Impact of overlapping in the radio coverage areas of multiple Wi-Fi access points on detecting encounters

Karim Keramat Jahromi, Md Aktaruzzaman & Mehdi Jalili

Published online: 24 Jan 2019.

Submit your article to this journal ⎘

Article views: 72

View Crossmark data ⎘

Taylor & Francis
Taylor & Francis Group

🔓 OPEN ACCESS | Check for updates

Impact of overlapping in the radio coverage areas of multiple Wi-Fi access points on detecting encounters

Karim Keramat Jahromi[a], Md Aktaruzzaman [b] and Mehdi Jalili[c]

[a]Computer Science Department, University of Milan, Milan, Italy; [b]Department of Computer Science and Engineering, Islamic University, Kushtia, Bangladesh; [c]Department of Computer Engineering, Soufian Branch, Islamic Azad University, Soufian, Iran

ABSTRACT

Understanding the potential role of Smartphones and other portable wireless devices as relay nodes in message dissemination and content delivery in Delay Tolerant and Opportunistic networks depend on the knowledge about patterns and the number of encounter events among mobile nodes. One of the main challenges for extracting encounters is overlapping in the radio coverage areas of nearby access points (APs). Data about the usage of Wi-Fi networks can be used to perform an analysis of encounters among mobile devices. A realistic estimation of the number of encounters among mobile nodes is now a big challenge. In this paper, the effects of overlapping of radio coverage area among multiple APs for extracting realistic encounters among mobile devices has been discussed, and also an analytical approach has been proposed for extracting realistic encounters from overlapping in the coverage areas of multiple nearby APs. A significant difference was observed between the number of encounters by considering and ignoring overlapping. Our study finds that Wi-Fi datasets are not reliable source to estimate the number of encounters when there are overlapping in radio coverage areas of multiple APs.

1. Introduction

Human mobility studying (Karamshuk, Boldrini, Conti, & Passarella, 2011) has become the topic of interest among the wide group of researchers in the last decade, mainly due to the its impact on so many different theoretical and applicative researches (Gonzalez, Hidaglo, & Barabasi, 2008; Keramat Jahromi Zignani, Gaito, & Rossi, 2016; Papandrea et al., 2016; Pirozmand, Wu, Jedari, & Xia, 2014; Riascos & Mateos, 2017). Wireless devices such as smartphones, tablets, and other Wi-Fi enabled devices are mainly carried by humans, thus exhibiting movement patterns that are related to the human mobility patterns and behaviours. Such mobility patterns affect the operation and performance of wireless networks. On the other hand, mobility and nodal

---

CONTACT Karim Keramat Jahromi ✉ karim.karamat@gmail.com 🖂 Computer Science Department, University of Milan, Milan, Italy

encounters are utilized for routing and forwarding data to deliver messages in intermittently connected delay tolerant (DTN) and opportunistic networks (Barbosaa et al., 2018; Cuttone, Lehmann, & Gonzalez, 2018; Hui, Crowcroft, & Yoneki, 2011; Karamshuk et al., 2011; Lindgren, Doria, & Schelén, 2004; Onghena & Milano, 2015; Pirozmand et al., 2014). Encounter event means meeting face-to-face which implies physical proximity among involving objects. In communication perspective, an encounter among mobile devices occurs when they are in communication range of each other or within the coverage area of WLAN infrastructures that devices are associated with (it will be discussed with more details in Section 3). Understanding the nodal encounter patterns is a critical basis for the success of protocols and deployment in opportunistic networks because delivery mechanism depends on the nodal encounters. Exploiting the nodes encounter patterns could be used to design better protocols or applications in future. In order to understand human mobility patterns, we need to observe human mobility. In situations where continuous and direct observation of human mobility and encounter patterns is difficult or even against privacy rights, knowledge about the usage of Wi-Fi networks can be used to perform an analysis of encounters among mobile nodes.

Usually Wi-Fi datasets which are collected by access points (APs) for mobility analysing, consisting of the log of association and dis-association of smartphones, tablets, and other wireless devices with the APs, including Mac address of associated AP and mobile devices and time stamp of association and dis-association. These Wi-Fi datasets mostly don't include any information about radio signal (finger print) and geolocation (GPS) coordinates of APs. Some researchers in (Hsu & Helmy, 2010; Moon & Helmy, 2010; Wang, Nascimento, & MacGregor, 2012) have discussed encounter events among mobile nodes and some of their distributions, and they also observed some regularities. However, these works have ignored some issues for extracting realistic encounters such as ping-pong events in Wi-Fi datasets and overlapping in radio coverage areas among nearby access points. Ignoring these issues might lead to an unrealistic encounter (Keramat Jahromi, Meneses, & Moreira, 2014). In addition to this, the real number of encounters extracted from Wi-Fi datasets might be underestimated, first due to the overlapping in the radio coverage areas of APs and secondly, APs deployment in the real world tend to concentrate covering a small fraction of a region such as a campus or part of a city. Hence, within this limited areas, some of the encounters between devices occur out of APs coverage areas. It means in some situations there are realistic encounters but not detectable through the Wi-Fi datasets. Despite these clear limitations, if collected datasets of APs are used carefully (i.e, accounting for the effects of ping-pong events and missed encounters), it would appear to be a good source of empirically derived data on human encounters, since a large amount of data can be gathered easily at low cost, allowing longitudinal comparisons of encounter patterns.

The main contribution of this paper is to propose an approach for calculating the number of encounters in the situation where there are overlapping in the coverage areas of multiple nearby APs. Also we will perform a controlled experiment involving smartphones and laptops to assess how Wi-Fi logs are reliable for extracting real encounters. In this study, we address the main challenges in extracting realistic encounter events, and we propose an estimate of the number of pair encounters analytically.

