Introducing benchmarks for evaluating user-privacy vulnerability in WiFi

Abhishek Kumar Mishra, Aline Carneiro Viana, Nadjib Achir

Introducing benchmarks for evaluating user-privacy vulnerability in WiFi

Abhishek Kumar Mishra*, Aline Carneiro Viana*, Nadjib Achir*†
* *Inria*, France † *University Sorbonne Paris Nord*, France

***Abstract*—WiFi-based crowdsensing is a major source of data in a variety of domains such as human-mobility, pollution-level estimation, and, opportunistic networks. MAC randomisation is a backbone for preserving user-privacy in WiFi, as devices change their identifiers (MAC addresses). MAC association frameworks in the literature are able to associate randomized MAC addresses with a device. Such frameworks facilitate the continuation and validity of works based on device-based identifiers. In this paper, we first question and verify the reliability of these frameworks with respect to the datasets (scenarios) used for their validation. Indeed, we observe a substantial discrepancy between the performances obtained by these frameworks when confronting them with different contextual environments. We identify that the device heterogeneity in the input scenario is privacy-preserving. Henceforth, we propose a novel metric: *randomization complexity*, capable of successfully catching the *degree of randomization* in evaluated datasets. Existing and new frameworks can thus be *benchmarked* using this metric to ensure their *reliability* for any datasets with similar or lower randomization complexities. Finally, we open discussions on the potential impact of the benchmarks in the domain of MAC randomization.**

***Index Terms*—Crowdsensing, MAC randomisation, MAC association frameworks, Privacy-preserving**

I. Introduction

The growth and ubiquity of smart devices supporting WiFi brought meaningfulness and higher reliability to crowdsensing and consequently, improved effectiveness of related services and applications [1]–[4]. In this context, passive sniffing consisting of deployed wireless and static sensing devices – i.e., referred here as sniffers – in a geographical area is considered as a powerful crowdsensing technique for users' mobility data collection. Indeed, passive wireless measurements are non-intrusive and cheaper to be implemented. Such passive measurements rely on the collection of WiFi frames, called probe-requests, sent by mobile smart devices looking for close-by networks (e.g., WiFi access points), also known as active scanning. Probe-requests serve as the source of data in WiFi-based crowdsensing.

In order to preserve user-privacy, WiFi standards force mobile devices to periodically change (randomize) their MAC addresses [5] announced during their active scanning. The MAC randomization process hardens the users' device identification in the crowd, i.e., association between MAC addresses broadcasted in probe-requests and the emitting device.

This work has been partially funded by the ANR MITIK project, French National Research Agency (ANR), PRC AAPG2019.

MAC randomization affects the continuation and validity of important works in domains relying on device-based identifiers, such as user trajectory inference [6], [7] and crowd flow estimation [8], from public packets. Recent literature heavily investigates the *MAC address association*, i.e. co-relating randomised MAC addresses emitted by a particular device [9]–[13].

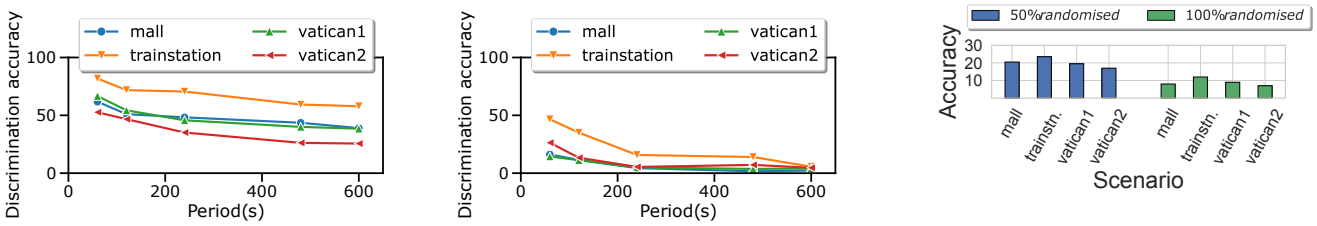
MAC association opens critical issues on users privacy and henceforth bringing the necessity for the understanding of such association, giving insights on the potential improvement to currently vulnerable WiFi standards, thus bringing stronger user privacy protection. The challenging issue concerning association strategies is identifying their weaknesses and effectiveness, which are not available. This paper deals with this lack, which for the best of our knowledge, is the first work in literature.

