How much can Sniffer Redundancy Improve Wi-Fi Traffic?

Mohammad Imran Syed, Anne Fladenmuller, Marcelo Dias de Amorim

Copyright Notice

How much can Sniffer Redundancy Improve Wi-Fi Traffic?

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*—Sniffing is a cost-efficient and trouble-less method for capturing Wi-Fi traces within an area of interest. However, a sniffer is likely to miss Wi-Fi packets because of the inherent characteristics of the wireless medium, resulting in incomplete traces. In this paper, we investigate the *completeness* of Wi-Fi capture using sniffers and conduct experimental measurements to support our analyses. We collect Wi-Fi traces using ten co-located sniffers while moving the source node away in steps of ten meters. As expected, individual sniffers lead to low completeness. This motivates the study of completeness using redundant sniffers. Our experiments reveal that a certain gain in completeness is achieved even when the source node is far from the sniffers.**

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

I. Introduction

Monitoring wireless traffic represents an opportunity for researchers to do measurements for network analysis, diagnosis, and as well studying the user behavior [1]–[5]. However, monitoring the wireless medium poses many challenges because of its inherent characteristics [6].

Active measurements are challenging because they require deploying capturing mechanisms on different entities. One has to either request permissions for deployment on access points in the target area or create an application and ask users to install it on their mobile devices. In the latter case, it might result in an insufficient sample to be representative of the target area and hence lead to inaccurate results.

Passive measurements, which consist in placing several *sniffers* (devices collecting wireless packets in monitor mode) throughout the target area, are a productive alternative [7]–[10].[1] It does not require to bother the infrastructure administrators or users.

As a part of the ANR-MITIK project, the challenge we need to address is to infer contacts through non-intrusive methods like passive sniffing [11]. However, to deploy a passive measurement infrastructure, we first need to evaluate whether a single sniffer can properly capture the traffic in its vicinity. We know that for multiple reasons such as collisions, multi-path propagation, or interference, packets may be missed by a sniffer. An incomplete trace would potentially lead to a distorted view of the wireless activity. The problem we address is threefold. Firstly, we experimentally evaluate the proportion of packets that a single sniffer may miss in an outdoor environment. We define a completeness metric to evaluate the quality of the trace capture. We propose then to co-locate several sniffers (i.e. to introduce redundancy) to evaluate whether individual sniffers are complementary to each other. Finally, we evaluate the minimum number of co-located sniffers (which we call a super-sniffer) that we should adopt to obtain *reasonable* capture of a given area while maintaining a low cost and ease of deployment because redundancy also comes at a cost.

We conduct experiments in an outdoor environment with few obstacles to reduce the multi-path effect but surrounded by concurrent Wi-Fi traffic to be representative of an urban environment. We propose to monitor the packets that a controlled source sends, both to address privacy concerns and to determine the completeness of the collected trace. We make several observations:

- A single sniffer is insufficient to fully capture its wireless environment even if the source is very close to the sniffer.
- Each extra sniffer added to the super-sniffer improves the completeness of the trace, meaning that they all miss different packets.
- The gain obtained by increasing the number of co-located sniffers depends also on the distance between the source and the sniffer.

In Section II, we present the experiment that we perform with a single sniffer. Section III explains the synchronization tool we developed to combine traces of individual sniffers and formalizes the metric of absolute completeness that we use to evaluate the improvement of combining several sniffers. We show the experimentation and results in Section IV. Section V discusses the related work. We finally conclude the paper and mention some open issues in Section VI.

II. Is a single sniffer enough?

In this section, we detail the experiments of our measurement campaign and experimentally show that a single sniffer is not enough to capture its environment.

A. Explaining the sniffer and its trace

A sniffer is a device that can be set to monitor mode to capture Wi-Fi traffic passively. It generates a trace of the captured traffic in a certain format containing the header fields, and importantly the timestamps.