2. Related work

Wireless mobile devices are becoming more and more ubiquitous and popular. Analysing wireless networks formed over these devices is becoming an important field of research. Since encounter events among mobile nodes provide communication opportunities in DTN and opportunistic networks, knowledge about node encounter events, extracted from datasets is important for designing DTNs and opportunistic networks, and also understanding the diffusion of data in these networks (Hsu & Helmy, 2010; Karamshuk et al., 2011; Pirozmand et al., 2014). The number and patterns of encounters among mobile nodes have a vital role in communication and data exchange in infrastructure-less networks. The majority of works on empirical analysis of WLANs are focused on the extracting behavioural patterns of individual users. Although the understanding of individual behaviour is important, it does not reveal how mobile nodes interact with each other and how information can be diffused through relay nodes. Hsu and Helmy (2010) studied the encounters between mobile nodes and introduced the small world approach to explain the encounters relationship graph. They empirically analysed multiple WLAN datasets collected in universities campus environments and extracted some distributions about interactions among mobile nodes and also investigated information diffusion through encounters. In Moon and Helmy (2010), Wang et al. proposed a generic methodology to extract encounter patterns through an auto persistence function, and they also investigated whether the network formed by periodic encounters has a small world structure to provide communications to a large-scale network. In Keramat Jahromi, Zignani, Gaito, and Rossi (2017) authors propose an encounter and colocation predictive model which could improve the prediction of user's encounter /colocation events and their features by exploiting the spatio-temporal regularity in the history of these events through applying weighted features Bayesian predictor. To the best of our knowledge, the process and issues of extracting encounters have been considered only in the literature, for instance, overlapping in the radio coverage areas of APs. Ignoring such kinds of important issues Keramat Jahromi et al. (2014), in some cases might lead to errors in extracting realistic encounters, and even significant underestimation of the real number of encounters that might cause improper prediction about distribution and speed of diffusion (SOD) of data Nguyen, Senac, and Diaz (2012) in the network. In our previous study Keramat Jahromi et al. (2014), we discussed the issue of ping-pong events, and proposed an algorithm for smoothing and extracting encounters from Wi-Fi datasets. Also the impact of this issue on connectivity properties of node encounters have been discussed. Authors in Mitchell et al. (2012) have reported the extraction of direct Bluetooth encounters through Bluetooth scanners. However, they ignored the overlapping of coverage areas of the nearby Bluetooth scanners, and they observed that just 51% of the extracted encounters were corresponding to the actual encounters. They also observed that statistical properties of scanned encounters differ from statistics of actual encounters. Authors in Vanderhulst, Mashhadi, Dashti, and Kawsar (2015) have proposed a framework for detecting human spontaneous encounter with social interaction. In fact, they were interested in detecting those social interactions that are short-lived in nature and spontaneous between a small set of individuals, through leveraging existing Wi-Fi infrastructure and radio Wi-Fi signal radiating from mobile devices ignoring the overlapping in the coverage areas of nearby APs. In our previous work Keramat Jahromi et al. (2015), the impact of overlapping

**Table 1.** Description of the notations used for estimation of encounters.

Symbol	Descriptions
R_i	Radius coverage area of AP_i.
d_{ij}	Euclidean distance between AP_i and AP_j in 2-D plane.
d_{th}	Distance threshold.
S_{ij}	Overlap area between AP_i and AP_j.
S_{ijk}	Common intersection area among AP_i, AP_j and AP_k.
γ	Ratio of overlapping region.
T_l	Snapshot time duration.
$n_i(T_l)$	The number of mobile nodes associated with AP_i during snapshot T_l.
$n_{ix}(T_l)$	The number of mobile nodes that are only in the coverage area of AP_i and outside of the intersection area.
$n_{ij}(T_l)$	The number of mobile nodes in intersection coverage area of AP_i and AP_j during the snapshot T_l.
$n_{ijk}(T_l)$	The number of mobile nodes in the intersection area common among AP_i, AP_j and AP_k during the snapshot T_l.
$\phi_i(T_l)$	Densities of nodes associated with AP_i during snapshot T_l.
$n_i'(T_l)$	The total number of mobile nodes in the coverage area of AP_i during the snapshot T_l.
N_p	The number of nodes(Smartphones and Laptops) participated in experiment.
MN_{pair}	The maximum number of possible pair encounters in real situation.
N_{pair}	The number of distinct encounters extracted from the Wi-Fi logs.
MNB_{pair}	The maximum number of possible pair encounters extracted from log of Bluetooth.
NB_{pair}	The number of encounters extracted from logs of Bluetooth.
$R_W = \frac{N_{pair}}{MN_{pair}};$	The Ratio of the number of extracted encounters through Wi-Fi log over the maximum possible number of encounters in the real situation.
$R_B = \frac{NB_{pair}}{MNB_{pair}};$	The Ratio of the number of extracted encounters through Bluetooth log rather the maximum possible number of extracted encounters from Bluetooth.

on extracting encounters has been studied just by considering the simple case of overlapping between two nearby APs. In this work as the main contribution, we extended our previous work Keramat Jahromi, Meneses, and Moreira (2015) to generalize the estimation of the number of encounters by considering the overlapping of coverage areas of multiple APs. All notations used for estimation of the number of encounters have been described in Table 1.

3. Encounter event

Nowadays, the majority of short-range wireless devices such as smartphones, tablets are carried by humans. These devices can be used to observe human mobility behaviours and also to extract information about the physical proximity among them in the real world and as a result encounter events among humans. Therefore, in communication perspective, an encounter among mobile devices occurs when they are in communication range of each other or within the coverage area of WLAN infrastructures that devices are associated with.

