

Sniff'eirb, sniffer intelligent

# Rapport final

Nicolas Verdier      Nicolas Retrain      Gabrielle Schmidt  
Benjamin Vandenberghe      Fatima Iddir

**Responsable :** M. Pierre Lalet  
M. Mathieu Blanc

8 janvier 2013

# Table des matières

<b>Introduction</b>	<b>2</b>
<b>1 Les besoins et l'état de l'art</b>	<b>3</b>
1.1 L'état de l'art . . . . .	3
1.2 Nos objectifs . . . . .	3
1.2.1 Premières étapes . . . . .	3
1.2.2 La valeur ajoutée de Sniff'airb . . . . .	4
1.2.3 Les idées abandonnées ou non implémentées . . . . .	4
<b>2 L'architecture</b>	<b>5</b>
2.1 L'architecture générale . . . . .	5
2.2 La base du sniffer : la récupération des paquets et leur stockage .	5
2.3 La reconstruction... . . . .	5
2.3.1 ... des flux . . . . .	5
2.3.2 ... puis des documents . . . . .	5
2.4 L'affichage . . . . .	5
2.5 Choix des outils . . . . .	5
<b>3 L'implémentation</b>	<b>6</b>
3.1 La base du sniffer : la récupération des paquets et leur stockage .	6
3.2 La reconstruction . . . . .	6
3.2.1 Recontruction d'un flux . . . . .	6
3.2.2 Recontruction d'un Document . . . . .	7
3.3 L'affichage . . . . .	8
<b>Conclusion</b>	<b>8</b>
<b>Annexes</b>	<b>9</b>

# Introduction

# Chapitre 1

## Les besoins et l'état de l'art

### 1.1 L'état de l'art

Il existe un grand nombre de logiciels et de bibliothèques qui proposent des fonctionnalités de captures réseaux de bas niveau et d'analyse de paquets (d'un point de vue utilisateur final). Nous pouvons notamment citer libpcap, Scapy, Winpcap ou encore tcpdump. Ces outils ont été un modèle d'inspiration pour tenter de réaliser notre propre *Sniffer*<sup>1</sup> réseau.

### 1.2 Nos objectifs

#### 1.2.1 Premières étapes

L'objectif premier était tout d'abord de retrouver les fonctionnalités principales de Wireshark, comme par exemple afficher une liste des paquets circulant sur le réseaux ainsi que les informations s'y rapportant (timestamp, adresse IP source, adresse IP destination, port, protocole...), tout en apportant une solution différente et plus simple d'utilisation à un utilisateur néophyte. Ainsi nous avons défini trois premiers objectifs :

- Récupérer les communications<sup>2</sup> en clair d'un réseau.
- Stocker les paquets de cette communication.
- Afficher les différents flux<sup>3</sup> possibles de cette communication.

---

1. Terme anglicisé désignant un dispositif permettant d'analyser le trafic d'un réseau

2. Une communication est définie par un échange réseau entre deux paires (Adresse IP ; Port).

3. Un flux est défini comme un assemblage possible des paquets d'une communication.

### 1.2.2 La valeur ajoutée de Sniff'eirb

Le but de ce projet n'étant pas de reproduire un Wireshark, nous devons y apporter des atouts significatifs. Plusieurs fonctionnalités ont donc été implémentées :

**L'arbre des possibilités pour les flux TCP :** Lors de la capture des paquets circulant sur le réseau pour une même communication, des événements inattendus, comme la répétition de paquets ou encore la perte de certains, pouvaient survenir. La capture de tous les paquets permet donc de connaître les différentes façons de les réassembler pour obtenir un flux décrivant la communication.

**La gestion du protocole HTTP :** La reconstruction des pages webs visitées par un utilisateur sur le réseau était le fil conducteur du projet. L'analyse des paquets circulant sur le réseau est le principal outil pour arriver à cette fin.

### 1.2.3 Les idées abandonnées ou non implémentées

En parallèle, beaucoup d'idées durent être abandonnée du fait d'un manque de temps et d'énergie pour les traiter. Voici une liste non exhaustive de ces idées :

## Chapitre 2

# L'architecture

### 2.1 L'architecture générale

### 2.2 La base du sniffer : la récupération des paquets et leur stockage

### 2.3 La reconstruction...

#### 2.3.1 ... des flux

#### 2.3.2 ... puis des documents

### 2.4 L'affichage

### 2.5 Choix des outils

-python -javascript

## Chapitre 3

# L'implémentation

### 3.1 La base du sniffer : la récupération des paquets et leur stockage

- patch de scapy pour pouvoir sniffer infiniment (jusqu'à l'appel d'une fonction)
- mongodb, no sql, index, -pymongo

### 3.2 La reconstruction

Un des principaux objectifs de Sniff'eirb était de rendre l'exploitation des données capturées la plus évidente possible. Nous avons donc décidé de permettre l'affichage des échanges HTTP, notamment celui des pages HTML et des images. Pour ce faire, nous avons commencé par reconstruire les différents flux de données.

