



**HAL**  
open science

# Assessing the Completeness of Passive Wi-Fi Traffic Capture

Mohammad Imran Syed, Anne Fladenmuller, Marcelo Dias de Amorim

► **To cite this version:**

Mohammad Imran Syed, Anne Fladenmuller, Marcelo Dias de Amorim. Assessing the Completeness of Passive Wi-Fi Traffic Capture. 2022 International Wireless Communications and Mobile Computing (IWCMC), May 2022, Dubrovnik, Croatia. pp.961-966, 10.1109/IWCMC55113.2022.9824970 . hal-03721138

**HAL Id: hal-03721138**

**<https://hal.science/hal-03721138v1>**

Submitted on 4 Nov 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Assessing the Completeness of Passive Wi-Fi Traffic Capture

Mohammad Imran Syed, Anne Fladenmuller, and Marcelo Dias de Amorim  
 Sorbonne Université, CNRS, LIP6, Paris, France  
 {mohammad-imran.syed, anne.fladenmuller, marcelo.amorim}@lip6.fr

**Abstract**—The passive capture of Wi-Fi traces using sniffers is a cost-efficient and disturbance-free technique to assess the wireless activity of a target area. However, because of the inherent characteristics of the wireless medium, a sniffer is likely to miss Wi-Fi packets leading to incomplete traces. In this paper, we formulate the notion of *relative completeness* and investigate it experimentally. We consider anonymized Wi-Fi traces from 10 co-located sniffers in residential and office areas (different intensities of Wi-Fi traffic). We observe that individual sniffers lead to low completeness. Consequently, it is necessary to increase redundancy by packing several sniffers together (which we call a super-sniffer) to gather more complete traces. We observe that the results depend not on the hardware but on the environment. The results improve by increasing the size of the super-sniffer irrespective of the scenario.

**Index Terms**—wireless, passive measurements, completeness

## I. INTRODUCTION

Air is the preferred and winning medium of choice because of portability, affordability, and ever-increasing data rates. Wireless networks are everywhere, and understanding their behavior to improve their performance is of utmost importance [1, 2, 3, 4]. Nevertheless, measuring wireless traffic (wirelessly) is challenging because of the intrinsic volatile nature of the wireless links [5]. Actively collecting traffic is burdensome because it requires deploying probes at different nodes, which are likely under the control of various administrative entities. For example, to gather traffic from smartphones, one can either deploy probes at all access points the user associates with or create a measurement application and ask users to install it on their devices. The users that volunteer might be an insufficient sample of the population, leading to inaccurate results.

An efficient alternative is to run passive measurements by deploying several *sniffers* (devices collecting wireless packets in monitor mode) throughout the target area [6, 7, 8].<sup>1</sup> It is a low-cost and scalable measurement strategy that does not require bothering users with intrusive services. The concept of the ANR-MITIK project, which we are part of, is to infer contact traces through non-intrusive methods like passive sniffing [9]. Nevertheless, due to wireless transmission constraints like multi-path, fading effects, or collisions, there is no guarantee a single sniffer can capture all the packets, therefore, leading to incomplete traces. In Figure 1, we illustrate a

<sup>1</sup>It is, however, essential to know which data one can sniff depending on the location of the measurement campaign while preserving the privacy of the users.

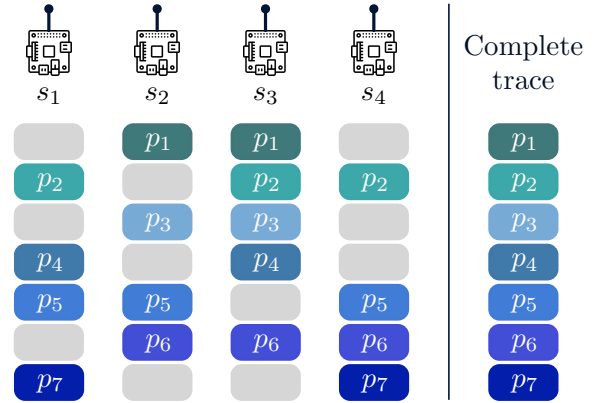


Fig. 1: Trace completeness. Because of the nature of the wireless medium, sniffers miss some packets. We need to combine individual traces to get as close as possible to the complete trace.

typical scenario where four sniffers ( $s_1, \dots, s_4$ ) do not have the same “view” of the wireless traffic because of capture misses. It leads to discrepancies in the measurements, and further analyses relying on such incomplete traces are likely to be flawed.