In wireless networks two kinds of encounters are defined: (i) direct and (ii) indirect (Hsu & Helmy, 2010). In case of direct encounter, an encounter occurs when devices come within the radio communication of each other. For example, Bluetooth datasets track direct encounters between mobile nodes. WLAN datasets record associations between mobile nodes and APs. Even though APs are stationary, they can link mobile nodes that never directly encountered. Therefore, in this case, we have indirect encounters since communication opportunity between nodes is established through APs. Most researchers (Hsu & Helmy, 2010; Moon & Helmy, 2010; Wang et al., 2012) define an encounter event occurred in a WLAN when two or more nodes are associated to the same AP during an overlapping time interval and classify this as an indirect encounter type. Anyway, the mentioned definition or condition for the occurrence of encounters in WLAN is a sufficient
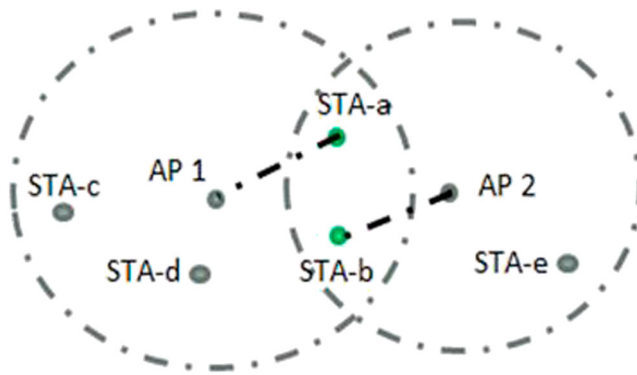
Figure 1. Overlapping between coverage areas of the neighbourhood APs.

condition, not a necessary condition. It means that if two or more mobile nodes are associated with the same AP during an overlapping time interval, they are considered encounters, but also some other encounters might be exist, even if they are not associated to the same AP. There is some possibility that mobile nodes be in the physical proximity of each other, for instance be in the overlapping coverage areas of nearby APs. As shown in Figure 1, although **STA-a** and **STA-b** are in physical proximity of each other (**STA-** is a mobile node), they are associated with different APs. It means that an encounter event is occurring in the real world, but it is not straightforward to be detected from the Wi-Fi dataset. So, due to this overlapping issue, sometimes the estimated number of the encounters would be underestimated.

Other issue related to the extracting the real encounters from Wi-Fi datasets is ping-pong events (Henderson, Kotz, & Abyzov, 2004; Kim & Kotz, 2005) that mobile devices change their associations from one AP to another frequently while nodes are stationary has been discussed in detail in Keramat Jahromi et al. (2014), and an algorithm for smoothing ping-pong events was proposed. The low aggressive behaviour of mobile devices in changing their associations with APs is another issue that leads to errors in estimating their location and in extracting real encounters from Wi-Fi datasets. This issue is mostly related to the hand off algorithm, called relative signal strength with hysteresis and thresholds, used by many manufacturers to define the conditions that trigger hand offs (González, Pérez, & Zárate, 2005).

4. Overlapping coverage area

In the previous section, we mentioned that ping-pong events and overlapping coverage areas among nearby APs are the main issues in extracting node encounters properly. After detecting and smoothing ping-pong events in Wi-Fi datasets (more details can be found in Keramat Jahromi et al., 2014), the changing in APs associations become mostly dependent on the actual movement of mobile nodes. We can now consider the issue of overlapping coverage areas among nearby APs for extracting real encounters. If data about the spatial distribution of APs (at least distances between APs) and also the estimation of their coverage ranges are available, it is possible to roughly specify which pairs of APs have been overlapped. With this knowledge, we can go one step ahead and update the definition

of an encounter in a WLAN: an encounter event between two mobile nodes occurs when they are associated with the same AP, or with the two nearby APs with overlapping coverage areas during an overlapping time interval. We use this modified definition for extracting encounters from Wi-Fi dataset after smoothing ping-pong events.

Wi-Fi datasets usually just contain information about association and disassociation of mobile nodes with APs and do not include any information about positions of APs or coordinates of mobile nodes within the coverage area of associated APs. Therefore, relying on Wi-Fi datasets, we do not know among the mobile nodes which are associated with an AP, which of them are just within the coverage area of that AP, and which of them are within the overlapping coverage areas with other nearby APs. It is obvious that the spatial distribution of APs and mobility patterns of mobile nodes have main roles in this regard. In the next subsection, we propose two approaches for calculating the number of encounters for two cases of overlapping radio coverage areas of pair of APs and multiple APs.

4.1. Encounters in the overlapped coverage area of two access points

In real situation, most of the times, there are overlapping in the coverage areas of APs. Here we consider the coverage areas of APs as disks in 2-D space and also R_i and R_j as estimated radius of the coverage areas of AP_i and AP_j, respectively. If d_{ij} denote the Euclidean distance CCI (2018) between AP_i and AP_j, and the condition $R_i + R_j > d_{ij}$ is held, then AP_i and AP_j have overlapping in their coverage areas. The overlapping coefficient (γ) between two nearby APs is defined as ratio of overlapping region over the total area covered by two APs:

$$\gamma = \frac{S_{ij}}{\pi \times (R_i^2 + R_j^2) - S_{ij}} \tag{1}$$

Thus, the overlapping coefficient of two nearby APs depends on the overlapping area (S_{ij}) and the radius of coverage areas of APs.

The overlapping area (S_{ij}) between AP_i and AP_j can be calculated according to CCI (2018) as:

$$S_{ij} = R_j^2 \arccos\left(\frac{d_{ij}^2 + R_j^2 - R_i^2}{2R_j d_{ij}}\right) + R_i^2 \arccos\left(\frac{d_{ij}^2 + R_i^2 - R_j^2}{2R_i d_{ij}}\right)$$
$$- 1/2 \times \left(\sqrt{(-d_{ij} + R_j + R_i) \times (d_{ij} - R_j + R_i) \times (d_{ij} + R_j + R_i) \times (d_{ij} + R_j - R_i)}\right); \tag{2}$$

