NS2 simulations. The main purpose of this section is to present how we choose a metric and under which scenario each metric

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Centralized/ distributed</th>
<th>Route maint. support</th>
<th>Mobility support</th>
<th>Common channel control</th>
</tr>
</thead>
<tbody>
<tr>
<td>[5], [6]</td>
<td>Distributed</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>[13]</td>
<td>Distributed</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>SEARCH [17], CAODV [19]</td>
<td>Distributed</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>STOD-RP [14]</td>
<td>Distributed</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DORP [15], [16]</td>
<td>Distributed</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>SAMER [18]</td>
<td>Distributed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MWRP [7], [21]</td>
<td>Distributed</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Coolest Path [39], [33], [34], [36]</td>
<td>Distributed</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>[32], [37], [35], LAUNCH [38]</td>
<td>Distributed</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>[40] Two versions</td>
<td>Distributed</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[41] Two versions</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[10] Distributed</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**TABLE III**  
CHARACTERISTICS OF CR SINGLE PATH ROUTING PROTOCOLS.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Local/ global</th>
<th>Focus region type</th>
<th>Forwarding metric</th>
<th>Neigh. discovery</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEARCH [17]</td>
<td>Global</td>
<td>Angular focus region</td>
<td>Nearest neigh. to dst</td>
<td>Beacon-based</td>
</tr>
<tr>
<td>[32]</td>
<td>Local</td>
<td>Angular focus region</td>
<td>Spectrum availability</td>
<td>Beacon-based</td>
</tr>
<tr>
<td>[33]</td>
<td>Local</td>
<td>Set of neighs. closer to dst</td>
<td>Spectrum availability</td>
<td>Beacon-based</td>
</tr>
<tr>
<td>CoRoute [34],[35]</td>
<td>Local</td>
<td>Set of neighs. closer to dst</td>
<td>Delay, stability</td>
<td>Beacon-based</td>
</tr>
<tr>
<td>LAUNCH [38]</td>
<td>Local</td>
<td>Set of neighs. closer to dst</td>
<td>Delay, stability</td>
<td>Beacon-based</td>
</tr>
<tr>
<td>[36]</td>
<td>Local</td>
<td>Set of neighs. closer to dst</td>
<td>Nearest or farth. neigh. to dst based on mode</td>
<td>Beacon-based</td>
</tr>
<tr>
<td>[37]</td>
<td>Local</td>
<td>Set of neighs. closer to dst</td>
<td>Spectrum availability</td>
<td>Beaconless</td>
</tr>
</tbody>
</table>

**TABLE IV**  
CHARACTERISTICS OF CR LOCATION-BASED ROUTING PROTOCOLS.

<table>
<thead>
<tr>
<th>Technique</th>
<th>Incremental</th>
<th>Initially</th>
<th>Optimization - centralized</th>
<th>Hop by hop</th>
<th>Coop. relaying/net. coding</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMR [45]</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NDMR [46]</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NDM_AODV [47]</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MR-ZHSR [60]</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLT [42], [44]</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urban-X [43]</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MRSA [48]</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPEAR [49]</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[9], [57]</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[50]</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[51], [52], [53], [54], [55], [56]</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gymkhana [55]</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**TABLE V**  
HOW THE MULTI ROUTES ARE SELECTED IN MULTI PATH ROUTING PROTOCOLS FOR CRNS. SOME METRICS MAY BE MORE SUITABLE FOR A SPECIFIC ROUTE SELECTION TECHNIQUE THAN OTHER METRICS.
would be suitable.

Our case study is based on the AODV routing protocol [20] with modification to cope with CRNs. In particular, each SU stops sending data when a PU becomes active in its vicinity until the PU stops sending.

We used three classes of routing metrics:

1) Stability-based: which reflects the activity of the PUs on the route. In particular, the probability that a PU will not affect a specific path ($P_{\text{idle}}$) can be calculated as:

$$P_{\text{idle}} = \prod_i (1 - P(i))$$

Where $P(i)$ is the probability that the PU will be active on link $i$ of the path. Therefore, the link metric can be taken as $-\log(1 - P(i))$, e.g. as in [38], [39].

2) Queueing delay: which reflects the delay on the route. In particular, we estimate the queueing delay as $\text{Num} P/B$, where $\text{Num}$ represents the number of contending nodes, $P$ is the packet length and $B$ is the bandwidth [15], [16].