The solution to circumvent the problem relies on the use of *super-sniffers*. It consists of introducing redundancy in the system by tying two or more sniffers together to increase the probability that at least one of the sniffers captures a packet. In Figure 2, we show a three-redundant super-sniffer, where each individual sniffer is composed of a Raspberry Pi computing unit and an Alfa antenna. The main question that we address in this paper is *how the level of redundancy helps improve the quality of the measure*. To provide an answer to this question, we propose a definition of a trace’s *relative completeness* and evaluate it through real-world experiments. Although we focus on Wi-Fi traces, our methodology is general and can apply to other technologies.

We evaluate the quality of capture of individual sniffers as well as super-sniffers of up to ten-redundancy. We use sniffers composed of two models of Raspberry Pi (3B and 4B) to ensure diversity in the capturing devices. We take into consideration two different scenarios, office and residential, with varying traffic loads. The individual sniffers achieve relatively low completeness (between 30% and 54% depending on the

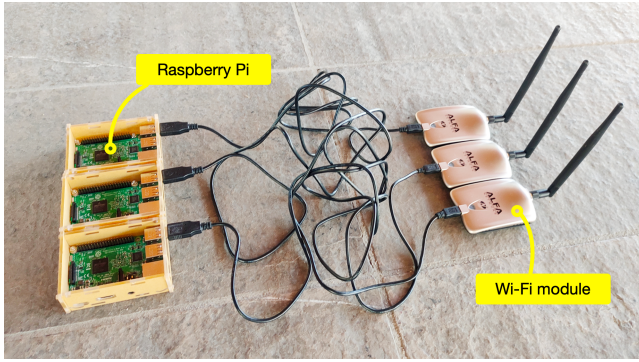


Fig. 2: Three sniffers forming a super-sniffer.

scenario), which confirms the need for redundancy. Secondly, commercial off-the-shelf (COTS) devices such as Raspberry Pi are powerful enough to play the role of a sniffer in each scenario. The sniffers uniformly miss packets, indicating that the environment essentially dictates the quality of the sniffing process.

As summary, the contributions of this paper are:

- **Metric for completeness.** We propose a formal definition of completeness that incorporates the notion of redundancy. We also define a measurement methodology to help network analysts better characterize their wireless environment.
- **Experimental evaluation.** We adopt an experimental approach to evaluate the behavior of the completeness metric under multiple network conditions. We confront the completeness with the network conditions, in particular in terms of traffic load.
- **Environment dependence.** We show that the results are not hardware but environment/scenario dependent.

The rest of the paper is organized as follows. In Section II, we define the *relative completeness* in passive measurements. We describe the experimental methodology in Section III and the measurement campaign that we have conducted in Section IV. In Section V, we provide practical evidence of the need for redundancy when sniffing traffic. We assess such redundancy in Section VI. We postpone the related work to Section VII so that the reader has enough material to understand how our work contributes to the state of the art. We conclude the paper and list some open issues in Section VIII.

## II. COMPLETENESS IN PASSIVE MEASUREMENT

The quality of a passive measurement system improves as the redundancy level of a super-sniffer increases. In Figure 1, for example, it is possible to obtain the complete trace through the combinations  $s_1 + s_2$  or  $s_3 + s_4$ , but not  $s_2 + s_3$ . We need to estimate how much the level of redundancy of the super-sniffer impacts the number of packets captured because adding more and more sniffers to a super-sniffer comes at a financial as well as management cost. To this end, we propose the notion of *relative completeness*.

**Relative completeness.** We define the *relative completeness* as the amount of traffic that we capture relative to the union of all traces from all the individual sniffers that participated in the capture (which explains why we use the term “relative”).

Before formally defining relative completeness (in Section II-B), let us first introduce the notion of super-sniffers.

### A. Super-sniffers

Let  $S = \{s_1, s_2, \dots, s_M\}$  be the set of  $M$  sniffers that we have at our disposal to compose a super-sniffer,  $T_{s_i}$  be the trace (i.e., set of packets) captured by sniffer  $s_i \in S$ , and  $\mathcal{T} = \{T_{s_1}, T_{s_2}, \dots, T_{s_M}\}$ .

We define  $\pi^m$  as a subset of  $m$  elements of  $\mathcal{T}$  and  $\Pi^m$  be the set of all instances of different combinations of  $\pi^m$ :

$$\Pi^m = \{\pi_1^m, \pi_2^m, \dots, \pi_{\binom{M}{m}}^m\} = \{X = \{x_1, x_2, \dots, x_m\}, x_1, x_2, \dots, x_m \in \mathcal{T}, x_1 \neq x_2 \neq \dots \neq x_m\} \quad (1)$$

where  $\binom{M}{m}$  is the number of combinations of super-sniffers of size  $m$  that can be built out of  $M$  sniffers.