Broad strategies for association utilise from probe-requests: i) their sequence numbers (i.e., SEQ) [9], [10], ii) their fields (e.g., information elements(IE)) [10]–[12], iii) parameters identified from the communication patterns (e.g., inter-burst time (IBT)) [13], and iv) their RSSI values [10].

All these frameworks claim to associate randomized MAC addresses to the same device with high accuracy measures in their respective evaluation datasets. For instance, [10] obtain a discrimination accuracy $> 80\%$ inside a shopping mall while [13] get accuracy of up to 75% in laboratory settings. [12] construct transmitter fingerprint that is 80% to 67.6% percent unique for 50 to 100 observed devices in a music festival and laboratory scenarios with a varied number of devices.

Despite the promise shown by these address association frameworks, we show in this paper their performances are highly sensitive to the input datasets, also showing there is space for WiFi standard's improvement. In this context, the ***unreliable*** *nature of association frameworks call for the need of* ***benchmarks*** *for measuring their real ability in providing a certain level of performance with respect to any new datasets (scenarios). This ensures WiFi users' privacy in a given scenario.*

We introduce the benchmarks for WiFi privacy by giving a general characterization for the effectiveness of MAC address association. We illustrate the overview of the process of obtaining benchmarks in Fig. 2. *First*, we explore and verify this unreliability and identify their causes. *Second*, we show the impact of such causes on MAC address randomization. *Third*, we introduce a novel metric (conflict size) to ensure



(a) *Information elements (IE)* (b) *Sequence numbers (SEQ)* (c)

Fig. 1: Case studies: (a-b) Infocom2021 [10], (c) WiSec16 [13].

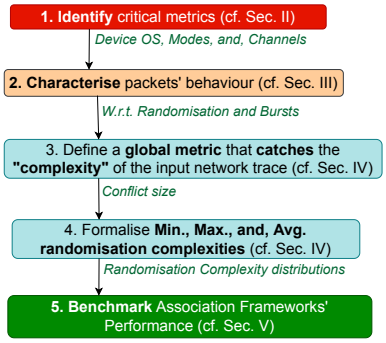


Fig. 2: Obtaining benchmarks for MAC address association

the reliability of frameworks. *Fourth*, we use this metric to formalise the *complexity* that a framework is handling with respect to an input dataset. *Finally*, we discuss and validate the potential of randomisation complexities on acting as an association framework's performance-benchmarks.

The introduced randomisation complexity distributions in the last step act as a *yardstick* for the level of user-privacy breach, when an adversary utilises a particular address association framework. Benchmarks enable the understanding of the usefulness of MAC association frameworks, while knowing their limitations, which is privacy preserving for users.

II. Identifying the Unreliability

We identify and validate the unreliability in current MAC association frameworks.

A. Case Studies

We consider two literature works as case studies for testing the resilience of address association frameworks which we refer as: 1) Infocom2021 [10] and ii) WiSec16 [13], against various data-collection scenarios. The case studies cover all association strategies: usage of SEQ, IE, packet-reception timings, and RSSI - based signatures (cf. Sec. I).

Datasets used in the literature: Infocom2021 uses probe-requests captured inside a shopping mall for evaluation by deploying multiple sensors. It consists of around 5000 unique MACs per hour. However, only the devices transmitting their true (physical) MAC address are considered for evaluating the framework. On the other hand, authors evaluate WiSec16 on a dataset collected inside a laboratory consisting of probe-requests with non-randomized addresses. They partially randomized their dataset (100 out of 550 captured devices) by manually changing MAC addresses four bursts (cf. Section III-A) of probe-requests.

Evaluation with a common dataset: Both frameworks consider different datasets, and to compare the two case studies, we have to compare them with the same input probe-request dataset. Moreover, we need a *ground truth* of randomized MAC addresses emitted from the same device. Hence, we use the public anonymized *Sapienza* trace [1], consisting of about 11 million probe-requests, passively collected in eight different contextual scenarios. We choose datasets from five scenarios: *trainstation*, *themall*, *vatican 1*, and, *vatican 2*. The first two showcase highly frequented public spaces, while the remaining denote dense outdoor environments. The dataset consists of the true MAC addresses of devices. We randomize the dataset by changing the addresses every 4 bursts, while considering various percentages of captured MAC addresses and keeping the *ground truth* of original addresses.

