Practical Spectrum Aggregation for Secondary Networks with Imperfect Sensing

Haeyoung Lee, Youngwook Ko, Member, IEEE, Seiamak Vahid, and Klaus Moessner, Senior Member, IEEE

Abstract—We investigate a collision-sensitive secondary network that intends to opportunistically aggregate and utilize spectrum of a primary network to achieve higher data rates. In opportunistic spectrum access with imperfect sensing of idle primary spectrum, secondary transmission can collide with primary transmission. When the secondary network aggregates more channels in the presence of the imperfect sensing, collisions could occur more often, limiting the performance obtained by spectrum aggregation. In this context, we aim to address a fundamental query, that is, how much spectrum aggregation is worthy with imperfect sensing. For collision occurrence, we focus on two different types of collision: one is imposed by asynchronous transmission; and the other by imperfect spectrum sensing. The collision probability expression has been derived in closed-form with various secondary network parameters: primary traffic load, secondary user transmission parameters, spectrum sensing errors, and the number of aggregated sub-channels. In addition, the impact of spectrum aggregation on data rate is analysed under the constraint of collision probability. Then, we solve an optimal spectrum aggregation problem and propose the dynamic spectrum aggregation approach to increase the data rate subject to practical collision constraints. Our simulation results show clearly that the proposed approach outperforms the benchmark that passively aggregates sub-channels with lack of collision awareness.

Index Terms—Spectrum aggregation, collision probability, opportunistic spectrum access, secondary networks

I. INTRODUCTION

Enabling secondary users (SUs) to utilize the unoccupied channels of the primary users (PUs), opportunistic spectrum access under the interweave paradigm [1] is regarded as a promising solution for resolving spectrum underutilization [2]. Through spectrum sensing, secondary users are able to identify and utilize the unoccupied primary channels.

In opportunistic spectrum access, secondary transmission can collide with primary transmission. Such collision may happen in a case when PUs asynchronously access own channels with SUs. That is, having the priority over SUs in accessing the channels, PUs can dynamically access the channel even when a SU is accessing the same channel. In addition, there is another collision case when SUs identify the channels’ status incorrectly; SUs incorrectly find the channels to be unused and access the ones on which a PU is already using. When SUs have high mobility feature in mobile networks operating under highway condition (e.g., vehicular networks [3]), collisions can happen more frequently [4]. If a collision happens between PU and SU transmission, the SU has to vacate the channel immediately and dynamically access a new available one. In particular, the SU communication has to be interrupted, packets must wait in the transmission buffer. The communication can be resumed when a connection is successfully established on a new channel. Such spectrum handoff caused by the presence of the collision results in additional latency that affects SU performance in addition to causing short-term interference to PUs [5].

When a SU needs the high data rate, the SU can use multiple channels simultaneously through spectrum aggregation [6]. However, the use of an aggregate channel comprising of multiple channels could incur more collisions. The improvement of the SU’s performance can be obtained by spectrum aggregation, while frequent collisions degrade the performance. Spectrum aggregation is challenged by the uncertainty in the collision events which in turn results from the lack of information on the relationship between the collision probability and aggregated spectrum. To the best of our knowledge, however, the collision events have not been taken into account with the design of spectrum aggregation in the literature. In [7] [8], the mathematical modelling of spectrum aggregation of SUs and the channel capacity obtained through spectrum aggregation are investigated, respectively. In [9] [10], the research focuses on theoretical analysis of spectrum aggregation for the dynamic traffic, in order to accommodate more secondary users and to improve the data rates. In [11] [12], while considering the hardware constraint for spectrum aggregation, aggregation algorithms with only the aim to achieve higher network throughput have been proposed. In addition, in [13], the impact of the secondary packets’ service time over multiple channels on the transmission latency is investigated. In [14], a given number of total/idle channels, the optimal sensing time and power allocation to maximize energy efficiency is investigated. Even though the collision probability of the SU [15] and spectrum allocation strategies for the secondary network considering the SU’s collision probability [16] [17] [18] have been investigated, they are limited to the single channel use by the SU.

In this paper, a collision-sensitive secondary network opportunistically accessing multiple primary channels is considered in the presence of non-zero collision. Two collision cases
are focused: one is caused by asynchronous transmission; and the other is by imperfect spectrum sensing. This work aims to address a fundamental query, how much spectrum for aggregation is worthy with less collision in the presence of imperfect sensing. In the concerned collision cases, a closed-form expression for the collision probability is derived, developing the relationship between spectrum allocation with aggregation and collision occurrence. By using the collision probability, an optimisation problem is formulated and we propose a new method for dynamic spectrum aggregation to increase data rate while the impact of collision occurrence on network performance can be managed. In addition, we show that imperfect spectrum sensing increases the collision happening six times greater for a given sensing accuracy and impacts on the network performance in terms of the allowed number of sub-channels for aggregation. In this paper, we extend our former work [19], where we consider collision happening only from asynchronous transmission. When SUs are in mobile networks under high-speed mobility, the high mobility feature can also influence collision occurrence. However, we focus on collisions caused by the change of PUs’ temporal channel usage with imperfect sensing and leave the analysis of high mobility for future work.

The rest of this paper is organized as follows. Section II presents the model of primary and secondary network, describing the traffic pattern and the dynamic spectrum access scheme. The spectrum aggregation problem to maximize the data rate considering the collision probability is formulated in Section III. The data rate of an aggregate channel and the collision probability of the multiple channels used by a SU are analysed in Section IV. Then, the optimal solution to aggregate multiple channels is addressed in Section V. After analysing some extreme cases in Section VI, the analytical and simulation results are shown in Section VII. Finally, our conclusions and future work are provided in Section VIII.

II. NETWORK AND CHANNEL MODEL

A. Network model

We consider the downlink transmission of a secondary network coexisting with a primary network. In the primary network, a base station (BS_p) communicates with the PUs through multiple licensed channels in the same band. The PUs, as the licensed users, have priority over the SUs in accessing the licensed (primary) channels. Thus, the secondary network opportunistically detects and aggregates idle primary channels. Then, the aggregate channel (consisting of multiple primary channels) is allocated for data transmission with the SU.

