IEEE INFOCOM 2021 - IEEE Conference on Computer Communications | 978-1-6654-0325-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/INFOCOM42981.2021.9488769

Efficient Association of Wi-Fi Probe Requests under MAC Address Randomization

Jiajie Tan   S.-H. Gary Chan
Department of Computer Science and Engineering
The Hong Kong University of Science and Technology, Hong Kong, China
Email: {jtanad, gchan}@cse.ust.hk

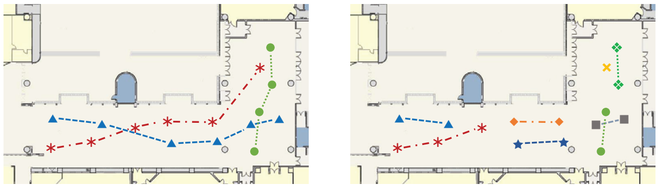
***Abstract*—Wi-Fi-enabled devices such as smartphones periodically search for available networks by broadcasting probe requests which encapsulate MAC addresses as the device identifiers. To protect privacy (user identity and location), modern devices embed random MAC addresses in their probe frames, the so-called MAC address randomization. Such randomization greatly hampers statistical analysis such as people counting and trajectory inference. To mitigate its impact while respecting privacy, we propose *Espresso*, a simple, novel and efficient approach which *es*tablishes *p*robe *re*quest a*sso*ciation under MAC address randomization. Espresso models the frame association as a flow network, with frames as nodes and frame correlation as edge cost. To estimate the correlation between any two frames, it considers the multimodality of request frames, including information elements, sequence numbers and received signal strength. It then associates frames with minimum-cost flow optimization. To the best of our knowledge, this is the first piece of work that formulates the probe request association problem as network flow optimization using frame correlation. We have implemented Espresso and conducted extensive experiments in a leading shopping mall. Our results show that Espresso outperforms the state-of-the-art schemes in terms of discrimination accuracy ($> 80\%$) and V-measure scores (> 0.85).**

***Index Terms*—Wi-Fi, probe request, virtual MAC address, MAC address randomization, frame association**

I. Introduction

Wi-Fi-enabled devices periodically search for nearby networks by broadcasting probe requests. In the past, the request frames carry media access control (MAC) addresses as the unique device identifiers to communicate with other devices such as access points (APs). The fixed and unique MAC address of a device is a piece of personal information (may link to personal identity). By collecting the probes over time and integrating them with positioning technology [1], one may track the trajectory of the user holding the device, even though it is not connected with any network. This has raised grave privacy concerns on identity and location.

To protect privacy against Wi-Fi sensing, MAC address randomization has been proposed and implemented in modern commercial devices [2], [3], including smartphones working with iOS [4] and Android [5]. Instead of using *real* physical MAC addresses in probe frames, the device generates randomized *virtual* addresses at random times. In other words, the probe requests emitted from a single device no longer share the same MAC address, but change to some random addresses over time at unpredictable intervals.



(a) Complete trajectories for the case without MAC address randomization

(b) Fragmented trajectories for the case of MAC address randomization

Fig. 1. An illustration showing the influence of MAC address randomization in a tracking application. The markers represent the positions where probe requests are emitted. Each type of marker corresponds to a MAC address.

A direct consequence of MAC address randomization is that it breaks the continuity and semantics of probe requests and leads to fragmentation in data acquisition and analytics. While protecting privacy, this adversely affects some statistic gathering such as people counting, crowd flow estimation and trajectory inference [6], [7]. As an illustration of the impact, we show in Figure 1 that three users carrying smartphones walking through an area. Figure 1(a) depicts the three complete trajectories without changing addresses. However, as each address is considered to be an individual device, these trajectories would be fragmented into more pieces due to the randomization mechanism. Figure 1(b) demonstrates a possible sensing result with eight fragmented trajectories despite there being only three devices.

In this paper, we study how to *associate* probe requests which have common emitters under MAC address randomization. By identifying the correlated frames, the association recovers Figure 1(a) from Figure 1(b) for our example above. It would also enable more fruitful and meaningful (anonymized) data analytics under MAC address randomization.

A major challenge of probe request association lies in the dense case, where many simultaneous devices are within the same area. This implies significant ambiguity or conflict in frame attributes of probe requests from different devices. Traditional association techniques based on location do not perform well under such scenarios due to the large transmitting interval between adjacent frames and localization errors. Other work uses frame payloads such as information elements or sequence numbers to determine the probe sequence [4], [8],

This work was supported, in part, by Hong Kong General Research Fund under grant number 16200120.

978-1-6654-0325-2/21/$31.00 ©2021 IEEE

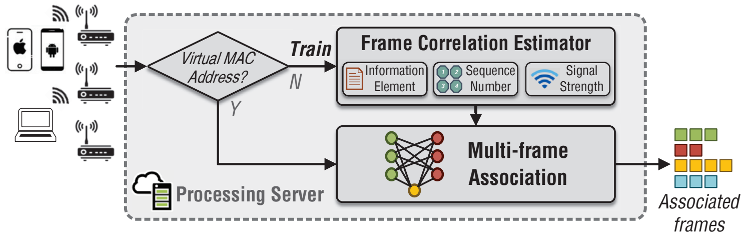


Fig. 2. The system framework of Espresso.

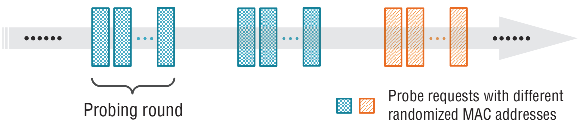


Fig. 3. An illustration of active scan under MAC address randomization over time. The bars represent the probe requests emitted from a single device. Different colors and patterns indicate different randomized MAC addresses in the frames.

[9]. However, none of them consider the association in the frame context (i.e., the association decision is made only based on the two frames). Furthermore, the single-modality nature is not sufficient enough to differentiate between senders with high accuracy, especially in a crowded region.

To address the above, we propose ***Espresso***, a novel and efficient approach that ***es***tablishes ***p***robe ***re***quest a***sso***ciation under MAC address randomization. Espresso works well even in crowded cases and is designed to ride on existing Wi-Fi infrastructure without specially-designed hardware (only normal Wi-Fi sensors or APs are needed). It requires neither external localization systems nor manual calibration/labeling beforehand. Note that Espresso fully respects user privacy, because it simply connects the probe requests; it does not, and cannot, reverse engineer the actual physical MAC addresses from the virtual MAC addresses, or link the virtual addresses to any identifiable individual.

Espresso consists of two major modules:

1) *Frame correlation estimator*: A basic question in the frame association problem is, given two probe frames, how likely they are associated. Espresso employs a probabilistic model to estimate the correlation between any two frames. The estimator considers the multimodality of frame attributes, including *information element* (IE), *sequence number* and *signal strength*. It uses the captured probe requests with real MAC addresses as the labeled dataset to train the estimator.
2) *Multi-frame association*: Espresso constructs the association among a batch of probe requests, which is modeled as a minimum-cost flow problem. In the flow network, the nodes correspond to the frames and edges represent the potential associations in between. The cost on an edge is determined by the correlation between the frames. By seeking the flow with the minimum association cost, Espresso obtains the optimal association among the frame batch. Furthermore, Espresso improves the efficiency of the large frame set by utilizing a mini-batch mechanism. To the best of our knowledge, this is the first piece of work that has considered associating multiple simultaneous frames using network flow optimization.

Figure 2 shows the system framework of Espresso. Wi-Fi sensors deployed in the monitoring venue capture probe requests from nearby devices and measure their signal strengths. The payload in frames, along with their corresponding received signal strengths (RSS), are transmitted to the processing server for analysis through Ethernet or mesh networks [10]. On the server, Espresso recognizes whether the MAC addresses in the input frames are real or virtual by checking the 7th bit (0 for real and 1 for virtual) [11]. We leverage the frames with real addresses to train the *frame correlation estimator* where the unchanged addresses are treated as labels. The association among the frames with virtual MAC addresses are constructed by the *multi-frame association* module.

Espresso is simple to implement. We have implemented it and conducted extensive experiments in a leading shopping mall for more than three months. The results show that Espresso greatly outperforms the state-of-the-art schemes [4], [8], [9] in terms of discrimination accuracy ($> 80\%$) and V-measure score of the associated frame sequences (> 0.85).