Nodes. We have eleven Raspberry Pi 4B nodes in our measurement set-up [12], ten as sniffers and one as the source

[1]It is, however, essential to clearly define which data one can sniff depending on the location of the measurement campaign to preserve the privacy of the users.

Fig. 1: Co-located sniffers and the source at a distance of 10m.

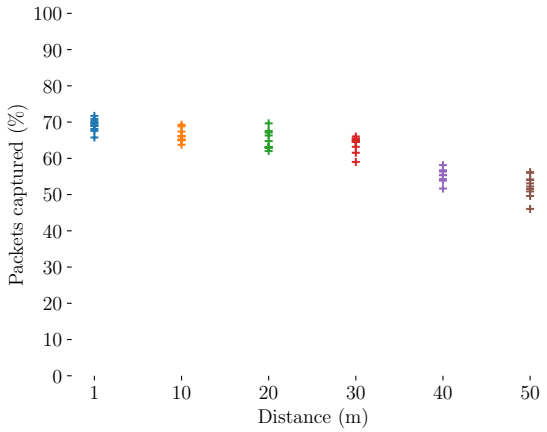


Fig. 2: Completeness of 10 individual sniffers at different distances (in meters) from the source.

node for generating traffic. We use an external Wi-Fi module, Alfa AWUS051NH, one per sniffer [13]. This specific external Wi-Fi module can be easily set to monitor mode to passively capture the Wi-Fi traffic. For the given results, the source and sniffers are set to channel 1 of the 2.4 GHz band.

Trace generation. We use `scapy` [14] at the sender node to send beacons. The average sending rate is 9 beacons per second. This is the maximum possible sending rate to be able to differentiate the packets at the sniffers using sequence numbers.

Trace capture. Sniffers run `tcpdump` to collect traces [15]. We configure filters to gather only the traffic generated from the controlled source node. The final captured trace is one `pcap` file per individual sniffer. The `pcap` file is then split into six different files to extract traces for each distance.

B. Experimental set-up

We know that there is no guarantee that a single sniffer can capture 100% of the traffic because of losses due to inherent characteristics of the wireless medium. We run experiments in the outdoor scenario to examine the traces collected by the individual sniffers. We place the source node at distances of 1, 10, 20, 40, and 50 meter(s) from the sniffers for these experiments and we get 5 runs for each scenario. Figure 1 shows the experimental set-up when the source is placed at a distance of 10 meters from the sniffers. We show the percentage of packets captured by individual sniffers in Figure 2 for each run. For the best case, the value is around 70% for 10m and only between 40% and 60% at 50m. It shows that a single sniffer ends up missing a lot of packets and the rate of this loss increases as the source moves away from the sniffer. These results underline the need of improving the sniffing capacity of our sniffer and therefore, we propose to introduce redundancy and evaluate the potential gain of coupling the sniffers.

III. Introducing Redundancy

We propose to study whether the quality of a passive measurement system improves as the redundancy level of the super-sniffer increases. A super-sniffer is a collection of sniffers that are co-located and operate side-by-side to increase the completeness of the trace. The number of co-located sniffers give the redundancy level of the super-sniffer. Figure 3 illustrates a scenario with 3 co-located sniffers. A super-sniffer merging the traces from each of the 3 sniffers would obtain the complete trace. But if the super-sniffer is only composed of 2 sniffers, it may or may not capture all the packets : s_1+s_2 would, but s_2+s_3 or s_1+s_3 would not.

A. Combining the traces of multiple sniffers

Generation of a combined trace. The notion of a super-sniffer involves the procedure of merging traces collected by its individual sniffers by removing duplicate packets. It, however, requires the individual traces to be synchronized so that a packet that is captured by multiple individual sniffers is accurately identified. We developed a Python tool called `PyPal` that performs such an operation[2].

Steps involved in synchronization. The beacons are the closest representatives of real-time clocks. We use these frames as a base for the synchronization of traces. Two traces are used as input, one as a reference trace and the second trace is the one which has to be synchronized. The first step is to independently extract the beacons that are common in both traces. Hence, the coverage areas of the sniffers capturing these traces should overlap to perform this step. The common frames are referred to as reference frames. In the next step, 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.

