TELECOM nancy
Ingénieurs du numérique • Inspiring your digital future

École associée
INSTITUT
Mines-Télécom

UNIVERSITÉ DE LORRAINE

Export de NetFlows 9 sous Android et **Localisation d'un utilisateur de smartphone** à partir de ses flux réseaux

5 septembre 2013

Julien VAUBOURG

Soutenance de stage 3^{e} année

Loria
Laboratoire lorrain de recherche en informatique et ses applications

Inria
INVENTEURS DU MONDE NUMÉRIQUE

Introduction

Contexte

- Stage 3^e année de TELECOM Nancy

Contexte

- Stage 3^e année de TELECOM Nancy
- 4 mois

Contexte

- Stage 3^e année de TELECOM Nancy
- 4 mois
- Orienté recherche

Contexte

- Stage 3^e année de TELECOM Nancy
- 4 mois
- Orienté recherche
- Partie ingénierie

Madynes

- Stage Inria

Madynes

- Stage Inria
- Établissement public de recherche

Madynes

- Stage Inria
- Établissement public de recherche
- Sciences et technologies de l'information et de la communication

Madynes

- Stage Inria
- Établissement public de recherche
- Sciences et technologies de l'information et de la communication
- LORIA : Unité Mixte de Recherche

Madynes

- Stage Inria
- Établissement public de recherche
- Sciences et technologies de l'information et de la communication
- LORIA : Unité Mixte de Recherche
- Équipe Madynes (Abdelkader LAHMADI)

Objectifs

1. Développer un **exportateur NetFlow** 9 modulaire pour Android

2. Travailler sur la **localisation d'un utilisateur de smartphone** à partir de ses flux réseaux

Plan

Plan

1. Android et NetFlow 9
2. Exportateur de NetFlows

Plan

1. Android et NetFlow 9
2. Exportateur de NetFlows
3. Inférence de la localisation

Plan

1. Android et NetFlow 9
2. Exportateur de NetFlows
3. Inférence de la localisation
4. Résultats

Android et NetFlows 9

Android

- Système d'exploitation

Android

- Système d'exploitation
- Smartphones et tablettes

Android

- Système d'exploitation
- Smartphones et tablettes
- Distribution GNU/Linux

Android

- Système d'exploitation
- Smartphones et tablettes
- Distribution GNU/Linux
- Libre (licence Apache)

Android

- Système d'exploitation
- Smartphones et tablettes
- Distribution GNU/Linux
- Libre (licence Apache)
- Système mobile le plus populaire

NetFlows

- Synthétisation de flux **unidirectionnels**

NetFlows

- Synthétisation de flux **unidirectionnels**
- Paquets avec propriétés communes (IP et ports)

NetFlows

- Synthétisation de flux **unidirectionnels**
- Paquets avec propriétés communes (IP et ports)
- Version 9 : fonctionnement par **templates**

Fonctionnement par templates

- Ajout dynamique de champs

Fonctionnement par templates

- Ajout dynamique de champs
- Sémantique détachée du format

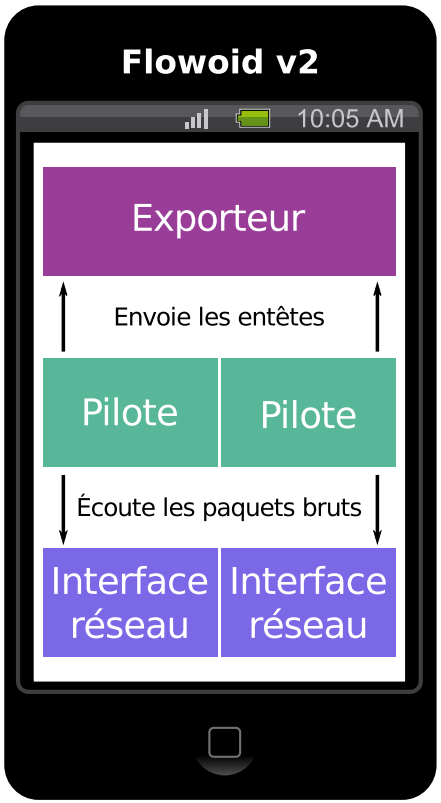
Fonctionnement par templates

- Ajout dynamique de champs
- Sémantique détachée du format
- Optimisation des envois avec plusieurs templates