The overlapping area depends on both the radius of coverage areas and the Euclidean distance () between APs (being Null if $R_i + R_j <= d_{ij}$ and maximum if $d_{ij} = 0$). If distance (d_{ij}) between AP_i and AP_j is less than a predefined distance threshold (d_{th}) i.e, $d_{ij} < d_{th}$, where the condition $d_{th} \ll R_i + R_j$ is held and hence the overlapping area tends to be so large and $\gamma \longrightarrow 1$. Under this situation the two APs would be assumed to be merged as an unique AP, and all mobile nodes which were associated with AP_i and AP_j will be associated with the merged AP. Under this situation, the number of encounters in the coverage area of merged AP during time snapshot T_l can be represented by

$$En(T_l) = \frac{[n_i(T_l) + n_j(T_l)] \times [n_i(T_l) + n_j(T_l) - 1]}{2} \tag{3}$$

where, $n_i(T_l)$ and $n_j(T_l)$, respectively are the number of associated nodes with AP_i and AP_j in time snapshot T_l. This approach is more applicable and reliable where $d_{th} \ll R_i + R_j$ which implies high overlapping in the coverage areas of APs ($\gamma \longrightarrow 1$).

On the other hand, if $\gamma \longrightarrow 0$, means that the intersection area between two APs is negligible *i.e.*, this case can be considered as non-overlapping, and hence the number of encounters is calculated by

$$En(T_l) = \frac{n_i(T_l) \times (n_i(T_l) - 1)}{2} + \frac{n_i(T_l) \times (n_i(T_l) - 1)}{2} \tag{4}$$

4.2. Encounters in the overlapped coverage areas of multiple access points

If $n_i(T_l)$ denotes the number of mobile nodes associated with AP_i during the time snapshot T_l; and $n_{ix}(T_l)$ as a portion of $n_i(T_l)$ mobile nodes which are just in the coverage area of AP_i (in non overlapping area), then the portion of $n_i(T_l)$ mobile nodes which are in the intersection areas of nearby APs during snapshot T_l is represented by $n_i(T_l) - n_{ix}(T_l)$.

The number of possible combinations of mobile nodes which are associated with AP_i and also remain in the intersection area can be represented by $\begin{pmatrix} \boldsymbol{n_i(T_l)} \\ n_i(T_l) - n_{ix}(T_l) \end{pmatrix}$.

If the coverage area of just two nearby APs are overlapping then the total number of choices for selecting

$$n_{ij}(T_l) = n_i(T_l) + n_j(T_l) - (n_{ix}(T_l) - n_{jx}(T_l)) \tag{5}$$

nodes in the intersection coverage areas of two nearby access points AP_i and AP_j are given by

$$\begin{pmatrix} n_i(T_l) \\ n_i(T_l) - n_{ix}(T_l) \end{pmatrix} \times \begin{pmatrix} n_j(T_l) \\ n_j(T_l) - n_{jx}(T_l) \end{pmatrix} \tag{6}$$

So, there would be a large number of different combinations for occurrence of pair encounters.

For simplicity, and to avoid the complexity of different combinations of encounters due to movement and commuting of mobile nodes between intersection and non-intersection areas, we exploit the granularity time interval (snapshot). It means that we compare the number of encounters in different snapshots. So the total duration of the collection of datasets will be divided into non-overlapping time intervals with specified granularity (T), then mobile nodes will be considered as stationary during each time interval (snapshot duration).

Considering the coverage areas of APs as a disk in 2-D space and also exploiting GRAMM (Mitsche, Resta, & Santi, 2014) or Spatial-RWP (Mitsche et al., 2014) mobility model for movement of mobile nodes, then the mobile nodes that are associated with the AP during each snapshot, according to the (Chen, Wu, & Ke, 2013; Mitsche et al., 2014), can be considered that spatially distributed uniformly within the AP coverage area as disk point picking (DPP, 2018).

If $\phi_i(T_l)$ denotes the density of nodes associated with AP_i within its coverage area during the snapshot T_l, then we can write

$$\phi_i(T_l) \simeq \frac{n_i(T_l)}{\pi \times R_i^2} \tag{7}$$

The number of mobile nodes that are in the coverage area of AP_i but outside of the intersection area, can be expressed as:

$$n_{ix}(T_l) = Round[\phi_i(T_l) \times (\pi \times R_i^2 - S_{ij})] \qquad (8)$$

where S_{ij} is the intersection area between AP_i and AP_j. So, the number of nodes in the intersection of coverage areas is given by

$$n_{ij}(T_l) = n_i(T_l) + n_j(T_l) - [n_{ix}(T_l) + n_{jx}(T_l)] \qquad (9)$$

Then, the number of encounters in snapshot T_l can be analytically calculated by

$$En(T_l) = \frac{(n_{jx}(T_l) + n_{ij}(T_l)) \times (n_{jx}(T_l) + n_{ij}(T_l) - 1)}{2} +$$
$$\frac{(n_{ix}(T_l) + n_{ij}(T_l)) \times (n_{ix}(T_l) + n_{ij}(T_l) - 1)}{2} - \frac{n_{ij}(T_l) \times (n_{ij}(T_l) - 1)}{2} \qquad (10)$$

where $(n_{ix}(T_l) + n_{ij}(T_l))$ and $(n_{jx}(T_l) + n_{ij}(T_l))$ are the total number of mobile nodes in the coverage areas of AP_i and AP_j, respectively.

It should be highlighted that the above calculated number of encounters is relying on this assumption that during each time snapshot, mobile nodes are stationary and don't change their positions within the coverage area of APs. This is obvious that snapshots with short time duration lead to more accurate and realistic results compare to longer ones, since the probability that mobile nodes change their association during a short snapshot is less than the long ones. However a short snapshot imposes more processing load on the processor. So, choosing time granularity of snapshot is very important for the trade-off between accuracy and processing load.

Apart from the above-mentioned, where we considered the case of just overlapping between two nearby APs, in the real situation several APs may be overlapped in their coverage areas, and in this situation we have to consider the intersection areas between several (three or more) nearby pairs of APs to count the number of encounters. Depending on the distances and radius of coverage areas of APs, very different taxonomy about overlapping intersection areas may be appeared (Fewell, 2006) that should be considered in counting the encounters.