We assume that in the primary network, the packet arrival rate at each channel in the same band follows a Poisson process with a rate \( \lambda_p \). The service time of the packets, \( 1/\mu_p \), has the general distribution [18]. The packet arrival rate and service time at each channel are assumed to be independent and identically distributed (i.i.d.) [20].

Suppose in the secondary network, its base station (BS_s) is able to sense the availability of primary channels during a sensing time through spectrum sensing (e.g., energy detection, cyclostationary feature detection, and so on) [21]) [22]. Particularly, every sensing period, BS_s will identify a subset of primary channels unoccupied by PUs. In practice, we have a noisy (imperfect) spectrum sensing that may cause imperfect estimate in the status of the channels. With miss-detection probability (denoted by \( P_m \)), the primary signal could be missed and the busy channel could be detected as available erroneously. If the channel is not occupied, according to a false alarm probability (denoted by \( P_f \)), the idle channel can be identified as occupied. For given (noisy) estimate of \( N \) unoccupied channels, the BS_s opportunistically sends the data to a SU during the data transmission interval, \( T_d \). The data transmission interval of the secondary network, \( T_d \), is assumed to be the same length as the sensing period, since the sensing time is in general much shorter than the data transmission interval. \( T_d \) is the maximum time interval over which a SU remains unaware of any changes in channel occupancy.

In opportunistic spectrum access, we consider two realistic collision cases: (i) PUs return to the channels that are used for data transmission with the SU; (ii) the channel occupancy by PUs is miss-detected and the SU starts to use the primary channels for data transmission. Notice that such collision in (i) is inherent in realistic systems where no synchronization between the primary and secondary network is assumed. Thus, even in analysis of collisions from the case (ii) (i.e., imperfect sensing), collisions from the case (i) (i.e., asynchronous transmission) are included. To measure such collision, we define the collision probability, \( P_c \), as the probability that the secondary transmission collides with the primary transmission and it can be given by [15]

\[
P_c = \lim_{T \to \infty} \frac{\text{the number of collided SU transmissions in } [0, T]}{\text{the number of SU transmissions in } [0, T]}
\]

In the presence of the certain level of collisions, let us consider the secondary network can perform effective communication while \( P_c \) remains below the pre-defined threshold level (denoted by \( \xi_{th} \)). When collisions occur, data to transmit is assumed to be lost but a certain level of lost data can be recovered in the secondary network [23]. That is, the collision probability’s threshold indicates the maximum collision probability tolerated by the secondary network [24]. Notice that the value of \( \xi_{th} \) can influence the performance of the secondary network as well as that of the primary network. That is, until PUs are discovered from the time they collide with the SU, PUs can suffer from short-term interference. Thus, the quality of service of PUs could be also considered in the setting of \( \xi_{th} \).

B. Channel model

For given \( N \) available primary channels in the same band, we propose that the BS_s of the secondary network aggregates only a subset of \( n(\leq N) \) primary channels before every transmission interval \( T_d \), properly selecting the value of \( n \). Particularly, following a Uniform distribution, BS_s randomly selects \( n \) among \( N \) idle channels since the channel quality and PU traffic loads of all channels in the same band are assumed homogeneous. This random selection leads to each
one of the $N$ channels available being equally likely to be selected before each transmission interval [25]. In addition, let the maximum transmission power be denoted by $P_{\text{max}}$, and $P_{\text{max}}$ is equally distributed among the $n$ randomly selected channels [26]. Since the $n$ primary channels now comprise an aggregate channel, hereafter, the term ‘primary channel’ can be interchangeable with ‘sub-channel (of an aggregate channel)’.

Suppose that available sub-channels between the $BS_s$ and the SU are independent and Rayleigh flat fading. The channel coefficient of sub-channel $i$ (denoted by $h_i$, for all $i$) is complex Gaussian random variable with zero mean and variance $\sigma^2_i$, i.e., $h_i \sim CN(0, \sigma^2_i)$. Such channel state information (CSI) is assumed to be perfectly known at the receiver. Then, the received signal to noise ratio (SNR) at sub-channel $i$ for $i = 1, 2, \ldots, N$ can be represented as

$$\rho_i = \frac{|h_i|^2 P_i}{\sigma^2} = \frac{g_i P_i}{\sigma^2}$$  \hspace{1cm} (2)

where $g_i = |h_i|^2$, for all $i$. $g_i$ is a Chi-squared distributed random variable with $2k$ degrees of freedom where $k$ denotes the number of receiver antennas along with multi-antenna techniques (e.g., maximal ratio combining) [27]. $P_i$ is the transmit power, i.e., $P_i = P_{\text{max}}/n$ for equal power allocation. Notice in (2) that $\sigma^2$ is the variance of the complex-valued zero-mean additive white Gaussian noise (AWGN).

We can express the data rate, as the capacity of sub-channel $i$ of bandwidth $B$, $C_i = B \log(1 + \rho_i)$, by using Shannon’s capacity theorem [24]. Using (2), $C_i$ can be further given by

$$C_i(n, g_i/\sigma^2) = B \log_2 \left( 1 + \frac{g_i \cdot P_{\text{max}}}{n \cdot \sigma^2} \right).$$  \hspace{1cm} (3)

From (3), the average data rate for the selected sub-channel $i$ can be obtained by taking the expectation of (3) with respect to $g_i/\sigma^2$ which can be expressed as

$$E[C_i] = \int_0^\infty B \cdot \log_2 \left( 1 + \frac{g_i \cdot P_{\text{max}}}{n \cdot \sigma^2} \right) \frac{x^{\frac{1}{2} - 1} e^{-x}}{2^\frac{k}{2} \Gamma(\frac{k}{2})} \, dx,$$  \hspace{1cm} (4)

where $E[\cdot]$ stands for the expectation operator and $p_x(x)$ is the probability density function of $x = g_i/\sigma^2$, for all $i$, which can be given by

$$p_x(x) = \frac{1}{2^k/\Gamma(\frac{k}{2})} x^{\frac{1}{2} - 1} e^{-x},$$

where $\Gamma(\cdot)$ denotes the gamma function [24].