B. Completeness

We introduce here the notion of *absolute completeness*. It gives the share of the packets that a sniffer captures for the actual traffic from a specific source.

[2]https://gitlab.lip6.fr/syed/pypal.

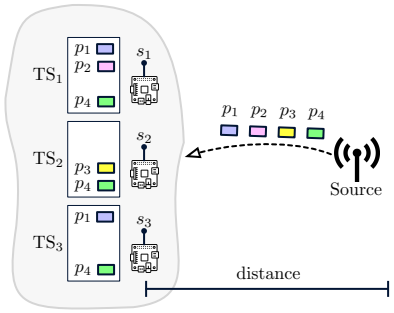


Fig. 3: Using redundancy to improve trace completeness. Because of the nature of the wireless medium, sniffers miss some packets. Combining traces of co-located sniffers could help get closer to the complete trace.

Composing Traces for a Super-sniffer of size m. Let $S = \{s_1, s_2, \ldots, s_M\}$ be the set of M sniffers that we have at our disposal, T_{s_i} be the trace (i.e., set of packets) captured by sniffer $s_i \in S$, and $\mathcal{T} = \{T_{s_1}, T_{s_2}, \ldots, T_{s_M}\}$. So, the question is to estimate the improvement that we get when we increase the redundancy of a super-sniffer.

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

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

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 merging of individual sniffers' traces. We refer to such a trace as $A^{\pi_i^m}$, i.e., as the union of the traces $\pi_i^m \in \Pi^m$, $i = 1, 2, \ldots \binom{M}{m}$:

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

and

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

Completeness. Let A_{abs} be the set of packets that actually circulated in the network at the time of the capture. The absolute completeness is:

$$C(A^{\pi_i^m}) = \frac{|A^{\pi_i^m}|}{|A_{\text{abs}}|}. \quad (4)$$

Number of traces per size of super-sniffer. We need to build traces π_i^m for all combinations of sniffers of different sizes. If we consider $m = 4$, then Π^m in Equation 1 is equivalent to $\{\pi_1^4, \pi_2^4, \ldots, \pi_{\binom{M}{4}}^4\}$ which means that we need to build traces for all combinations of $m = 4$ sniffers out of the total M sniffers. It represents all combinations of sniffer s_1 with combinations of three sniffers other than s_1, similarly combinations of sniffer s_2 with sniffers other than s_2 itself, and so on.

IV. Evaluation: Up to what extent can redundancy help?

In this section, we experimentally evaluate the importance of building a super-sniffer to improve the quality of traces in the process. We capture traces in the outdoor scenario. We run new experiments by combining sniffers in groups of $[2, 3, \ldots, 10]$ sniffers. The experiments took place at the same locations as the one presented in Section II. The sniffers are co-located and remain stationary for the duration of the experiments. The source node is again placed at distances of 1, 10, 20, 30, 40, and 50 meters from the sniffers. We run the test five times at each distance. We present the results for all combinations of $m = \{1, 2, \ldots, 10\}$ sniffers for each distance. Table I shows the number of super-sniffers of size m that we can obtain from our 10 sniffers.

TABLE I: Number of combinations of super-sniffers of size m for our experimental scenario of 10 sniffers.