Considering the case of only three APs, where addition to the intersection areas between each pair of nearby APs (S_{ij}, S_{ik} and S_{jk}), there is a common intersection area for the three APs; (S_{ijk}). Then the overlapping coefficient for this case with three APs (AP_i, AP_j and AP_k) is defined as

$$\gamma = \frac{(S_{ij} + S_{ik} + S_{jk}) - S_{ijk}}{\pi \times (R_i^2 + R_j^2 + R_k^2) - (S_{ij} + S_{ik} + S_{jk}) + S_{ijk}} \qquad (11)$$

The authors in Fewell (2006) have proposed an algorithm for calculating the intersection area of three circles (S_{ijk}). The number of nodes that are only in the coverage area of AP_i, AP_j and AP_k (outside of intersection areas) is given by the following equations:

$$n_{ix}(T_l) = n_i(T_l) - Round[(S_{ij} + S_{ik} - S_{ijk}) \times \phi_i(T_l)] \qquad (12)$$

$$n_{jx}(T_l) = n_j(T_l) - Round[(S_{ij} + S_{jk} - S_{ijk}) \times \phi_j(T_l)] \qquad (13)$$

And

$$n_{kx}(T_l) = n_k(T_l) - Round[(S_{ik} + S_{jk} - S_{ijk}) \times \phi_k(T_l)]$$ (14)

Also the number of nodes in the intersection areas between each pair of APs; AP_i, AP_j and AP_k are given respectively by

$$n_{ij}(T_l) = Round[(\phi_i(T_l) + \phi_j(T_l)) \times S_{ij} + \phi_k(T_l) \times S_{ijk}]$$ (15)

$$n_{ik}(T_l) = Round[(\phi_i(T_l) + \phi_k(T_l)) \times S_{ik} + \phi_j(T_l) \times S_{ijk}]$$ (16)

And

$$n_{jk}(T_l) = Round[(\phi_j(T_l) + \phi_k(T_l)) \times S_{jk} + \phi_i(T_l) \times S_{ijk}]$$ (17)

where $\phi_i(T_l)$ is the density of mobile nodes in the coverage area of AP_i. Also the number of nodes in the common intersection area among three APs coverage areas is equal to:

$$n_{ijk}(T_l) = Round[(\phi_i(T_l) + \phi_j(T_l) + \phi_k(T_l)) \times S_{ijk}]$$ (18)

Then the total number of nodes in coverage areas of AP_i, AP_j and AP_k are respectively equal to:

$$n_i'(T_l) = [n_{ix}(T_l) + n_{ij}(T_l) + n_{ik}(T_l) - n_{ijk}(T_l)]$$ (19)

$$n_j'(T_l) = [n_{jx}(T_l) + n_{ij}(T_l) + n_{jk}(T_l) - n_{ijk}(T_l)]$$ (20)

And

$$n_k'(T_l) = [n_{kx}(T_l) + n_{ik}(T_l) + n_{jk}(T_l) - n_{ijk}(T_l)],$$ (21)

where $n_{ij}(T_l)$ and $n_{ijk}(T_l)$ are the number of nodes in intersection area of AP_i, AP_j and AP_i, AP_j, AP_k, respectively, in snapshot T_l. Then the total number of encounters extracted in snapshot T_l in coverage areas of these three APs, is:

$$En(T_l) = n_i'(T_l) \times \frac{(n_i'(T_l) - 1)}{2} + n_j'(T_l) \times \frac{(n_j'(T_l) - 1)}{2} +$$

$$n_k'(T_l) \times \frac{(n_k'(T_l) - 1)}{2} - n_{ij}(T_l) \times \frac{(n_{ij}(T_l) - 1)}{2} - n_{ik}(T_l) \times \frac{(n_{ik}(T_l) - 1)}{2} -$$ (22)

$$n_{jk}(T_l) \times \frac{(n_{jk}(T_l) - 1)}{2} + n_{ijk}(T_l) \times \frac{(n_{ijk}(T_l) - 1)}{2};$$

Thus, if there are N (where $N \geq 3$) APs, with pair by pair intersections in their coverage areas, the overlapping coefficient and also the total number of encounters in their coverage areas are given by:

$$\gamma = \frac{(\sum_{i \neq j;i=1}^{\binom{n}{2}} S_{ij} - \sum_{i \neq j \neq k;i=1}^{\binom{n}{3}} S_{ijk} + \sum_{i \neq j \neq k \neq m;i=1}^{\binom{n}{4}} S_{ijkm} - \cdots)}{\pi \times (\sum_{i=1}^{\binom{n}{2}} R_i^2) - \sum_{i \neq j;i=1}^{\binom{n}{2}} S_{ij} + \sum_{i \neq j \neq k;i=1}^{\binom{n}{3}} S_{ijk} - \sum_{i \neq j \neq k \neq m;i=1}^{\binom{n}{4}} S_{ijkm} + \cdots}$$ (23)

$$En_{(T_l)} = \sum_{i=1}^{n} n_i'(T_k) \times \frac{(n_i'(T_k) - 1)}{2} - \sum_{i \neq j; i=1}^{\binom{n}{2}} n_{ij}(T_k) \frac{(n_{ij}(T_k) - 1)}{2}$$

$$+ \frac{\sum_{i \neq j \neq k; i=1}^{\binom{n}{2}} n_{ijk}(T_k) \times (n_{ijk}(T_k) - 1)}{2} - \frac{\sum_{i \neq j \neq k \neq m; i=1}^{\binom{n}{4}} n_{ijkm}(T_k) \times (n_{ijkm}(T_k) - 1)}{2} + \cdots$$

(24)

where

$$n_{ix}(T_l) = n_i(T_l) - Round\left[\left(\sum_{i \neq j; j=1}^{\binom{n}{2}} S_{ij} - \sum_{i \neq j \neq k; j=1}^{\binom{n}{3}} S_{ijk} + \right.\right.$$

$$\left.\left. \sum_{i \neq j \neq k \neq m; j=1}^{\binom{n}{4}} S_{ijkm} - \cdots \right) \times \varphi_i(T_l) \right]$$