3) Energy-based: which refers to using the remaining energy of the nodes in the selected routes into account in the routing decision, which directly impacts the life time of network. In particular, the link metric is taken as the used battery capacity of the source node of the link, e.g. as in [47].

Our main performance measure is the packet loss rate, which also reflects other metrics, such as throughput.

Figure 6 shows the topology of the scenario and Table VII summarizes its parameters. There is only one PU in the network and only nodes $B$ and $D$ lie in its range. In addition, there is one main connection between Node $A$ and Node $F$ so the queuing delay is not an issue. In addition, there are ample energy in the nodes battery, so energy is not the bottleneck.

Under the given topology and parameters, the stability metric chooses the route $A-C-E-F$ as its $P_{\text{idle}} = 1$, which is lower than all other routes. On the other hand, the queueing delay metric can choose any of the available routes, including the route $A-B-D-F$. The energy metric, however, alternates between the two top and bottom routes as the remaining battery of the nodes on the routes evolves over time.

Figure 7 shows that as we increase the PU’s activity, the loss ratio of the queueing delay and energy metrics increases dramatically. On the other hand, the performance of the stability-based metric, by avoiding the paths affected by the PU, is not affected. The figure also shows that the energy-based metric is better than the delay-based metric for high PU’s activity due to its alternating behaviour over the different routes. However, when the PU becomes less active, the delay-based metric becomes better than the energy-based metric as the overhead of switching between routes become more than its benefit. To summarize, the stability metric is inevitably needed in CRNs, especially in networks with high PUs activity.

### B. Energy Scenario

This scenario highlights the case when the energy-based metric provides better performance. We reduce the activity of the PU so that it is not the bottleneck and maintain only one connection so that the queuing delay does not affect the route selection. Figure 8 shows the effect of varying the initial energy of the nodes on the loss ratio using the three metrics. As the energy increases, the life time of the network increases and so the loss ratio of all protocols decreases. The energy-based metrics, taking the path remaining energy into account can balance the energy consumption of the entire network and hence can further extend the network lifetime which reflects on

<table>
<thead>
<tr>
<th>Technique</th>
<th>Primary - backup</th>
<th>Round Robin</th>
<th>Per packet - dynamic</th>
<th>Flow splitting</th>
<th>Coop. relaying/net. coding/</th>
<th>Coop. beam-forming</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMR [45]</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NDMR [46]</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NDM_AODV [47]</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MR-ZHSR [60]</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OLT [42], [44]</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urban-X [43]</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MRSA [48]</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPEAR [49]</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[9]</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[51], [52], [53], [54], [58], [59]</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gymkhana [55]</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[57]</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[50]</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**TABLE VI**

Traffic distribution techniques used by multi-path routing protocols for CRNs. These techniques affect the performance of the entire network.
TABLE VII
EXPERIMENT PARAMETERS.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value range</th>
</tr>
</thead>
<tbody>
<tr>
<td>PU interference range (m)</td>
<td>125</td>
</tr>
<tr>
<td>SU transmission range (m)</td>
<td>125</td>
</tr>
<tr>
<td>Data rate (Kbps)</td>
<td>100</td>
</tr>
<tr>
<td>PU idle time (ms)</td>
<td>10, 20, 40, 80, 160, 320</td>
</tr>
<tr>
<td>Initial energy (J)</td>
<td>2, 4, 6, 8, 10</td>
</tr>
<tr>
<td>Number of active connections</td>
<td>1, 2, 3, 4, 5</td>
</tr>
<tr>
<td>Number of channels</td>
<td>1</td>
</tr>
<tr>
<td>Frequency (GHz)</td>
<td>2.4</td>
</tr>
<tr>
<td>Effective bandwidth (Mbps)</td>
<td>2</td>
</tr>
<tr>
<td>Packet size (KB)</td>
<td>1.5</td>
</tr>
<tr>
<td>Running time (s)</td>
<td>200</td>
</tr>
<tr>
<td>Area size</td>
<td>1000m × 1000m</td>
</tr>
</tbody>
</table>

In summary, in case of limited energy networks, which is the case in many mobile networks, the routing protocol needs to fairly share the data across different routes.

C. Delay Scenario

This scenario highlights the case when the delay-based metric provides better performance. We set the PUs’ activity to a very low level and the initial node battery capacity to a high level so that both are not the bottleneck on performance. We maintain the main connection between A and F and introduce other secondary connections between nodes C and F.