The remainder of this paper is organized as follows. We introduce the preliminaries of Espresso in Section II. Then we present in Section III the multi-modal frame correlation estimator, followed by the novel association algorithm for multiple frames in Section IV. Section V illustrates the experimental results. We finally review the related work in Section VI and conclude in Section VII.

II. Preliminaries

In this section, we introduce the preliminaries of probe requests (Section II-A) and the modeling of frame association (Section II-B).

A. Preliminaries of Probe Request

Active scan is one of the major methods that Wi-Fi devices use to discover nearby wireless networks [12]. During the scanning, devices initiate a network search by broadcasting management frames known as *probe requests*.

The scan is generally triggered on a periodic basis in order to reduce energy consumption. Figure 3 illustrates the process of active scan over time. To discover all the networks, devices need to probe every available channel. We refer to the probe request emission of one scan as a *probing round*. The duration of a probing round is about $1\,\mathrm{s}$ – $4\,\mathrm{s}$ subjected to the number of scanned channels. In most cases, the MAC addresses remain unchanged during a probing round (whether it is real or virtual). Besides that, the trigger of randomization is not necessarily synchronized with probing rounds, and a device may use the same MAC address in different consecutive rounds. The time interval between two consecutive probing rounds is subject to the factory configurations of devices.

In Espresso, we represent the i-th captured probe request as the tuple $\boldsymbol{P}_i = \langle \boldsymbol{I}_i, s_i, \boldsymbol{R}_i, t_i \rangle$, where $\boldsymbol{I}_i$ is the IE vector, s_i is the sequence number, $\boldsymbol{R}_i$ is the RSS vector, and t_i is the transmission time. The IE vector $\boldsymbol{I}_i = \langle I_i^h \,|\, 1 \leq h \leq |\boldsymbol{I}| \rangle$, where I_i^h is the h-th field in frame i and $|\boldsymbol{I}|$ is the total number of fields. Note that some IE fields (e.g., *Vendor Specific*) may appear multiple times in a frame. We concatenate them by their presenting order and regard the combination as the content of the field. The sequence number s_i is a 12-bit counter indicating the transmission order on the device. It is bounded between 0 and 4095. The RSS vector of frame i is denoted as $\boldsymbol{R}_i = \langle r_i^u \,|\, 1 \leq u \leq |\boldsymbol{R}| \rangle$, where r_i^u is the signal strength measured by sensor u, and $|\boldsymbol{R}|$ is the total number of sensors.

B. Modeling of Frame Association

As the frames in the same probing round usually share the same MAC address [9], we are more interested in associating frames between different probing rounds. In Espresso, we consider frames to be in the same round if they have identical MAC addresses and their adjacent time intervals are less than 1 s. We randomly pick one frame from each probing round to represent it. In the following, we work on the set of selected frames $\boldsymbol{\mathcal{P}} = \{\boldsymbol{P}_i \,|\, 1 \leq i \leq M\}$, where M is the number of frames (rounds).

Let $\boldsymbol{\mathcal{D}}$ be the set of transmission devices in $\boldsymbol{\mathcal{P}}$. We use $\boldsymbol{\mathcal{S}}_k = \{\boldsymbol{P}_l^k \,|\, 1 \leq l \leq M_k, \boldsymbol{P}_l^k \in \boldsymbol{\mathcal{P}}\}$ to denote the sequence of probe requests from a device $d_k \in \boldsymbol{\mathcal{D}}$, where M_k is the number of frames in the sequence. In the frame sequence $\boldsymbol{\mathcal{S}}_k$, we name $\boldsymbol{P}_{l-1}^k$ to be the *predecessor* frame of $\boldsymbol{P}_l^k$ $(2 \leq l \leq M_k)$, and $\boldsymbol{P}_{l+1}^k$ as the *successor* frame of $\boldsymbol{P}_l^k$ $(1 \leq l \leq M_k - 1)$.

Given above, we define that two probe requests $\boldsymbol{P}_i$ and $\boldsymbol{P}_j$ $(1 \leq i < j \leq M)$ are *associated* if $\boldsymbol{P}_j$ is the successor of $\boldsymbol{P}_i$ (or $\boldsymbol{P}_i$ is the predecessor of $\boldsymbol{P}_j$). We use a binary indicator $\boldsymbol{\Phi}_{ij}$ to denote the association between $\boldsymbol{P}_i$ and $\boldsymbol{P}_j$, where

$$\boldsymbol{\Phi}_{ij} = \begin{cases} 1, & \boldsymbol{P}_j \text{ is associated to } \boldsymbol{P}_i, \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

Note that the association $\boldsymbol{\Phi}_{ij} = 1$ also implies that $\boldsymbol{P}_i$ and $\boldsymbol{P}_j$ are transmitted from the same device. In this work, we aim to find the associations in the probe request set $\boldsymbol{\mathcal{P}}$.

III. Frame Correlation Estimator

In this section, we present a multimodal correlation estimator between probe requests. We first show in Section III-A that the association probability can be decomposed as the correlations of different frame attributes. Then we discuss in detail the correlation estimation of IE, sequence number and signal strength in Sections III-B to III-D, respectively.

A. Model Decomposition

The association correlation between probe frames $\boldsymbol{P}_i$ and $\boldsymbol{P}_j$ is defined as their association probability $p(\boldsymbol{\Phi}_{ij} \,|\, \boldsymbol{P}_i, \boldsymbol{P}_j)$. By applying Bayes' theorem and assuming the independence among different frame attributes, we factorize the posterior as

$$\begin{aligned} & p(\boldsymbol{\Phi}_{ij} \,|\, \boldsymbol{P}_i, \boldsymbol{P}_j) \\ & = p(\boldsymbol{\Phi}_{ij} \,|\, \boldsymbol{I}_i, s_i, \boldsymbol{R}_i, \boldsymbol{I}_j, s_j, \boldsymbol{R}_j, \Delta t_{ij}) \\ & \propto p(\boldsymbol{I}_j, s_j, \boldsymbol{R}_j \,|\, \boldsymbol{I}_i, s_i, \boldsymbol{R}_i, \boldsymbol{\Phi}_{ij}, \Delta t_{ij})\, p(\boldsymbol{\Phi}_{ij} \,|\, \boldsymbol{I}_i, s_i, \boldsymbol{R}_i, \Delta t_{ij}) \\ & = p(\boldsymbol{I}_j \,|\, \boldsymbol{I}_i, \Delta t_{ij}, \boldsymbol{\Phi}_{ij})\, p(s_j \,|\, s_i, \Delta t_{ij}, \boldsymbol{\Phi}_{ij})\, p(\boldsymbol{R}_j \,|\, \boldsymbol{R}_i, \Delta t_{ij}, \boldsymbol{\Phi}_{ij}) \\ & \quad\; p(\boldsymbol{\Phi}_{ij} \,|\, \Delta t_{ij}), \end{aligned} \tag{2}$$

where $\Delta t_{ij} = t_j - t_i$ is the time interval in between. In the decomposition, $p\,(\boldsymbol{I}_j \,|\, \boldsymbol{I}_i, \Delta t_{ij}, \boldsymbol{\Phi}_{ij})$, $p\,(s_j \,|\, s_i, \Delta t_{ij}, \boldsymbol{\Phi}_{ij})$ and $p\,(\boldsymbol{R}_j \,|\, \boldsymbol{R}_i, \Delta t_{ij}, \boldsymbol{\Phi}_{ij})$ can be interpreted as the likelihoods of IE, sequence number and signal strength, respectively. $p(\boldsymbol{\Phi}_{ij} \,|\, \Delta t_{ij})$ describes prior knowledge of $\boldsymbol{\Phi}_{ij}$ given the time interval. In particular, we consider that the prior $p(\boldsymbol{\Phi}_{ij} \,|\, \Delta t_{ij})$ follows an exponential decay distribution, i.e.,

$$p(\boldsymbol{\Phi}_{ij} \,|\, \Delta t_{ij}) = \begin{cases} \exp(-\lambda \Delta t_{ij}), & \Delta t_{ij} > 0, \\ 0, & \Delta t_{ij} \leq 0, \end{cases} \tag{3}$$

where λ is the decay rate. In the following, unless otherwise stated, we only discuss the case where $\Delta t_{ij} > 0$.