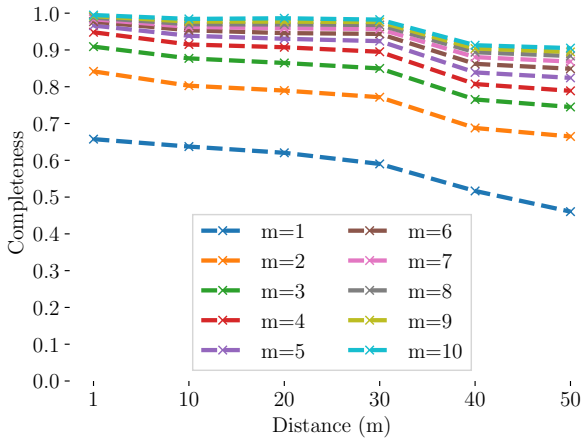
	Combinations of super-sniffers								
Size	2	3	4	5	6	7	8	9	10
Number	45	120	210	252	210	120	45	10	1

A. Completeness and Super-sniffer

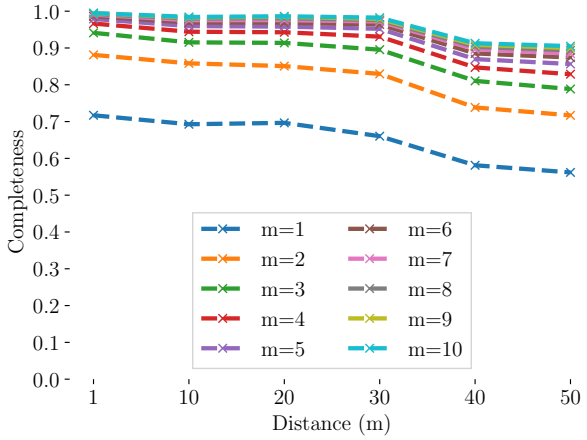
We know that the farther the source, the fewer chances of its traffic being captured by the sniffers. It means there are more chances of a sniffer missing packets if the source of traffic is far. Figure 4 shows the impact of *distance* of source from the sniffers on the *minimum* and *maximum* completeness of all super-sniffers of sizes $m = \{1, 2, \ldots, 10\}$. The dark blue line at the bottom represents the results for m=1 for each distance of the source and the sky blue line at the top for m=10.

Figures 4a and 4b present respectively the minimum and maximum completeness of super-sniffers up to size ten for each distance. The improvement in completeness stands out for super-sniffers of sizes up to 5. The results are closely spaced for super-sniffers of size 5 and more. These observations hold for both cases. The minimum completeness for 50m distance is 46% when $m = 1$ but it reaches 90% for a super-sniffer of size $m = 10$. We see that the completeness improves for each distance as the size of the super-sniffer increases. It is also evident from the bottom two lines that completeness improves massively just by increasing the number of sniffers from $m = 1$ to $m = 2$. This result holds for all distances.

Table II shows the maximum completeness (C_{max}^m) gain for the super-sniffers of each size at distances of 1m and 50m (note the trend is similar for the remaining distances). We get a noteworthy improvement of 16.5% by adding just one sniffer ($m = 2$) for 1m. The rate of improvement keeps falling as we keep adding a sniffer to the super-sniffer. It eventually becomes zero for $m = 10$. We see a similar trend for 50m where there is an improvement of 15.5% for $m = 2$. The value of improvement again keeps decreasing for increasing m. The improvement in completeness is stagnant for super-sniffers of



(a) Minimum Completeness.



(b) Maximum Completeness.

Fig. 4: Minimum and maximum completeness of super-sniffers of the same sizes at each distance. The completeness for each m decreases as the distance increases.

TABLE II: Maximum gain of completeness for combinations of different number of sniffers compared to a single sniffer at 1m and 50m.

Number of sniffers	1m	50m
2	16.4	15.5
3	22.4	22.7
4	24.9	26.7
5	26.1	29.5
6	26.8	31.3
7	27.2	32.3
8	27.4	33.2
9	27.6	33.8
10	27.6	34.3

size greater than 5 and 7 for 1m and 50m respectively. It implies that more number of sniffers are needed if the source is far from the sniffers. The redundancy in the number of sniffers, therefore, improves the quality of the traces captured as it increases completeness.

