Jusqu'où la redondance peut aider dans la capture passive de trafic Wi-Fi

Mohammad Imran Syed, Anne Fladenmuller, Marcelo Dias de Amorim

Jusqu'où la redondance peut aider dans la capture passive de trafic Wi-Fi[†]

Mohammad Imran Syed, Anne Fladenmuller et Marcelo Dias de Amorim

Sorbonne Université, CNRS, LIP6, Paris

La capture passive de trames Wi-Fi est une méthode prometteuse pour caractériser la dynamique d'un environnement sans fil (p.ex. pour estimer l'activité des utilisateurs ou leurs déplacements). Cette méthode repose sur une mesure exhaustive ou tout du moins représentative du trafic de la zone considérée. Cependant, la question de la complétude de la trace obtenue se pose alors, car elle conditionne la qualité de la caractérisation à mener. Dans cet article, nous formalisons la notion de complétude et étudions ce paramètre à partir de la collecte de trames Wi-Fi anonymisées obtenues simultanément par dix sniffeurs co-localisés dans une zone intérieure. Malgré le choix de sniffeurs de technologies identiques, chaque trace obtenue (une par sniffeur) ne contient qu'une fraction de l'ensemble des trames observées (entre 33% et 54% dans nos expérimentations). Nous observons ensuite l'importance de regrouper les sniffeurs tout en soulignant là encore une différence de complétude en fonction du choix de regroupement des sniffeurs.

Mots-clefs : Capture de trafic sans fil, Wi-Fi, mesures passives, complétude.

1 Introduction

La mise en place d'un système de monitorage de trafic sans fil doit prendre en compte la nature diffuse du médium sans fil [CLIU10]. Les mesures actives ne sont pas toujours possibles car elles exigent l'installation de sondes de mesure dans plusieurs équipements tels que les stations de base ou les terminaux des utilisateurs. Une alternative intéressante consiste à déployer des *sniffeurs* dans la région que l'on souhaite mesurer [GAOS14, MRWZ06, CBB+06]. Ces sniffeurs sont des dispositifs opérant en mode "moniteur", c'est-à-dire qu'ils sont capables de collecter les paquets qui occupent le médium sans fil quelles que soient leurs sources et destinations.

La contrepartie des collectes passives est que les sniffeurs, à cause de phénomènes tels que les collisions et les perturbations liées aux transmissions sans fil (obstructions, multi-trajets, etc.), sont susceptibles de manquer des paquets. L'analyse de traces incomplètes peut alors induire des résultats biaisés et voire erronés.

Une solution pour palier le problème des captures incomplètes réside dans l'utilisation de sniffeurs redondants, que nous appelons *super-sniffeurs*. La Figure 1 illustre un super-sniffeur constitué de quatre sniffeurs simples, introduisant ainsi une redondance de quatre. Dans cet exemple, chaque sniffeur manque un certain nombre de paquets (indiqués par p_i). En fusionnant la "vue" partielle de chacun des sniffeurs, il devient possible de reconstruire une trace plus complète.

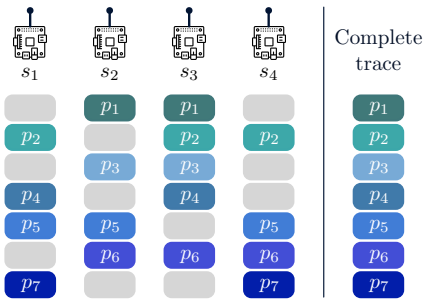


FIGURE 1 – Redondance pour améliorer le taux de capture.

La question que l'on se pose dans cet article est d'évaluer le gain en terme du nombre de paquets capturés en fonction du niveau de redondance d'un super-sniffeur. Afin d'apporter des éléments de réponse à cette question, nous menons une étude expérimentale pour comparer la *complétude* du trafic capturé entre des super-sniffeurs ayant des niveaux de redondance allant de un jusqu'à 10. Nous considérons deux scénarios

[†]Ce travail a été partiellement financé par le projet ANR MITIK, Agence Nationale de la Recherche (ANR), PRC AAPG2019.