Figure 9 shows the effect of increasing the number of secondary connections on the loss ratio of the main connection. As the network becomes more congested, the loss ratio increases for the three metrics. Note that even though the queuing-delay based metric selects the non-congested route A-B-D-F, it intersects with the secondary connections at the destination and hence is still affected by the congestion, though in a much less way than the other two metrics. The alternating behavior of the energy-based metric makes its performance better than the stability-based metric at a low number of connections. However, this is switched as the number of connections increases.

D. Discussion

In this section, we showed that in a simple topology and operation conditions that different metrics can be better under different scenarios. Due to the nature of CRNs, the stability-based metrics should be integrated in metrics that cover realistic scenarios, especially those with high PUs’ activity. Energy-based metrics are needed for energy-constrained nodes, which is typical for many mobile nodes. Delay-based metrics become more important as the load on the secondary network increases.
A realistic dynamic CRN will combine all these scenarios on both the temporal and spatial domains. Therefore, combining such metrics in one global metric can produce better performance than using each of them separately. In the next section, we explain different techniques of achieving this.

V. FROM ATOMIC ROUTING METRICS TO A GLOBAL ONE

The previous sections provide an overview on various metrics used in multi-hop CRNs to calculate the appropriate routes, both single-path and multi-path. However, routes in multi-hop CRNs are seldom characterized by one single performance metric. Moreover, different routing metrics may lead to totally different routing solutions, e.g., minimizing hop number may require higher transmission power and often lead to poor transmission quality in terms of the bit error rate. Hence, it is often necessary to combine different atomic routing metrics studied previously to form a global metric to achieve a performance tradeoff among them.

This section is focused on the calculation of such high-level routing metric by presenting the relevant tools of multi-objective optimization (or multi-objective programming, also known as multi-criteria or multi-attribute optimization) and its application in the context of our focus. For a more detailed mathematical survey on the multi-objective optimization problem, readers are directed to [62].

As indicated by its name, multi-objective optimization is the process of simultaneously optimizing two or more conflicting objectives subject to certain constraints. Mathematically, the general multi-objective optimization problem is posed as follows (the constraints are omitted here for brevity):

\[
\min_{\mathbf{x}} \mathbf{F}(\mathbf{x}) = [F_1(\mathbf{x}), F_2(\mathbf{x}), \ldots, F_n(\mathbf{x})]^T, \tag{1}
\]

where \( n \) is the number of objective functions, \( \mathbf{F}(\mathbf{x}) \) is a vector of objective functions \( F_i(\mathbf{x}) \).

In contrast to its peer, the single-objective optimization, a solution to a multi-objective problem is more of a concept than a definition. Typically, there is no single global solution, and it is often necessary to determine a set of points that all fit a predetermined definition for an optimum. This point is particularly important in the context of routing in multi-hop CRNs as different objectives (characterized by respective metrics) often lead to different solutions (routes). A natural solution is to form a global metric that combines the different individual metrics. Mathematically, in terms of multi-objective optimization, we use an individual utility function \( M_i \) to quantify each atomic metric \( i \) (assume there are \( n \) atomic metrics, we have \( 1 \leq i \leq n \)). We define a global utility function \( M \) as an amalgamation of the atomic metrics in order to mathematically model the route preferences for a given node. In the language of multi-objective optimization, \( M_i \) can be expressed as a function of the route \( r \) and corresponds to the atomic objective function \( F_i \) in (1). \( M(r) \) corresponds to the global objective function \( F \) in (1). In the following part of this section, we provide a review of major methods of constructing the global metric and solving the resulting routing problem.

A. Weighted Global Routing Metric

One commonly used method to form the global utility metric \( M \) is to combine the atomic metrics in the form of weighted exponential sum as follows [63]:

\[
M(r) = \sum_{i=1}^{n} w_i [M_i(r)]^p,
\]

where \( r \) is the chosen route(s).

In the above formulas, \( \mathbf{w} \triangleq \{w_i\} \) satisfying \( w_i > 0, \forall 1 \leq i \leq n \) and \( \sum_{i=1}^{n} w_i = 1 \) is a vector of weights parameterized by the individual node. Different nodes may have different value of \( \mathbf{w} \) depending on its own evaluation.

A specifically form of the weighted global metric is the weighted sum metric by setting \( p = 1 \) in the above formulas, shown as follows:

\[
M(r) = \sum_{i=1}^{n} w_i M_i(r).
\]

A desired property of the weighted sum metric is that by optimizing this metric, we can reach a Pareto-optimal point\(^2\). However, the weighted sum method, like any method of selecting a single solution as preferable to all others, is essentially subjective, in that a decision maker needs to quantify the weights \( \mathbf{w} \).

