Mobile Content Offloading in Database-Assisted White Space Networks

Suzan Bayhan[1(✉)], Gopika Premsankar[2], Mario Di Francesco[2], and Jussi Kangasharju[1]

[1] Department of Computer Science, University of Helsinki, Helsinki, Finland
bayhan@hiit.fi, jakangas@cs.helsinki.fi

[2] Department of Computer Science, Aalto University, Espoo, Finland
{gopika.premsankar,mario.di.francesco}@aalto.fi

Abstract. Mobile data offloading leverages more affordable or even free network capacity to reduce the traffic experienced by cellular operators through their limited over-the-air resources. One way to harvest free capacity is to employ the white space, namely, frequencies that are assigned to licensed users but are not actively utilized, as long as no harmful interference is generated. In this article, we characterize the benefits of harnessing node contacts for mobile content offloading through dynamic spectrum access assisted by a white space database (WSDB). We take a content-centric approach and model the selection of distributors among the subscribers of each content served through a base station. We formulate an optimization problem to maximize the offloading gain based on realistic settings. We show that such a problem is NP-hard and devise efficient heuristics for practical mobile data offloading. Our results show that the offloading gain allowed by white space is significant even when WSDB data are inaccurate.

Keywords: White spaces · Dynamic spectrum access · Mobile opportunistic offloading · Content delivery · White space database · WSDB

1 Introduction

Mobile *data offloading* is a method to move traffic from the cellular network through other means, such as local area networks or device-to-device communications. It has emerged as a promising solution to decrease the load on mobile networks [1]. As WiFi is densely deployed, on-the-spot offloading to local wireless networks when the user is under coverage provides a significant decrease in the mobile operator traffic [2]. Mobile communications could also be postponed until users reach an area covered by a WiFi access point through the so-called delayed WiFi offloading. This is an option for delay-tolerant traffic, as long as the time spent without WiFi connectivity is short [2]. However, WiFi offloading may be restricted by the capacity of the backhaul [3] which is often subject to data caps for private WiFi networks. Motivated by these concerns, offloading to

D. Noguet et al. (Eds.): CROWNCOM 2016, LNICST 172, pp. 129–141, 2016.
DOI: 10.1007/978-3-319-40352-6_11

mobile opportunistic networks has been proposed [4,5]. It leverages the capacity of short-distance communications without relying on any infrastructure and entails almost no monetary cost. As this mode is driven by contacts between mobile nodes, it may fail to provide guaranteed delays, making it a better fit for delay-tolerant traffic. However, unlicensed bands such as the Industrial Scientific and Medical (ISM) are already congested, thus possibly incurring in a low transmission capacity for opportunistic offloading.

A different approach to address the spectrum capacity crunch is to employ the *white space*, namely, the spectrum that is licensed to *primary users* (PU) while being spatiotemporally unused. Offloading mobile data to unused PU channels is called *white space offloading* [6,7] and some existing solutions have explicitly targeted proximity-based communications in such a context. Among them, Cui et al. [7] presented a model to leverage WiFi and white spaces instead of cellular communications, with focus on power efficiency and channel assignment under delay constraints. Ding et al. [8] proposed using TV bands for device-to-device communications by creating location-specific white space databases with the help of "big spectrum data" collected by the mobile crowd.

In this article, we characterize the benefits of mobile data offloading through dynamic spectrum access assisted by a white space database (WSDB). We take a content-centric approach and model the selection of distributors among the subscribers of content served through a base station. We then formulate an optimization problem to maximize the offloading gain based on realistic settings. We show that such a problem is NP-hard and devise efficient heuristics for practical mobile data offloading. Our results show that the offloading gain allowed by white space is significant even when WSDB data are inaccurate.

The key contributions of this article are the following.

- We consider database-assisted white space access in realistic settings. While WSDBs are expected to provide accurate and up-to-date information on the incumbents, some flaws (e.g., bogus entries and incorrect device locations) have been discovered due to several reasons (e.g., unsynchronized WSDBs and manual entry of device information) [9]. We explicitly include the factors affecting the availability and the reliability of the WSDB in our model.
- We provide a general and flexible framework for content-driven mobile data offloading. Our model supports two different options: mobile opportunistic offloading through ISM bands and white space offloading via unoccupied PU channels retrieved from the WSDB.
- We propose several heuristics with different levels of complexity to improve the offloading gain. Some of them focus on the number of distributors for each content based on its size and popularity, whereas more sophisticated ones aim at identifying nodes with high offloading potential. Our experiments using realistic user mobility demonstrate that white space enables offloading 67 % more capacity compared to a purely opportunistic approach when the information in the WSDB is accurate. Even in the presence of inaccuracies, the offloading gain is still higher than 47 %.