Fonctionnement par templates

- Ajout dynamique de champs
- Sémantique détachée du format
- Optimisation des envois avec plusieurs templates
- Templates envoyés régulièrement mais pas systématiquement

Format des NetFlows



Exemple de paquet d'export

Set de modèles

Modèle (1): src is IPv6; dst is IPv6; size is int
Modèle Record (2): src is IPv4; dst is IPv4; size is int

Set de données (1)

NetFlow: src=fe80::1; dst=fe80::2; size=1337
NetFlow: src=fe80::2; dst=fe80::1; size=512

Set de données (2)

NetFlow: src=192.88.99.1; dst=192.88.99.2; size=42
NetFlow: src=192.88.99.2; dst=192.88.99.1; size=53

Envoie paquet d'export →

Exportateur de NetFlows

Exportateur NetFlow pour Android

- **Inexistant** sur le marché

Exportateur NetFlow pour Android

- **Inexistant** sur le marché
- **Nécessaire** pour l'analyse des flux réseau

Exportateur NetFlow pour Android

- **Inexistant** sur le marché
- **Nécessaire** pour l'analyse des flux réseau
- **Scalabilité** et modularité

Bibliothèque abstraite

- Bibliothèque d'export **générique**

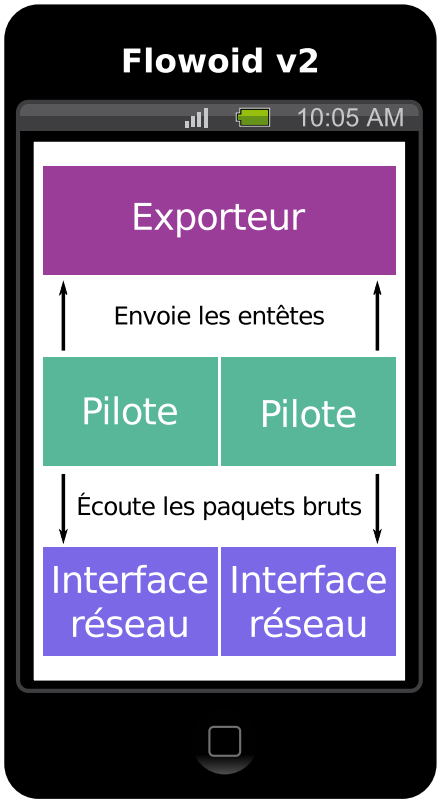
Bibliothèque abstraite

- Bibliothèque d'export **générique**
- **Implémentation stricte** de la RFC 3954 (NetFlow 9)

Bibliothèque abstraite

- Bibliothèque d'export **générique**
- **Implémentation stricte** de la RFC 3954 (NetFlow 9)
- **Interfaces claires** et souples

Pilotes de communication



Exemple de paquet d'export

Set de modèles

Modèle (1): src is IPv6; dst is IPv6; size is int
Modèle Record (2): src is IPv4; dst is IPv4; size is int

Set de données (1)

NetFlow: src=fe80::1; dst=fe80::2; size=1337
NetFlow: src=fe80::2; dst=fe80::1; size=512

Set de données (2)

NetFlow: src=192.88.99.1; dst=192.88.99.2; size=42
NetFlow: src=192.88.99.2; dst=192.88.99.1; size=53

Envoie paquet d'export →

Pilotes de communication

- Capturent les paquets du réseau

Pilotes de communication

- Capturent les paquets du réseau
- **Bas niveau**

Pilotes de communication

- Capturent les paquets du réseau
- **Bas niveau**
- Nécessite des privilèges

Pilotes de communication

- Capturent les paquets du réseau
- **Bas niveau**
- Nécessite des privilèges
- **Code natif** (C)

Pilotes de communication

- Capturent les paquets du réseau
- **Bas niveau**
- Nécessite des privilèges
- **Code natif** (C)
- Pilote **IPv4 et IPv6** fourni

Flowoid v2

- Exportateur de NetFlows pour Android

Flowoid v2

- Exportateur de NetFlows pour Android
- v1 par Madynes avec un exportateur externe en C

Flowoid v2

- Exportateur de NetFlows pour Android
- v1 par Madynes avec un exportateur externe en C
- Difficile à maintenir / faire évoluer