(25)

And

$$n_i'(T_l) = \left[n_{ix}(T_l) + \sum_{i \neq j; j=1}^{\binom{n}{2}} n_{ij}(T_l) - \sum_{i \neq j \neq k; j=1}^{\binom{n}{3}} n_{ijk}(T_l) + \right.$$

$$\left. \sum_{i \neq j \neq k \neq m; j=1}^{\binom{n}{4}} n_{ijkm}(T_l) - \cdots \right]$$

(26)

And $n_{ijkm}(T_l)$ is the number of nodes in the intersection area of AP_i, AP_j, AP_k and AP_m in snapshot T_l. The algorithm for calculating the intersection areas among four and more APs has been presented in Librino, Levorato, and Zorzi (2009).

5. Validation

Encounter investigations are built on top of mobility traces. In this section, we study the impact of overlapping areas of APs for estimation of the number of encounters computed using synthetic mobility traces. The synthetic mobility traces were obtained by adopting a common mobility model such as modes have been proposed in Chen et al. (2013); Mitsche et al. (2014). We compare encounters obtained by the usual method of considering two nodes encountering if they are associated with the same AP and by considering the overlapping correction.

The simulations were performed by considering that the number of APs ranged from 2 to 3 and radio coverage areas of APs were supposed to be a disc with a radius of 30 m. The distance between the centres of each pair of APs was 15m, nodes were instantiated to APs and uniformly spreaded inside the covering area of the related APs. Movement of all nodes in each snapshot was synchronized and each simulation was run along 100 snapshots. During each snapshot, the number of encounters with (and without) overlapping were computed. The simulations were performed by varying the experimental set-up and obtained results are reported in Table 2. Considering n_1, n_2 and n_3 as the number of mobile nodes associated with AP_1, AP_2 and AP_3, respectively in the experimental setup (DPP, 2018).

Table 2. Comparison of simulation results (represented by their Average:Avg, standard deviation: Std) and analytical results.

Scenarios	Number of encounters extracted by simulation result		Analytical results of number of extracted encounters	
	Avg	S.t.d	considering overlapping	without considering overlapping
$n_1 = 30; n_2 = 20$	1210	11.37	1171	625
$n_1 = 40; n_2 = 40$	3120	18.60	2991	1560
$n_1 = 50; n_2 = 30$	3121	23.40	3016	1660
$n_1 = 30; n_2 = 30; n_3 = 30$	3936	26	3771	1305
$n_1 = 40; n_2 = 40; n_3 = 40$	7017	35.8	6828	2340
$n_1 = 50; n_2 = 50; n_3 = 50$	10,981	66	10,938	3675

In Table 2, the average and standard deviations of the number of encounters computed on the synthetic traces with (and without) considering overlapping for 100 consecutive time snapshots were computed. The experimental results indicate that when the amount of overlapping is high ($\gamma = 0.52$ and $\gamma = 0.99$ for two and three APs; the estimated values are comparable against the theoretical ones, obtained using Equation (22) for the case of two APs and Equation (24) for more general cases of three APs.), ignoring the overlapping leads to a significant underestimation of the number of encounters. The number of encounters is almost double in case of two APs and is triple in the case of three APs.

6. Experiment

An experiment has been conducted at the library of the Minho university in Guimaraes, Portugal to assess how Wi-Fi datasets are reliable for extracting proximity among people (who use the Wi-Fi network to access the Internet for data communication). The library was covered with several APs with overlapping within their radio coverage areas. In all area of the library, signals of three APs specified in Figure 2 with significant signal level were accessible. In this experiment, 13 controlled mobile devices have been used simultaneously, six smartphones and seven laptops. Smartphones used active Wi-Fi and Bluetooth network interfaces (Bluetooth in discovery mode) and just Wi-Fi network interfaces (most of the laptops were not equipped with Bluetooth interfaces) were used in laptops. All devices had access to the eduroam Wi-Fi network. Here, we ignored other available mobile devices (laptops and smartphones) were not under our control. Bluetooth logs were collected through the Geoanuncious (Geo, 2018) (also it is possible to use AWARE (AWA, 2018)) applications, which was previously installed on the smartphones.This application collects logs of Bluetooth in every 30 Sec and uploads the collected data to a server when users have access to the Internet. Bluetooth sensors were used to capture direct encounters between mobile nodes, and therefore, the physical proximity in the real world. Since our goal was to assess how reliable are the collected Wi-Fi datasets for estimating the encounters among mobile devices. To achieve this goal, we used the Bluetooth logs collected from physical nearby smartphones to extract the direct encounters among mobile nodes, and then these direct Bluetooth encounters were compared with Wi-Fi indirect encounters extracted from the logs of APs in the same place. The use of smartphones with activated Bluetooth

Figure 2. Distribution of participant mobile nodes at library in different scenarios with different colours.

and Wi-Fi interfaces, allows getting knowledge about physical proximity of other equipped Bluetooth mobile nodes, and simultaneously communicates with the APs for collecting APs logs.

The experiment was repeated for 6 different spatial configurations (scenarios) of the mobile nodes to study the impact of spatial distribution of nodes on extracting encounters, as shown in Figure 2. The distributions of participant mobile nodes in each experiment scenario have been presented with different colours. Each experiment scenario was run in a specified time snapshot for several minutes to assure that mobile nodes are located in specified locations, and avoiding encounters which may occur during the transition from one scenario to another. Table 3 indicates different defined scenarios, time interval running duration of each scenario and also the symbol colour in each scenarios.

In each experiment, we compare the real number of encounters observed visually by us (Ground Truth) with encounters extracted from the Bluetooth logs and the number of distinct encounters extracted from the Wi-Fi logs (APs logs) after smoothing ping-pongs (Keramat Jahromi et al., 2014).