B. Information Element

In probe requests, IEs contain device specification and configuration information. Prior work has explored their device-dependent nature and treated IEs as unique device fingerprints [4]. However, such methods are not reliable in practice because a device may emit probe frames with diverse IEs. To show this, we compile statistics from 100,000 pairs of probe requests collected from the same device but at different times. Figure 4 depicts the percentage of distinct IE fields despite the same device. This confirms the fact that traditional IE-fingerprinting approaches result in a large number of false negatives. In Espresso, we propose a probabilistic estimation of IE correlation to enhance the generality and robustness.

We assume that IEs are independent of time, i.e.,

$$p(\boldsymbol{I}_j \,|\, \boldsymbol{I}_i, \Delta t_{ij}, \boldsymbol{\Phi}_{ij}) = p(\boldsymbol{I}_j \,|\, \boldsymbol{I}_i, \boldsymbol{\Phi}_{ij}). \tag{4}$$

Moreover, note that estimating $p(\boldsymbol{I}_j \,|\, \boldsymbol{I}_i, \boldsymbol{\Phi}_{ij})$ directly is challenging in reality since the sample space of $\boldsymbol{I}_j$ is exponentially large ($3^{|\boldsymbol{I}|}$). By assuming that IEs are uniformly distributed, the correlation is proportional to the association posterior, i.e.,

$$p(\boldsymbol{I}_j \,|\, \boldsymbol{I}_i, \boldsymbol{\Phi}_{ij}) \propto p(\boldsymbol{\Phi}_{ij} \,|\, \boldsymbol{I}_i, \boldsymbol{I}_j). \tag{5}$$

Since we are only interested in whether $\boldsymbol{I}_i$ and $\boldsymbol{I}_j$ are similar, we further introduce a feature vector $\boldsymbol{\Delta I}_{ij}$ to represent the difference between $\boldsymbol{I}_i$ and $\boldsymbol{I}_j$, i.e., $\boldsymbol{\Delta I}_{ij} = \langle \Delta I_{ij}^h \,|\, 1 \leq h \leq |\boldsymbol{I}| \rangle$, where

$$\Delta I_{ij}^h = \begin{cases} 1, & (I_i^h = \emptyset \land I_j^h = \emptyset) \lor I_i^h = I_j^h, \\ -1, & (I_i^h \neq \emptyset \land I_j^h \neq \emptyset) \land I_i^h \neq I_j^h, \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

In particular, $\Delta I_{ij}^h = 0$ indicates the case where the h-th IE field is missing in either frame (but not both). In that case, we have

$$p(\boldsymbol{\Phi}_{ij} \,|\, \boldsymbol{I}_i, \boldsymbol{I}_j) = p(\boldsymbol{\Phi}_{ij} \,|\, \boldsymbol{\Delta I}_{ij}). \tag{7}$$

We apply logistic regression to estimate $p(\boldsymbol{\Phi}_{ij} \,|\, \boldsymbol{\Delta I}_{ij})$. To train the model, we obtain training data from the sensed frames with real MAC addresses. Specifically, we randomly choose N_{IE}^{+} pairs of frames with identical MAC addresses and label them as positive. Similarly, the negative samples come from the N_{IE}^{-} frame pairs with different real addresses. To mitigate the bias in the model, we suggest to use the equal size of positive and negative samples i.e., $N_{\text{IE}}^{+} = N_{\text{IE}}^{-}$.

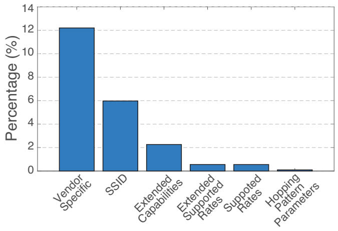


Fig. 4. The ratio of distinct IE fields in the probe request pairs.

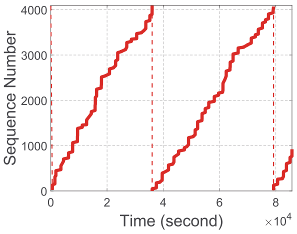


Fig. 5. The growth of sequence numbers over time.

C. Sequence Number

Sequence numbers are used to record the transmission order of frames. Wi-Fi chips normally increase the sequence number by 1 when emitting a new frame (the number will be reset to 0 when reaching the maximum value 4096). When devices do not connect to any wireless network, they transmit nothing but probe requests for periodical active scans. To illustrate this, we record the probe requests emitted from an iOS device for 24 hours. The device does not connect to any network and keeps its screen off. Frames are captured by a sensor operating on channel 1 of $2.4\,$GHz Wi-Fi. Figure 5 shows the growth of sequence numbers during the period. Despite the imperfect linearity due to channel switching and frame loss, the continuity of sequence numbers can be clearly observed.

Let Δs_{ij} denote the difference of sequence numbers between s_i and s_j, and

$$\Delta s_{ij} = \begin{cases} s_j - s_i, & s_i < s_j, \\ 4096 - s_j + s_i, & s_i \geq s_j. \end{cases} \tag{8}$$

As Δs_{ij} does not depend on the current sequence number or the time interval, the memoryless model can be represented as

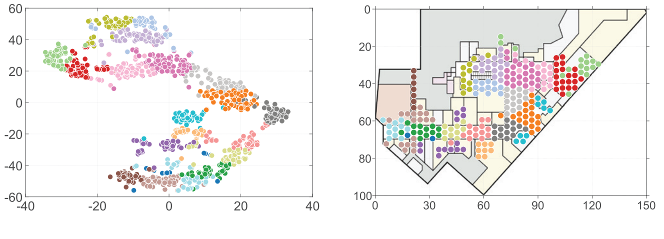
$$p(s_j \mid s_i, \Delta t_{ij}, \mathbf{\Phi}_{ij}) = p(\Delta s_{ij} \mid \Delta t_{ij}, \mathbf{\Phi}_{ij}). \tag{9}$$

In particular, we are interested in the case of $\mathbf{\Phi}_{ij} = 1$ since the continuity of sequence numbers is only valid among the sequences of associated frames. Additionally, we discretize Δt_{ij} with the granularity g_{seq} to reduce the model complexity while preserving its generality ($g_{\text{seq}} = 30$s in the paper).

Given a discretization of Δt_{ij}, we describe the modeling of Δs_{ij} as follows. Recall that probe requests are usually broadcast via all the channels in succession (e.g., channel 1 to 14 in $2.4\,$GHz Wi-Fi). Ideally, a sensor sticking on a fixed channel would capture the sequence numbers with a regular gap. In practice, however, Δs_{ij} exhibits multiple modes mainly due to the frame loss and the heterogeneity among devices. We hence represent the distribution of Δs_{ij} as a Gaussian mixture, i.e.,

$$p(\Delta s_{ij} \mid \Delta t_{ij}, \mathbf{\Phi}_{ij} = 1) = \sum_{r=1}^{R} \pi_r \mathcal{N}(\Delta s_{ij} \mid \mu_S(r), \sigma_S^2(r)), \tag{10}$$

where R is the number of mixture components, π_r is the probability of the r-th component, $\mathcal{N}(\cdot)$ denotes a Gaussian distribution, and $\mu_S(r)$ and $\sigma_S^2(r)$ are the mean and variance of the r-th component, respectively.



(a) Signal space

(b) Physical space

Fig. 6. The partition of signal and physical spaces using K-means ($K = 20$). Different colors indicate different partition labels. The signal space is visualized in 2D using t-distributed Stochastic Neighbor Embedding (t-SNE).

The model parameters π_r, $\mu_S(r)$ and $\sigma_S^2(r)$ are determined by applying expectation maximization (EM). We obtain the training dataset by randomly selecting N_{seq} pairs of associated frames in a period of frames with real MAC addresses. They are segregated into multiple groups according to g_{seq}. Independent mixture models are then trained from the data in the different groups.

D. Signal Strength

Due to the location-dependence nature of RSS, the transitions of signals reflect user movements in the physical space. For instance, a device should not move far away in a short time period (say, several seconds), resulting in similar signals. By contrast, if a dramatic signal difference is detected in close time, we have reasons to speculate that the frames are not associated. Inspired by the observation, we propose to leverage the transitions of RSS to infer frame association.

A straightforward solution is to convert RSS into exact locations via localization techniques such as trilateration [13] and fingerprinting [14], [15], [16]. However, the reliance on location limits the system flexibility and raises privacy concerns. Considering the factor that similar signals are generally located in the close region, Espresso applies clustering techniques to divide the signal space and directly infers the association from the transitions among signal partitions.