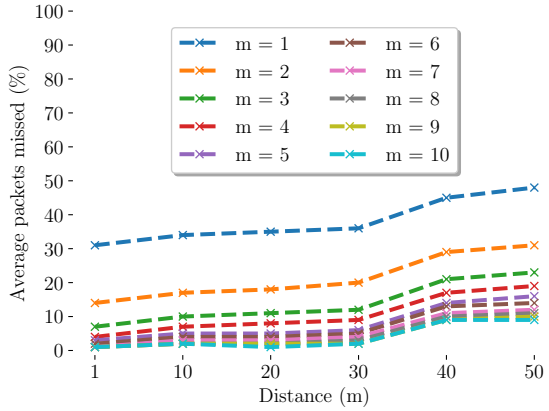


Fig. 5: Missed packets. Average packets missed by super-sniffers of all sizes ($m = 1, 2, \ldots, 10$) for different distances between the source and super-sniffer.

B. Distribution of Packets Captured

Figure 5 shows the average percentage of packets lost by super-sniffers of all sizes ($m = 1, 2, \ldots, 10$) at each distance i.e. not a single sniffer captured these packets. The dark blue line at the top and teal line at the bottom represent the results for the minimum and maximum size of the super-sniffer respectively. As expected, hardly any packets are lost at 1m for $m = 10$ but the values are higher for a lower number of m especially 1 and 2. For $m = 10$, around 1.75% packets are lost up to the distance of 30m, and then there is a sharp rise at 40m where the percentage of packets lost reaches 9.5%. The maximum amount of packets are lost at a 50m distance which is the expected behavior. If we consider the individual sniffer, around 31% packets are lost at a distance of 1m. The percentage increases as the distance of the source increases. Nearly 48% packets are lost by the single sniffer for a distance of 50m. Whereas the percentage falls to 9.5% for the same distance when $m = 10$. We notice that the completeness of individual sniffers ($m = 1$) is not identical for all tests which means the lowest and the highest packet loss is not always from the same sniffer. This observation coupled with the fact that the packets are not missed by all sniffers simultaneously, indicates that it is likely to be due to poor reception and not necessarily collisions at the receiver(s).

It is hard to differentiate between the average percentage of packets missed for super-sniffers of size 5 and more in Figure 5. To address the matter in question, Table III presents 95% Confidence Interval (CI) in the form of range for super-sniffers of size up to 9 for distances 1, 30, and 50 meters (the results are similar for 10, 20, and 40 meters). We see that the difference between the ranges is quite small for super-sniffers of size greater than 5 for all distances and that is why they are very closely spaced in the graph. The CI ranges reduce with increasing m which means we are more confident about our results with the introduction of the super-sniffer. However, there is no overlap in the intervals. This means that

each new sniffer still brings new information to the super-sniffer, albeit minimal. Therefore, it is up to the designer which level of completeness is desirable. It could be interesting to run a preliminary test of 5 minutes to have an idea of how to visualize the measuring system.

TABLE III: Confidence interval of the percentage of missed packets.

size of the super-sniffer	**Distance** 10m	30m	50m
1	32.46 - 34.97	34.50 - 37.59	45.24 - 49.82
2	16.49 - 17.23	19.10 - 19.92	30.60 - 31.50
3	10.07 - 10.36	12.21 - 12.56	23.04 - 23.41
4	6.85 - 7.01	8.44 - 8.64	18.60 - 18.81
5	4.99 - 5.10	6.13 - 6.26	15.70 - 15.85
6	3.79 - 3.89	4.60 - 4.71	13.68 - 13.81
7	2.96 - 3.07	3.54 - 3.65	12.19 - 12.32
8	2.34 - 2.48	2.75 - 2.88	11.04 - 11.20
9	1.82 - 2.07	2.13 - 2.32	10.10 - 10.35

V. Related Work

There is some work on the concept of trace completeness in the literature. Schulman et al. estimate the number of missed packets using sequence numbers and re-transmission bit [16]. Mahanti et al. examine the MAC-layer sequence numbers and placement of sniffers to address the incomplete traces [17]. Garcia et al. develop a passive monitoring system called EPMOSt which focuses on election to choose the sniffers but more in terms of energy consumption which reduces the number of packets captured by 0.62% [9]. PMSW is a passive monitoring system that relies on sequence numbers to infer the missing packets. However, it only captures data and acknowledgment packets, leading to a complex synchronization solution [18]. Our work stands distinctive as we focus on redundancy for trace completeness and to the best of our knowledge, no one has done such an exhaustive analysis based on the traffic generated from a controlled source.