The outcome trace of a super-sniffer is a single trace resulting from the combination of the individual traces of the sniffers composing the super-sniffer. We refer to such a trace as  $A^{\pi_i^m}$ , i.e., as the union of the traces  $\pi_i^m \in \Pi^m$ ,  $i = 1, 2, \dots, \binom{M}{m}$ :

$$A^{\pi_i^m} = T_a \cup T_b \cup \dots \cup T_m, \quad T_a, T_b, \dots, T_m \in \pi_i^m, \quad (2)$$

and

$$T_a \neq T_b \neq \dots \neq T_m. \quad (3)$$

### B. Relative completeness

As underlined earlier, the *maximum reachable quality* is obtained when the super-sniffer is  $M$ -fold redundant (i.e., it is composed of all  $M$  individual sniffers):

$$A_{\max} = A^{\pi^M} = T_{s_1} \cup T_{s_2} \cup \dots \cup T_{s_M}. \quad (4)$$

We need to make two observations now. Firstly, note from Equation 1 that  $\Pi^M$  has a single element, which is  $\pi^M$ . Therefore, the quality of a capture is denoted by  $A^{\pi_i^m}$ . The value of this measure quality is obtained by taking its ratio with the result of maximum value when all  $M$  sniffers are considered. Secondly,  $A_{\max}$  is the best result that we can obtain. That is why we consider it as the reference number to define the “relative” completeness:

$$C(A^{\pi_i^m}) = \frac{|A^{\pi_i^m}|}{|A_{\max}|}. \quad (5)$$

There are multiple super-sniffers of size  $m$ , each one resulting from a different combination of  $m$  out of  $M$  sniffers. Each of the  $\binom{M}{m}$  super-sniffers leads to a different value of completeness. We can then define two special cases, which come respectively, from the super-sniffer that leads to the

largest completeness and the super-sniffer that leads to the smallest completeness:

$$C_{\max}^m = \max_{i=1,2,\dots,(M)} C(A^{\pi_i^m}) \quad (6)$$

and

$$C_{\min}^m = \min_{i=1,2,\dots,(M)} C(A^{\pi_i^m}). \quad (7)$$

### III. SNIFFING WI-FI PACKETS

We provide in this section information about the experiments we run to assess the *relative completeness* of passive wireless capture of Wi-Fi traces.

**Individual sniffing nodes.** We have ten sniffers in our measurement set-up, out of which five are Raspberry Pi model 3B (RPi3 hereafter) and the other five are Raspberry Pi model 4B (RPi4 hereafter) [10, 11]. We use an external Wi-Fi module, Alfa AWUS051NH, one per sniffer [12]. The advantage of this specific external Wi-Fi module is that it can be easily set to monitor mode. The monitor mode is a radio mode that makes it possible for the Wi-Fi card to passively listen to all Wi-Fi traffic in the wireless medium. We choose the 2.4 GHz band and channel 1 for our measurements.

**Trace capture.** Sniffers run `tcpdump` to collect traces [13]. We configure some filters to gather only the data we need for this work (for example, to avoid capturing personal data as discussed below). The outcome of the capture process is one `pcap` file per individual sniffer.

**Privacy preserving.** The privacy of the users is a top priority for us. We anonymize the traces by running several protection techniques on the packets. Firstly, we do not disclose the geographic locations of our measurements. Secondly, we configure the sniffers to capture only the headers of the packets. In our work, we need the header as it brings the necessary information to combine traces from different sniffers. But, since headers contain MAC addresses of the devices, which are considered personal information, we need to provide extra privacy guarantees. To this end, we hash the packet headers.