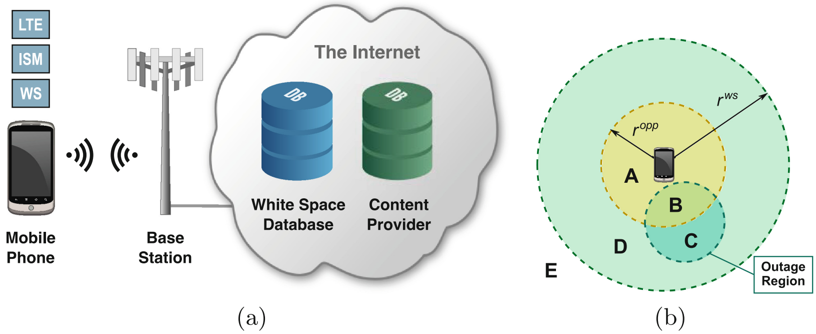


Fig. 1. (a) Reference architecture and (b) offloading regions: opportunistic offloading takes place in zones A and B, white space offloading in zone D, while no offloading is possible in zones C and E.

To the best of our knowledge, our work is the first one focusing on white space offloading for mobile content delivery. In fact, existing solutions in the literature addressed either wireless capacity of white space networks [8] or distributed dynamic access schemes [10]. In contrast, we propose a content delivery framework for white space offloading in database-assisted networks that jointly utilizes opportunistic contacts for offloading. Our solution also distinguishes itself from the state of the art on mobile opportunistic offloading. For instance, Li et al. [5] presented optimal offloading in mobile opportunistic networks through careful selection of distributors. Even though our approach is somewhat similar, we focus on the offloading gain rather than on the distribution delay and buffer constraints.

2 System Model

Our reference architecture is the cellular network illustrated in Fig. 1a. A base station (BS) is connected to both a content provider and a white space database (WSDB). A set $\mathcal{N}$ of mobile users (through their respective devices) is also part of the network and requests content as well as spectrum availability through the BS. We assume that the overhead[1] associated with such requests (and the related responses) is negligible. Each mobile device is equipped with three radio transceivers: one for cellular connectivity, one for communications in the ISM bands, and one for white space access. We denote the ranges of the white space and the ISM radios by r^{ws} and r^{opp}, respectively (Fig. 1b). These ranges may differ as they depend on the actual frequencies employed. In the following, we consider TV bands for white space offloading; as a consequence, $r^{ws} > r^{opp}$. Moreover, we assume that the connection to the BS can be used independently

[1] We analyze the WSDB querying delay and its impact on offloading capacity in Sect. 4, based on our measurements from the Google Spectrum Database.

from the others, while only one of the white space and the ISM interfaces can be active at a given time.

The WSDB stores the information related to white spaces. In contrast to opportunistic offloading, nodes cannot immediately start offloading to white spaces after the discovery of a peer node. Instead, nodes first consult the WSDB by using a database communication protocol (e.g., PAWS [11]) for the permitted operation parameters, including the list of available channels. Even though the WSDB is assumed to have perfect information about white space utilization, unregistered primary users (PU) may still access certain frequencies, thus resulting in interference. To this end, we model the probability of unsuccessful communications due to PU collisions on a given frequency as p^{un}. We assume that the availability of white space is such that all requests can be accommodated without competition between users. We also assume that mobile devices may not be able to reach the BS from certain regions in the nominal coverage range of the BS, e.g., due to shadowing. We call these regions *outage areas* and denote the probability that a mobile device lies in such a region with p^{sh}.

We describe the different content in the network through the set $\mathcal{C} = \{c_1, \cdots, c_k, \cdots, c_K\}$. Each specific content c_k is characterized by its size l_k and its delivery deadline T_k. The set of mobile nodes subscribing to c_k (i.e., the *subscribers*) is denoted as $\mathcal{S}_k$. Without loss of generality, we assume that c_1 is the most popular and c_K is the least popular content. Each mobile user subscribes to one content only. The BS serves the content requests by the users in its coverage area through the *distributors* $\mathcal{D}_k$, each responsible for content c_k. The BS transfers the allocated content to distributors through its F frequencies. The distributors, in turn, deliver the cached content to rest of the nodes (i.e., $\mathcal{S}_k \setminus \mathcal{D}_k$) as long as it is valid. Mobile nodes request the content directly from the BS as soon as the related deadline expires. The BS selects the distributors and announces the association between them and the cached content to the network. As a consequence, subscribers know from which node to request their content.