The data rate of the aggregate channel is upper bounded by the sum of the data rate of the $n$ active sub-channels, as $BS_s$ aggregates $n$ sub-channels to transmit its data. With no loss of generality, let $B$ be normalized (i.e., $B = 1$). Since each of the $N$ sub-channels has the same probability to be randomly selected, the average data rate of an aggregate channel, $C_{\text{total}}$, for a given $N$ is calculated, weighting the average data rate in (4) by its own aggregation probability and adding them all. Thus, $C_{\text{total}}$ can be expressed as

$$C_{\text{total}}(n) \leq \frac{1}{N} \sum_{i=1}^{N} \left[ \log_2 \left( 1 + \frac{g_i \cdot P_{\text{max}}}{n \cdot \sigma^2} \right) \right],$$  \hspace{1cm} (5)

where $n/N$ denotes the probability that each sub-channel is chosen as one of $n$ active sub-channels for aggregation.

### III. Problem Formulation

We study the problem of the spectrum aggregation, to maximize the achievable data rate $C_{\text{total}}$ during the data transmission interval while satisfying a collision probability requirement. While intra-band aggregation where average channel quality and PU traffic loads are same for different channels in the same band is considered [6], the problem of spectrum allocation with aggregation is simplified to properly select only a subset of $n$ among $N$ total available sub-channels for the secondary network. Using (5), this problem can be posed as

$$\max C_{\text{total}}(n, \xi_{\text{th}}) = \max \frac{1}{N} \sum_{i=1}^{N} \left[ \log_2 \left( 1 + \frac{g_i \cdot P_{\text{max}}}{n \cdot \sigma^2} \right) \right],$$  \hspace{1cm} (6)

subject to $P_{c,n} \leq \xi_{\text{th}},$  \hspace{1cm} (7)

where $P_{c,n}$ denotes the probability that the secondary transmission exploiting the $n$ sub-channels collides with the primary transmission and $\xi_{\text{th}}$ stands for the maximum tolerable collision probability.

Notice that collision by the secondary transmission on the $n$ sub-channels includes all the events of collision across all possible subsets of the $n$ sub-channels. Thus, for a given $n$, $P_{c,n}$ can be derived, utilising the collision probability of an individual sub-channel, as follows.

$$P_{c,n} = \Pr \{ \text{Collision in any of } n \text{ channels} \} = 1 - \Pr \{ \text{No collision in all } n \text{ channels} \}.$$  \hspace{1cm} (8)

To solve the problems (6)-(8), we raise and address a fundamental question, “how many sub-channels must be aggregated for higher data rate under the collision requirement?”. To that end, we first start by analysing the impact of the number of aggregate sub-channels on the performance, followed by the optimal solution using analytical results.

### IV. Analysis on the Data Rate and Collision Probability

#### A. Impact of the channel aggregation on $C_{\text{total}}$

Since the log function is a concave function, by using Jensen’s inequality [24], the data rate of an aggregate channel composed of $n$ sub-channels for a given $N$ can be obtained by

$$C_{\text{total}}(n) = \frac{1}{N} \sum_{i=1}^{N} \left[ \log_2 \left( 1 + \frac{g_i \cdot P_{\text{max}}}{n \cdot \sigma^2} \right) \right] \leq \frac{1}{N} \sum_{i=1}^{N} \log_2 \left( 1 + \frac{\gamma_i}{n} \right),$$  \hspace{1cm} (9)

where $\gamma_i$ denotes the average received SNR level with the maximum transmit power on sub-channel $i$, i.e., $\gamma_i = E[g_i] \cdot P_{\text{max}}/\sigma^2$.

Let us figure out how the data rate behaves with $n$, the number of sub-channels used for aggregation. For this, the
first and the second derivatives of the data rate with respect to \( n \) are computed as follows
\[
\frac{\partial C_{\text{total}}}{\partial n} = \frac{1}{N(\log 2)} \sum_{i=1}^{N} \left[ \log \left( 1 + \frac{\gamma_i}{n} \right) - \frac{\gamma_i}{n + \gamma_i} \right].
\]
(10)
\[
\frac{\partial^2 C_{\text{total}}}{\partial^2 n} = \frac{1}{N(\log 2)} \sum_{i=1}^{N} \frac{-\gamma_i^2}{n(n + \gamma_i)^2}.
\]
(11)

Consider the homogeneous channel setup where \( \gamma_i = \gamma \), for all \( i \). Noticing the fact that \( \partial^2 C_{\text{total}}/\partial^2 n \) in (11) is not positive for all possible \( n \), \( C_{\text{total}} \) is a concave function of \( n \) for a given \( N \). [28]. Then the value of \( n \) to maximize the data rate, \( n^* \), must be the one, satisfying
\[
n^* = \arg \min_n \left| \log \left( 1 + \frac{\gamma}{n} \right) - \frac{\gamma}{n + \gamma} \right|.
\]
(12)

In (12), for a given \( \gamma \) where \( \gamma > 0 \), the argument \( \cdot \) from \( | \cdot | \) in (12) is always positive for any \( n \) where \( n > 0 \). That is, while \( C_{\text{total}} \) is being strictly concave (i.e., \( \partial^2 C_{\text{total}}/\partial^2 n < 0 \)), the first derivative is always positive (i.e., \( \partial C_{\text{total}}/\partial n > 0 \)). This reveals that the data rate monotonically increases with \( n \) which is the number of sub-channels selected for aggregation. Therefore, more \( n \), higher \( C_{\text{total}} \) is obtained.

### B. Impact of the channel aggregation on collision probability

Let us consider the homogeneous case where the PU traffic intensity on sub-channel \( i \), \( \forall i \), are i.i.d. The collision probability of each sub-channel is equally likely being \( P_{c} \) in (1). Using this, (8) can be further simplified to
\[
P_{c,n}(n) = 1 - \prod_{i=1}^{n} (1 - P_{c}) = 1 - (1 - P_{c})^n.
\]
(13)

It can be shown in (13) that the collision probability, \( P_{c,n} \), can be derived as an increasing function of the number of sub-channels, \( n \), once the collision probability by a single sub-channel use, \( P_{c} \), is given.

To further analysis on \( P_{c,n} \), we now consider two realistic cases for collision occurrence. One collision is in the case when collision of PU transmission at the SU transmission happens only due to no synchronization between primary and secondary network. The other is in the presence of noisy estimate of the occupancy on primary channels, that is, the imperfect sensing case when collision happens due to wrong identification of sub-channels’ occupancy state as well as asynchronous transmission.