Flowoid v2

- Exportateur de NetFlows pour Android
- v1 par Madynes avec un exportateur externe en C
- Difficile à maintenir / faire évoluer
- **Implémentation de la bibliothèque**

Difficultés

- Pas de reprise de code

Difficultés

- Pas de reprise de code
- Cycle de vie de Android

Difficultés

- Pas de reprise de code
- Cycle de vie de Android
- Nom de l'interface réseau active

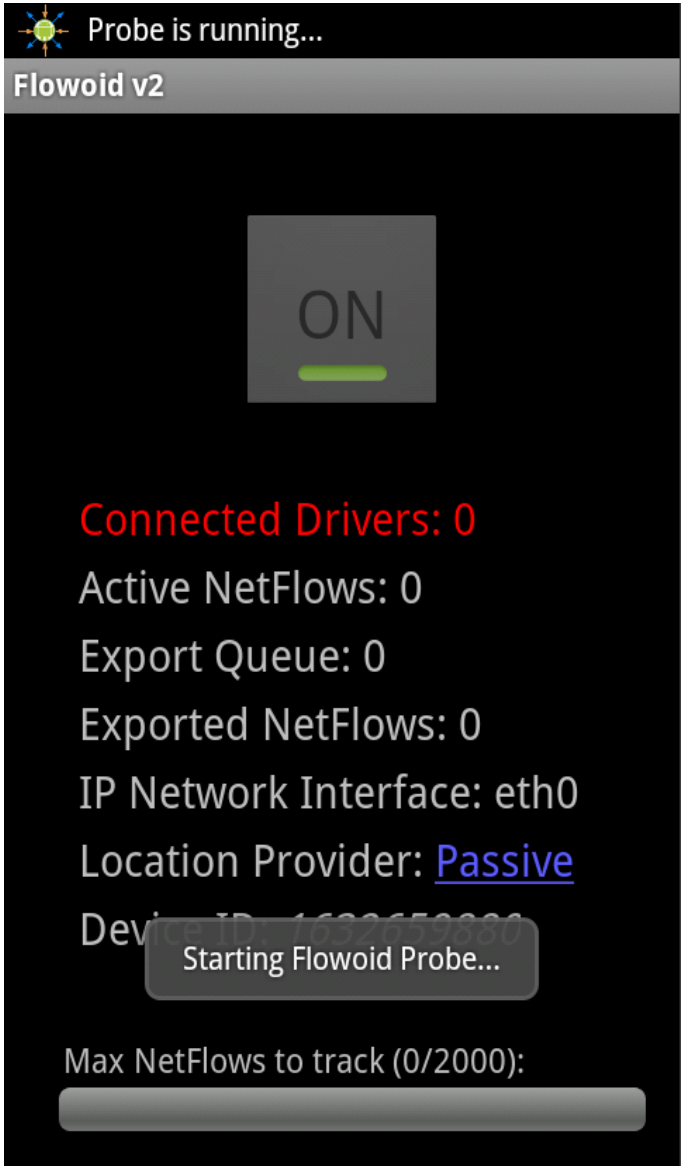
Difficultés

- Pas de reprise de code
- Cycle de vie de Android
- Nom de l'interface réseau active
- Application associée au flux

Difficultés

- Pas de reprise de code
- Cycle de vie de Android
- Nom de l'interface réseau active
- Application associée au flux
- Contraintes du collecteur

Aperçu de Flowoid v2



Aperçu de Flowoid v2

Inférence de la localisation

Recherche

≪ Inférence de la localisation d'un utilisateur de smartphone uniquement à partir de ses flux réseaux et sans données géotagguées ≫

Recherche

≪ Inférence de la localisation d'un utilisateur de smartphone uniquement à partir de ses flux réseaux et sans données géotagguées ≫

- **Hypothèse 0 :**
  1 type de lieu = 1 activité spécifique

Recherche

≪ Inférence de la localisation d'un utilisateur de smartphone uniquement à partir de ses flux réseaux et sans données géotagguées ≫

- **Hypothèse 0 :**
  1 type de lieu = 1 activité spécifique
- **Hypothèse 1 :**
  1 séquence de NetFlows récurrente (*pattern*) = 1 type de lieu

Recherche

≪ Inférence de la localisation d'un utilisateur de smartphone uniquement à partir de ses flux réseaux et sans données géotagguées ≫

- **Hypothèse 0 :**
  1 type de lieu = 1 activité spécifique
- **Hypothèse 1 :**
  1 séquence de NetFlows récurrente (*pattern*) = 1 type de lieu
- **Objectif :**
  Trouver de quand à quand l'utilisateur était au lieu A, au lieu B, etc.