Table 3. Explanation of different experimental scenarios.

Scenarios	Experiment running time interval	Symbolic colour
First	12:00–12:10	Yellow
Second	12:11–12:16	Blue
Third	12:17–12:22	Green
Fourth	12:23–12:28	Red
Fifth	12:29–12:34	Purple
Sixth	12:35–12:40	Brown

Table 4. Details of the extracted encounters in each scenario.

Scenario	N_p	MN_{Pair}	N_{Pair} with smoothing	MNB_{Pair}	NB_{Pair}	R_W	R_B
1	12	66	27		14	41%	93.33%
2			22		15	28%	100%
3			36		14	46%	93.33%
4	13	78	26	15	12	33%	80%
5			20		14	26%	80%
6			25		14	32%	93.33%

The results of this experiment are presented in Table 4. We observe that for all scenarios, the ratio between the number of extracted encounters through Wi-Fi and the maximum possible number of encounters in the real situation (R_W) is less than 50%.

7. Discussion

Although the ratio between the number of extracted encounters through Wi-Fi and the maximum possible number of encounters in the real situation is influenced by the spatial distribution of the nodes in each experiment scenario and also by the aggressiveness behaviour of mobile devices for changing association with APs (González et al., 2005). We observed a significant differences between the numbers of encounters in real situations (Ground Truth) and the number of encounters extracted from the Wi-Fi logs. One of the reasons for this large difference is the overlapping in the coverage areas of APs. We could not find visually clear correlation between spatial distribution of mobile nodes and the number of extracted encounters, for instance in scenario 2 (blue colour) and 3 (green colour), although in both scenarios, nodes almost are spread in the wide area of library, but there is a significant difference in extracted encounters in these two scenarios, while for scenario 4 (purple colour), the extracted encounters are lowest when the mobile nodes are collected in a limited area of library.

On the other hand, we can see that R_B in all scenarios has a value equal or higher than 80%, it means that Bluetooth logs are more reliable for estimating encounters between devices. The value of R_B is influenced by the number of mobile devices equipped with Bluetooth sensor, coverage radius area of Bluetooth sensor, and also by the distribution of participants in the area of experiment.

The ratio between maximum possible extractable pair encounters of Bluetooth and Wi-Fi ($\frac{15}{78} = 19.2\%$) in this experiment is very small since just half of the participating devices were equipped with activated Bluetooth sensor. This means that although direct Bluetooth is more reliable compared to Wi-Fi logs for extracting proximity, but not necessarily most of the mobile devices (which use Wi-Fi to access the Internet) are equipped with Bluetooth sensors. Even in equipped devices, might be in-activated or not being in discovery mode, encounters extracted from Bluetooth logs may not reflect the real number of encounters as we observed from our experiment (We had 13 controlled mobile users in our experiment, means at most 78 encounters could be existed but in best cases 15 encounters were recognized because just half of them were equipped and activated Bluetooth sensor).

8. Conclusion

We discussed the issue of overlapping in the radio coverage areas of APs for estimating the real number of encounters. A new approach for estimating the number of encounters from

APs logs under different configurations in overlapping coverage areas of multiple APs has been proposed. The number of encounters obtained through the analytical approach and extracted from the simulations are very close to each other (with more than 95% agreement), and the difference between the analytical and simulation results decreases with increasing the nodes density of the associated APs. There are significant differences between the number of encounters in real situation and number of encounters estimated from Wi-Fi APs logs. Although the spatial distribution of participants, aggressiveness behaviour of mobile devices and also the number of participants may influence this difference, one of the main reasons for this big error is overlapping in the APs coverage area. So, it implies that for calculating encounters through APs logs, ignoring the overlapping coverage areas among APs causes big underestimation of the number of encounters.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Dr Karim Keramat Jahromi obtained his Ph.D. degree from Computer Science Department of Universita degli Studi di Milano, Milan, Italy in 2017. He received his B.Sc. degree in Electronic Engineering and Master degree in Automatic Control Engineering in Tehran Polytechnic University of Technology (Amirkabir University), Tehran, Iran in 2005. He was a research fellow in Ubiquitous and Mobile Computing Group in Algoritmi Research Center in Minho University, Guimaraes, Portugal for two years. His research interests include Mobility Analysing through Smartphones, Mobility Modelling and Prediction, Machine Learning, Networked Control Systems (NCS), and Wireless Sensor Network (WSN).

Dr Md Aktaruzzaman has completed his M.Sc. degree in Computer Science and Engineering from Islamic University, Kushtia, Bangladesh. Dr Aktaruzzaman obtained his Ph.D. in Computer Science from the University of Milan, Italy in 2015. He was also employed as a Postdoctoral fellow for one year after completing his Ph.D. at the same institute. Dr Aktaruzzaman is currently teaching as an Associate Professor in the Department of Computer Science and Engineering in Islamic University, Bangladesh. His field of research interest includes biomedical signal processing, mobile health informatics, pattern recognition, machine learning, and deep learning.

Dr Mehdi Jalili received his B.Sc. degree in Software Engineering from Kharazmi University, Iran, in 2003, and M.Sc. degree in Software Engineering from IA University, Iran, in 2009. He has obtained his Ph.D. in Computer Science from Università degli Studi di Milano, Italy, in 2018. He has joined IA University of Iran as a lecturer at Computer Engineering Department 10 years ago. His field of research interest includes software evolution, machine vision, Data and Mobility analysing.

ORCID

Md Aktaruzzaman http://orcid.org/0000-0003-0349-1434

References

2018. http://geoanuncios.dsi.uminho.pt/download/geoanuncios.apk.
2018. http://mathworld.wolfram.com/circle-circleintersection.html.
2018. http://mathworld.wolfram.com/diskpointpicking.html.
2018. http://www.awareframework.com/.