#### 1) Collision due to asynchronous transmission: The collision happens only if PUs reappear on the sub-channels on which the SU is still transmitting. As shown in [15], the collision probability of a single sub-channel due to asynchronous transmission between primary and secondary network, \( P_{c}^{AT} \), is equivalent to the probability that at least one PU’s packet arrives during a SU’s transmission period \( T_d \). As PU’s packets arrive with a rate of \( \lambda_p \) according to a Poisson process, \( P_{c}^{AT} \), is given as
\[
P_{c}^{AT} = 1 - \exp(-\lambda_p \cdot T_d).
\]
(14)

From (13) and (14), \( P_{c,n} \) due to asynchronous transmission (denoted by \( P_{c,n}^{AT} \)) can be given, for given \( n \) aggregated sub-channels, by
\[
P_{c,n}^{AT}(n) = 1 - (1 - P_{c}^{AT})^n = 1 - \exp(-\lambda_p \cdot T_d \cdot n).
\]
(15)

In (15), \( P_{c,n}^{AT} \) indicates the probability that at least one primary packet arrives during a secondary transmission period in \( n \) sub-channels. It can be seen from (15) that \( P_{c,n}^{AT} \) increases with the product of \( \lambda_p \) and \( T_d \), for given \( n \) aggregated sub-channels. For given \( \lambda_p \) and \( T_d \), it can be shown from (15) that \( P_{c,n}^{AT} \) increases with \( n \).

#### 2) Collision due to imperfect sensing and asynchronous transmission: Collision events may happen not only in the asynchronous case (when the primary transmission starts on the channels before the secondary transmission terminates) but also in the case when noisy sensing estimate occurs such that the occupied channels are incorrectly identified to be unoccupied, assigning the SU. As shown in [15], we can define \( P_{idle} \) as the probability that the channel is identified unoccupied by the secondary network after sensing, and \( P_{busy} \) as the probability that the channel is identified occupied. Then, we have
\[
\begin{align*}
P_{idle} &= \frac{\lambda_p}{\mu_p} P_m + (1 - \frac{\lambda_p}{\mu_p})(1 - P_f), \\
P_{busy} &= \frac{\lambda_p}{\mu_p} (1 - P_m) + (1 - \frac{\lambda_p}{\mu_p}) P_f.
\end{align*}
\]
(16)

Let us consider the case of a single channel use. BSs will transmit with the SU with a channel only if spectrum sensing result indicates an idle channel, with probability \( P_{idle} \). The secondary transmission may be collided due to collision with primary transmission whenever errors occur in spectrum sensing or primary transmission occurs during the transmission interval. Thus, the collision probability under imperfect sensing denoted by \( P_{IS} \) contains not only the term for asynchronous transmission case, but also the term for misdetection case. In addition, since channels only identified as idle are accessed, \( P_{IS} \) includes the effect of \( P_{idle} \) and is expressed as follows
\[
\begin{align*}
P_{IS} &= \frac{(\lambda_p/\mu_p) P_m + (1 - \lambda_p/\mu_p)(1 - P_f) P_{c,n}^{AT}}{P_{idle}} \\
&= \frac{\frac{\lambda_p}{\mu_p} (1 - P_m) + (1 - \lambda_p/\mu_p)(1 - P_f) [1 - \exp(-\lambda_p T_d)]}{\frac{\lambda_p}{\mu_p} (1 - P_m) + (1 - \lambda_p/\mu_p)(1 - P_f)}.
\end{align*}
\]
(17)

From (13) and (17), therefore, \( P_{c,n} \) based on the imperfect sensing and asynchronous transmission (denoted by \( P_{IS,c,n} \)) can be given by (18). From (18), it is shown that the SU transmission via the aggregate channel could collide only except for the case that the selected \( n \) sub-channels are really idle and no PU transmission reclaim those selected sub-channels. As per the impact of \( n \), it can be seen from (18) that \( P_{IS,c,n} \) increases with \( n \) for given \( \lambda_p, \mu_p, P_f, P_m, \) and \( T_d \). Unlike in the asynchronous transmission case, for a given sensing method, \( P_{IS,c,n} \) is impacted by \( P_f, P_m \) and \( \mu_p \) as well as \( \lambda_p \) and \( T_d \).
\[
P_{c,n}^{I_S}(n) = 1 - (1 - P_c^{I_S})^n = 1 - \left[1 - \frac{(\lambda_p/\mu_p)P_m + (1 - \lambda_p/\mu_p)(1 - P_f)[1 - \exp(-\lambda_p T_d)]}{(\lambda_p/\mu_p)P_m + (1 - \lambda_p/\mu_p)(1 - P_f)}\right]^n = 1 - \left[\frac{(1 - \lambda_p/\mu_p)(1 - P_f)[\exp(-\lambda_p T_d)]}{(\lambda_p/\mu_p)P_m + (1 - \lambda_p/\mu_p)(1 - P_f)}\right]^n. \tag{18}
\]

V. Optimal solution

We consider the collision-sensitive secondary network where it is desired to provide a realistic optimal solution that maximizes the data rate $C_{\text{total}}$, while simultaneously limiting the collision probability $P_{c,n}$ below the desired level. For this, we solve the maximization problem described in (6) and (7).

As per $C_{\text{total}}$ and $P_{c,n}$, the analytical results in Section IV presented that the both are functions of the number $n$ of active sub-channels used for aggregation. Particularly, notice the fact that the monotonicity of $C_{\text{total}}$ with $n$ presented in Section IV. Then, it can be stated that, for a given collision probability threshold $\xi_{th}$, the optimal value of $n(\leq N)$ maximizing $C_{\text{total}}$ may exist and can be found as the largest among possible integer values satisfying the collision probability requirement ($P_{c,n} \leq \xi_{th}$). With respect to the collision probability requirement, moreover, we focus on the two cases of collision (due to solely asynchronous transmission or both imperfect sensing and asynchronous transmission).

To the universal framework of the two cases, the requirement (7) can be rewritten, using (13), as

\[
1 - (1 - P_c)^n \leq \xi_{th}. \tag{19}
\]

Taking the logarithm of both sides in (19), the necessary condition on being possible candidate $n_s$ for the optimum can be obtained for a given $\xi_{th}$ as

\[
n \leq \left\lfloor \frac{\log(1 - \xi_{th})}{\log(1 - P_c)} \right\rfloor \tag{20}
\]

where $\lfloor x \rfloor$ denotes the largest integer not larger than $x$, and $P_c$ denotes the collision probability being (14) and (17) for two cases, respectively. Here, note that $n$ is the number of active sub-channels for aggregation, and thus the valid range of $n$ in (20) must be given to $1 \leq n \leq N$.