**Generation of a combined trace.** The principle behind a super-sniffer is its ability to merge traces collected by its individual sniffers. The merging process requires that input traces be synchronized so that a packet that appears in multiple individual traces is identified unambiguously. We developed a Python tool called `PyPal` that performs such an operation.<sup>2</sup>

**Steps involved in synchronization.** The beacon and probe response frames are the closest representatives of real-time clocks. These frames lay the foundation for the synchronization process. The tool can only synchronize two traces at a time. Therefore, a reference trace and as well as the trace which has to be synchronized is the input to the tool. The first step is to extract the beacon and non-re-transmitted probe response frames from both traces independently. These frames

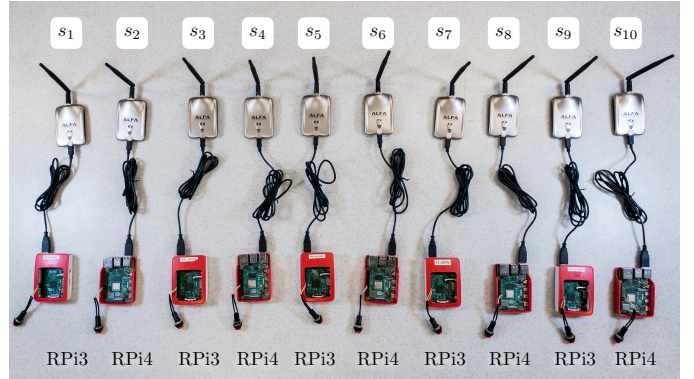


Fig. 3: Composition of the super-sniffer we used in our experiments.

are called unique frames. The next step is to extract the unique frames that are common in both traces. The coverage areas of the sniffers capturing these traces must overlap to execute this step. The common frames are referred to as reference frames. Next, the timestamps of reference frames are synchronized using linear regression over a sliding window of 3 frames. The synchronized reference frames are then used to synchronize the complete trace. The tool provides an additional option of concatenating or merging the synchronized traces.

### IV. MEASUREMENT CAMPAIGN

**Super-sniffer setup.** In all experiments, we deploy a super-sniffer by arranging all ten individual sniffers in the way depicted in Figure 3. The sniffers are placed at a distance of  $\sim 20$  cm from each other. Note that  $s_i$  refers to RPi3 nodes if  $i$  is odd and to RPi4 if  $i$  is even.

**Scenarios.** We capture traces in two different scenarios. The goal is to test the behavior of the sniffing system for different intensities of wireless traffic. The first scenario corresponds to a *residential* area while the second scenario involves *offices*. As we will see later in the paper, the second scenario is much denser and stresses the sniffers more. In Figure 4, we show the traffic that we observe in the two scenarios. As we see in Figure 4a, the traffic in the office area is dense with an average of roughly 1,000 packets per second. The measurement location in the residential area (Figure 4b) is isolated and has less Wi-Fi activity in the surroundings. Therefore, the traffic in the residential area is sparse – an average of around 100 packets per second.

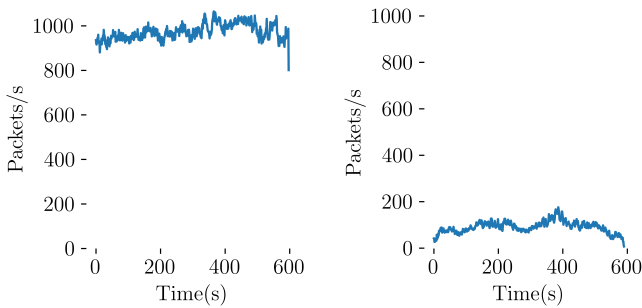
**Execution.** We run each test 10 times at three different spots in the target scenarios to rule out anomalies. The duration of each data collection is 10 minutes, and the sniffers remain stationary for the whole capture period.

### V. EXPERIMENTAL EVIDENCE OF THE NEED FOR SUPER-SNIFFERS

To evaluate the impact of the environment on the capture, we co-locate ten sniffers to collect wireless traffic. The merge of all these traces provides a *full* capture of the environment,

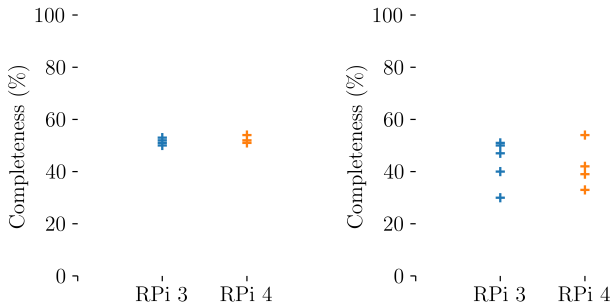
<sup>2</sup><https://gitlab.lip6.fr/syed/pyपाल>.





(a) Office area. (b) Residential area.

Fig. 4: Difference of traffic in the two scenarios.



(a) Office area. (b) Residential area.

Fig. 5: Average relative completeness. Each dot represents an individual sniffer

which is then used as the reference to compute the individual relative completeness for each sniffer trace. To evaluate the proportion of traffic a single sniffer can capture, we average the relative completeness for each sniffer using the 30 different runs of each scenario. In the following, we present an analysis of the relative completeness per individual sniffer.