We model the inter-contact time between pair of nodes through an exponential distribution with parameter $\lambda_{i,j}^{opp}$ for the opportunistic radio and $\lambda_{i,j}^{ws}$ for the white space interface. We assume that contacts are long enough to completely transfer a content item.

3 White Space Offloading

A subscriber node n_i fetches content c_k in one of the three modes detailed next.

– *Mobile opportunistic* (or *ISM) offloading:* Let n_j be a distributor for c_k: $n_j \in \mathcal{D}_k$, and $d(n_i, n_j)$ denote the Euclidean distance between n_i and n_j. Node n_i receives the content from n_j if it is in the range of n_j's ISM radio interface, i.e., the $d(n_i, n_j) \leqslant r^{opp}$. Given that the inter-contact time between n_i and n_j is exponentially distributed with parameter $\lambda_{i,j}^{opp}$, opportunistic offloading is possible if n_i has a contact with n_j during the lifetime of c_k. More formally, we state opportunistic offloading probability as:

$$p_{i,j,k}^{opp} = 1 - \mathrm{e}^{-\lambda_{i,j}^{opp} T_k}. \tag{1}$$

– *White space offloading:* Node n_i receives the content from n_j through the white space only if it cannot get the content by opportunistic offloading. This case is possible only if $r^{opp} < d(n_i, n_j) \leqslant r^{ws}$. In other words, opportunistic offloading is preferred over white space offloading due to the entailed cost and possibly poorer performance, e.g., inaccurate WSDB data. In this case, the distributor first consults the BS to get an available white space channel for offloading. Therefore, this mode is possible only when n_j has an uplink channel to the BS. In Fig. 1b, n_j sends the content to n_i using white space offloading only in region D. Recall that a node may be under outage with probability p^{sh} due to shadowing or other channel impairments. Moreover, even if the BS assigns a channel for its use, the transmission may fail as the assigned channel may be occupied by an unregistered PU. By considering all these cases, we calculate the probability of white space offloading $p^{ws}_{i,j,k}$ as follows:

$$p^{ws}_{i,j,k} = (\mathrm{e}^{-\lambda^{opp}_{i,j} T_k} - \mathrm{e}^{-\lambda^{ws}_{i,j} T_k})(1 - p^{sh})(1 - p^{un}), \tag{2}$$

where the first term represents the probability that the distributor node is in the white space offloading range but not close enough for opportunistic offloading. Under our assumption that $r^{ws} > r^{opp}$, the contact rates are such that $\lambda^{ws}_{i,j} > \lambda^{opp}_{i,j}, \forall i, j$. Based on that, we calculate $p_{i,j,k}$ which is the probability that n_i gets c_k from n_j in one of the two offloading modes before T_k as:

$$p_{i,j,k} = p^{opp}_{i,j,k} + p^{ws}_{i,j,k}. \tag{3}$$

– *No offloading:* Node n_i receives the content directly from the BS in two cases: (i) it is selected as a distributor (i.e., $n_i \in \mathcal{D}_k$) and gets the content just after the related request, or (ii) it could not receive the content from any of the distributors during time T_k (precisely, the BS serves the content just after T_k). We then express the probability of getting the content from the BS as:

$$p_{i,k} = 1 - \prod_{j \in \mathcal{D}_k} (1 - p_{i,j,k}). \tag{4}$$

We define *offloading gain* for a content item c_k as the traffic saved by offloading which would otherwise be delivered by the BS through the cellular network. Let $\mathbf{Y} = [y_{i,k}]$ denote the subscriber matrix where $y_{i,k} = 1$ indicates that n_i requests c_k. The BS decides which nodes to select as distributors for each item c_k based on $\mathbf{Y}$, $\Lambda = [\lambda_{i,j}]$, p^{sh}, and p^{un}. Let $\mathbf{X} = [x_{j,k}]$ represent the decision variables where $x_{j,k} = 1$ stands for n_j being selected as distributor for c_k. We define the set of subscribers as $\mathcal{S}_k = \{n_i \,|n_i \in \mathcal{N}\}$ and the set of distributors as $\mathcal{D}_k = \{n_j \,|n_j \in \mathcal{S}_k\}$, where $S_k = |\mathcal{S}_k|$ and $D_k = |\mathcal{D}_k|$. Note that $D_k \subseteq S_k, \forall k$.