avec différentes intensités de trafic et densités de nœuds. Nos observations principales sont : (i) toutes les expériences nous ont montré qu'un sniffeur seul (pas de redondance) n'arrive à capturer qu'une part limitée du trafic (entre 30% et 54%), ce qui confirme le besoin de redondance ; (ii) à chaque sniffeur que l'on ajoute au super-sniffeur, la complétude de la trace augmente, ce qui veut dire que chaque sniffeur capture des paquets que d'autres sniffeurs ne capturent pas ; et (iii) le choix des sniffeurs formant un super-sniffeur impacte plus ou moins la complétude en fonction de l'environnement retenu. Nous obtenons une grande homogénéité des résultats dans notre environnement où le trafic est le plus dense et une forte disparité lorsque le trafic est moindre. Le choix de la taille idéale du super-sniffeur dépendra donc très vraisemblablement de l'environnement d'expérimentation à considérer.

2 Mesures passives et complétude des traces

La qualité d'une mesure passive tend à croître à mesure que la redondance des sniffeurs augmente. Nous proposons une métrique, que nous appelons *complétude relative* pour quantifier la qualité de la mesure en fonction du niveau de redondance d'un super-sniffeur. Nous définissons cette métrique comme la quantité de trafic capturé par un super-sniffeur d'une certaine redondance relativement au trafic capturé par le super-sniffeur de redondance maximale (10 dans notre cas).

Soient $S = \{s_1, s_2, \ldots, s_M\}$ l'ensemble de tous les sniffeurs individuels à notre disposition pour composer un super-sniffeur, T_{s_i} la trace (c'est-à-dire, un ensemble de paquets) capturée par le sniffeur $s_i \in S$ et $\mathscr{T} = \{T_{s_1}, T_{s_2}, \ldots, T_{s_M}\}$. Nous définissons $\pi^m_\cdot$ comme un sous-ensemble de m éléments de $\mathscr{T}$ et Π^m comme l'ensemble de toutes les instances des différentes combinaisons de $\pi^m_\cdot$:

$$\Pi^m = \{\pi^m_1, \pi^m_2, \ldots, \pi^m_{\binom{M}{m}}\}, \tag{1}$$

où $\binom{M}{m}$ est le nombre de combinaisons de super-sniffeurs de taille m (de redondance m, de façon équivalente) que nous pouvons former à partir des M sniffeurs individuels.

La trace capturée par un super-sniffeur résulte de la fusion des traces des sniffeurs le composant. Nous nous référons à cette trace comme $A^{\pi^m_i}$, c'est-à-dire l'union des traces $\pi^m_i \in \Pi^m$, $i = 1, 2, \ldots \binom{M}{m}$:

$$A^{\pi^m_i} = T_a \cup T_b \cup \ldots \cup T_m, \quad T_a, T_b, \ldots, T_m \in \pi^m_i \quad \text{et} \quad T_a \neq T_b \neq \ldots \neq T_m. \tag{2}$$

La qualité maximale atteignable $A_{\max}$ est obtenue lorsque le super-sniffeur est composé de tous les sniffeurs mis à disposition (10 dans le cas de nos expérimentations) :

$$A_{\max} = A^{\pi^M} = T_{s_1} \cup T_{s_2} \cup \ldots \cup T_{s_M}. \tag{3}$$

La *complétude relative* d'un super-sniffeur de redondance m est obtenue par :