Organisation du travail

1. État de l'art

Organisation du travail

1. État de l'art

2. Pertinence des informations à disposition

Organisation du travail

1. État de l'art
2. Pertinence des informations à disposition
    - Adresses IP de destination

Organisation du travail

1. État de l'art
2. Pertinence des informations à disposition
    - Adresses IP de destination
3. Construction d'un modèle

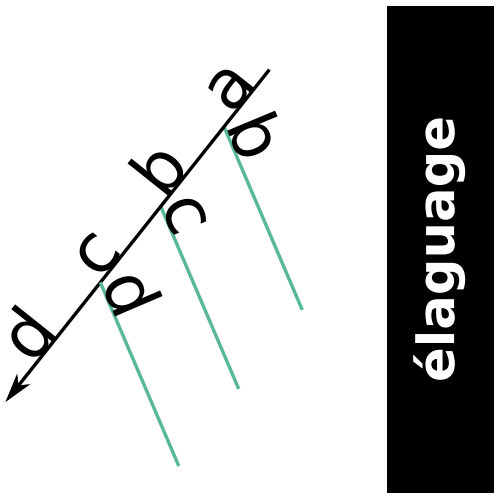
Organisation du travail

1. État de l'art
2. Pertinence des informations à disposition
   - Adresses IP de destination
3. Construction d'un modèle
4. Expérimentations

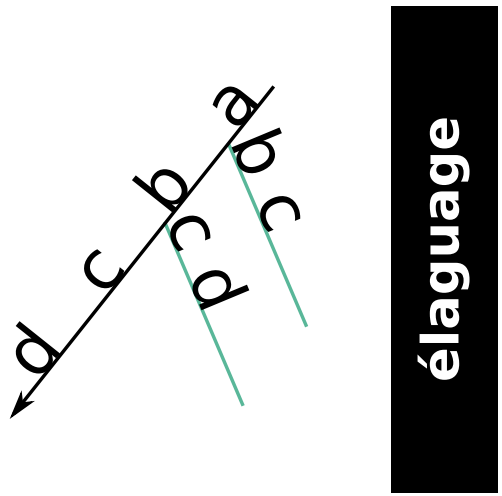
Organisation du travail

1. État de l'art
2. Pertinence des informations à disposition
    - Adresses IP de destination
3. Construction d'un modèle
4. Expérimentations
5. Conclusions