We can now formulate the offloading gain (i.e., the traffic saved by either opportunistic or white space offloading) maximization problem as:

$$\max_{\mathbf{X}} \sum_{k=1}^{K} l_k \left(S_k - D_k - \sum_{i \in \mathcal{S}_k \setminus \mathcal{D}_k} (1 - p_{i,k}) \right) \tag{5}$$

subject to the following constraints:

$$S_k = \sum_{i=1}^{N} y_{i,k} \qquad \forall k \in \mathcal{C} \tag{6}$$

$$D_k = \sum_{i=1}^{N} x_{i,k} \qquad \forall k \in \mathcal{C} \tag{7}$$

$$p_{i,k} = 1 - \prod_{j \in \mathcal{D}_k} (1 - p^{opp}_{i,j,k} - p^{ws}_{i,j,k}) \quad \forall i \in \mathcal{S}_k, k \in \mathcal{C} \tag{8}$$

$$p^{opp}_{i,j,k} = 1 - \mathrm{e}^{-x_{j,k} \lambda^{opp}_{i,j} T_k} \quad \forall i, j \in \mathcal{N}, k \in \mathcal{C} \tag{9}$$

$$p^{ws}_{i,j,k} = (\mathrm{e}^{-x_{j,k} \lambda^{opp}_{i,j} T_k} - \mathrm{e}^{-x_{j,k} \lambda^{ws}_{i,j} T_k})(1 - p^{sh})(1 - p^{un}) \quad \forall i, j \in \mathcal{N}, k \in \mathcal{C} \tag{10}$$

$$p_{i,k} = 0 \qquad \forall i \in \mathcal{N} \setminus \mathcal{S}_k, k \in \mathcal{C} \tag{11}$$

$$x_{i,k} \leqslant y_{i,k} \qquad \forall i \in \mathcal{N}, k \in \mathcal{C} \tag{12}$$

$$\sum_{k \in \mathcal{C}} D_k \leqslant F \tag{13}$$

$$x_{i,k} \in \{0, 1\} \qquad \forall i \in \mathcal{N}, k \in \mathcal{C}. \tag{14}$$

The number of subscribers and distributors are described by Eqs. (6) and (7), respectively. Distributors can offload content only when in contact with the related subscribers. As contacts are stochastic, each user receives the content with a certain probability from the selected distributors. In detail, the probability that n_i receives content c_k is expressed by Eq. (8), while the probability for each offloading mode is described by Eqs. (9) and (10). The constraint in Eq. (11) ensures that only offloading to the subscribers of a given content is taken into account. The constraint in Eq. (12) guarantees that distributors are selected only from the set of subscribers of that content. As the BS has only F frequencies, Eq. (13) ensures that the number of selected distributors is smaller than or equal to F. Finally, Eq. (14) signifies that the decision variables are binary.

3.1 Heuristics

The optimization problem introduced earlier is a variant of the 0–1 knapsack problem: the total number of frequencies F corresponds to the knapsack capacity while nodes are items to be packed. The utility of each node depends on its capacity to deliver content to the other unselected nodes. Precisely, our optimization formulation is a more general version of the target set selection problem shown to be NP-hard in [4]. As a consequence, we introduce several heuristics – with varying computational complexity – to leverage node and (or) content diversity and obtain a high offloading gain.

Random selection (RAND). The BS randomly selects F nodes. Let $p(n_i)$ denote the probability that n_i is selected as distributor. In this case, $p(n_i)$ is set

Algorithm 1. IBOS

$\mathbb{D} = \emptyset$ and set $p_{j,i,k} = 0$ for all i, j, k
for $n_i \in \mathcal{N}$ **do**
  Get the content id k where $n_i \in \mathcal{S}_k$
  for $n_j \in \mathcal{S}_k$ **do**
    Calculate $p_{j,i,k}$ as in Eq. (3)
for $f = 1$ to F **do**
  $U(n_i) = l_k \sum_j p_{j,i,k}$ for all $n_i, n_j \in \mathcal{N} \setminus \mathbb{D}$
  Select $n_o = \arg\max U(n_i)$ where $n_i \in \mathcal{N} \setminus \mathbb{D}$
  if $U(n_o) > 0$ **then**
    $\mathbb{D} = n_o \cup \mathbb{D}$ and assign f to n_o for content delivery
    Set $p_{o,i,k} = 0$ for all $n_i \in \mathcal{N} \setminus \mathbb{D}$
  else
    return $\mathbb{D}$
return $\mathbb{D}$

to $\min(1, F/N)$. This heuristic has a complexity of $O(F)$ and does not employ content diversity or node diversity. We use RAND for comparison purposes only.