Unlike (12), in practice, we need to consider the requirement in (19), in order to obtain the optimum $n^*$. Then, $n^*$ must be the largest integer satisfying the inequality of (20). Thus, $n^*$ subject to $P_{c,n} \leq \xi_{th}$ can be expressed using (20).

\[
n^* = \begin{cases} \left\lfloor \frac{\log(1 - \xi_{th})}{\log(1 - P_c)} \right\rfloor, & \text{if } 1 \leq \left\lfloor \frac{\log(1 - \xi_{th})}{\log(1 - P_c)} \right\rfloor \leq N, \\
N, & \text{elseif } \left\lfloor \frac{\log(1 - \xi_{th})}{\log(1 - P_c)} \right\rfloor > N, \\
0, & \text{otherwise} \end{cases} \tag{21}
\]

As the universal framework solution, it can be shown from (21) that $n^*$ relies on both $P_c$ and $\xi_{th}$ for a given $N$. Notice that $P_c$ from (21) is indicating the collision probability of each sub-channel for aggregation, resulting from any given sensing technique concerned.

To specify $n^*$ of the aforementioned two collision cases, $P_c$ in (14) and (17) are now taken into account, respectively, further solving (21). To that end, in following subsections, we need to derive the expression for $n^*$ with respect to the primary traffic intensity $\lambda_p/\mu_p$, the SU transmission interval $T_d$ and spectrum sensing error metrics (by either $P_f$ or $P_m$, or both). For simplicity in analysis, hereinafter, we consider only when all the candidates satisfying (20) are below $N$, i.e., $1 \leq \lfloor x \rfloor \leq N$, for a given $N$. This leads us to consider only the first inequality condition in (21). Since in the other conditions in (21) the corresponding optimal solutions can be provided straightforwardly, being beyond our interests.

A. The case of asynchronous transmission

Inserting (14) into (21), i.e., $P_c = P_c^{AT}$, $n^*$ can be found as follows

\[
n^* = \left\lfloor \frac{-\log(1 - \xi_{th})}{\lambda_p \cdot T_d} \right\rfloor \text{ for } ns \text{ in } \{1, \ldots, N\}. \tag{22}
\]

In (22), it is observed for a given collision threshold $\xi_{th}$, the optimal number of sub-channel for aggregation should be chosen inversely with a product of $\lambda_p \cdot T_d$. Particularly, it can be found from (22) that lower $\lambda_p$ (or smaller $T_d$), more $n^*$ can be.

This reveals that as the product of $\lambda_p \cdot T_d$ decreases, the optimum number $n^*$ of the aggregate sub-channels are allowed to increase. Intuitively, the low rate of primary packets arrival in a given secondary transmission interval results in the large active number of sub-channels for the aggregation, leading to the increase in the data rate of an aggregated channel. Moreover, it is worth mentioning that when $\xi_{th}$ increases (towards 1) for a given $\lambda_p \cdot T_d$, $n^*$ in (22) also does. This indicates that large sub-channels can be exploited for the aggregation for a collision-tolerated secondary network with high $\xi_{th}$. Similarly, as $\xi_{th}$ decreases for the collision-sensitive network, $n^*$ should decreases for a given $\lambda_p \cdot T_d$.

For example, let us consider the homogeneous channel environments where $\gamma_i = \gamma, \forall i$. Using (9) and (22), the maximum data rate, $C_{AT}^{\text{TOTAL}}$, can be derived as follows

\[
C_{\text{total}}^{\text{AT}} \leq \left[\frac{-\log(1 - \xi_{th})}{\lambda_p \cdot T_d}\right] \log_2 \left(1 + \left[\frac{\gamma}{-\log(1 - \xi_{th})}\right]\right) \tag{23}
\]

where recall that $\lfloor x \rfloor$ equals $n^*$ in (22).

It can be shown from (23) that $C_{\text{total}}^{\text{AT}}$ increases monotonically whatever $n^*$ grows, resulting from either 1) low $\lambda_p \cdot T_d$ or 2) high $\xi_{th}(\leq 1)$. From (23), it is also worth mentioning that for a given $\xi_{th}$, $C_{\text{total}}^{\text{AT}}$ can remain at a certain desired level, adjusting $T_d$ to the variations of $\lambda_p$. That is, when PU packet arrives more frequently (i.e., with large $\lambda_p$), it is desired to reduce $T_d$, enabling more frequent spectrum sensing. This is in order to accurately detect the PU traffic arrival, leading $C_{\text{total}}^{\text{AT}}$ to the desired level.
B. The case of imperfect sensing and asynchronous transmission

Inserting (17) into (21), i.e., $P_e = P_c^{IS}$, $n^*$ in the case considering both imperfect sensing and asynchronous transmission can be derived as a function of $P_m$, $P_f$ and $\mu_p$ as well as $\lambda_p$ and $T_d$ in (24). It can be shown from (24) that the larger $\xi_{th}$, larger $n^*$ is obtained. Intuitively, while the collision probability threshold $\xi_{th}$ is large, the collision-tolerable secondary network will be able to aggregate many sub-channels. However, when $\xi_{th}$ is small for the collision-sensitive secondary network, spectrum aggregation with many sub-channels would not be suitable to guarantee the collision probability requirement.

Unlike (22), it is secondly worth mentioning from (24) that $n^*$ can be dependent of $\lambda_p/\mu_p$ for given $T_d$, $P_m$ and $P_f$. Accordingly, using (9) and (24), the maximum data rate (denoted by $C_{total}^{IS}$) of the imperfect sensing case can be obtained with respect to $\lambda_p$, $\mu_p$, $T_d$, $P_m$ and $P_f$.

VI. ASYMPOTIC ANALYSIS

In this section, we consider two extreme cases: (i) asynchronous transmission with very high and low SNRs; (ii) imperfect sensing transmission with very large and small sensing thresholds. This intends to examine how achievable $n^*$ and the data rate approximately behave.