Infocom2021 case-study: Using diverse scenarios, we evaluate two major components of the signature introduced by authors in [10]: the IE and the SEQ. We do not consider the third component: RSSI, as it requires multiple geographically-known sniffers around each transmitting device while contributing very less (drops up to 10%) to the effectiveness of the signature (Fig. 13, [10]). The effectiveness metric utilised in Infocom2021 is *discrimination accuracy*. Authors define discrimination accuracy as the ratio of the correct association, estimated from 1000 randomly selected probes (P_i) and their previous probes in the *Period* ($[t_i - \tau, t_i]$). They vary the period between (0, 600s).

WiSec16 case-study: Similarly, we proceed with the performance analysis of the timing-based signature introduced in [13], under different scenarios. They propose distance metrics based on timing (inter-frame arrival time) and use them together with an incremental learning algorithm to group probes. It claims up to 75% partially randomized traces collected in a laboratory environment.

Observations: Regarding Infocom2021, we can observe in Fig. 1a and 1b that discrimination accuracy of IE and SEQ

[1] https://crawdad.org/sapienza/probe-requests/20130910

based signatures vary significantly across input datasets. For instance, we can see that in relatively more static environments (*train-station*) the performances are good for small periods. However, when the environment starts to be more dynamic the performances falls considerably. The accuracy of IE in *vatican2* drops to 23 percent while it goes close to zero when considering probes in the period of 600s. The reported discrimination accuracy for the same period, in Infocom2021 is 44% and 19%, for IE and SEQ-based signatures, respectively. This shows that the parameter period, highly impacts the results in all scenarios.

Considering WiSec16, we notice in Fig. 1c that the proposed association framework performs variably to the difference in scenarios and the degree of MAC randomization present in the collected probe-requests trace. Also, the achieved accuracy is significantly lower than the claimed best performance (75%) in fully randomized datasets.

Awareness requirement: We observe that the performance of the studied association frameworks is considerably variable. We argue this variability induces the unreliability in the obtained accuracy. Such observation leads us to the essential need for phenomena inference in input datasets, i.e., contextual inference of scenarios.

B. No heterogeneity characterisation

We identify a major issue in the validation of current association frameworks: **a lack of *characterization* of the *heterogeneity*** in terms of the burst's and the randomisation's behaviour (cf. Sec. III-A) in utilised probe-request datasets. The *model* and the *user-usage patterns* of a device introduces this heterogeneity that influences the generation of probe-request packets (cf. Sec. III). The *model* of devices denote the class of devices along with its manufacturer (e.g., smart watches, mobile phones with Android, iOS) while the *usage patterns* of users refer to the *mode* of the device when emitting a probe-request (e.g., screen ON/OFF).

The reason for the variation in the probe-request sending rates and the MAC randomization is that every manufacturer independently decides these behaviors to ensure a good trade-off between the network experience for users and the device's performance (e.g., battery life).

The main issues we identify are: **First**, the literature does not use the same dataset (to ensure *homogeneity*) or any kind of *benchmark* to show the trustworthiness of the obtained framework's performance (cf. Sec. IV). **Second**, there is general lack of *ground truth* of randomized MAC addresses with respect to the sending device. The first issue causes looking only at a *selective view* to the framework's performance. In contrast, the second issue leads to the usage of indirect accuracy metrics such as discrimination accuracy (as in Infocom2021). **Third**, there is a lack of analysis of the performance with respect to varying *intensity* of MAC address changes that a data-collection scenario can capture (cf. Sec. V). The more the *intensity* of address swaps, the more difficult, in general, it is for a framework to associate MACs.

The need: We claim the need for a metric to *characterize* the device *heterogeneity* in terms of the burst's and randomization's behaviors. Hence, we introduce the *randomisation complexity* to bring the notion of *benchmarks* for validating the reliability of association frameworks.

III. Impact of Heterogeneity

We investigate the variability of MAC randomization in WiFi probe-requests with respect to *heterogeneity* introduced by the pool of current mobile devices. The three identified factors that make the emission of probe-requests heterogeneous and subsequently have an impact on the address randomization in WiFi are: i) the device's *model*, ii) the device's *mode*, and, iii) the transmission channel. While the *model* captures the device *heterogeneity* with respect to randomisation behaviour, we observe that the *mode* is the one that has the most impact on the probe-request generation, and hence in the following, we will focus mainly on the device *mode* for characterising probe-requests.