Content diversity (CD). This two-step approach explicitly considers both content size and popularity, different from RAND. In the first step, the BS determines the number of distributors for each content, i.e., $D_k \propto S_k l_k$. After D_k is decided, D_k nodes are randomly selected from $\mathcal{S}_k$ in the second step. This approach ignores the differences among nodes and it does not consider content lifetimes. We calculate $p(n_i)$ for $n_i \in \mathcal{S}_k$ as $p(n_i) = \min(1, D_k/S_k)$, where

$$D_k = F \frac{S_k l_k}{\sum_{m \in \mathcal{C}} S_m l_m}.$$

The complexity of this approach is $O(K)$.

Content and node diversity (CND). CND differs from CD in the second step to better harness the diversity among nodes. For a given c_k, nodes in $\mathcal{S}_k$ are evaluated according to their capacity to offload that content item to the other subscribers. The probability of selection is proportional to the utility of n_i, defined as $U(n_i) = \sum_{j \in \mathcal{S}_k} p_{j,i,k}$. Specifically, the probability $p(n_i)$ is:

$$p(n_i) = \min\left(1, F \frac{S_k l_k}{\sum_{m \in \mathcal{C}} S_m l_m} \frac{U(n_i)}{\sum_{j \in \mathcal{S}_k} U(n_j)}\right).$$

The complexity of this approach is $O(N^2)$.

Iterative Best Offloader Selection (IBOS). This heuristic does not select all F distributors at once, but rather applies the iterative approach detailed in Algorithm 1. Let $\mathbb{D} = \{\mathcal{D}_k\}$ be the set of distributors at the current iteration. First, IBOS initializes $\mathbb{D} = \emptyset$. At each iteration, it calculates the utility of each node as $U(n_i) = l_k \sum_{j \in \mathcal{S}_k} p_{j,i,k}$. After sorting the nodes according to $U(n_i)$, IBOS adds the node with the highest utility, i.e., $n_o = \arg\max U(n_i)$ to the distributors set: $\mathbb{D} = n_o \cup \mathbb{D}$. Next, it sets $p_{o,i,k} = 0$ for all n_i that are candidates to be selected as distributors in the next iteration. This iteration is necessary to better identify the contribution of candidate nodes to offload data to the remaining unselected nodes. IBOS re-calculates $U(n_i)$ according to the updated $p_{j,i,k}$ and follows the same iterations until F number of nodes are selected as distributors or the maximum utility equals to zero. In fact, selecting new nodes as distributors is not expected to increase the offloaded traffic even if there are still some unassigned frequencies. The resulting complexity is $O(FN^2)$.

Improved IBOS (IBOS+). Different from IBOS, this heuristic stores a vector $P(\mathbb{D}) = [p_{i,k}]$ to keep track of each node's probability of receiving the content from the current set of distributors $\mathbb{D}$. Then, after a new node is selected as distributor, it updates $P(\mathbb{D})$ according to Eq. (4) and adds nodes with $p_{i,k}$ higher than some predefined probability (*safety threshold*) to the set of *safe nodes* $\mathbb{A}$. Nodes in $\mathbb{A}$ are then excluded in the calculation of the utility, i.e., in line 7 of Algorithm 1: $n_j \in \mathcal{N} \setminus (\mathbb{D} \cup \mathbb{A})$. Hence, IBOS+ selects nodes that can reach those without a high probability of getting the content from the already selected distributors. The complexity of IBOS+ is the same as that of IBOS, namely, $O(FN^2)$.