We first discuss the partition of signal space. Given a set of RSS vectors in a venue, we apply K-means clustering to construct a partition. The entire signal space is thus divided into K clusters, and each signal cluster corresponds to a physical region. Figure 6 demonstrates the resulting partition of both signal and physical spaces in a shopping mall. Note that the clustering does not require any location label. The RSS vectors can be collected by the deployed sensors automatically.

The region where an RSS vector $\boldsymbol{R}_i$ lies can be determined by seeking its closest cluster in the signal space, i.e.,

$$\mathbf{\Gamma}_i = \arg\min_k \|\boldsymbol{R}_i - \boldsymbol{C}_k\|_2, \tag{11}$$

where $\mathbf{\Gamma}_i$ is the index of partition that $\boldsymbol{R}_i$ belongs to, and $\boldsymbol{C}_k$ is the center of the k-th signal cluster. RSS transitions can be thus represented as their region transitions, i.e.,

$$p(\boldsymbol{R}_j \mid \boldsymbol{R}_i, \Delta t_{ij}, \mathbf{\Phi}_{ij}) = p(\mathbf{\Gamma}_j \mid \mathbf{\Gamma}_i, \Delta t_{ij}, \mathbf{\Phi}_{ij}). \tag{12}$$

Similar to the discussion in Section III-C, we discretize the dataset according to the time granularity g_{trans} ($g_{\text{trans}} = 30$s in the paper). In the case of $\boldsymbol{\Phi}_{ij} = 1$, given the time interval Δt_{ij}, the transition probability between a pair of associated frames can be estimated by the frequency of transitions, i.e.,

$$p(\boldsymbol{\Gamma}_j \,|\, \boldsymbol{\Gamma}_i, \Delta t_{ij}, \boldsymbol{\Phi}_{ij} = 1) = \frac{N_{\text{trans}}(\boldsymbol{\Gamma}_i, \boldsymbol{\Gamma}_j)}{N_{\text{trans}}(\boldsymbol{\Gamma}_i)}, \tag{13}$$

where $N_{\text{trans}}(\boldsymbol{\Gamma}_i, \boldsymbol{\Gamma}_j)$ denotes the number of transitions leaving region $\boldsymbol{\Gamma}_i$ for region $\boldsymbol{\Gamma}_j$ within the period, and $N_{\text{trans}}(\boldsymbol{\Gamma}_i) = \sum_{k=1}^{K} N_{\text{trans}}(\boldsymbol{\Gamma}_i, \boldsymbol{\Gamma}_k)$ is the total number of transitions departing from region $\boldsymbol{\Gamma}_i$.

IV. Multi-frame Association

In this section, we present our novel algorithm which efficiently associates multiple frames. Section IV-A describes the multi-frame association problem. In Section IV-B, we show that the association problem can be viewed as a minimum-cost flow problem. Finally, we present in Section IV-C the mini-batch adaptation for processing a large dataset.

A. Multi-frame Association Formulation

We first present the formulation of the multi-frame association problem. Recall that $\boldsymbol{\mathcal{P}} = \{\boldsymbol{P}_i \,|\, 1 \le i \le M\}$ denotes the set of M frames obtained in a period. Our goal is to find the predecessor frame (i.e., the previous frame sent by the same device) and the successor frame (i.e., the next frame sent by the same device) of each frame $\boldsymbol{P}_i$ in $\boldsymbol{\mathcal{P}}$.

Let $\boldsymbol{\Phi}$ be the $M \times M$ association matrix where the element $\boldsymbol{\Phi}_{ij}$ is defined in Equation 1. In case the successor and/or predecessor of a frame is not in $\boldsymbol{\mathcal{P}}$, we further introduce two indicator vectors of length M, denoted by $\boldsymbol{A}$ and $\boldsymbol{B}$, where

$$\boldsymbol{A}_i = \begin{cases} 1, & \boldsymbol{P}_i \text{ has no successor in } \boldsymbol{\mathcal{P}}, \\ 0, & \text{otherwise,} \end{cases} \tag{14}$$

and

$$\boldsymbol{B}_i = \begin{cases} 1, & \boldsymbol{P}_i \text{ has no predecessor in } \boldsymbol{\mathcal{P}}, \\ 0, & \text{otherwise.} \end{cases} \tag{15}$$

Given a set of probe requests $\boldsymbol{\mathcal{P}}$ and their pairwise association probability given by the correlation estimator (Section III), the multi-frame association problem seeks an association with the highest joint probability over the entire set. That is,

$$\begin{aligned} \underset{\boldsymbol{\Phi}}{\arg\max} \quad & \prod_i \prod_j p(\boldsymbol{\Phi}_{ij} \,|\, \boldsymbol{P}_i, \boldsymbol{P}_j)^{2\boldsymbol{\Phi}_{ij}} \prod_i \gamma^{\boldsymbol{A}_i} \prod_j \gamma^{\boldsymbol{B}_j}, & (16) \\ \text{subject to:} \quad & \sum_{j=1}^{M} \boldsymbol{\Phi}_{ij} + \boldsymbol{A}_i = 1, \quad \forall i : 1 \le i \le M, & (17) \\ & \sum_{i=1}^{M} \boldsymbol{\Phi}_{ij} + \boldsymbol{B}_j = 1, \quad \forall j : 1 \le j \le M, & (18) \\ & \boldsymbol{\Phi}_{ij} \in \{0,1\}, \boldsymbol{A}_{ij} \in \{0,1\}, \boldsymbol{B}_{ij} \in \{0,1\}, & \\ & \forall i : 1 \le i \le M, \ \forall j : 1 \le j \le M, & (19) \end{aligned}$$

where γ is the probability that a frame has no predecessor (or successor) in $\boldsymbol{\mathcal{P}}$. The objective function (16) is the joint probability of all the association decisions over $\boldsymbol{\mathcal{P}}$. The constraints (17) and (18) require that each frame has at most one predecessor and at most one successor, respectively. The value of γ can be empirically determined according to historical data.

B. Minimum-cost Network Flow Solution

We show that the above multi-frame association problem can be solved efficiently by viewing it as a minimum-cost network flow problem.

Figure 7 illustrates the graph structure of the corresponding flow network $\mathcal{G}$. For each frame $\boldsymbol{P}_i \in \boldsymbol{\mathcal{P}}$, two nodes are added into the network $\mathcal{G}$ — a *sender* node u_i and a *receiver* node v_i. A sender node has a supply of 1 and a receiver node demands 1. Edge $(u_i, v_j) \in E$ ($1 \le i, j \le M, i \ne j$) indicates that $\boldsymbol{P}_j$ is a possible successor frame of $\boldsymbol{P}_i$. The cost of edge (u_i, v_j) is assigned by the negative logarithm of the frame correlation, i.e., $-\log p(\boldsymbol{\Phi}_{ij} \,|\, \boldsymbol{P}_i, \boldsymbol{P}_j)$. We further set the edge capacities to 1 since the association between frames is a binary decision.

Moreover, we add in $\mathcal{G}$ an auxiliary node w to represent the case where frames have no successor and/or predecessor. A flow from a sender node to the auxiliary node indicates that the frame has no successor, while a flow from it to a receiver implies the case of no predecessor. The edge costs to/from the auxiliary node are $-\log \gamma/2$, and their capacities are 1. The supply on the auxiliary node is 0 due to the following theorem.