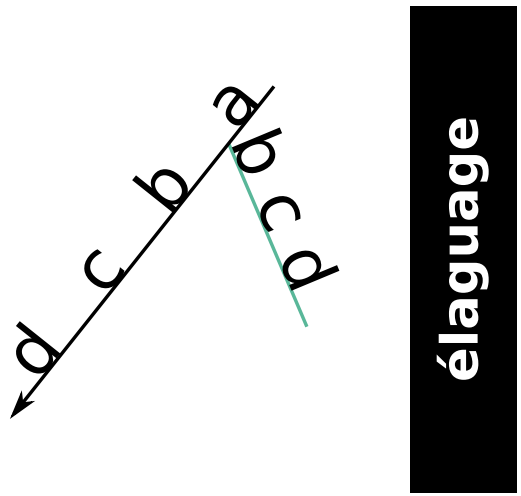
Recherche des patterns



Itération n°1



Itération n°2



Iteration n°3

Regroupement des patterns

1. **Nombre cohérent** de patterns

Regroupement des patterns

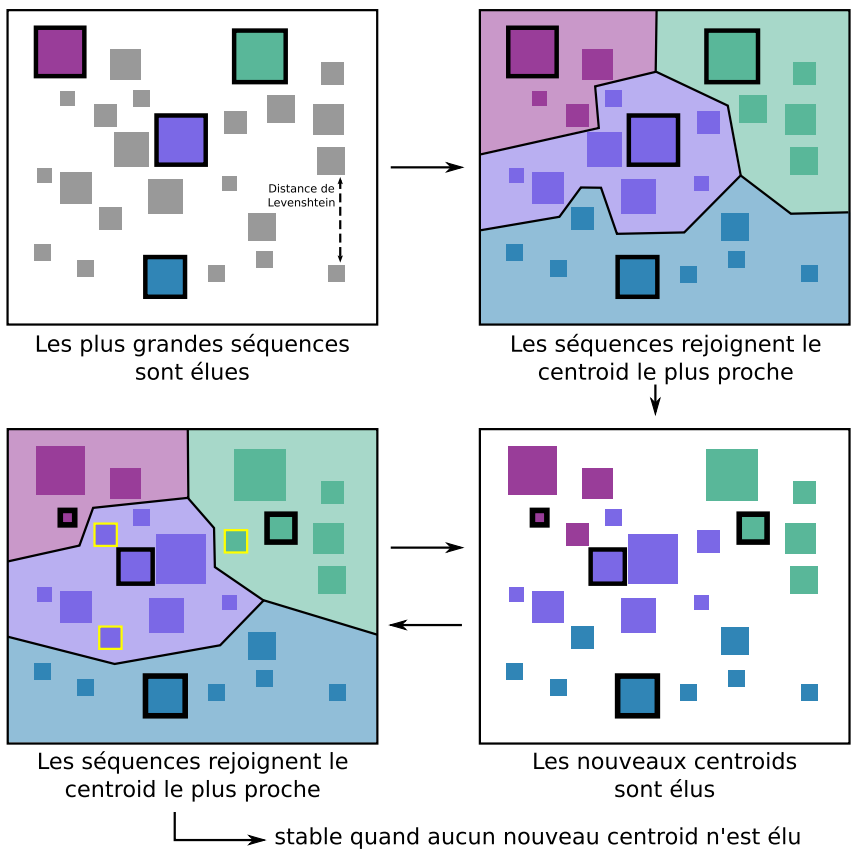
1. **Nombre cohérent** de patterns
2. K-means avec un **nombre imposé**

Regroupement des patterns

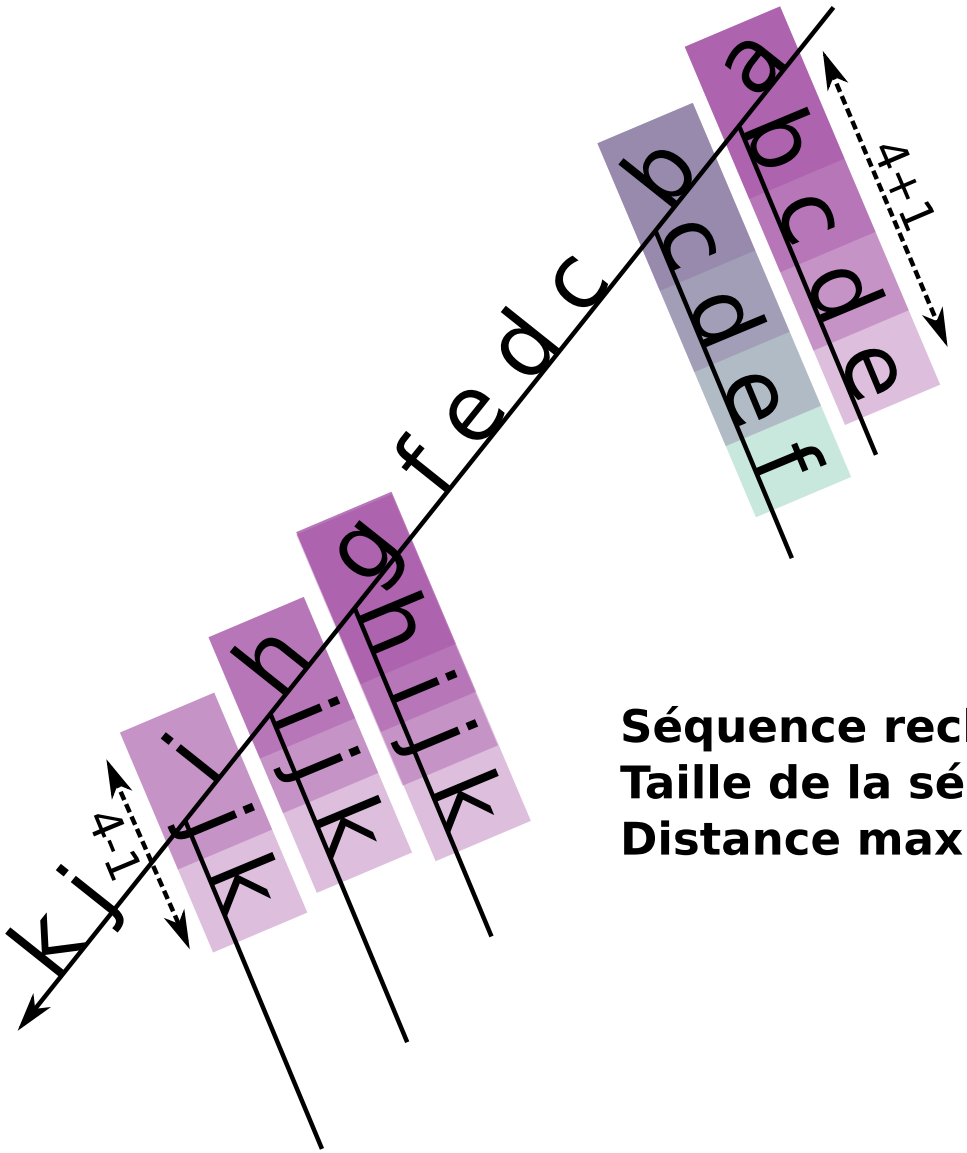
1. **Nombre cohérent** de patterns
2. K-means avec un **nombre imposé**
3. Domicile, Déplacements, Travail, Loisirs