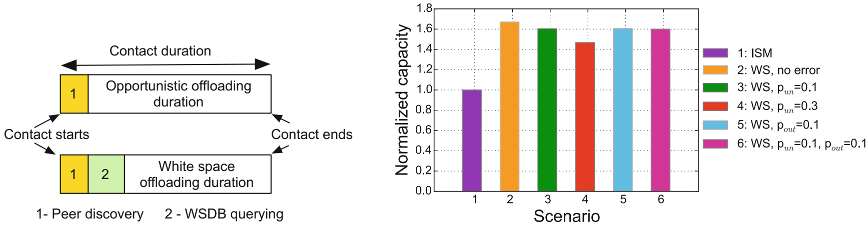
4 Performance Evaluation

We developed a custom simulator in Python to carry out our experiments. We used as input mobility traces generated through the ONE simulator [12], which are based on pedestrian paths extracted from real roads in the city of Helsinki. Pedestrians walk with a speed of $[0.5, 1.5]$ m/s and wait at a reached location for $[1, 4]$ minutes before moving towards their next destination. We recorded the contacts among $N = 200$ pedestrians using two transmission ranges $r^{opp} = 20$ m and $r^{ws} = 100$ m, according to the relatively higher range of white spaces [7]. Node contacts (e.g., their start and end times) lasted for three hours. We derived the average pairwise contact rate $\lambda_{i,j}$ from the trace and assumed that the related information is available to the BS for CND, IBOS, and IBOS+.

We drew content popularity from a Weibull distribution with parameters $k = 0.513$ and $\lambda = 6010$ according to [13]. We considered content sizes and lifetimes uniformly distributed between $[2, 5]$ MB and $[1, 3]$ hours, respectively.

4.1 Offloading Capacity

In our first set of experiments, we study the impact of white space availability on the offloading capacity irrespective of the offloading algorithm employed. To this end, we define *effective offloading capacity* as the product between the channel bandwidth and the time during which a node can transmit to its peer. As Fig. 2a shows, this period is equal to the remaining time after peer discovery and before the contact ends for opportunistic offloading. In case of white spaces, offloading starts only after the WSDB returns the data and it is successful only



(a) (b)

Fig. 2. (a) Offloading duration and (b) effective offloading capacity.

if the transmission does not collide with an unregistered (active) PU. As peer discovery is common to both modes, we ignore the related time loss in the results.

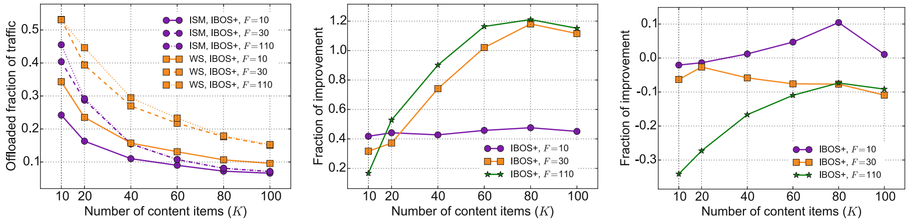
To obtain a realistic estimate for WSDB query overhead, we developed a mobile application that connects to the Google Spectrum Database to query the spectrum availability at some predetermined (urban as well as rural) locations in the US. We recorded the round trip time of the WSDB requests/responses and subtracted the related values from the contact duration to derive the effective offloading capacity in the white spaces. According to our observations from 10,000 queries, 90 % of delays are below 4.2 s. This is consistent with the delays below 3 s reported in [14], especially considering that our queries were run from a mobile device located in Helsinki. Regarding contact durations, we used the contact trace generated by the ONE simulator. At the beginning of each contact, we simulated the chance of outage as well as that of PU collision.

As Fig. 2b shows, exploiting white spaces in addition to ISM channels significantly improves the effective offloading capacity: 67 % when there is no outage and the WSDB information is reliable; 47–60% under different levels of unreliability, i.e., the outages and PU collision probabilities associated with the scenarios 3–6 reported in the figure.

4.2 Content Offloading

We now investigate the impact of the number of content items K on the offloaded traffic for different values of the frequency bands F available at the BS (Fig. 3). For clarity, we only describe[2] the results obtained for IBOS+ when $p^{sh} = p^{un} = 0.1$. Figure 3a shows that when K increases, the offloaded traffic decreases for all schemes. For a low number of contents, opportunistic offloading is very efficient. For instance, if there is a single item, ten nodes ($F = 10$) deliver 86 % of the traffic through opportunistic offloading. For white spaces, the offloaded traffic reaches 92 %. For $K = 100$, ISM-only mode can offload only 6 % of the traffic, as opposed to the 11 % value obtained when white spaces are employed for offloading. This behavior is due to the increasing content diversity. In other words, the probability that two random nodes subscribe to the same content is

[2] A detailed comparison of the different heuristics is provided in the next subsection.