To allow functions with different orders of magnitude to have similar significance and to avoid having to transform objective functions, we can construct the global metric in the following multiplicative way:

\[
M(r) = \prod_{i=1}^{n} [M_i(r)]^{w_i},
\]

where \( w_i \) are weights indicating the relative significance of the \( i \)th atomic metric. This metric is essentially similar to the weighted sum metric noticing that the above formula can be transformed as

\[
\log M(r) = \sum_{i=1}^{n} w_i \log M_i(r).
\]

With the global routing metric \( M \), the routing problem is to find the route(s) that optimizes \( M \).

B. Lexicographic Global Routing Metric

Another commonly used method is the lexicographic metric [64], in which the atomic routing metrics are arranged in order of importance such that \( M_1 \) is the most important metric. The following optimization problems are then solved one at a time from \( i = 1 \) to \( i = n \):

\[
\min_r \quad M_i(r) \\
\text{S. t.} \quad M_j(r) \leq M_j(r^*), \quad 1 \leq j \leq i - 1.
\]

\(^2\)In our context, a Pareto-optimal route (or route set in case of multi-path routing) is defined as a route that we cannot find any other route that outperforms it in terms of every routing metric.
In the above formulas, \( i \) represents a metric’s position in the preferred order, \( M_j(r^*) \) represents the optimum of the \( j \)th metric, found in the \( j \)th iteration. As a more tolerated version of the lexicographic global metric, we can add a tolerance parameter \( \epsilon \) in the constraint such that it becomes \( M_j(r) \leq M_j(r^*) + \epsilon \) where \( \epsilon \) is configured by the node itself.

The lexicographic method is adapted in the scenario where the node have an ordered preference (not necessarily quantitative) among the different atomic metrics.

**C. Weighted Min-max Criteria**

The weighted min-max global [64] metric is defined as follows:

\[
M(r) = \max_{1 \leq i \leq n} \{ w_i [M_i(r) - C_i] \},
\]

where \( C_i, i = 1, \ldots, n \) are constants. The intrinsic motivation of the weighted min-max metric is to achieve certain balance and fairness among the atomic metrics. Mathematically, we can transform the above optimization problem into the following one by introducing an auxiliary variable \( \lambda \):

\[
\begin{align*}
\min_{r, \lambda} & \quad \lambda \\
\text{S. t.} & \quad w_i [M_i(r) - C_i] \leq \lambda, \quad i = 1, \ldots, n.
\end{align*}
\]

**D. Constraint-based Routing Metric**

Another method to form the global metric, without introducing weights, is to regard one of the atomic metrics as the global metric to optimize while transforming others into constraints, formally characterized as follows:

\[
\begin{align*}
\max & \quad M_i^\ast(r) \\
\text{S. t.} & \quad A_i^{\text{min}} \leq M_i(r) \leq A_i^{\text{max}},
\end{align*}
\]

where the \( i \)th atomic metric is regarded as the principal metric, \( A_i^{\text{min}}, A_i^{\text{max}} \) denote the constraints on the other atomic metrics.

**E. Application in Cognitive Radio Routing**

We now provide an illustrative example on a CR routing scenario to show how to combine multiple routing metrics to a global one, as analyzed in this subsection. Specifically, we consider two routing metrics: the first metric is the throughput, denoted as \( d(r) \) for route \( r \); the second metric is the cost when the PUs become active during the routing process, denoted as \( C(r) \) (e.g., in terms of channel switching delay and/or energy). The process of choosing an appropriate route in this scenario consists of striking a tradeoff between the two routing metrics. Following the analysis in this subsection, a natural way to find the best route is to combine the two metrics into a global one, denoted as \( M(r) \), as:

\[
M(r) = \omega d(r) + (1 - \omega) C(r),
\]

where \( \omega \) is a relative weight set by the SU balancing the two atomic metrics indicating its preference. For example, a small \( \omega \) signifies that the SU prefers a stable and robust route less impacted by the primary traffic.

Another way of finding an operational route is to solve the following optimization problem:

\[
\begin{align*}
\max_r & \quad d(r) \\
\text{S. t.} & \quad C(r) \leq C_{th},
\end{align*}
\]

where \( C_{th} \) denotes a threshold modeling the limit on the cost of the PU impact that the SU can tolerate.