Theorem 1. *Let N denote the number of devices that transmit probe requests $\boldsymbol{\mathcal{P}}$ ($N \le M$). The number of frames with no successor is equal to the number of frames with no predecessor, i.e., $\sum_{i=1}^{M} \boldsymbol{A}_i = \sum_{i=1}^{M} \boldsymbol{B}_i = N$.*

Proof. Let $\boldsymbol{\mathcal{S}}_k = \{\boldsymbol{P}_l^k \,|\, 1 \le l \le M_k\}$ be the frame sequence emitted from the k-th device. In $\boldsymbol{\mathcal{S}}_k$, all the frames except for the last one $\boldsymbol{P}_{M_k}^k$ have successors, i.e., $\boldsymbol{A}_{M_k}^k = 1$ and $\boldsymbol{A}_l^k = 0$ ($1 \le l \le M_k - 1$); all the frames except for the first one $\boldsymbol{P}_1^k$ have predecessors, i.e., $\boldsymbol{B}_1^k = 1$ and $\boldsymbol{B}_l^k = 0$ ($2 \le l \le M_k$). Because $\boldsymbol{\mathcal{P}}$ is the union of frame sequences from the M devices, the number of devices N can be represented by the total number of frames with no successor ($\sum_{i=1}^{M} \boldsymbol{A}_i = \sum_{k=1}^{N} \boldsymbol{A}_{M_k}^k = N$) or the total number of frames with no predecessor ($\sum_{i=1}^{M} \boldsymbol{B}_i = \sum_{k=1}^{N} \boldsymbol{B}_1^k = N$). □

In the flow network above, we can find a maximum flow f with the minimum cost to obtain the optimal association, i.e.,

$$\begin{aligned} \underset{f}{\arg\min} \quad & -\sum_i \sum_j f(u_i, v_j) \log p(\boldsymbol{\Phi}_{ij} \,|\, \boldsymbol{P}_i, \boldsymbol{P}_j) & \\ & -\frac{1}{2} \log \gamma \sum_i f(u_i, w) - \frac{1}{2} \log \gamma \sum_j f(w, v_j), & (20) \\ \text{subject to:} \quad & \sum_{j=1}^{M} f(u_i, v_j) + f(u_i, w) = 1, \ \forall i : 1 \le i \le M, & (21) \\ & \sum_{i=1}^{M} f(u_i, v_j) + f(w, v_j) = 1, \ \forall j : 1 \le j \le M, & (22) \\ & \sum_{i=1}^{M} f(u_i, w) = \sum_{j=1}^{M} f(w, v_j), & (23) \\ & f(u_i, v_j) \in \{0,1\}, \quad \forall (u_i, v_j) : (u_i, v_j) \in E, & (24) \end{aligned}$$

where $f(u_i, v_j)$ represents the flow on the edge (u_i, v_j). Equations (21–23) are the constraints of flow conservation, and Equation (24) specifies the capacities on edges.

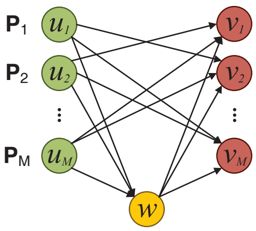


Fig. 7. The graph structure of a flow network.

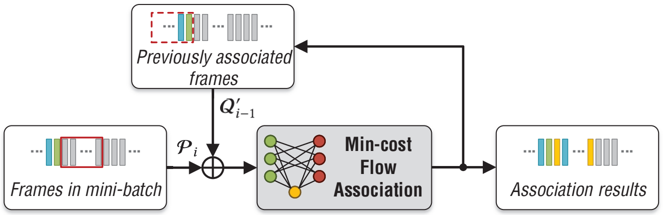


Fig. 8. The workflow of frame association for a large dataset.

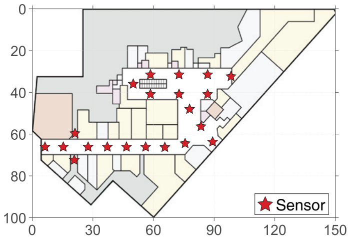


Fig. 9. The floor plan of the experimental site in a large shopping mall (in meter).

The minimum-cost flow problem can be solved efficiently by using algorithms such as Cycle Canceling [17], [18] and Network Simplex [19]. Given the optimal flow f, the corresponding frame association can thus be obtained by

$$\mathbf{\Phi}_{ij} = f(u_i, v_j), \tag{25}$$

$$\boldsymbol{A}_i = f(u_i, w), \tag{26}$$

$$\boldsymbol{B}_j = f(w, v_j). \tag{27}$$

Last but not least, we prove the correctness of the proposed minimum-cost flow modeling by the following theorem.

Theorem 2. *The proposed minimum-cost flow problem is equivalent to the multi-frame association problem.*

Proof. The theorem is proven by showing that both the objective functions and the constraints in the two problems are equivalent. We first discuss the objective functions. Let O_{MFA} denote the objective function (16) in the original multi-frame association problem, and O_{MCF} be the objective function (20) of the minimum-cost flow formulation. We can easily verify that $O_{\text{MCF}} = -\log O_{\text{MFA}}/2$. Hence maximizing the objective function (16) is equivalent to minimizing the objective function (20). We then show the equivalence of constraints. By applying the conversion in Equations (25–27), constraints (21–24) can be rewritten to the format of constraints (17–19), respectively. Constraint (23) corresponds to the characteristic of the multi-frame association as described in Theorem 1. Therefore, we conclude that the proposed minimum-cost flow problem is equivalent to the multi-frame association problem. □

C. Mini-batch Processing for Large Dataset

Although the above algorithm can construct the association on a given frame set, the efficiency would be affected when processing a large body of frames in a single batch (say, a set of frames in several hours). To address this, we present in the following an efficient processing scheme based on mini-batch.

The basic idea is to divide the dataset into multiple smaller batches and sequentially construct associations for each. Apart from the association within each mini-batch, Espresso also considers the association across different mini-batches in order to obtain the complete association over the entire period. Figure 8 illustrates the workflow. The frame set $\mathcal{P}$ is partitioned into multiple mini-batches of the same interval T, i.e., $\mathcal{P} = \mathcal{P}_1 \cup \mathcal{P}_2 \cup \cdots$. Let $\mathcal{Q}_i$ denote the set of frames without successors after processing the i-th mini-batch (initially, $\mathcal{Q}_0 = \emptyset$). To avoid the size exploding of $\mathcal{Q}_i$, Espresso only keeps the frames in the last time period of length V, denoted by $\mathcal{Q}'_i$. V is usually the possible maximum interval between two consecutive probe requests ($V = 600$s empirically). In the i-th iteration, given the mini-batch frames $\mathcal{P}_i$ and the pruned set of previously associated frames $\mathcal{Q}'_{i-1}$, we solve the minimum-cost flow problem over $\mathcal{P}_i \cup \mathcal{Q}'_{i-1}$. The ultimate association over the entire set $\mathcal{P}$ can be obtained by the union of the association in all the iterations. Note that the scheme also enables Espresso to operate in an online manner with a delay of T.

V. Experimental Evaluation and Results

We present in this section the experimental evaluation of Espresso. We introduce the experimental settings in Section V-A and discuss the illustrative results in Section V-B.

A. Experimental Settings

We have implemented Espresso and conducted extensive experiments to validate its performance. The system consists of multiple Wi-Fi sensors and a centralized server. The sensors are implemented on commercial Wi-Fi APs (GL-AR150 with OpenWrt 18.06). They capture probe requests via *libpcap*. Captured frames are then transmitted to the server through Ethernet for storage and processing. The server is built on a PC equipped with Intel Core i7 3.6 GHz CPU and 16 GB RAM. The algorithm is implemented in Python, and we employ *Google OR-Tools* [20] as the solver to the minimum-cost flow problem.

The experiments are conducted on an entire floor of a large shopping mall in Hong Kong. Figure 9 shows its floor plan (~8000 m^2) and sensor placement (21 sensors in total). Wi-Fi sensors are installed on the ceiling. They operate on channel 1 of 2.4 GHz. The system has been running for more than three months. We use the frames collected in one day for training the frame correlation estimator and the data in another day for evaluation. During a business hour on a typical weekend day, there are ~20,000 frames captured by the system with ~5000 unique MAC addresses.

Obtaining the actual frame association (as ground truth) is a major challenge in the experiments. We introduce two methods to address this. The first solution is to use the probe frames of physical MAC addresses. The real addresses are naturally the device identifiers and hence indicate the true associations of frames. Although the frames come from the devices without

MAC address randomization, we can still leverage them to evaluate the overall performance under a large number of simultaneous frames. The second way is to manually collect the emitted MAC addresses. We attach external Wi-Fi sensors close to the transmitting devices, and the frames emitted can be captured with strong signal strength. In particular, we consider the frames with RSS greater than $-40\,\mathrm{dBm}$ coming from the targeting device and thus obtain its frame association. In the experiment, we invite 6 volunteers to carry smartphones (one for each user; the Wi-Fi function is switched on but does not connect to any network) and roam in the site. This can be used to validate the system performance on real devices with MAC address randomization. Unless otherwise stated, the dataset labeled with physical MAC addresses is our default test set.