**Relative completeness of individual sniffers.** We show in Figure 5 the average relative completeness of each sniffer trace for both the residential and office scenarios. To make a fair comparison between both types of devices, we plot, for each scenario, the values obtained with RPi3 and RPi4 sniffers.

We observe that with a low traffic load in the residential environment, the average relative completeness ranges from 50% in the best case to 30% in the worst case with RPi3 sniffers while the same values for RPi4 sniffers read 54% and 33% respectively. In the heavy-traffic environment of the office scenario, completeness appears slightly better, ranging between 50% and 54%.

The most striking element is the low value of the relative completeness for both scenarios. The best individual sniffers only 54% of the packets in both scenarios. By comparing both scenarios, we deduce that our sniffers are powerful enough to handle the traffic load. Thus, it seems that capture misses are due to the conditions of the wireless medium.

We show that the wireless environment is challenging to

TABLE I: Jaccard similarity: office area.

| Rel. compl. | $s_1$<br>50% | $s_2$<br>52% | $s_3$<br>53% | $s_4$<br>51% | $s_5$<br>53% | $s_6$<br>52% | $s_7$<br>51% | $s_8$<br>54% | $s_9$<br>52% | $s_{10}$<br>51% |
|-------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|-----------------|
| $s_1$       | –            | 0.62         | 0.62         | 0.64         | 0.62         | 0.63         | 0.63         | 0.63         | 0.63         | 0.64            |
| $s_2$       | 0.62         | –            | 0.62         | 0.63         | 0.62         | 0.63         | 0.63         | 0.62         | 0.63         | 0.63            |
| $s_3$       | 0.62         | 0.62         | –            | 0.63         | 0.62         | 0.63         | 0.63         | 0.62         | 0.62         | 0.63            |
| $s_4$       | 0.64         | 0.63         | 0.63         | –            | 0.63         | 0.63         | 0.64         | 0.63         | 0.63         | 0.64            |
| $s_5$       | 0.62         | 0.62         | 0.62         | 0.63         | –            | 0.62         | 0.63         | 0.62         | 0.62         | 0.63            |
| $s_6$       | 0.63         | 0.62         | 0.63         | 0.63         | 0.62         | –            | 0.63         | 0.63         | 0.63         | 0.63            |
| $s_7$       | 0.63         | 0.63         | 0.63         | 0.64         | 0.63         | 0.63         | –            | 0.63         | 0.63         | 0.63            |
| $s_8$       | 0.63         | 0.62         | 0.62         | 0.63         | 0.62         | 0.62         | 0.63         | –            | 0.62         | 0.63            |
| $s_9$       | 0.63         | 0.62         | 0.62         | 0.63         | 0.62         | 0.62         | 0.63         | 0.62         | –            | 0.63            |
| $s_{10}$    | 0.63         | 0.63         | 0.63         | 0.64         | 0.63         | 0.63         | 0.63         | 0.63         | 0.63         | –               |

capture, and signal strength plays a substantial role in the quality of the captured traces. Redundancy is, therefore, a strategy worth exploring.

**Jaccard similarity.** We use the Jaccard index to measure the similarity of the traces in a pairwise way. The higher the percentage, the more similar the traces. The goal is to verify whether sniffers capture the same packets or not. If so, there would be no interest in deploying super-sniffers. If not, the conclusion is that we can improve the quality of the capture by adding redundancy.

Results for the office scenario are given in Table I. We recall from Section II that the relative completeness of each sniffer trace indicates the relative proportion of packets captured by each sniffer. As mentioned in Section IV,  $i$  in  $s_i$  reflects the position of each sniffer in our testbed; we note that the completeness has no direct correlation with the relative position of sniffers.

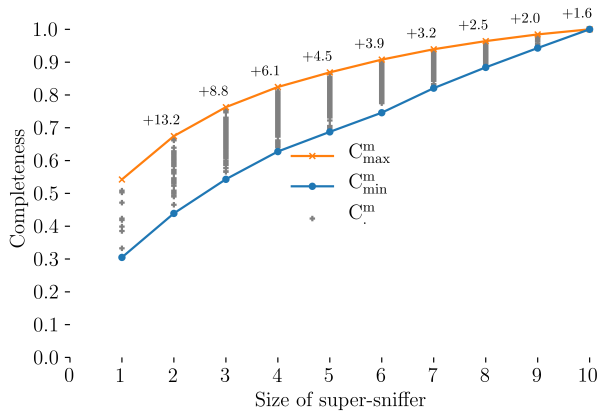
We observe that the similarity indexes remain quite stable as they are bounded between 62% and 64%. So, whichever combination of two sniffers, the two respective traces differ by 36% – 38%. This strong stability combined with a low similarity index value denotes each collected trace is complementary to any other one and any of the sniffers can bring useful information to the global trace. We can therefore conclude the need to combine sniffers to capture the traffic in a wireless environment adequately. The results for the residential environment are nearly identical, so we do not present the table to save some space. The following section presents the benefits of building a super-sniffer of up to 10 sniffers and analyses the benefits of all potential combinations.