As another example of formulating the routing problem in CRNs as a multi-objective optimization problem and seeking efficiency routing solutions adapted in such environments, the authors of [65] develop a multi-objective reinforcement learning based routing protocol where the spectrum statistics are assumed to be unknown and should be learned by cognitive users. The performance of the routing process is characterized by multiple metrics like average delay and packet loss rate. To address the challenges of randomness, uncertainty and multiple metrics, the multi-objective reinforcement learning algorithm is developed. The core idea of the developed routing algorithm is to use reinforcement learning, a technique widely applied in the field of artificial intelligence. In terms of multi-objective optimization, two metrics are considered, with one transformed into a constraint.

**VI. FUTURE AND OPEN ISSUES**

In this section, we discuss open issues in the design and implementation of routing metrics for CRNs and future directions.

**A. Metrics for QoS Routing and Realtime Applications**

A natural evolution for routing metrics in CRNs is providing support for realtime applications for SUs. Traditional approaches for QoS support in wireless networks [66] have to be extended to meet the unique challenges of CRNs. In particular, new multi-objective metrics that combine route stability with other QoS metrics, such as delay and bandwidth, are good candidate metrics for QoS in CRNs. For this goal, the approaches discussed in Section V can be leveraged. Combining these metrics with multi-path routing metrics, to select backup routes or to route concurrently on different paths, provide a good framework for realtime applications for SUs in CRNs. For example, the work in [67] proposed QoS routing protocol in CR wireless mesh networks. They formulated the problem as an optimization problem and presented an Integer Linear Programming (ILP) formulation to provide optimal solutions. The optimization problem is to minimize the hop count of the route subject to the constraints are within the admissible bandwidth. However, all the existing work proposed only theoretical approaches that lack practicality.

**B. Context-aware Metrics**

With mobile phones becoming ubiquitous computing devices, it becomes attractive everyday to leverage the CRN paradigm to enhance their performance. Smart phones come with an array of sensors that provide information about the context surrounding a CRN node such as location, direction, motion state, sound, etc. Designing new routing metrics that
leverage this context information to enhance the performance of the CRN is an important direction for future research. Available approaches for adhoc networks, for example location-aided routing metrics [68], [69], can be extended to fit the nature of CRNs. In addition, new routing metrics that leverage other context information can also be developed [70].

C. Security-based Metrics

Another untapped direction in CRNs routing metrics is the security-based metrics that address route selection in case malicious nodes exist in the network. One possible metric can be related to the route closeness metric described in Section III-B5 as it separates the paths as far as possible to minimize the effect of a malicious node attacking all nodes at the same time. Another possibility is to model malicious nodes as PUs whose activity should trigger the SUs to switch channels.

D. Combining with Other Technologies

CRNs are designed to work mainly with wireless networks and inherit from their characteristics. Other evolving networks have similar operational environments and can be combined with CRNs to enhance their performance and extend their application domain. Of particular interest is combining delay tolerant networks (DTNs) [71], [72] with CRNs as they share the challenge of tolerating the delays introduced by the interruption of communication due to the activity of the PU. Combining the design characteristics to obtain a new metric that reflects both domains can lead to better performance in a number of application domains that can tolerate delays.

E. Realistic PU Modeling

Current routing metrics for CRNs, e.g. [41], are mainly based on simple models for the PU (e.g. an On/Off model). However, this is far from realistic in a typical CRN, where PUs’ activity are correlated both in time and space. Predictive activity models that estimate future PUs activity are also of importance for better performance.

F. Spatial Reuse-based Metrics

Current CRNs target temporal reuse, where the SUs access the spectrum when the PUs are not active. Another possibility to increase the gain of CRNs is to leverage spatial reuse. Directional antennae is a way to achieve spatial reuse in wireless networks, where multiple co-located nodes can communicate together using independent sectors of directional antennae. Such an opportunity of increasing the spectrum utilization comes with its own challenges of designing routing metrics that capture the nature of directional antennae, including the time to synchronize the transmitter and receiver antennae, increased deafness problem, among others.

G. Mobility Metrics

Most of the current metrics are designed for stationary nodes. However, mobile users, both primary and secondary, are becoming the norm in wireless networks. Mobility increases the disconnection rate of routes in wireless networks. Designing routing metrics that address different mobility models ranging from the mobility of PUs alone to the more challenging case of both mobile SUs and PUs is another direction for future work in the area of designing routing metrics for CRNs.