Our experiment uses the following performance metrics:

- *Discrimination accuracy*: Given a probe request $\boldsymbol{P}_i$ at time t_i and the set of previous frames captured in $[t_i - \tau, t_i)$ that contains at least one predecessor of $\boldsymbol{P}_i$, we regard it as a correct association if the frame of the highest correlation to $\boldsymbol{P}_i$ is emitted from the same device transmitting $\boldsymbol{P}_i$. Discrimination accuracy is defined as the ratio of the correct cases among all the tests. The metric reflects the effectiveness of the frame correlation estimator. In our experiments, discrimination accuracy is estimated from 1000 randomly selected frames and their previous frames.
- *Homogeneity*, *completeness* and *V-measure*: Homogeneity and completeness are widely used in clustering performance analysis by leveraging normalized conditional entropy. We borrow its concept to evaluate the goodness of associated frame sequences by regarding them as clusters. Readers may refer to [21] for a detailed explanation. Homogeneity reflects whether generated sequences contain only frames from the same devices, while completeness implies whether all frames from a device are assigned to the same frame sequence. V-measure gives a comprehensive score which is defined as the harmonic mean of homogeneity and completeness. All the scores are between 0 and 1, and a higher value indicates better performance.

We compare Espresso with the following prior schemes:

- *Frame fingerprinting (FrameFpr)* [8]: The work uses a bit-level fingerprinting approach to distinguish between devices. It performs entropy-based analysis (i.e., variability and stability) on each frame bit to extract the informative ones as the devices' fingerprints.
- *Sequence number thresholding with IE clustering (SeqThresh)* [4]: The scheme proposes a two-stage association based on IEs and sequence numbers. It first clusters probe requests according to their IE fields. Within each cluster, it then links frames if the difference between sequence numbers is below a specified threshold η. We follow the original paper to select IE fields and set $\eta = 64$.
- *Time-based signature (TimeSig)* [9]: The scheme treats the distribution of the inter-frame arrival time (i.e., the time interval of pairwise frames) in probing rounds as the devices' signatures. A distance function is also designed to compare the similarity between two interval distributions.

Unless otherwise specified, the following baseline parameters are used in the paper: in the frame correlation estimator, the number of sequence number modes $R = 3$ and the number of region partition $K = 30$; in the multi-frame association, the mini-batch duration $T = 30\mathrm{s}$ and the probability of no predecessor or successor $\gamma = 0.00001$.

B. Illustrative Results

We first study the impact of parameters in the frame correlation estimator. Figure 10 demonstrates the discrimination accuracy of sequence number against different numbers of modes R. The parameter controls the model complexity of the Gaussian mixture model (Section III-C). We can observe that the accuracy peaks near $R = 3$. It outperforms the degraded case of Gaussian distribution ($R = 1$) significantly because a single-mode model cannot well-capture the actual distribution of sequence number differences. With the growth of R ($R > 3$), the discrimination accuracy declines gradually since the model becomes overfitting to the training data. To achieve the optimal performance with good generalization, we suggest $R = 3$ in the paper.

Figure 11 shows the discrimination accuracy of signal strength over different numbers of partitions K. The selection of parameter K depends on the scale and layout of sites. A coarse-grained partition (i.e., K is small) cannot distinguish between transitions and thus achieves low accuracy. On the contrary, when the granularity is too fine (i.e., K is large), the performance suffers from signal noises and hence leads to a huge classification error. Therefore, we use $K = 30$ in the paper for the modest granularity.

Figure 12 depicts the ROC curves of the different modalities of frame attributes, i.e., information element (*IE*), sequence number (*SeqNum*) and signal strength (*RSS*). We view the frame correlation estimator as a binary classifier to predict if a pair of probe requests are emitted from the same device. A frame pair from the same emitter is regarded as "positive"; it is "negative" otherwise. We can see that *IE* generally has a high true positive rate (TPR) while *SeqNum* has a low false positive rate (FPR). Although *RSS* itself does not perform as well as the others, it can work as a supplementary factor to enhance discriminability and robustness. Espresso considers the multimodality of frame attributes and hence gains the best performance in terms of high TPR and low FPR.

We further investigate the discrimination performance in the periods of various lengths τ. Figure 13 compares the discrimination accuracy of different frame modalities. Among the models with a single modality, *SeqNum* and *RSS* suffer from severe ambiguity among a large number of frames and hence achieve low accuracies. *IE* performs better than the others, but it alone still cannot provide satisfactory association (discrimination accuracy is around 0.5). By fusing multiple modalities, Espresso outperforms all the single-modal mode significantly (by at least 59%).

Figure 14 shows the discrimination accuracy of different model combinations over various periods. Espresso achieves

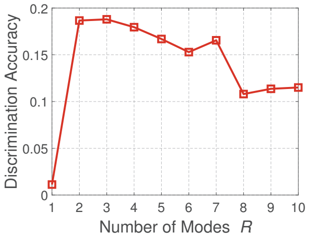


Fig. 10. Discrimination accuracy versus the number of sequence number modes R.

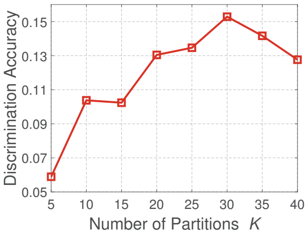


Fig. 11. Discrimination accuracy versus the number of signal partitions K.

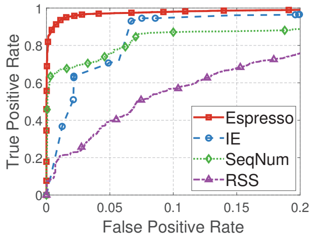


Fig. 12. ROC curves of model components.

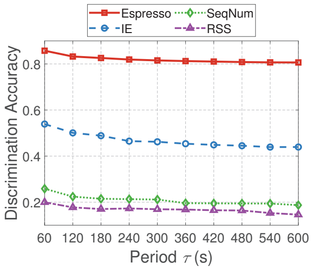


Fig. 13. Discrimination accuracy of model components in different lengths of periods.

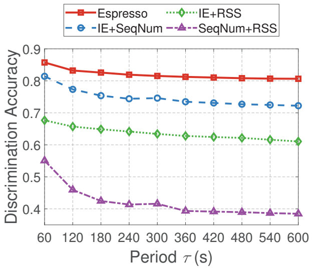


Fig. 14. Discrimination accuracy of the combinations of model components in different lengths of periods.

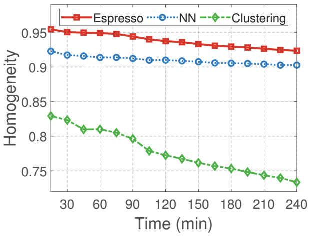


Fig. 15. Homogeneity of the frame sequences constructed by different association methods.

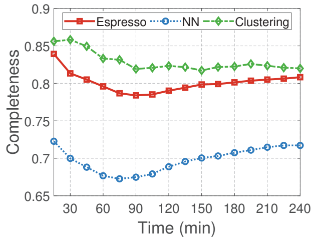


Fig. 16. Completeness of the frame sequences constructed by different association methods.

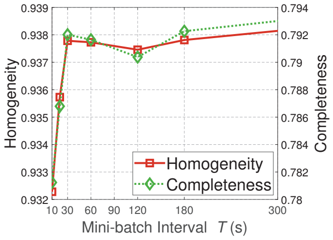


Fig. 17. Homogeneity and completeness against the mini-batch interval T.

the highest discrimination accuracy due to its multimodal nature. We also notice that the accuracy declines gradually with the increase of period τ. This can be explained by the fact that the increasing number of frames brings more ambiguity for the association. Espresso has the minimum drop of 5.9% from $\tau = 60$s to $\tau = 600$s, which further verifies its robustness.

Figures 15 and 16 illustrate the homogeneity and completeness of the frame sequences constructed by different association methods, respectively. *NN* refers to the nearest neighbor algorithm which associates a frame with the other one of the highest correlation. *Clustering* denotes the approach that applies DBSCAN [22] to cluster probe requests based on the pairwise correlations. We plot the dynamics of homogeneity (Figure 15) and completeness (Figure 16) under different time periods. Since each probe request can be exclusively associated with another frame (or none), a single mistaken decision may cause a domino effect on the remaining associations. *NN* considers only the local association optimality and hence tends to fragment the frame sequences, resulting in low completeness. By contrast, *Clustering* achieves high completeness but low homogeneity. The reason is that *Clustering* greedily associates frames. Such unconstrained association tends to merge frame sequences and leads to low homogeneity. Compared with other methods, Espresso employs the minimum-cost flow approach to capture the global optimality and thus achieves a balanced performance between homogeneity and completeness.