VI. Conclusion

In this paper, we present the notion of trace completeness. We present the analysis for traces captured simultaneously by ten co-located sniffers while the source node is placed at six different distances. We highlight that single sniffers miss on average between 30% to 50% of all packets sent depending on the distance from the source. To improve the performance, we propose to build a super-sniffer by combining several sniffers and merging their traces with the merging tool we have developed. We then show the benefits of grouping sniffers into super-sniffers. Combining 3 sniffers improves significantly the performances for every distance. For the short distance between the sniffers and the source, increasing the size of the super-sniffer may not be necessary as it comes with a cost. But to cover larger areas this may be necessary. We plan to do outdoor deployment of super-sniffers in different areas to do passive measurements in the real environment. To reduce the cost of deployment and based on our experimental results, we intend to adapt the size of our super-sniffers to the deployment areas. For zones smaller than 30m, we will build super-sniffers of size 3 and increase them to 5 to cover 50m areas.

VII. Acknowledgment

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

References

[1] P. A. K. Acharya, A. Sharma, E. M. Belding, K. C. Almeroth, and K. Papagiannaki, "Congestion-aware rate adaptation in wireless networks: A measurement-driven approach," in *2008 5th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks*, 2008.

[2] 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," in *Proceedings of the IMC*. New York, NY, USA: Association for Computing Machinery, 2018.

[3] A. Galanopoulos, V. Valls, G. Iosifidis, and D. J. Leith, "Measurement-driven analysis of an edge-assisted object recognition system," in *IEEE ICC*, 2020.

[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 *MobiCom*. New York, NY, USA: Association for Computing Machinery, 2019.

[5] W. Zhou, Z. Wang, and W. Zhu, "Mining urban WiFi QoS factors: A data driven approach," in *IEEE BigMM*, 2017.

[6] 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.

[7] 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*, ser. SIGCOMM '06. New York, NY, USA: Association for Computing Machinery, 2006.

[8] T. Claveirole and M. Dias de Amorim, "WiPal: Efficient offline merging of IEEE 802.11 traces," *SIGMOBILE Mob. Comput. Commun. Rev.*, p. 39–46, Mar. 2010.

[9] F. P. Garcia, R. M. C. Andrade, C. T. Oliveira, and J. N. De Souza, "EPMOSt: An energy-efficient passive monitoring system for wireless sensor networks," *Sensors*, pp. 10 804–10 828, 2014.

[10] R. Mahajan, M. Rodrig, D. Wetherall, and J. Zahorjan, "Analyzing the mac-level behavior of wireless networks in the wild," ser. SIGCOMM '06. New York, NY, USA: Association for Computing Machinery, 2006.

[11] "ANR MITIK," https://project.inria.fr/mitik/.

[12] "Raspberry Pi 4 model B," https://tinyurl.com/2p89uund.

[13] "AWUS051NH Wi-Fi adapter," https://tinyurl.com/yk8vk3vz.

[14] P. Biondi, "Scapy," https://scapy.net/.

[15] The Tcpdump Group, "Tcpdump and libpcap," https://tcpdump.org.

[16] A. Schulman, D. Levin, and N. Spring, "On the fidelity of 802.11 packet traces," in *Proceedings of the 9th PAM conference*. Berlin, Heidelberg: Springer-Verlag, 2008.

[17] A. Mahanti, M. Arlitt, and C. Williamson, "Assessing the completeness of wireless-side tracing mechanisms," in *2007 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks*, 2007.

[18] 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*, pp. 300–325, 2011.