$$C(A^{\pi^m_i}) = \frac{|A^{\pi^m_i}|}{|A_{\max}|}. \tag{4}$$

3 Un seul sniffeur ne suffit pas

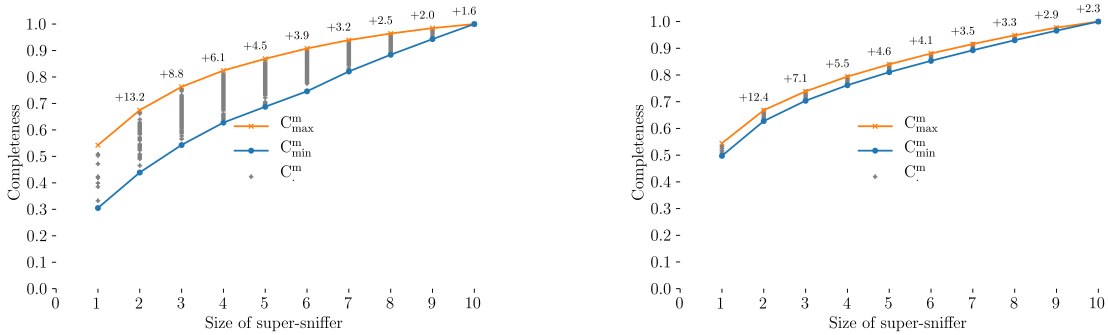
Nous allons maintenant montrer, de façon expérimentale, que l'utilisation d'un sniffeur seul ne peut suffire à capter fidèlement le trafic Wi-Fi dans une région. Nos sniffeurs, au nombre de 10, sont des Raspberry Pi couplés avec une antenne externe du modèle Alfa AWUS051NH. L'avantage de ce modèle de module Wi-Fi est qu'il dispose du mode "moniteur" par défaut. Dans nos expériences, nous nous focalisons sur la bande des 2.4 GHz et le canal 1 du protocole Wi-Fi (car le plus actif dans notre scénario).

Les sniffeurs capturent les paquets grâce à `tcpdump`. Nous avons configuré quelques filtres pour ne récupérer que les informations strictement nécessaires pour notre étude, et exclure les données personnelles des traces collectées. ‡ La trace capturée par chaque sniffeur individuellement est stockée dans un fichier du

‡. Plus précisément, nous ne capturons que l'entête de quelques paquets de contrôle. Ensuite, nous appliquons la fonction de hachage SHA512 sur les adresses MAC contenues dans les entêtes, ce qui les fait passer de 48 à 512 bits. Enfin, nous tronquons la version modifiée de l'adresse à 256 bits.

TABLE 1 – Complétude relative pour chacun des sniffeurs s_i (%).

	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}
Scénario résidentiel	47	50	40	30	51	42	33	42	54	39
Scénario bureaux	50	53	53	51	52	52	51	52	54	51



(a) Zone résidentielle.

(b) Zone de bureaux.

FIGURE 2 – Complétude relative pour chaque super-sniffeur de redondance m.

type `pcap`. Nous avons développé un outil pour synchroniser et fusionner hors-ligne les traces des sniffeurs individuels afin de générer la trace du super-sniffeur correspondant [SFA22].

Nous montrons dans le Tableau 1 la complétude relative des sniffeurs individuellement pour deux scénarios différents. Dans les deux cas, les sniffeurs sont regroupés et n'ont que quelques centimètres les séparant. Dans le premier scénario, résidentiel, le nombre d'équipements détectés est moins important et le trafic Wi-Fi plutôt léger, autour de 1 200 paquets/s. Dans le second scénario, dans une zone occupée par des bureaux, la densité spatiale est forte et le trafic Wi-Fi beaucoup plus intense, générant une charge sur les sniffeurs d'environ 10 000 paquets/s.

Nous pouvons observer que dans le scénario résidentiel peu dense, la complétude relative varie entre 30% et 54%. Dans le scénario composé de bureaux, la complétude se situe entre 50% and 54%. On pourrait s'attendre à ce que les captures soient meilleures dans l'environnement résidentiel où la densité de trafic est moindre. Cependant, ces valeurs pouvant paraître contre-intuitives de prime abord, sont en fait tout à fait naturelles. Alors que dans le cas résidentiel les paquets captés proviennent de sources éparses et plus ou moins éloignées des sniffeurs, dans le scénario des bureaux la plupart des sources de paquets sont à proximité. Or, un signal faible (source plus éloignée) donne lieu à une qualité de capture moindre. Par ailleurs, quel que soit le scénario considéré nous sommes surpris du faible taux de capture observé. Le meilleur des sniffeurs n'a capturé que 54% des paquets dans les deux scénarios.

4 Redondance

La Figure 2 montre la complétude relative d'un super-sniffeur dont la redondance varie entre 1 et 10, et ce dans les deux scénarios décrits plus haut. Pour chaque valeur de redondance, nous pouvons voir plusieurs valeurs de complétude, chacune correspondant à une combinaison particulière de sniffeurs (voir Section 2).