Figure 17 demonstrates the impact of the mini-batch interval T in terms of homogeneity and completeness. The mini-batch interval does not significantly affect the performance since the overall standard deviations of homogeneity and completeness are low (0.002 and 0.004, respectively). However, we can still observe that both homogeneity and completeness increase rapidly with the growth of T when T is small ($T < 30$s). With a large T ($T \geq 30$s), the dynamics of both metrics become gentle. On the other hand, the period T also affects the system's responsiveness (i.e., the delay of outputting association results). To balance the association performance and responsiveness, we choose $T = 30$s in the system.

Figure 18 demonstrates the V-measures of the frame sequences constructed by different schemes under various periods. V-measure can reflect the comprehensive performance of frame association. We can observe that Espresso outperforms the others in all the experimental periods. The score of *IEFpr* declines significantly as time goes on. This is because it cannot accurately distinguish between devices under a large number of ambiguous frames and thus results in low homogeneity. We have not included the curve of *TimeSig* in the figure since its V-measure is extremely low (< 0.1). The possible reason is that *TimeSig* cannot well-characterize the inter-frame interval feature from a limited number of frames in probing rounds since the sensors only monitor a single channel.

We finally evaluate the overall performance of the probe requests with virtual MAC addresses (the ground-truth association is captured by external sensors). Figure 19 depicts the homogeneity, completeness and V-measure of the associated frame sequences. Among all the schemes, Espresso achieves balanced and satisfactory homogeneity and completeness (0.874 and 0.811, respectively) with the highest V-measure (0.841). Although *IEFpr* and *SeqThresh* perform slightly better in homogeneity, they have much lower com-

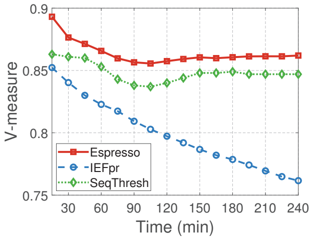


Fig. 18. V-measure of the frame sequences constructed by different schemes.

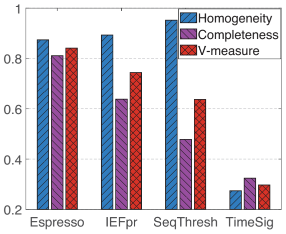


Fig. 19. Performance scores of the frame sequences constructed by different schemes (virtual MAC address).

pleteness and hence lower V-measure. This is because the deterministic methods applied are easily affected by frame ambiguities and noises, causing the incomplete association in the sequences. *TimeSig* again has the worst performance due to the lack of inter-frame interval features. The results are also consistent with the experiments with physical MAC addresses shown in Figure 18.

VI. Related Works

Prior works for frame association leverage device-specified IE contents to fingerprint devices [4], [8], [23], [24]. Specific fields such as transmission rate [23] and service set identifier (SSID) [8], [24] are used to distinguish between devices. Vanhoef *et al.* further explore using the combination of multiple IE fields as fingerprints. They analyze the discriminability of fields and utilize the most discriminative and stable ones to form fingerprints [4]. Though promising, IE alone is not robust against the diversity of devices since different devices (say, smartphones of the same model) may emit probe requests with identical IEs and the same devices may even produce different IEs. By contrast, our work uses the multimodality of probe requests to establish accurate frame association.

In contrast to content-based fingerprinting mentioned above, some works study inter-frame patterns to distinguish between devices. Clock skew is the inherent drifts of device clocks due to variations in the manufacturing process. The uniqueness of skews is explored to differentiate transmitting devices [25], [26]. The works in [9], [27] use the arrival time between frames as a unique pattern to identify devices. Bezawada *et al.* extract features from network traffic to form fingerprints [28]. However, these works require either specialized hardware or a large body of consecutive probe requests and hence are not feasible in commercialized scenarios. On the other hand, the continuity of sequence numbers is another effective indicator for consecutive frames. Frame association can be constructed by predicting the sequence number of the next emitting frame [4], [29], [30]. Espresso also utilizes the sequence number as one of the features in frame correlation estimation. Different from the deterministic methods in previous works, Espresso proposes a probabilistic scheme to adapt to heterogeneous devices and avoid dedicated calibration.

Some works take advantage of the leaks in protocols or system designs to obtain the device identifiers. Probe requests from certain devices may carry the universally unique identifier-enrollees (UUID-Es), which are derived from their MAC addresses. The works in [4], [31] reverse-engineer the UUID-E to recover the original MAC addresses via pre-computed hash tables. The work in [3] infers the Wi-Fi MAC address from the Bluetooth MAC address as lots of manufacturers assign consecutive MAC addresses for the two interfaces of the same phone. Vanhoef *et al.* propose to set up fake APs with usual SSIDs (e.g., “Starbucks” and “Airport”) [4]. Devices will expose their true MAC addresses when they auto-connect to these networks. These schemes rely on special assumptions and hence cannot be generalized for ubiquitous devices. Meanwhile, privacy concerns are also raised as they attempt to acquire the true MAC addresses of devices. On the contrary, Espresso uses only the universal information available on almost every device. Espresso respects user privacy. The constructed association is not able to link to user identities or exposes user locations.

Data association is to find matching between two sets of objects. It is conventionally used in correlating measurements with targets in multi-object tracking problems [32], [33], [34]. Later, the applications have been extended to wilder fields such as multi-sensor data fusion [35], simultaneous localization and mapping (SLAM) [36], [37], person re-identification in visual surveillance system [38], [39], etc. However, none of the prior works has considered the association of multiple probe requests. Espresso proposes a novel and efficient multi-frame association algorithm by formulating it as a minimum-cost flow problem. To the best of our knowledge, this is the first piece of work which applies data association techniques to probe request association.

VII. Conclusion

MAC address randomization has been deployed on modern devices, where randomly generated virtual MAC addresses are used in probe requests. As the MAC address from a single device changes at random intervals, statistical analysis such as people counting, crowd flow estimation and trajectory inference is defeated. In this paper, we present ***Espresso*** to establish the association among probe requests under MAC address randomization. Espresso works on existing Wi-Fi infrastructure without specially-designed hardware, external localization systems, or offline device calibration/training. The scheme consists of two important modules: frame correlation estimator and multi-frame association. The correlation estimator leverages the multimodality of frame attributes, such as information element, sequence number and signal strength, to estimate the association probability among frames. Based on that, Espresso models the multi-frame association as a minimum-cost network flow problem, where the nodes represent the frames to be associated and the edge weights correspond to the frame correlation. We have implemented Espresso and conducted extensive experiments to verify its performance. The experiment in a leading shopping mall shows that Espresso greatly outperforms the state-of-the-art works in terms of discrimination accuracy ($> 80\%$) and V-measure scores (> 0.85).

References

[1] S. He and S.-H. G. Chan, “Wi-Fi Fingerprint-Based Indoor Positioning: Recent Advances and Comparisons,” *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 466–490, 2016.

[2] C. J. Bernardos, J. C. Zúñiga, and P. O’Hanlon, “Wi-Fi internet connectivity and privacy: Hiding your tracks on the wireless Internet,” in *Proceedings of 2015 IEEE Conference on Standards for Communications and Networking*. Tokyo, Japan: IEEE, Oct. 2015, pp. 193–198.

[3] J. Martin, T. Mayberry, C. Donahue, L. Foppe, L. Brown, C. Riggins, E. C. Rye, and D. Brown, “A Study of MAC Address Randomization in Mobile Devices and When it Fails,” in *Proceedings of 2017 Privacy Enhancing Technologies Symposium*, vol. 2017, Oct. 2017, pp. 365–383.

[4] M. Vanhoef, C. Matte, M. Cunche, L. S. Cardoso, and F. Piessens, “Why MAC Address Randomization is Not Enough: An Analysis of Wi-Fi Network Discovery Mechanisms,” in *Proceedings of the 11th ACM on Asia Conference on Computer and Communications Security*, ser. ASIA CCS ’16. Xi’an, China: ACM, 2016, pp. 413–424.

[5] Google, “Privacy: MAC Randomization,” https://source.android.com/devices/tech/connect/wifi-mac-randomization, Jan. 2020.

[6] A. Di Luzio, A. Mei, and J. Stefa, “Mind your probes: De-anonymization of large crowds through smartphone WiFi probe requests,” in *Proceedings of the 35th Annual IEEE International Conference on Computer Communications*. San Francisco, CA, USA: IEEE, Apr. 2016, pp. 1–9.