The probe-requests are susceptible to fingerprinting due to: 1) the nature of its *bursts* and 2) the randomising strategy of its MAC address. In the following, we evaluate probe-requests' susceptibility with respect to heterogeneity introduced by device's *models* and *modes*.

A. Probe-request bursts

We can not analyze probe-requests' behavior using *Sapienza* datasets used for the case studies (Sec. II-A) as we need a wide range of current mobile devices with additional information such as device's *mode* when transmitting. Therefore, we used a probe-request dataset [14]. This dataset has 22 popular device models in practice, which were sniffed when present in various device *modes* (see Tab. I). This is the first open-source probe-requests dataset, *labelled* with the *ground truth* of randomised addresses. It contains 20-minute duration captures of known devices using a Raspberry Pi based sniffer.

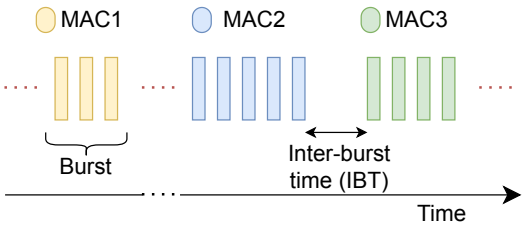


Fig. 3: A device's randomised probe-requests.

Mode	Screen ON	Power-saving ON	WiFi ON
A	Yes	No	Yes
S	No	No	Yes
PA	Yes	Yes	Yes
PS	No	Yes	Yes
WA	Yes	No	No
WS	No	No	No

TABLE I: Device's *modes* [14]

There are active-screen modes (A, PA, and WA) and inactive-screen modes (S, PS, and WS). In power-saving