A. Asynchronous transmission case

The impact of the SNR level on the maximum data rate will be analysed here, while the impact of various $\lambda_p$, $T_d$ and $\xi_{th}$ are analysed in Section V.A. To this end, we consider following two extreme situations in terms of SNR.

1) When SNR is very high, $\gamma \rightarrow \infty$: As the SNR increases for a given $n^*$, $C_{total}^{AT}$ in (23) can be approximated as

$$n^* = \left. \frac{-\log(1 - \xi_{th})}{\lambda_p T_d - \log(1 - \frac{\lambda_p}{\mu_p}) - \log(1 - P_f) + \log \left( \frac{2P_m}{\mu_p} + (1 - \frac{\lambda_p}{\mu_p})(1 - P_f) \right)} \right|$$

for $n^*$ in (20) $\in \{1, \cdots, N\}$. (24)

SNR environments, in Fig. 1(b), it is shown that aggregation of multiple sub-channels does not contribute to improving the data rate. As the SNR level is lower, the data rate saturates at the lower values with the lower $n^*$. For example, in Fig. 1(b), it is shown that the data rate at -10 dB SNR environment almost saturates the value with $n^* = 1$ and does not change for increase in $n^*$. That is, the data rate of an aggregate channel is significantly influenced by the SNR level, as confirmed in (26).

Although, in this paper, we consider aggregation for the downlink channel, spectrum aggregation in downlink can impact on the uplink transmission. When multiple sub-channels are exploited for downlink, the SU is expected to feedback channel quality information and (non-) acknowledgement indicating (un)succes of transmission per sub-channel [6]. Aggregation of many sub-channels, thus, could lead to increase in signalling overhead. Thus, in low SNR environments, as $n^*$ for the downlink transmission increases, the data rates for the
uplink transmission could be decreased. It is recommended to use a single sub-channel using the full transmit power rather than multiple sub-channels by aggregation at low SNR environments.

**B. Imperfect spectrum sensing case**

In this section, let us consider the impacts of sensing error parameters, $P_m$ and $P_f$, on the data rate with extreme situation of the energy detection based spectrum sensing method. In the energy detection method, the setting of a threshold $\eta_{th}$ to determine the presence of PU signals influences the sensing errors, $P_m$ and $P_f$, at the same time [21]. Two extreme cases of $\eta_{th}$ (being very low and high) are exploited to investigate the influence of spectrum sensing performance.

1) **When the spectrum sensing threshold is very low, $\eta_{th} \to 0$**:
   
   For very low $\eta_{th}$, it is straightforward that $P_m$ decreases and $P_f$ increases ($P_m \to 0$ and $P_f \to 1$) [21]. In this case, the optimum $n^*$ in (24) can be calculated as
   
   $\lim_{P_m \to 0, P_f \to 1} n^* \approx \frac{-\log(1 - \xi_{th})}{\lambda_p T_d - \log \left(1 - \frac{\lambda_p}{\mu_p}\right) + \infty} \approx 0. \quad (27)$

   In this case, $BS_s$ can detect PU presence accurately due to low $P_m$, but identify unoccupied sub-channels as occupied due to high $P_f$. This results in a reduction in detectable spectrum opportunities for the secondary network. That is, although there are many available sub-channels, $BS_s$ uses the smaller number of sub-channels, leading to lower $CIS_{total}$.

2) **When the spectrum sensing threshold is very high, $\eta_{th} \to \infty$** [21]: For very high $\eta_{th}$, $P_m$ converges 1 and $P_f$ is towards 0 ($P_m \to 1$ and $P_f \to 0$) In this context, while spectrum opportunities are guaranteed to the secondary network (due to lower $P_f$), the challenge is that the secondary network might miss-detect the occupancy of sub-channels by PUs (due to higher $P_m$). $n^*$ in (24) can be approximated as follows

   $\lim_{P_m \to 1, P_f \to 0} n^* \approx \frac{-\log(1 - \xi_{th})}{\lambda_p T_d - \log \left(1 - \frac{\lambda_p}{\mu_p}\right)} \quad (28)$

   In (28), it is worth noting that in the case that $1/\mu_p$ is short, (28) can be simplified to (22) in asynchronous transmission case. That is, when the service time of primary packets ($1/\mu_p$) decreases, collisions occurring due to miss-detection will be reduced and most collisions will happen due to asynchronous transmission. Thus, for the short service time of primary packets, secondary transmissions with aggregation are not sensitive for $P_m$.

   From (27) and (28), it is analysed that for a low PU traffic, higher threshold setting could be better than lower threshold because it guarantees spectrum opportunities for the SU and does not increase the collision probability due to low PU traffic. Thus, the patterns of primary traffic can be utilised to set the spectrum sensing threshold.

### Table I

**Simulation parameters**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>The number of total sub-channels</td>
<td>30</td>
</tr>
<tr>
<td>Bandwidth of a sub-channel</td>
<td>200 kHz</td>
</tr>
<tr>
<td>PU packet arrival rate, $\lambda_p$</td>
<td>[0.01-0.6] /sec</td>
</tr>
<tr>
<td>PU packet service rate, $1/\mu_p$</td>
<td>[0.1-6.0] sec</td>
</tr>
<tr>
<td>SU Transmission interval, $T_d$</td>
<td>10 ms</td>
</tr>
<tr>
<td>Total transmit power, $P_{max}$</td>
<td>1 (Equal power allocation)</td>
</tr>
<tr>
<td>Average SNR, $E[g]/\sigma$</td>
<td>20 dB (Rayleigh fading)</td>
</tr>
<tr>
<td>False alarm probability, $P_f$</td>
<td>0.05</td>
</tr>
<tr>
<td>Miss-detection probability, $P_m$</td>
<td>0.05</td>
</tr>
<tr>
<td>Collision probability threshold, $\xi_{th}$</td>
<td>0.005-0.030</td>
</tr>
<tr>
<td>Simulation time</td>
<td>$10^5 T_d$</td>
</tr>
</tbody>
</table>

**VII. SIMULATION RESULTS**

In this section, we present simulation results for the proposed dynamic spectrum aggregation algorithm in a secondary network that intends opportunistically to access the idle channels of the primary network. We consider total 30 primary channels. Each primary channel has the bandwidth of 200 kHz [29]. The primary packets arrive at the rate of $\lambda_p = [0.01 - 0.6]$ /sec and leave after the service time of $1/\mu_p = [0.1 - 6]$ sec. The average number of idle sub-channels, depending on the primary traffic load, is $E[N] = 30(1 - \lambda_p/\mu_p)$ for a given $\lambda_p$ and $\mu_p$. In the secondary network, $BS_s$ senses the spectrum before every transmission interval ($T_d = 10$ ms) and identifies idle sub-channels with the sensing accuracy of $P_f = 0.05$ and $P_m = 0.05$. The collision probability threshold allowed is set to $\xi_{th} = [0.005 - 0.03]$. For the channel between $BS_s$ and the SU, a rayleigh fading channel of average 20 dB SNR is assumed. In order to guarantee the reliability of the simulation result, the simulation results are averaged for the obtained results during $10^5 T_d$. The parameters used for simulation are described in Table I.