Regroupement des patterns

1. **Nombre cohérent** de patterns
2. K-means avec un **nombre imposé**
3. Domicile, Déplacements, Travail, Loisirs



Application des patterns



Séquence recherchée : c d e f
Taille de la séquence : 4
Distance max. : 1

Résultats

Expérimentation

Périodes témoins :

- **Coordonnées géographiques**

Expérimentation

Périodes témoins :

- **Coordonnées géographiques**
- Nom des **applications associées**

Expérimentation

Périodes témoins :

- **Coordonnées géographiques**
- Nom des **applications associées**
- **Fusion à la main** des flux et lieux similaires

Expérimentation

Périodes témoins :

- **Coordonnées géographiques**
- Nom des **applications associées**
- **Fusion à la main** des flux et lieux similaires
- **Classification** des types lieux et réduction à 4 différents

Expérimentation

Périodes témoins :

- **Coordonnées géographiques**
- Nom des **applications associées**
- **Fusion à la main** des flux et lieux similaires
- **Classification** des types lieux et réduction à 4 différents
- **Interrogation de l'utilisateur**

Expérimentation

Périodes témoins :

- **Coordonnées géographiques**
- Nom des **applications associées**
- **Fusion à la main** des flux et lieux similaires
- **Classification** des types lieux et réduction à 4 différents
- **Interrogation de l'utilisateur**

Expérimentation :

Expérimentation

Périodes témoins :

- **Coordonnées géographiques**
- Nom des **applications associées**
- **Fusion à la main** des flux et lieux similaires
- **Classification** des types lieux et réduction à 4 différents
- **Interrogation de l'utilisateur**

Expérimentation :

- Découverte des **patterns**

Expérimentation

Périodes témoins :

- **Coordonnées géographiques**
- Nom des **applications associées**
- **Fusion à la main** des flux et lieux similaires
- **Classification** des types lieux et réduction à 4 différents
- **Interrogation de l'utilisateur**

Expérimentation :

- Découverte des **patterns**
- Détection des **périodes**

Expérimentation

Périodes témoins :

- **Coordonnées géographiques**
- Nom des **applications associées**
- **Fusion à la main** des flux et lieux similaires
- **Classification** des types lieux et réduction à 4 différents
- **Interrogation de l'utilisateur**

Expérimentation :

- Découverte des **patterns**
- Détection des **périodes**
- **Comparaison** avec les périodes témoins