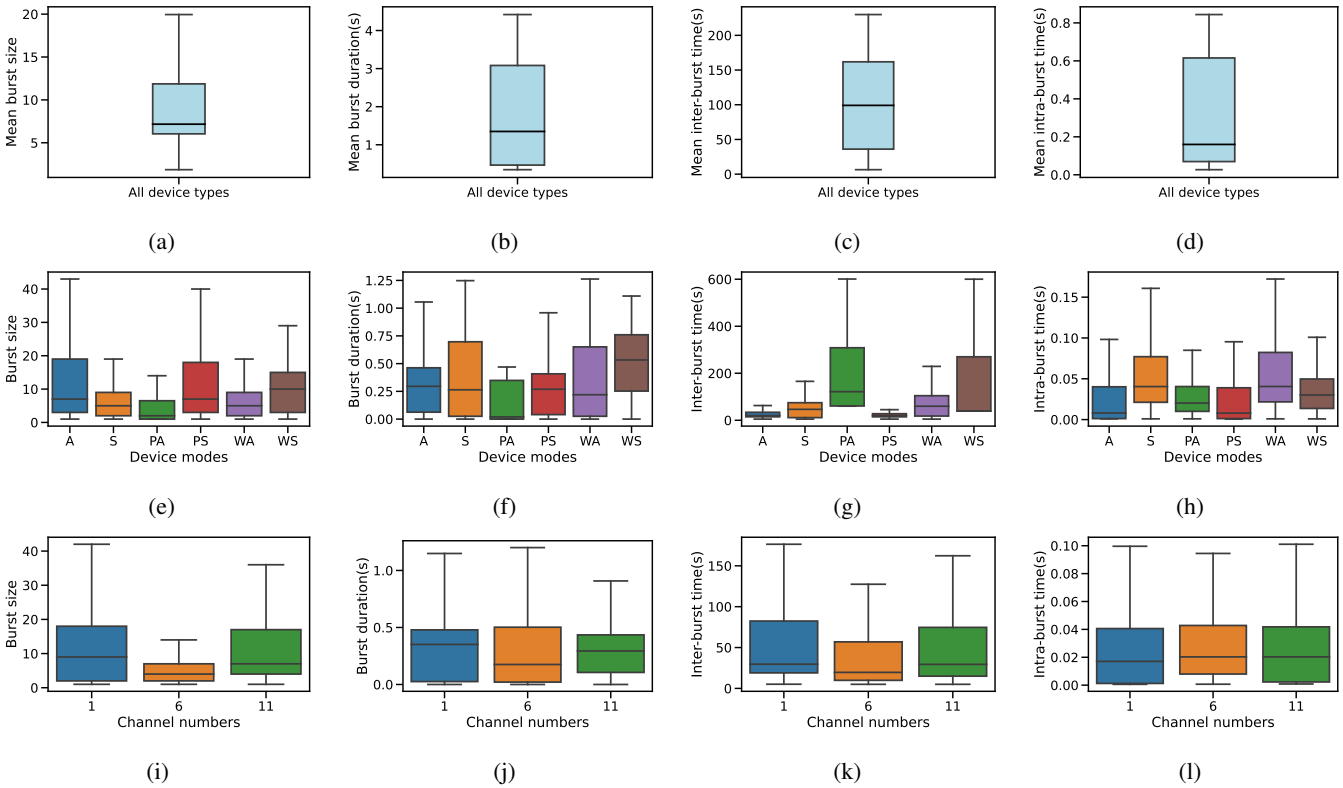


Fig. 4: Analysis of probe-requests' burst behaviours – i.e., size, duration, Inter-burst time, and Intra-burst time – under varying: (a-d) *Device model's*, (e-h) *Device modes*, and, (i-l) *Channels*.

modes (PA and PS), the device additionally keeps the power-saving setting active, while in WA and WS modes, the device also has the Wi-Fi interface switched off. Each device configuration is observed in the three non-overlapping channels (1,6, and, 11) of the 2.4 GHz frequency band. The details of capture conditions for each device modes can be found in [14]. In the following, we discuss the metrics concerning Wi-Fi probe-request emissions that facilitate the association.

Fig. 3 illustrates the probe-requests sent from a device over time in the active scanning process. Devices send probes on every available channel to discover available networks in proximity. Each probing round consists of a *burst* of frames. The burst size is variable, depending mainly on the number of available channels. The period between successive bursts, inter-burst time (IBT), is variable to device models and its operating *modes*. In essence, the temporal behavior of WiFi probe-requests can be characterized by four metrics:

Bursts' size: The average number of frames in a burst when considering all device types (models) varies considerably with the mean value of 7 (see Fig. 4a). We further analyze a device's behavior per mode in Fig. 4e. We observe that devices send many frames in modes (A and PS). Intuitively, devices reduce the burst sizes to the minimum in the mode with active-screen and power saving mode (PA).

Bursts' duration: The average duration of a probe-request burst, across device types varies around $1s - 3s$, as shown in Fig. 4b. It is interesting to note in Fig. 4f that burst duration is relatively lower in active-screen modes (A and PA) than in static-screen modes (S and WS). This can be attributed to devices conserving energy when under constraints.

Inter-burst time (IBT): The inter-burst time is typical to device manufacturer's factory configurations. Hence, across all metrics we consider to characterize probe-requests, it shows the most variation across the device types. Fig. 4c illustrates that IBT alters from a few seconds to more than 200 seconds. Moreover, we notice in Fig. 4g that IBT is the lowest in active-screen mode (A) while the highest when the power-saving mode is additionally turned ON (PA).

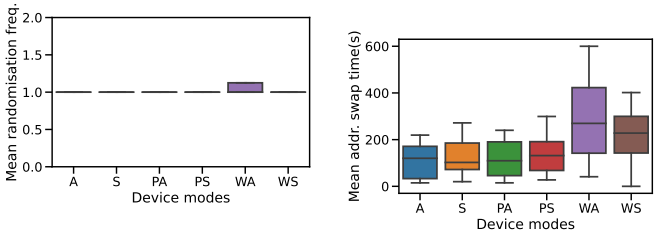
Intra-burst time: Finally, we investigate the time between frames in a single burst. We observe in Fig. 4d that it goes up to $0.8s$. When looking at the effect of device modes on the intra-burst time, we could notice in Fig. 4h that devices tend to send frames quickly in a burst in the active-screen mode (A) than in the static-screen mode (S).

We also observe a slight variation in the burst's size and duration as well as IBT and intra-burst time with respect to transmission channels (see Fig. 4i - 4l).