## VI. EVALUATING REDUNDANCY WITH SUPER-SNIFFERS

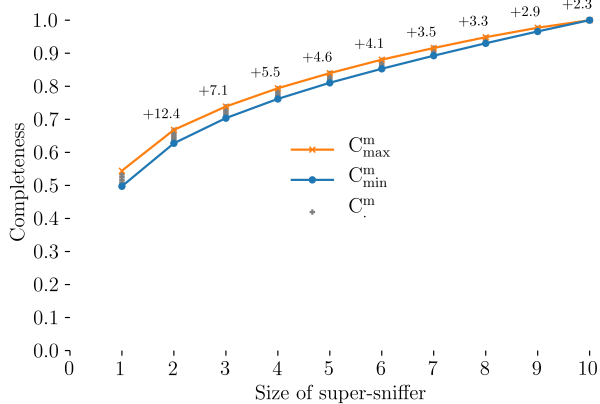
We investigate the importance of grouping individual sniffers to build a super-sniffer. Firstly, we present the impact for all combinations of  $m = \{1, 2, \dots, 10\}$  sniffers on trace completeness. Secondly, we compare all combinations of up to 5 sniffers of the same model, either RPi3 or RPi4, to rule out the influence of the hardware.

### A. General results

Figure 6 shows the completeness for all combinations of  $m$  sniffers. The  $x$ -axis represents the size  $m$  of the super-sniffer,



(a) Residential area.



(b) Office scenario.

Fig. 6: Relative completeness of each super-sniffer aggregating  $m$  single sniffers.

while the  $y$ -axis gives the joint completeness for a combination of up to 10 sniffers in both the residential and office areas.

The orange and blue curves inform respectively about  $C_{\max}^m$  and  $C_{\min}^m$ , the upper and lower bounds of the completeness for each combination of  $m$  sniffers. All other intermediate values of  $C^m$  are indicated as gray crosses on the plot. When  $m$  equals 1, the completeness values are identical to the ones given in Figure 5.

It is interesting to note that, in both curves, the completeness keeps increasing as the size of the super-sniffers grows. For each value of  $m$  greater than 2, the gain between  $C_{\max}^{m-1}$  and  $C_{\max}^m$  is explicitly given in the figures. We observe that this gain is higher as  $m$  is low, but it is never null. We obtain the maximum relative completeness only as  $m=10$  in both scenarios. This means a super-sniffer always benefits from combining an extra sniffer.

As we discussed in Section V, the completeness varies notably in the residential environment while the values are pretty stable in the office area. The analysis of completeness for a given size  $m$  confirms this, as for the residential setting we observe up to 20% difference between  $C_{\min}^m$  and  $C_{\max}^m$  depending on the choice of the  $m$  sniffers composing the

super-sniffer. This variation is, as expected, less and less visible as the size of the super-sniffer increases.

In the office environment, we note that the variation of completeness for a given size of super-sniffer does not go beyond 4%. This stability between the lower and higher completences is welcome. Still, it would be hazardous to conclude on the better performance of the office scenario. For  $m$  greater than 1, we observe a higher  $C_{\max}^m$  in the residential scenario, and this value increases as  $m$  grows.

A super-sniffer combining five sniffers can capture 80% of the packets in the office environment, and a combination of four to nine sniffers offers the same performance in the residential area. In both scenarios, we need to get the maximum size super-sniffer to reach maximum completeness. Reaching it for a lower value of  $m$  would be a good indicator that the capture is effectively complete. Although we observe a convergence, we have no guarantees to capture all packets even with the maximum size super-sniffer.