#### 3.2.1 Reconstruction d'un flux

Comme nous ne pouvions pas implémenter la reconstruction de flux de la même façon pour tous les protocoles de couche 4 (UDP et TCP) et que nous voulions présenter un résultat démonstratif même pour le néophyte, nous nous sommes consacrés au protocole\* TCP, puis au média\* HTTP.

Il a donc tout d'abord fallu reconnaître le protocole qui nous intéresse et le média, avant d'en reconstruire le flux. Comme nous l'avons vu dans la partie précédente, le protocole est facilement retrouvable directement dans la base de données (puisqu'il s'agit d'un champ du paquet IP). Pour le média, cela a été plus technique. Au départ, nous l'avons simplement déterminé grâce au port (source ou destination). En effet, traditionnellement, certains ports sont utilisés pour certains médias (par exemple, 80 pour HTTP). Mais en ne nous concentrant que sur cette information, nous aurions manqué toutes les communications d'un média sur d'autres ports (comme 8080 pour HTTP), qui, d'un point de

vue de la sécurité, pouvaient être les plus intéressantes. Nous avons donc décidé de parser la charge utile du paquet TCP à la recherche d'expression régulière trahissant le HTTP(S), comme les en-têtes *HTTP/numero\_de\_version* et les mots clés *GET* ou *POST* ou des URLs.

À ce moment-là, nous disposons donc de l'ensemble des paquets appartenant à la même communication\* ainsi que du média qui correspond. Comme le protocole est TCP, il paraît assez facile de reconstruire le flux : on classe les paquets par numéro de séquence croissant et on concatène les charges utiles. Cependant, cette démarche n'est pas possible. En effet, il arrive parfois que plusieurs paquets arrivent avec le même numéro de séquence, voire qu'il manque des paquets au milieu du flux. C'est pourquoi nous avons dû prévoir de reconstruire tous les flux possibles.

Pour ce faire, nous avons créé une structure de liste doublement chaînée que nous avons ensuite parcouru pour en tirer tous les flux possibles. Pour remplir cette chaîne doublement chaînée, baptisée *LianaTree*, nous avons simplement parcouru la liste des paquets d'une communication en commençant par celui qui a le plus petit numéro de séquence et nous cherchions tous ses successeurs. Le successeur d'un paquet ayant pour numéro de séquence la somme de la taille de sa charge utile et du numéro de séquence de son prédécesseur (modulo  $2 * 32$ ). Cependant, nous avons un problème dans certains cas, car le numéro de séquence TCP est pris au hasard parmi  $2 * 32$  possibilités et il était donc possible d'avoir la fin du flux avant le début (si le numéro de séquence du début de la communication était proche de  $2 * 32 - 1$ ). Finalement, nous recherchons simplement les successeurs et prédécesseurs de chaque paquet via le numéro de séquence et la taille de la charge utile.

Une fois cette liste créée, nous devons la parcourir en entier en partant du paquet qui n'a pas de prédécesseur. Si plusieurs paquets n'ont pas de prédécesseur, nous recommençons le parcours autant de fois que nécessaire. Nous réalisons un parcours en largeur de la liste qui découle de ce paquet et nous concaténons les charges utiles au fur et à mesure. À la fin, nous éliminons les doublons et renvoyons le résultat.

Dans le cas où nous aurions plusieurs flux, nous avons décidé de les afficher tous, mais dans l'ordre où ils nous paraissent le plus cohérent. Cette cohérence est définie spécifiquement pour le type de document. Dans le cas de document HTML en texte, nous calculons le pourcentage de balises bien fermées par rapport au nombre total.

### 3.2.2 Reconstruction d'un Document

Reconnaissance des protocoles : HTTP, FTP, HTTPS, IMAP, SNMP, ...  
-reconstruction des pages html (.html et .gzip!) -reconstruction des images



### 3.3 L'affichage

-création d'un serveur web minimaliste -template html -jquery datatables  
bootstrap -Ajax

# Conclusion

# Annexes