B. Behavior of WiFi MAC addresses

The randomization strategy of a device's MAC address is equally critical to the temporal behavior of probe-request that we just discovered. In WiFi, most of the current devices change their MAC identifiers after a period of time. We notice that

18 of 22 considered major device types in the dataset [14] do randomize their MAC addresses. Fig. 5a illustrates that most devices change their MAC address after a burst, irrespective of device modes. Moreover, the average address swap times is low and almost similar across device modes, except when the WiFi interface is switched OFF (WA and WS), as shown in Fig. 5b.



(a) Mean randomisation frequency (num. of bursts)

(b) Mean address swap time

Fig. 5: Behaviour of MAC addresses

We have two key conclusions:

1st conclusion: The device *heterogeneity* does play a major role in WiFi MAC address randomisation. We could see that all broad types of signatures are based on metrics that vary with device *types* and *modes*. Signatures based on temporal-information of probe-requests [13] rely upon the behavior of bursts, while those based on frame fields [10] depend upon device-specific frame fields. Finally, sequence numbers-based signatures [9], [10] also subsequently depends upon the burst's size and IBT of individual devices, as it dictates the probability of sniffers recording a device's frames.

2nd conclusion: MAC association frameworks need a representative evaluation. As we just discussed, we must *catch* device's population (models) as well as modes present at the time of dataset collection. We have to *ensure* that the this dataset used for the evaluation of a address association framework, should maintain the performance in *similar* scenarios. We need a metric that quantifies this *similarity* between data collection scenarios. In the following section, we introduce *randomisation complexity*, a metric that enables us to compare two datasets with respect to the *difficulty* that a framework has to face, while associating contained MAC addresses. This metric hence allows the evaluation of various association frameworks to be comparable.

IV. Introducing Benchmarks

We build upon our conclusions to introduce the formalization on *benchmarks* for the evaluation of MAC association frameworks. The proposed benchmarks are *generic* and are *representative* of the *complexity* in the used input dataset.

We believe such formalization: (i) Brings new ways of interpreting devices' randomisation behavior, (ii) Opens paths for designing adaptive randomisation techniques by the standard. For instance, devices can randomise more frequently in situations with sparse nearby device population to maintain a high *complexity* for a adversary, and, (iii) And off-course allow association frameworks to be more robust and reliable.

MAC association as resolution of *conflicts:* *Conflicts* ($\mathcal{C}$) refer to changing MAC addresses in a given time period. They are either caused by devices which stop emitting during the MAC changing process or due to their entry/exit from the sniffing range. We refer to this time period as *conflict period* (T_c). We illustrate conflict periods in Fig. 6. Dotted lines in different colors represent different appearing and disappearing MAC addresses sent by devices, in a T_c.

Any address association framework in essence has to resolve these conflicts and perform correct assignment between the disappearing and appearing MAC from individual devices. Each MAC address M is a bunch of probe-requests with a start (M^{start}) and a stop (M^{stop}). For instance, a disappearing MAC, M_j is said to be in conflict (illustrated in Fig. 6) with an appearing MAC, M_k if:

$$\mathcal{C} : M_k, M_j \mapsto (i-1)T_c < M_j^{stop}, M_k^{start} \leq iT_c \quad (1)$$