(a) (b) (c)

Fig. 3. (a) Offloaded traffic fraction, improvement in (b) offloaded traffic and (c) mean content delivery delay of white spaces compared to ISM.

lower when K increases. Consequently, the chance of offloading decreases too. Next, we observe better performance under white space offloading. Figures 3b and c illustrate the related fraction of improvement in performance compared to ISM-only offloading. Considering all three sub-plots, we can see that for low K, white spaces approximately allow a 20–40 % improvement in offloaded traffic as opportunistic offloading finds sufficiently many contacts for the distribution of say $K = 10$ contents. However, Fig. 3c shows that white spaces speed up the delivery very significantly in this operating region. With increasing content diversity, the benefit of white spaces becomes more apparent. For instance, the relative improvement is above 100 % for $K = 100$ and $F = 30$. For $F = 10$, the improvement is around 45 % due to the low number of distributors.

In summary, white spaces provide the most significant gains when: there are many diverse content items and sufficient frequencies; content items are less diverse but the number of frequencies available at the BS is limited, i.e., only a small fraction of the nodes can be selected as distributors. The improvement lies in the offloaded traffic in the first case, whereas it consists of reduced delivery delay in the latter case.

4.3 Comparison of Heuristics

We finally evaluate and compare our heuristics as a function of the fraction of distributors (Fig. 4). We define the fraction of distributors as the ratio of number of distributors that the BS can select to the total number of subscribers. Note that the BS may select less nodes than the allowed fraction for IBOS and IBOS+, as these algorithms stop when the remaining nodes are not expected to further increase the offloaded traffic. The number of selected nodes is indeed equal to the number of available frequencies for the rest of the algorithms.

First, we note that with increasing F, all schemes can initially offload more traffic by employing more distributors. However, after a certain number of frequencies is reached (e.g., $F = 50$ corresponding to 0.25 fraction of distributors in Fig. 4a), CD, CND, and RAND redundantly select nodes as distributors. Since IBOS and IBOS+ stop allocating distributors when the maximum utility of a

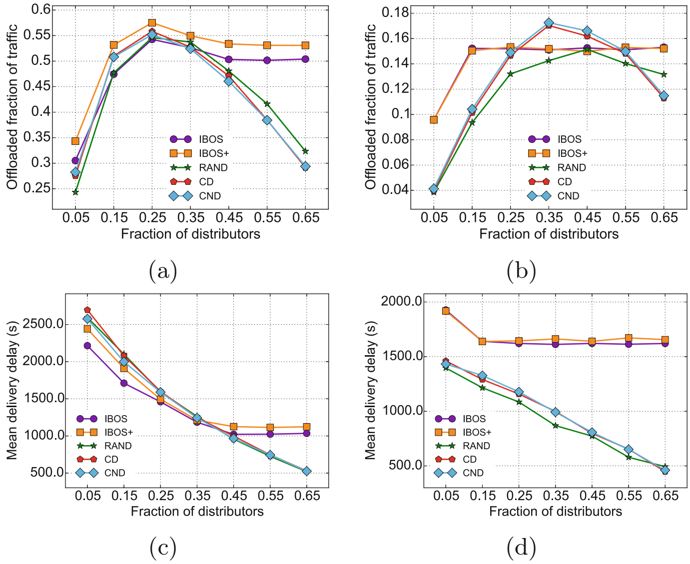


Fig. 4. Offloaded traffic for (a) 10 and (b) 100 content items. Delivery delay for (c) 10 and (d) 100 content items.

selected node is zero, these schemes can still maintain a high offloading performance. Second, IBOS+ outperforms others for $K = 10$ while IBOS is the second best scheme only when F is either low or high. In between, other naive schemes have higher offloading capacity. We expect a typical setting to have low to moderate F; hence, the best scheme is IBOS+ in these conditions. However, CD is a sensible choice when contact statistics are inaccurate or not even available.

In contrast to Fig. 4a, b shows that there is almost no difference in the performance of IBOS and IBOS+. This result can be explained through the impact of the safety threshold in IBOS+ (set to 0.9 in our experiments). As content diversity is higher in this scenario (i.e., $K = 100$), the probability that a random contact results in content offloading is much lower too. Hence, IBOS+ cannot add nodes to the safety set, which effectively reduces IBOS+ to IBOS. Moreover, our model assumes exponentially distributed inter-contact times and calculates the expected offloading probabilities based on such a model. However, the contact traces do not necessarily exhibit this property.