[7] J. Weppner, B. Bischke, and P. Lukowicz, “Monitoring crowd condition in public spaces by tracking mobile consumer devices with wifi interface,” in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, ser. UbiComp ’16. Heidelberg, Germany: ACM, Sep. 2016, pp. 1363–1371.

[8] P. Robyns, B. Bonné, P. Quax, and W. Lamotte, “Noncooperative 802.11 MAC Layer Fingerprinting and Tracking of Mobile Devices,” *Security and Communication Networks*, vol. 2017, pp. 1–22, May 2017.

[9] C. Matte, M. Cunche, F. Rousseau, and M. Vanhoef, “Defeating MAC Address Randomization Through Timing Attacks,” in *Proceedings of the 9th ACM Conference on Security & Privacy in Wireless and Mobile Networks*, ser. WiSec ’16. Darmstadt, Germany: ACM, 2016, pp. 15–20.

[10] T. He, S.-H. G. Chan, and C.-F. Wong, “HomeMesh: A low-cost indoor wireless mesh for home networking,” *IEEE Communications Magazine*, vol. 46, no. 12, pp. 79–85, Dec. 2008.

[11] IEEE, “Guidelines for Use of Extended Unique Identifier (EUI), Organizationally Unique Identifier (OUI), and Company ID (CID),” IEEE, Tech. Rep., Sep. 2017.

[12] IEEE Standards Association, “802.11-2016 - IEEE Standard for Information technology–Telecommunications and information exchange between systems Local and metropolitan area networks–Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications,” 2016.

[13] F. Zafari, A. Gkelias, and K. K. Leung, “A Survey of Indoor Localization Systems and Technologies,” *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2568–2599, Apr. 2019.

[14] P. Bahl and V. N. Padmanabhan, “RADAR: An In-Building RF-based User Location and Tracking System,” in *Proceedings of the 19th Annual Joint Conference of the IEEE Computer and Communications Societies*, vol. 2. Tel Aviv, Israel: IEEE, 2000, pp. 775–784.

[15] M. Youssef and A. Agrawala, “The Horus WLAN Location Determination System,” in *Proceedings of the Third International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys ’05. Seattle, Washington, USA: ACM, 2005, pp. 205–218.

[16] S. He and S. H. G. Chan, “Sectjunction: Wi-Fi indoor localization based on junction of signal sectors,” in *Proceedings of 2014 IEEE International Conference on Communications*, Jun. 2014, pp. 2605–2610.

[17] W. H. Cunningham, “A network simplex method,” *Mathematical Programming*, vol. 11, no. 1, pp. 105–116, Dec. 1976.

[18] J. B. Orlin, “A polynomial time primal network simplex algorithm for minimum cost flows,” *Mathematical Programming*, vol. 78, no. 2, pp. 109–129, Aug. 1997.

[19] A. V. Goldberg and R. E. Tarjan, “Finding minimum-cost circulations by canceling negative cycles,” *Journal of the ACM*, vol. 36, no. 4, pp. 873–886, Oct. 1989.

[20] Google, “OR-Tools,” https://developers.google.com/optimization/flow/mincostflow, Oct. 2018.

[21] A. Rosenberg and J. Hirschberg, “V-Measure: A conditional entropy-based external cluster evaluation measure,” in *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*. Prague, Czech Republic: Association for Computational Linguistics, Jun. 2007, pp. 410–420.

[22] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, “A Density-based Algorithm for Discovering Clusters in Large Spatial Databases with Noise,” in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, ser. KDD ’96. Portland, OR, USA: AAAI, 1996, pp. 226–231.

[23] J. Pang, B. Greenstein, R. Gummadi, S. Seshan, and D. Wetherall, “802.11 User Fingerprinting,” in *Proceedings of the 13th Annual ACM International Conference on Mobile Computing and Networking*, ser. MobiCom ’07. Montréal, Québec, Canada: ACM, 2007, pp. 99–110.

[24] M. Cunche, M.-A. Kaafar, and R. Boreli, “Linking wireless devices using information contained in Wi-Fi probe requests,” *Pervasive and Mobile Computing*, vol. 11, pp. 56–69, Apr. 2014.

[25] T. Kohno, A. Broido, and K. Claffy, “Remote physical device fingerprinting,” *IEEE Transactions on Dependable and Secure Computing*, vol. 2, no. 2, pp. 93–108, Apr. 2005.

[26] C. Arackaparambil, S. Bratus, A. Shubina, and D. Kotz, “On the Reliability of Wireless Fingerprinting Using Clock Skews,” in *Proceedings of the Third ACM Conference on Wireless Network Security*. Hoboken, NJ, USA: ACM, 2010, pp. 169–174.

[27] J. Franklin, D. McCoy, P. Tabriz, V. Neagoe, J. Van Randwyk, and D. Sicker, “Passive Data Link Layer 802.11 Wireless Device Driver Fingerprinting,” in *Proceedings of the 15th Conference on USENIX Security Symposium*, ser. USENIX-SS ’06, vol. 15. Vancouver, BC, Canada: USENIX, Jul. 2006, pp. 1–12.

[28] B. Bezawada, M. Bachani, J. Peterson, H. Shirazi, I. Ray, and I. Ray, “Behavioral Fingerprinting of IoT Devices,” in *Proceedings of 2018 Workshop on Attacks and Solutions in Hardware Security*, ser. ASHES ’18. Toronto, Canada: ACM, 2018, pp. 41–50.

[29] F. Guo and T.-c. Chiueh, “Sequence Number-Based MAC Address Spoof Detection,” in *Proceedings of 2005 International Workshop on Recent Advances in Intrusion Detection*, A. Valdes and D. Zamboni, Eds. Berlin, Heidelberg: Springer, 2005, pp. 309–329.

[30] G. Chandrasekaran, J.-A. Francisco, V. Ganapathy, M. Gruteser, and W. Trappe, “Detecting Identity Spoofs in IEEE 802.11e Wireless Networks,” in *Proceedings of 2009 IEEE Global Telecommunications Conference*. Honolulu, HI, USA: IEEE, Nov. 2009, pp. 1–6.

[31] J. Martin, E. Rye, and R. Beverly, “Decomposition of MAC Address Structure for Granular Device Inference,” in *Proceedings of the 32nd Annual Conference on Computer Security Applications*, ser. ACSAC ’16. Los Angeles, CA, USA: ACM, 2016, pp. 78–88.

[32] Subhash Challa, Mark R. Morelande, and Robin J. Evans, *Fundamentals of Object Tracking*. Cambridge, UK: Cambridge University Press, 2011.

[33] M. Bredereck, X. Jiang, M. Körner, and J. Denzler, “Data association for multi-object Tracking-by-Detection in multi-camera networks,” in *Proceedings of the Sixth International Conference on Distributed Smart Cameras*. Hong Kong, China: IEEE, Oct. 2012, pp. 1–6.

[34] L. Fan, Z. Wang, B. Cail, C. Tao, Z. Zhang, Y. Wang, S. Li, F. Huang, S. Fu, and F. Zhang, “A Survey on Multiple Object Tracking Algorithm,” in *Proceedings Ot 2016 IEEE International Conference on Information and Automation*. Ningbo, China: IEEE, Aug. 2016, pp. 1855–1862.

[35] B. Khaleghi, A. Khamis, F. O. Karray, and S. N. Razavi, “Multisensor data fusion: A review of the state-of-the-art,” *Information Fusion*, vol. 14, no. 1, pp. 28–44, Jan. 2013.

[36] J. Fuentes-Pacheco, J. Ruiz-Ascencio, and J. M. Rendón-Mancha, “Visual simultaneous localization and mapping: A survey,” *Artificial Intelligence Review*, vol. 43, no. 1, pp. 55–81, Jan. 2015.

[37] S. L. Bowman, N. Atanasov, K. Daniilidis, and G. J. Pappas, “Probabilistic data association for semantic SLAM,” in *Proceedings of 2017 IEEE International Conference on Robotics and Automation*. Singapore: IEEE, May 2017, pp. 1722–1729.

[38] R. Vezzani, D. Baltieri, and R. Cucchiara, “People reidentification in surveillance and forensics: A survey,” *ACM Computing Surveys*, vol. 46, no. 2, pp. 29:1–29:37, Dec. 2013.

[39] Q. Leng, M. Ye, and Q. Tian, “A Survey of Open-World Person Re-Identification,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 4, pp. 1092–1108, Feb. 2019.