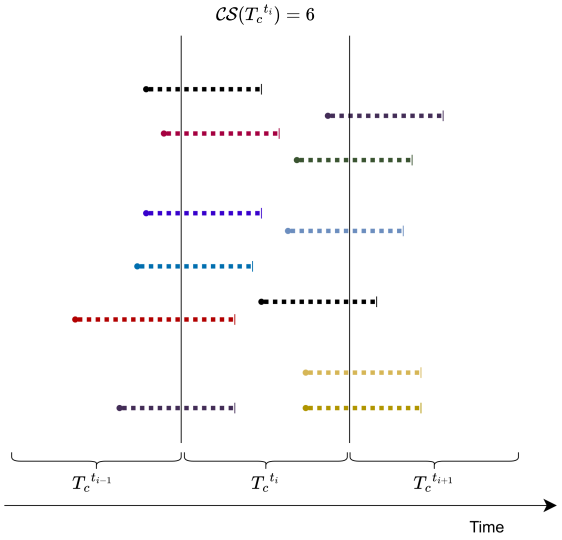


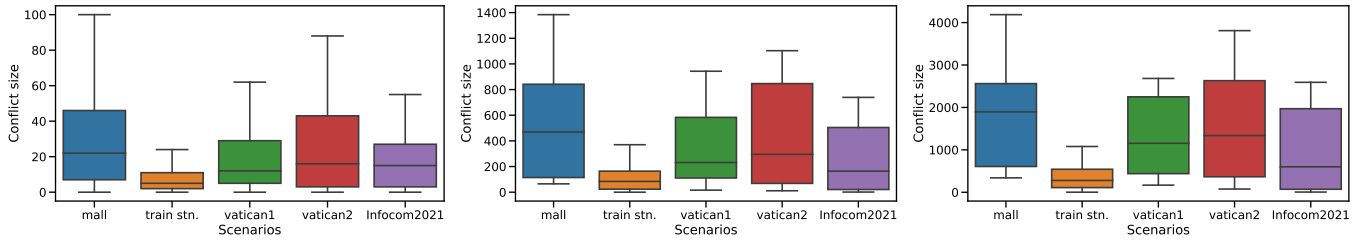
Fig. 6: An illustration of conflict periods (T_c)

We obtain benchmarks from the following three steps:

1. Determining the *conflict period*: We denote set of modes as $\mathcal{M}$, the set of device models $\mathcal{D}$, and the transmitting channels (frequency bands) as $\mathcal{F}$. The conflict period (T_c) should be such that we let the association framework consider possible associations. We observe in Sec. III-B that a device is likely to change its MAC after a burst. Hence, we must consider T_c to be at least of the inter-burst time (IBT).

First, we vary the device types (d_j) and channels (f) while keeping the mode fixed to obtain the min., max. and, avg. values of IBT per mode (IBT_m) [Eq. 2]. Then, we vary modes and take the corresponding min., max. and, avg. of the union set, to obtain the T_c's minimum, maximum and, average values [Eq. 3]. We find the $T_c^{(min.,avg.,max.)}$ to be $(5.00s, 95.72s, 365.42s)$ using the *labelled* dataset [14].

$$IBT_m{}^{(min.,avg.,max.)} = (min., avg., max.) \bigcup_{d_j \in \mathcal{D}, f \in \mathcal{F}} IBT_{(d_j,f)_m} \quad (2)$$



(a) Min. randomisation complexity (b) Avg. randomisation complexity (c) Max. randomisation complexity

Fig. 7: Randomisation complexities

$$T_c^{(min.,avg.,max.)} = (min., avg., max.) \bigcup_{m \in \mathcal{M}} {IBT_m}^{(min.,avg.,max.)} \tag{3}$$

2. Obtaining the *conflict size*: We introduce a global metric: *conflict size* ($\mathcal{CS}$), which is capable of catching the "complexity" of the any input network trace. We define conflict size as half the number of address trails (M^{start}'s and M^{stop}'s), in a conflict period (T_c). For instance, as shown in Fig. 6, the conflict size ($\mathcal{CS}$) is 6 for ${T_c}^{t_i}$. Higher $\mathcal{CS}$ makes it difficult for frameworks to *resolve* conflicts.

3. Inferring the *randomisation complexity*: We define the corresponding minimum, maximum, and, average: randomisation complexities ($\mathcal{RC}^{(min.,avg.,max.)}$) as the set of conflict sizes ($\mathcal{CS}(T_c^{(min.,avg.,max.)})$) when considering corresponding conflict periods (t_i), over all slots ($\mathcal{S}$) in the considered dataset [Eq. 4].

$$\mathcal{RC}^{(min.,avg.,max.)} = \bigcup_{i \in \mathcal{S}} \mathcal{CS}(T_c^{(min.,avg.,max.)})^{t_i} \tag{4}$$

Benchmarks: The randomisation complexities act as a *benchmark* for the lower, upper, and, average performance limits of address association frameworks. The calculated values of $T_c^{(min.,avg.,max.)}$ in this paper can be used to obtain randomisation complexities of any input dataset by just following the steps (2) and (3) above. New frameworks in literature, can accompany their frameworks with our benchmarks to ensure their *reliability* for any new scenarios with similar or lower complexities.

V. Discussing benchmarks

In Fig. 7, we report the min., max., and avg. values in complexity of Infocom2021 dataset and of the *Sapienza* datasets. We notice that the randomisation complexity and the framework's performance are inversely proportional, as expected. We observe that scenarios: *vatican2* and *mall* have relatively higher complexities than the other ones.