Firstly, we validate the formulated collision probability that
Fig. 3. Comparison of collision probabilities $P_{IS}^{c,n}$ obtained from simulation and from analytical analysis for various $1/\mu_p$ for the case of imperfect sensing and asynchronous transmission.

Fig. 4. Comparison of the proposed scheme with the three reference schemes in the asynchronous transmission case.
In Fig. 4(b), the collision probability $P_{c,n}^{AT}$ for the four schemes is depicted for various $\lambda_p$. Whilst $\xi_{th}$ is set to 0.005, the scheme should generate $P_{c,n}^{AT}$ lower than $\xi_{th}$ for a given $\lambda_p$ and $T_d$. In the case of $n = 1$ (labelled by ‘1 Sub-CH’), $P_{c,n}^{AT}$ remains below (or equal to) $\xi_{th}$ for various $\lambda_p$. However, for $n = 5$ (labelled by ‘5 Sub-CHs’) and $n = 10$ (labelled by ‘10 Sub-CHs’), aggregation schemes, the $P_{c,n}^{AT}$ becomes larger than $\xi_{th}$ and beyond $\lambda_p$ s marked by arrow 2 and 1, respectively, in Fig. 4(b). Since we consider the discrete range of $\lambda_p$ (from 0.01 to 0.27 with an interval of 0.02), under the collision probability constraint, the ‘5 Sub-CHs’ scheme can be exploited only for $\lambda_p = [0.01 - 0.09]$ and the ‘10 Sub-CHs’ scheme is possible to be used only for $\lambda_p = [0.01 - 0.05]$. Unlike the reference schemes, Fig. 4(b) depicts that the proposed optimal aggregation scheme obtains the $P_{c,n}^{AT}$ for all $\lambda_p$ to remain below (or equal to) the $\xi_{th}$ by adaptively changing the number $n$ of sub-channels for a given $\lambda_p$.

Fig. 4(c) illustrates the achievable data rates of the aggregate channel, $C_{total}^{AT}$. For a range of $\lambda_p = [0.01 - 0.05]$ (marked by R1) in Fig. 4(c), the ‘10 Sub-CHs’ scheme is shown to obtain the highest $C_{total}^{AT}$ among the three reference schemes. Due to $P_{c,n}^{AT}$ larger than $\xi_{th}$, for the range of $\lambda_p (>0.05)$, the ‘10 Sub-CHs’ scheme no longer guarantees the collision constraint. The ‘5 Sub-CHs’ scheme shows better $C_{total}^{AT}$ than the ‘1 Sub-CH’ allocation scheme for a range of $\lambda_p = [0.05 - 0.09]$ (marked by R2). However, for $\lambda_p (>0.09)$, the $P_{c,n}^{AT}$ becomes larger than $\xi_{th}$. In a remaining range of $\lambda_p = (0.09 - 0.27)$ (marked by R3), only the ‘1 Sub-CH’ allocation scheme can be utilized. While the reference algorithms using a fixed number of sub-channel can be utilized with the limited range of $\lambda_p$, the optimal algorithm shows the highest $C_{total}^{AT}$ among four schemes for all regions, R1, R2 and R3. For a range of $\lambda_p = [0.27 - 0.51]$, although not shown due to lack of space, only one sub-channel is exploited by the optimal algorithm. For very frequent PU traffic arrivals ($\lambda_p > 0.51$), idle sub-channels cannot be utilized by the secondary network under a given collision probability constraint.

Fig. 5 depicts the performance of the optimal aggregation scheme with that is compared with the reference schemes for the case of imperfect sensing and asynchronous transmission. $\xi_{th}$ is set to 0.03.

In Fig. 5(a), it is shown that the number $n$ of sub-channels for aggregation of each scheme for various $1/\mu_p$. While the optimal algorithm changes $n$ for varying $1/\mu_p$ in Fig. 5(a), the $P_{c,n}^{AT}$ remains below $\xi_{th}$ for all $1/\mu_p$ as shown in Fig. 5(b). At $1/\mu_p$ marked by arrow 3 and 4, the $P_{c,n}^{AT}$ of the ‘10 Sub-CHs’ aggregation scheme and the ‘5 Sub-CHs’ aggregation scheme becomes larger than $\xi_{th}$, respectively.

Fig. 5(c) depicts the average data rate of the aggregate channel, $C_{total}^{IS}$. For $1/\mu_p = [0.1 - 0.3]$ (marked by R4), the ‘10 Sub-CHs’ aggregation scheme is shown to obtain the highest $C_{total}^{IS}$ among the three reference schemes. For $1/\mu_p = (0.3 - 0.7]$ (marked by R5), ‘5 Sub-CHs’ aggregation scheme shows better $C_{total}^{IS}$ than the ‘1 Sub-CH’ scheme. For a remaining range, $1/\mu_p = (0.7 - 2.7]$ (marked by R6), only the ‘1 Sub-CH’ allocation scheme can be used.

Similar to the asynchronous transmission case, the optimal
algorithm shows the best performance of $C^S_{\text{total}}$ for all ranges of $1/\mu_p$ while it adaptively changes the number $n$ of sub-channels for aggregation for a given $\xi_{\text{th}}$. However, for very high PU traffic intensity (with $1/\mu_p > 3.3$), although not shown, idle sub-channels cannot be utilized under the given collision probability constraint. We show that the optimal scheme can adaptively change the number of sub-channels for aggregation only for the change of the PU traffic intensity for a given collision probability threshold, but it is also expected that the optimal scheme can adaptively operate for the changes of the collision probability threshold and the sensing accuracy as well.