### B. Impact of the hardware

Our experimental platform is composed of sniffers based on RPi3 or RPi4 devices. External Wi-Fi adapters are identical, so the sensitivity of all devices should be equal. Table II displays the average completeness and the standard deviation for each combination of sniffers according to their model for the residential scenario. The rows and columns inform respectively about the number of RPi3 and RPi4 composing the super-sniffer. Thus, [row 4 ; column 0] gives the average completeness for all combinations of 4 RPi3, and [row 3 ; column 1] represents all super-sniffers composed of 3 RPi3 and 1 RPi4, and so on.

When comparing diagonals composed of the same number of sniffers, we note that the average completeness is quite similar. We do not see any performance improvement when using RPi4 rather than RPi3. Table II even shows sometimes some slightly better results with RPi3 than RPi4. The results for the office environment are nearly identical, so we do not present the table to save some space.

The position of each sniffer in our testbed does not affect the completeness of each trace, and the model of RPi used does not seem to play a significant role either. Thus, our off-the-shelf platform appears sufficient to hold the traffic load of both scenarios.

## VII. RELATED WORK

Xu et al. merge the individual traces into a single and then run an inference procedure to reconstruct the missing packets [14]. It needs at least one packet of a conversation in a trace to infer the missing packets and its accuracy also depends on the capture percentage. The evaluation is dependent on a simulation where the process removes packets from the trace randomly whereas, we keep the packet with the best RSSI value. Wit is a tool to merge multiple traces and then reconstruct the missing packets by inferring if they were received by the destination by making use of the frames like Association Request and Response [7]. PMSW is a passive

TABLE II: Residential area: average relative completeness and standard deviation for each combination of RPi.

| RPI3 | RPI4              |                   |                   |                   |                    |                   |
|------|-------------------|-------------------|-------------------|-------------------|--------------------|-------------------|
|      | 0                 | 1                 | 2                 | 3                 | 4                  | 5                 |
| 0    | –                 | <b>0.42</b> ±0.08 | <b>0.57</b> ±0.05 | <b>0.65</b> ±0.04 | <b>0.72</b> ±0.023 | –                 |
| 1    | <b>0.44</b> ±0.09 | <b>0.58</b> ±0.06 | <b>0.67</b> ±0.05 | <b>0.75</b> ±0.04 | <b>0.80</b> ±0.03  | <b>0.82</b> ±0.03 |
| 2    | <b>0.59</b> ±0.06 | <b>0.66</b> ±0.05 | <b>0.74</b> ±0.04 | <b>0.79</b> ±0.04 | <b>0.83</b> ±0.03  | <b>0.87</b> ±0.03 |
| 3    | <b>0.68</b> ±0.04 | <b>0.73</b> ±0.04 | <b>0.78</b> ±0.03 | <b>0.84</b> ±0.03 | <b>0.88</b> ±0.03  | <b>0.92</b> ±0.02 |
| 4    | <b>0.75</b> ±0.03 | <b>0.80</b> ±0.03 | <b>0.85</b> ±0.02 | <b>0.89</b> ±0.02 | <b>0.92</b> ±0.02  | <b>0.96</b> ±0.02 |
| 5    | –                 | <b>0.85</b> ±0.01 | <b>0.89</b> ±0.01 | <b>0.93</b> ±0.01 | <b>0.97</b> ±0.00  | –                 |

monitoring system that relies on sequence numbers to infer the missing packets in a wireless sensor network. However, it only captures data and acknowledgment packets, leading to a complex synchronization solution [15]. There are no conversation, data, or association frames as we rely on probe requests for contact traces.

Schulman et al. estimate the number of missed packets using sequence numbers and re-transmission bit [16] but they do not capture traffic of their own and rely on datasets available on CRAWDAD [17], whereas, we collect our own traces. The dependence on the re-transmission bit would create some bias because it is hard to infer how many packets are actually re-transmitted because they have the same sequence number.

Mahanti et al. examine the beacon and acknowledgment frames, MAC-layer sequence numbers, and placement of sniffer to address the incomplete traces [18]. They use the results from one sniffer to create a layout of four sniffers on three floors. The amount of packets captured in 24 hours is nearly the same as that captured by our sniffers in 10 minutes. Garcia et al. develop a passive monitoring system called EPMOS<sub>t</sub> which focuses on election to choose the nodes of the target area for their packets to be captured by the sniffers but more in terms of energy consumption which reduces the number of packets captured by 0.62% [6].

LiveNet provides a platform for monitoring and processing passive traces but the transfer of packets to the serial port seems to result in packet loss and the validation is also based on the data measured in a controlled environment [19]. Our work stands distinctive as we focus on redundancy for trace completeness based on real-world experiments in an uncontrolled environment and do an exhaustive analysis for different scenarios. Moreover, our solution is more oriented towards contact traces.