Figures 4c and d illustrate the mean content delivery delay for the nodes that receive the content before the deadline, i.e., the distributors and the nodes that are served by these distributors. IBOS and IBOS+ not only improve offloading capacity but also help faster content delivery for $K = 10$ and low F. IBOS outperforms IBOS+, which is followed by naive schemes until $F = 90$. After this point, CD, CND and even RAND perform better as they assign all frequencies to the distributors without considering the offloading capacity. For the same reason, Fig. 4d shows that CD, CND and RAND obtain lower delays while the delivery delay of IBOS and IBOS+ remains almost the same with increasing F. This could be seen as a trade-off between offloaded traffic fraction and delivery delay. However, we conclude that IBOS and IBOS+ are the best choices when timely delivery is needed to guarantee user satisfaction.

5 Conclusion

In this article, we leverage white spaces in addition to ISM bands for mobile content offloading. We specifically model database-assisted white space networks in which WSDB data may not be accurate. We then devise a general framework for content offloading through a content-centric approach. Specifically, we formulate an optimization problem to maximize the offloading gain through the selection of distributors. We show that solving such a problem is computationally hard, then propose several practical heuristics and analyze their offloading performance. Our results demonstrate that the availability of white space significantly increases the offloaded traffic, especially when there are many content items. When there are a few items, mobile opportunistic offloading provides a high gain, comparable to that for white spaces. In this case, white space offloading enables faster content delivery. As a future work, we seek to find an approximate solution to our optimization problem. We also plan to better analyze the impact of content and mobility characteristics on the offloading performance.

Acknowledgments. We thank Aleksandr Zavodovski for developing the mobile application for querying the WSDB. This work has been supported by the Academy of Finland under grant numbers 284806, 284807 and 278207.

References

1. Aijaz, A., Aghvami, H., Amani, M.: A survey on mobile data offloading: technical and business perspectives. IEEE Wirel. Commun. **20**(2), 104–112 (2013)
2. Lee, K., Lee, J., Yi, Y., Rhee, I., Chong, S.: Mobile data offloading: how much can WiFi deliver?. In: ACM CoNEXT (2010)
3. Golrezaei, N., Molisch, A.F., Dimakis, A.G., Caire, G.: Femtocaching and D2D collaboration: a new architecture for wireless video distribution. IEEE Commun. Mag. **51**(4), 142–149 (2013)
4. Han, B., et al.: Mobile data offloading through opportunistic communications and social participation. IEEE Trans. Mob. Comput. **11**(5), 821–834 (2012)
5. Li, Y., et al.: Multiple mobile data offloading through DTNs. IEEE Trans. Mob. Comput. **13**(7), 1579–1596 (2014)
6. Mancuso, A., Probasco, S., Patil, B.: Protocol to access white-space (PAWS) databases: Use cases and requirements. IETF, RFC 6953, May 2013
7. Cui, P., Tonnemacher, M., Rajan, D., Camp, J.: WhiteCell: energy-efficient use of unlicensed frequency bands for cellular offloading. In: IEEE DySPAN (2015)
8. Ding, G., Wang, J., Wu, Q., Yao, Y.D., Song, F., Tsiftsis, T.A.: Cellular-base-station assisted D2D communications in TV white space. IEEE J. Sel. Areas Commun. **34**(1), 107–121 (2016)

9. National Association of Broadcasters, Emergency motion for suspension of operations and petition for rulemaking, March 2015
10. Khoshkholgh, M.G.: Connectivity of cognitive D2D communications underlying cellular networks. IEEE JSAC **33**(1), 81–98 (2015)
11. Chen, V., Das, S., Zhu, L., Malyar, J., McCann, P.: Protocol to access whitespace (PAWS) databases, RFC 7545, May 2015
12. Keränen, A., Ott, J., Kärkkäinen, T.: The ONE simulator for DTN protocol evaluation. In: Simutools (2009)
13. Cheng, X., Dale, C., Liu, J.: Statistics and social network of YouTube videos. In: International Workshop on Quality of Service (IWQoS) (2008)
14. Majid, A.Y.: Better mobility support for radio spectrum white space-enabled devices, in MS Thesis, Delft University of Technology (2015)