By comparing Fig. 5 with Fig. 4, the impact of imperfect spectrum sensing on the collision probability can be analysed. Even though a higher level of collision probability threshold is assumed for the case of asynchronous transmission and imperfect sensing, the number of sub-channels for aggregation chosen by the proposed algorithm is not greater (i.e., equal or smaller) than one for the case of asynchronous transmission. For example, at 0.07 PU traffic intensity, the proposed algorithm aggregates 7 sub-channels with the constraint of $\xi_{\text{th}} = 0.005$ in the case of asynchronous transmission as shown in Fig. 4(a). However, when the effect of imperfect sensing is additionally considered, the proposed algorithm aggregates only 6 sub-channels with $\xi_{\text{th}} = 0.03$ at the same traffic load as shown in Fig. 5(a). In order to aggregate the same number of sub-channels in the case of imperfect sensing, the collision probability threshold not smaller than $0.0321 (\xi_{\text{th}} \geq 0.0321)$ should be allowed for the secondary network from (18). Similarly, at other points of PU traffic intensity in Fig. 5 (where less sub-channels are aggregated compared to one in Fig. 4), it is calculated that the collision probability threshold slightly greater than 0.03 is required to aggregate the same number of sub-channels with the case of the asynchronous transmission. When imperfect spectrum sensing is imposed on the secondary network, a six times larger collision probability threshold should be allowed in order to support the same level of data rates.

VIII. CONCLUSIONS

We have proposed the dynamic spectrum aggregation for the collision-sensitive secondary network so that the data rate of an aggregate channel is maximized subject to the constraint of the collision probability. The proposed method has been analysed with emphasis on the practical case in the secondary network when having the asynchronous transmission and the imperfect spectrum sensing for intra-band aggregation. The collision probability imposed by the proposed spectrum aggregation was mathematically analysed, taking into account the PU traffic intensity, SU transmission interval, the sensing performance parameters, and the number of sub-channels aggregated. Accordingly, the optimization problem of the spectrum aggregation was formulated in order to maximize the data rate under the collision probability constraint. Using our analysis and asymptotic results, it was clearly observed that properly aggregating only a fraction of the sub-channels outperforms the benchmark that is inflexible with a fixed number of sub-channels aggregation with taking into no consideration the collision sensitivity. Interestingly, our results showed that for low SNRs, the optimum spectrum aggregation performed towards round robin, meaning that the random selection of only one sub-channel is recommended at low SNRs. In this work, to focus the relationship between spectrum aggregation and collision occurrence, the channel quality and PU traffic load of all channels are assumed homogeneous. The generalisation to the heterogeneous channel case and consideration of uplink transmission in high mobile networks will be investigated as part of our future work.

ACKNOWLEDGMENT

This work was performed as part of ACROPOLIS (www.ict-acropolis.eu) projects supported by the European Communities Seventh Framework Program (FP7) and EPSRC/DST India-UK Advanced Technology Centre (IU-ATC) programme.

REFERENCES

Haeyoung Lee (Haeyoung.Lee@surrey.ac.uk) received her M.Sc. degree in information and communication engineering from Gwangju Institute of Science and Technology, Korea. From 2004 to 2007, she was a research engineer of mobile communication R&D centre at Samsung Electronics and had worked on mobile internet platforms in CDMA systems. In 2007, she has joined national Radio Research Agency in Korea. While she has worked on dynamic spectrum access and spectrum regulation frameworks for flexible spectrum use, she has been active in ITU activities for spectrum management and cognitive radios. She is currently pursuing her Ph.D. degree in the Institute for Communication Systems (ICS) at University of Surrey, U.K. Her research interests include spectrum aggregation, dynamic spectrum access and the regulatory implications of cognitive radios.

Youngwook Ko (Y.Ko@qub.ac.uk) received Ph.D. in 2006 and M.S. in 2002, in Electrical Engineering from Arizona State University, Tempe, AZ USA, and B.S.E. in Information and Communications Engineering with the outstanding award from Hanam University, South Korea. After his Ph.D., he worked at Samsung as a senior engineer. In 2008, he returned to academia as a postdoctoral fellow in Electrical and Computer Engineering at the University of Alberta, Canada. He worked in the CCSR, University of Surrey, UK, between 2010-2013 as a research fellow and then senior fellow. Since 2013, Dr. Ko has worked in the ECIT Institute at the Queen’s University of Belfast as a Lecturer in Digital Communications. He has authored or co-authored over 30 publications in major IEEE international journals, peer-reviewed international conferences and technical reports. His research is in the area of air-interface technology for the next generation wireless device-centric communications.

Seiamak Vahid (s.vahid@surrey.ac.uk) received his B.Sc. (hon.), M.Sc., and Ph.D. degrees in electrical and electronic engineering from Kings college, University of London, and University of Surrey (UK), in 1985, 1987, and 2000, respectively. From 1988 to 2000, he was a research fellow in the Centre for Communications Systems Research (CCSR), University of Surrey. From 2001 to 2010, he was with the GSM Networks division, Motorola UK, as team leader and later as 3G/LTE system architect with the GSM Networks division, Motorola UK, as team leader and later as 3G/LTE system architect and product security specialist. Since April 2010 he has been back at Institute for Communication Systems (ICS) as senior research fellow and his research interests include the areas of cognitive radio, communication systems modelling, multiple access, resource management, routing and optimization techniques and their applications to cellular and ad hoc wireless communication networks.

Klaus Moessner (k.moessner@surrey.ac.uk) is a Professor in the Institute for Communication Systems (ICS) at the University of Surrey, U.K. Klaus earned his Dipl.-Ing (FH) at The University of Applied Sciences in Offenburg, Germany, an MSc from Brunel University, UK and his PhD from the University of Surrey (UK). His research interests include reconfigurability of the different system levels, including reconfiguration management and scheduling in wireless networks and adaptability of multimodal user interfaces. He is involved in investigation and teaching of mobile service platforms, service oriented architectures, mobile service delivery and service enablers. He also works in the area of resource efficiency and on mechanisms for dynamic resource allocation. Klaus leads a research team investigating reconfigurable radio, dynamic spectrum access and the regulatory implications of DSA and Cognitive Radio Networks.