Barbosaa, H., Barthelemy, M., Ghoshal, G., James, C., Lenormand, M., Louai, T., … Simini, M., & Tomasini, F. (2018). Human mobility: Models and applications. *Physics Report*, *734*, 1–74.

Chen, C., Wu, H., & Ke, K. (2013). General ripple mobility model: A novel mobility model of uniform spatial distribution and diverse average speed. *IEICE Transcations on Communication*.

Cuttone, A., Lehmann, S., & Gonzalez, M. (2018). Understanding predictability and exploration in human mobility. *EPJ Data Science*, 1–17.

Fewell, M. P. (2006). *Area of common overlap of three circles*. Defense Science and Technology Organization of Australian Government. Retrieved from https://apps.dtic.mil/dtic/tr/fulltext/u2/a463920.pdf

Gonzalez, M., Hidaglo, C., & Barabasi, A. (2008). Understanding individual human mobility patterns. *Nature*, *453*, 779–782.

González, F., Pérez, J., & Zárate, V. (2005, April 3). *HAMS layer 2 Handoff accurate measurement strategy in WLANs 802.11*. Proceedings of the 1st IEEE International Workshop on Wireless Network Measurements (WiNMee 2005), Riva del Garda, Trentino, Italy.

Henderson, T., Kotz, D., & Abyzov, I. (2004). *The changing usage of a mature campus-wide wireless network*. Proceedings of the tenth annual international conference on mobile computing and networking, Philadelphia, PA (pp. 187–201). New York, NY: ACM.

Hsu, W.-j., & Helmy, A. (2010). On nodal encounter patterns in wireless lan traces. *IEEE Transaction on Mobile Computing*, *9*(11), 1563–1577.

Hui, P., Crowcroft, J., & Yoneki, E. (2011). Bubble rap: Social-based forwarding in delay-tolerant networks. *IEEE Transactions on Mobile Computing*, *10*(11), 1576–1589.

Karamshuk, D., Boldrini, C., Conti, M., & Passarella, A. (2011). Human mobility models for opportunistic networks. *IEEE Communications Magazine*, *49*(12), 157–165.

Keramat Jahromi, K., Meneses, F., & Moreira, A. (2014). *Impact of ping-pong events on connectivity properties of human mobility*. WMNC (pp. 1–8). IEEE.

Keramat Jahromi, K., Meneses, F., & Moreira, A. (2015). *On the impact of overlapping access points in detecting node encounters*. IEEE Ad Hoc Networking Workshop (MED-HOC-NET), 14th Annual Mediterranean (pp. 1–7).

Keramat Jahromi, K., Zignani, M., Gaito, S., & Rossi, G. (2016). Simulating human mobility patterns in urban areas. *Simulation Modeling Pratice and Theory*, *62*, 137–156.

Keramat Jahromi, K., Zignani, M., Gaito, S., & Rossi, G. (2017). Predicting encounter and colocation events. *Ad Hoc Networks*, *62*, 1–17.

Kim, M., & Kotz, D. (2005). *Modeling users' mobility among wi-fi access points*. Workshop on wireless traffic measurements and modeling.

Librino, F., Levorato, M., & Zorzi, M. (2009). *An algorithmic solution for computing circle intersection area and its application to wireless communications*. IEEE Proceeding of the 7th international conference on modeling and Optimization in Mobile Ad Hoc, and Wireless Networks (pp. 1672–1690).

Lindgren, A., Doria, A., & Schelén, O. (2004). Probabilistic routing in intermittently connected networks. In P. Dini, P. Lorenz, & J. N. de Souza (Eds.), *Service assurance with partial and intermittent resources (SAPIR 2004)* (Vol. 3126, pp. 239–254). Berlin: Springer.

Mitchell, J., O'Neill, E., Zyba, G., Voelker, G. M., Liljenstam, M., Mehes, A., & Johansson, P. (2012). *Limitations of scanned human copresence encounters for modelling proximity-borne malware*. IEEE communication systems and networks (COMSNETS).

Mitsche, D., Resta, G., & Santi, P. (2014). The random waypoint mobility model with uniform node spatial distribution. *Wireless Networks*, *20*(5), 1053–1066.

Moon, S., & Helmy, A. (2010). *Understanding periodicity and regularity of nodal encounters in mobile networks: A spectral analysis*. 2010 IEEE Global Telecommunications Conference GLOBECOM 2010, Miami, FL. IEEE.

Nguyen, A., Senac, P., & Diaz, M. (2012). *Understanding and modeling the small-world phenomenon in dynamic networks*. The 15th ACM international conference on modeling, analysis and simulation of wireless and mobile systems (MSWIM'12) (pp. 377–384).

Onghena, Y., & Milano, C. (2015). Certainties and uncertainties in tourism: Mobility and encounter. CIDOB (pp. 1–5).

Papandrea, M., Keramat Jahromi, K., Zignani, M., Gaito, S., Giordano, S., & Rossi, G. (2016). On the properties of human mobility. *Computer Communications*, *87*, 19–36.

Pirozmand, P., Wu, G., Jedari, B., & Xia, F. (2014). Human mobility in opportunistic networks: Characteristics, models and prediction methods. *Journal of Network and Computer Applications*, *42*, 45–58.

Riascos, A. P., & Mateos, L. (2017). Emergence of encounter networks due to human mobility. *PloS one*, *12*(10), 1–22.

Vanderhulst, G., Mashhadi, A., Dashti, M., & Kawsar, F. (2015). *Detecting human encounters from WiFi radio signals*. Proceedings of the 14th international conference on mobile and ubiquitous multimedia (MUM'15), Linz, Austria (pp. 97–108). New York, NY: ACM.

Wang, Z., Nascimento, M., & MacGregor, M. H. (2012). *On the analysis of periodic mobility behavior*. Mobile data challenge 2012 (by Nokia) workshop, Newcastle, UK. Retrieved from https://www.idiap.ch/project/mdc/publications/files/mdc-final220-wang.pdf