### VIII. CONCLUSION AND FUTURE WORK

In this paper, we introduce the notion of relative trace completeness. We present the analysis for traces captured simultaneously by ten co-located sniffers of two different types in low and high-activity scenarios. We highlight the importance of grouping sniffers into super-sniffers to improve completeness significantly. At the same time, we draw attention to the differences in trace completeness depending on the type of environment. We plan to study the impact of distance between the sniffers on completeness. We also plan to do a comparison of using multiple RPi nodes with one antenna each and one RPi node with multiple antennas. Lastly, we intend

to do measurements on different channels to study the impact of channel selection on completeness.

### IX. ACKNOWLEDGMENT

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

### REFERENCES

- [1] A. Galanopoulos, V. Valls, G. Iosifidis, and D. J. Leith, “Measurement-driven analysis of an edge-assisted object recognition system,” in *IEEE ICC*, 2020.
- [2] W. Zhou, Z. Wang, and W. Zhu, “Mining urban WiFi QoS factors: A data driven approach,” in *IEEE BigMM*, 2017.
- [3] P. De Vaere, T. Bühler, M. Kühlewind, and B. Trammell, “Three bits suffice: Explicit support for passive measurement of internet latency in QUIC and TCP,” New York, NY, USA, 2018.
- [4] J. Wang, Y. Zheng, Y. Ni, C. Xu, F. Qian, W. Li, W. Jiang, Y. Cheng, Z. Cheng, Y. Li, X. Xie, Y. Sun, and Z. Wang, “An active-passive measurement study of TCP performance over LTE on high-speed rails,” in *ACM Mobicom*, New York, NY, USA, 2019.
- [5] M. D. Corner, B. N. Levine, O. Ismail, and A. Upreti, “Advertising-based measurement: A platform of 7 billion mobile devices,” in *ACM Mobicom*, Snowbird, UT, USA, oct 2010.
- [6] F. Garcia, R. Andrade, C. De Oliveira, and J. Souza, “EPMOS<sub>t</sub>: An energy-efficient passive monitoring system for wireless sensor networks,” *Sensors (Basel, Switzerland)*, vol. 14, pp. 10804–10828, 06 2014.
- [7] R. Mahajan, M. Rodrig, D. Wetherall, and J. Zahorjan, “Analyzing the MAC-level behavior of wireless networks in the wild,” in *Proceedings of the 2006 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*. New York, NY, USA: Association for Computing Machinery, 2006.
- [8] Y.-C. Cheng, J. Bellardo, P. Benkö, A. C. Snoeren, G. M. Voelker, and S. Savage, “Jigsaw: Solving the puzzle of enterprise 802.11 analysis,” in *Proceedings of the 2006 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*. New York, NY, USA: Association for Computing Machinery, 2006.
- [9] “ANR MITIK,” <https://project.inria.fr/mitik/>.
- [10] “Raspberry Pi 3 model B,” <https://tinyurl.com/2p88aa94>.
- [11] “Raspberry Pi 4 model B,” <https://tinyurl.com/2p89uund>.
- [12] “AWUS051NH Wi-Fi adapter,” <https://tinyurl.com/yk8vk3vz>.
- [13] The Tcpdump Group, “Tcpdump and libpcap,” <https://tcpdump.org>.
- [14] X. Xu, C. Tong, and J. Wan, “Improve the completeness of passive monitoring trace in wireless sensor network,” in *2010 Asia-Pacific Services Computing Conference (APSCC 2010)*. Los Alamitos, CA, USA: IEEE Computer Society, dec 2010.
- [15] X. Xu, J. Wan, W. Zhang, C. Tong, and C. Wu, “PMSW: A passive monitoring system in wireless sensor networks,” *International Journal of Network Management*, vol. 21, no. 4, pp. 300–325, 2011.
- [16] A. Schulman, D. Levin, and N. Spring, “On the fidelity of 802.11 packet traces,” in *9th International Conference on Passive and Active Network Measurement*. Berlin, Heidelberg: Springer-Verlag, 2008.
- [17] “Crawdad website,” <https://crawdad.org/>.
- [18] A. Mahanti, M. Arlitt, and C. Williamson, “Assessing the completeness of wireless-side tracing mechanisms,” in *IEEE WoWMoM*, 2007.
- [19] B.-r. Chen, G. Peterson, G. Mainland, and M. Welsh, “LiveNet: Using passive monitoring to reconstruct sensor network dynamics,” in *Distributed Computing in Sensor Systems*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008.