Nous marquons respectivement par les courbes orange et bleue, les valeurs de complétude maximale et minimale. Toutes les valeurs intermédiaires sont représentées par un point gris. Lorsque $m = 1$ la complétude est identique à celle montrée dans le Tableau 1. Pour aider le lecteur à mieux analyser les résultats, nous montrons sur le graphe la différence entre C^m_{max} la valeur de la complétude obtenue par le meilleur des super-sniffeurs de taille m et C^{m-1}_{max} la valeur maximale parmi les super-sniffeurs de taille directement inférieure (nombres indiqués au-dessus de la courbe orange).

Assez naturellement, nous pouvons observer que la complétude croît à mesure que la redondance augmente. Plus surprenant cependant, la complétude ne cesse d'augmenter pour chaque augmentation de la

taille du super-sniffeur. Nous nous attendions à ce que tous les paquets soient capturés avec quelques sniffeurs, mais cette situation ne s'est jamais produite. Même avec l'ajout d'un dixième sniffeur, des nouveaux paquets sont collectés. Nous observons également que la complétude obtenue pour des super-sniffeurs de même taille varie de manière importante dans l'environnement résidentiel alors que les valeurs sont beaucoup plus stables dans l'environnement de bureaux. Le choix de la combinaison de sniffeurs retenue pour constituer chaque super-sniffeur impacte jusqu'à 20% la valeur de la complétude dans notre environnement résidentiel. Nous notons que l'impact du choix des sniffeurs sur la complétude s'amenuise à mesure que la taille du super-sniffeur augmente.

Dans l'environnement de bureau, la complétude entre le meilleur et le moins bon des super-sniffeurs de même taille ne varie pas de plus de 4%. Cette grande stabilité dénote une grande homogénéité des résultats quels ques soient les regroupements de sniffeurs considérés. Conclure sur de meilleurs performances dans l'environnement de bureau n'est cependant pas si évident car la complétude la plus élevée pour un super-sniffeur de taille donnée est obtenue dans l'environnement résidentiel.

La combinaison de 5 sniffeurs permet de capturer 80% des paquets dans l'environnement de bureau tandis que cette même proportion est obtenue en combinant entre 4 et 9 sniffeurs pour l'environnement résidentiel. Il est intéressant de noter que pour chaque scénario, la plus haute valeur de la complétude n'est atteinte qu'avec le super-sniffeur de taille maximum. Atteindre cette valeur avec moins de sniffeurs aurait représenté un bon indicateur que tous les paquets ont effectivement été capturés. Ces expérimentations ne nous permettent pas de conclure que nous sommes en mesure de capter tous les paquets transmis, même lorsque dix sniffeurs sont co-localisés.

5 Conclusion et travaux futurs

Dans cet article, nous menons des expérimentations pour évaluer si un sniffeur seul peut capturer une trace complète de son environnement. Le périmètre de détection d'un sniffeur étant difficile à évaluer, nous co-localisons 10 sniffeurs et nous définissons une trace complète comme étant celle obtenue après synchronisation et fusion de l'ensemble des traces collectées. Nos mesures montrent qu'un sniffeur seul ne capture qu'une fraction du trafic total. Nous concluons sur la nécessité de combiner plusieurs sniffeurs pour fiabiliser la mesure passive du trafic Wi-Fi. Par la suite, nous prévoyons d'étudier comment utiliser les informations de collecte d'un super-sniffeur (par exemple le nombre de sniffeurs ayant capturés un même paquet) pour caractériser le trafic.

Références

[CBB+06] Yu-Chung Cheng, John Bellardo, Péter Benkö, Alex C. Snoeren, Geoffrey M. Voelker, and Stefan 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*, 2006.

[CLIU10] 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, October 2010.

[GAOS14] 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*, 14 :10804–10828, 2014.

[MRWZ06] R. Mahajan, M. Rodrig, D. Wetherall, and J. Zahorjan. Analyzing the MAC-level behavior of wireless networks in the wild. *SIGCOMM Comput. Commun. Rev.*, 36, August 2006.

[SFA22] M. I. Syed, A. Flandenmuller, and M. Dias De Amorim. PyPal : Wi-Fi Trace Synchronization and Merging Python Tool. Technical report, LIP6 UMR 7606, UPMC Sorbonne Université, France, March 2022.