When considering Infocom2021, we can clearly observe in Fig. 1a and 1b that the discrimination accuracy obtained by both signature metrics: IE and SEQ, degrade with increasing randomisation complexities. The scenario *trainstation* for instance, obtains the best relative performance as it exhibits the lowest randomisation complexity in Fig. 7. We observe a similar trend of the high and low accuracy in WiSec16, for the scenarios: *trainstation* and *vatican2* as shown in Fig. 1c. These two scenarios show a relatively lower and greater randomisation complexities respectively, as expected.

This certifies our *benchmarks* of *adequately* catching the *complexity* of the dataset. Both frameworks are expected to have the same association accuracy when evaluated with another input dataset with similar randomisation complexity.

Future work: **First**, we plan to demonstrate the complexities of other address association frameworks in literature in order to *benchmark* its performance over different data-collection scenarios. **Second,** currently the *ground-truth* for MAC addresses emitted from the same device is also lacking for passively collected datasets, which hinders the usage of a direct association accuracy's metric. We need a data generation methodology to vary these devices' *modes* and *models*. We aim to demonstrate our ongoing solution to the issue of *ground-truth*, by linearly combining traces (in time) of different individual devices captured inside the Faraday cage, while considering realistic packet losses and sojourn time of devices. This will facilitate obtaining *direct* accuracy metric and benchmarks, over custom scenarios, by varying the linear combinations. **Finally**, we plan to investigate the rate of degradation of a association framework's performance with respect to randomisation complexities in future, when considering various classes of MAC association strategies.

References

[1] A. Das, K. Narayan, and S. Chakraborty, "Leveraging ambient sensing for the estimation of curiosity-driven human crowd," in *2022 IEEE International Systems Conference (SysCon)*, pp. 1–8, 2022.

[2] K. N. Choi, T. Dahanayaka, D. Kennedy, K. Thilakarathna, S. Seneviratne, S. S. Kanhere, and P. Mohapatra, "Poster abstract: Passive activity classification of smart homes through wireless packet sniffing," in *2020 19th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, pp. 347–348, 2020.

[3] Z. Koh, Y. Zhou, B. P. L. Lau, C. Yuen, B. Tuncer, and K. H. Chong, "Multiple-perspective clustering of passive Wi-Fi sensing trajectory data," *IEEE Transactions on Big Data*, pp. 1–1, 2020.

[4] B. Huang, G. Mao, Y. Qin, and Y. Wei, "Pedestrian flow estimation through passive WiFi sensing," *IEEE Transactions on Mobile Computing*, vol. 20, no. 4, pp. 1529–1542, 2021.

[5] A. K. Mishra, A. Carneiro Viana, N. Achir, and C. Palamidessi, "Public wireless packets anonymously hurt you," in *2021 IEEE 46th Conference on Local Computer Networks (LCN)*, pp. 649–652, 2021.

[6] M. Čavojskỳ, M. Uhlar, M. Ivanis, M. Molnar, and M. Drozda, "User trajectory extraction based on wifi scanning," in *2018 6th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW)*, pp. 115–120, IEEE, 2018.

[7] M. Kotaru and S. Katti, "Position tracking for virtual reality using commodity wifi," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 68–78, 2017.

[8] 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*, pp. 1363–1371, 2016.

[9] F. Guo and T.-c. Chiueh, "Sequence number-based mac address spoof detection," in *International Workshop on Recent Advances in Intrusion Detection*, pp. 309–329, Springer, 2005.

[10] J. Tan and S.-H. Gary Chan, "Efficient association of wi-fi probe requests under mac address randomization," in *IEEE INFOCOM 2021 - IEEE Conference on Computer Communications*, pp. 1–10, 2021.

[11] 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*, pp. 413–424, 2016.

[12] 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, 2017.

[13] C. Matte, M. Cunche, F. Rousseau, and M. Vanhoef, "Defeating mac address randomization through timing attacks," WiSec '16, (New York, NY, USA), Association for Computing Machinery, 2016.

[14] L. Pintor and L. Atzori, "A dataset of labelled device wi-fi probe requests for mac address de-randomization," *Computer Networks*, vol. 205, p. 108783, 2022.