H. Implicit Feedback Metrics

Current CRNs routing protocols use offline statistics or the local sensing information to estimate the PU behavior. However, in many situations, communication between nodes are full-duplex. This can be leveraged to provide feedback about the routing path and the PU activity to the transmitter. For example, this information can be piggybacked on the messages sent back from the destination to the source in a cheap way that minimizes the extra communication overhead in the already constrained SU network.

I. Realistic Testbeds for Routing Metrics

Evaluating routing metrics in CRNs is a challenging problem that is currently relaxed using simulations. However, simulations usually suffer from the inability to capture the realistic environment conditions, especially in CRNs where routing is a cross-layering problem. On the other hand, current testbeds usually focus on the MAC and PHY layers leading to prohibitive cost of large-scale deployment, and complexity of implementing the whole protocol stack up to the routing layer. New testbeds that address these issues would enable better evaluation of routing metrics. For example, a recent framework [73], [74] facilitates the development and evaluation of routing protocols in CRNs by abstracting the PHY and MAC layers and allowing the designers to focus on the CRNs routing protocols. This work uses standard computers and WiFi cards to reduce the cost while allowing integration with other special hardware for more flexibility.

VII. Conclusion

In this paper, we presented a survey and taxonomy of different routing metrics for CRNs, based on two main categories: single-path routing and multi-path routing. Different metrics were discussed as well as the routing protocols that employ them. We also listed the challenges of designing a routing metric for CRNs, both inherited from traditional wireless networks and those unique to CRNs. In addition, we provided a case study that compares the performance of different metrics in different scenario in CRNs.

An important direction for research in routing metrics for CRNs is hybrid metrics that combine a number of atomic metrics. Therefore, we presented different techniques for combining individual routing metrics to obtain a global one. We concluded the paper with a discussion of open issues in the design of new routing metrics.

VIII. Acknowledgements

This work has been supported in part by a grant from the Egyptian National Telecommunication Regulatory Authority (NTRA).

Moustafa Youssef is an associate professor at Egypt-Japan University of Science and Technology and Alexandria University. He has more than 10 issued and pending patents, is serving as an area editor of the ACM MC2R, served as a Co-Guest Editor of The IEEE Journal on Selected Areas in Communications - Special Issue on Mission-Critical Networking, and has served on the organizing and technical committees of numerous prestigious conferences. He has published more than 100 technical papers in the world’s top conferences and journals.

Mohamed Ibrahim is a research assistant at Egypt-Japan University of Science and Technology (E-JUST), Egypt. He received his M.Sc. in wireless technology from Nile University, Egypt in 2011 and a B.Sc. in computer science from Alexandria University, Egypt in 2009. His research interests include mobile wireless networks, sensor networks, location determination technologies, and pervasive computing.

Mohamed Abdelatif is a M.Sc. student at Alexandria University, Egypt. He received his B.Sc. in Computer Engineering from Alexandria University, Egypt in 2010 and worked as a research assistant in the Wireless Research Center at Egypt-Japan University of Science and Technology (E-JUST), Egypt. His research interests is focused on Cognitive Radio Networks and mobile wireless networks.
Lin Chen received his B.E. degree in Radio Engineering from Southeast University, China in 2002 and the Engineer Diploma from Telecom ParisTech, Paris in 2005. He also holds a M.S. degree of Networking from the University of Paris 6. He currently works as associate professor in the department of computer science of the University of Paris-Sud. His main research interests include modeling and control for wireless networks, security and cooperation enforcement in wireless networks and game theory.

Athanasios V. Vasilakos is currently a professor at the University of Western Macedonia, Greece. He has authored or co-authored over 200 technical papers in major international journals and conferences. He is the author/coauthor of five books and 20 book chapters in the areas of communications.

Prof. Vasilakos has served as General Chair, Technical Program Committee Chair, TPC member for many international conferences. He served or is serving as an Editor or/and Guest Editor for many technical journals, such as the IEEE Transactions on Network and Services Management, IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics, IEEE Transactions on Information Technology in Biomedicine, IEEE Transactions on Computers, ACM Transactions on Autonomous and Adaptive Systems, the IEEE JSAC special issues of May 2009, Jan 2011, March 2011, the IEEE Communications Magazine, ACM/Springer Wireless Networks(WINET), ACM/Springer Mobile Networks and Applications(MONET). He is founding Editor-in-Chief of the International Journal of Adaptive and Autonomous Communications Systems (IJAACS, http://www.inderscience.com/ijaacs) and the International Journal of Arts and Technology (IJART, http://www.inderscience.com/ijart). He is the General Chair of the Council of Computing of the European Alliances for Innovation.