Difficultés

- Périodes témoins peu fiables

Difficultés

- Périodes témoins peu fiables
- Sollication forte de l'utilisateur

Difficultés

- Périodes témoins peu fiables
- Sollication forte de l'utilisateur
- Immiscion dans sa **vie privée**

Résultats

- Changements supposés de lieu à des heures étranges

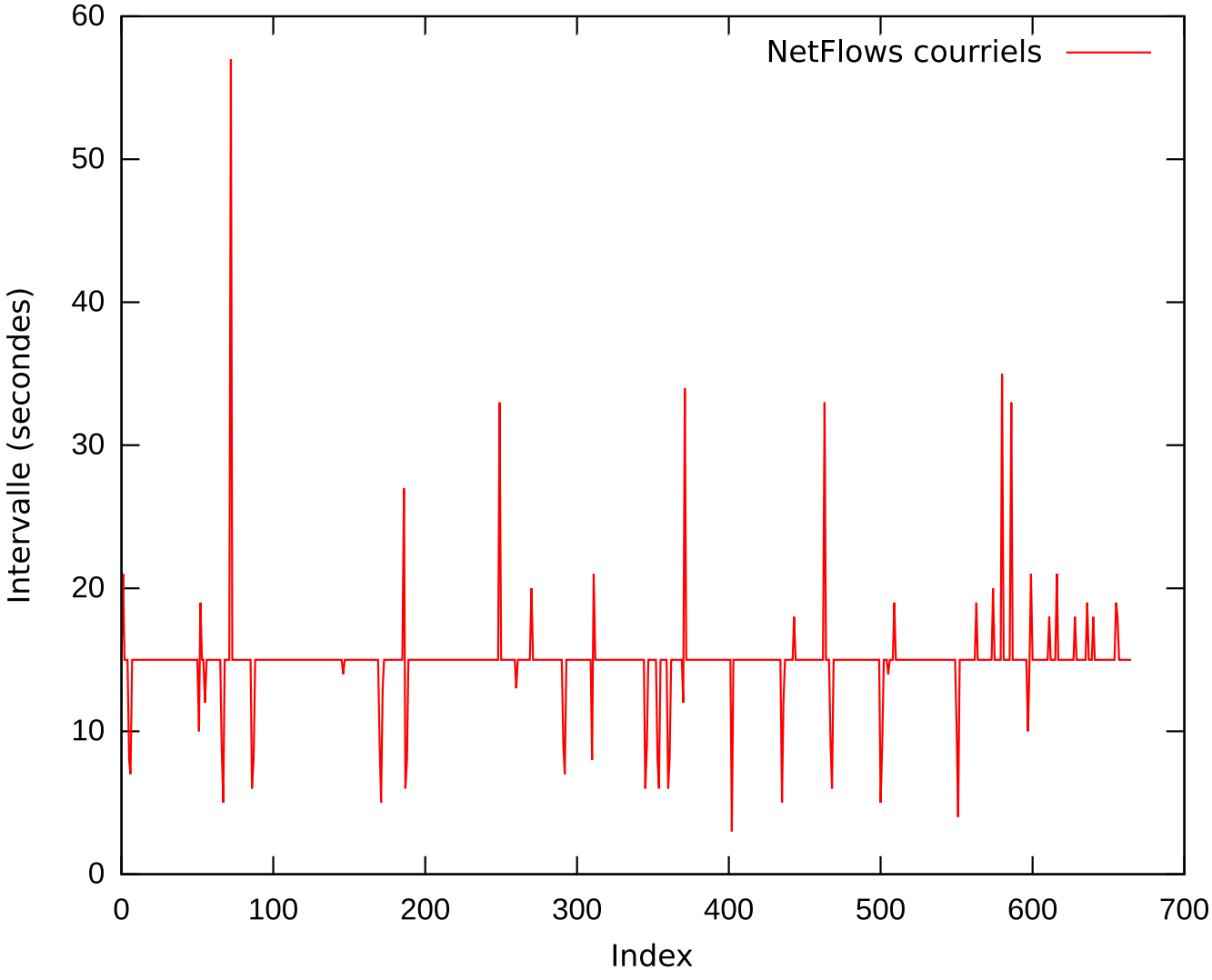
Résultats

- Changements supposés de lieu à des heures étranges
- Patterns pollués par des IP récurrentes

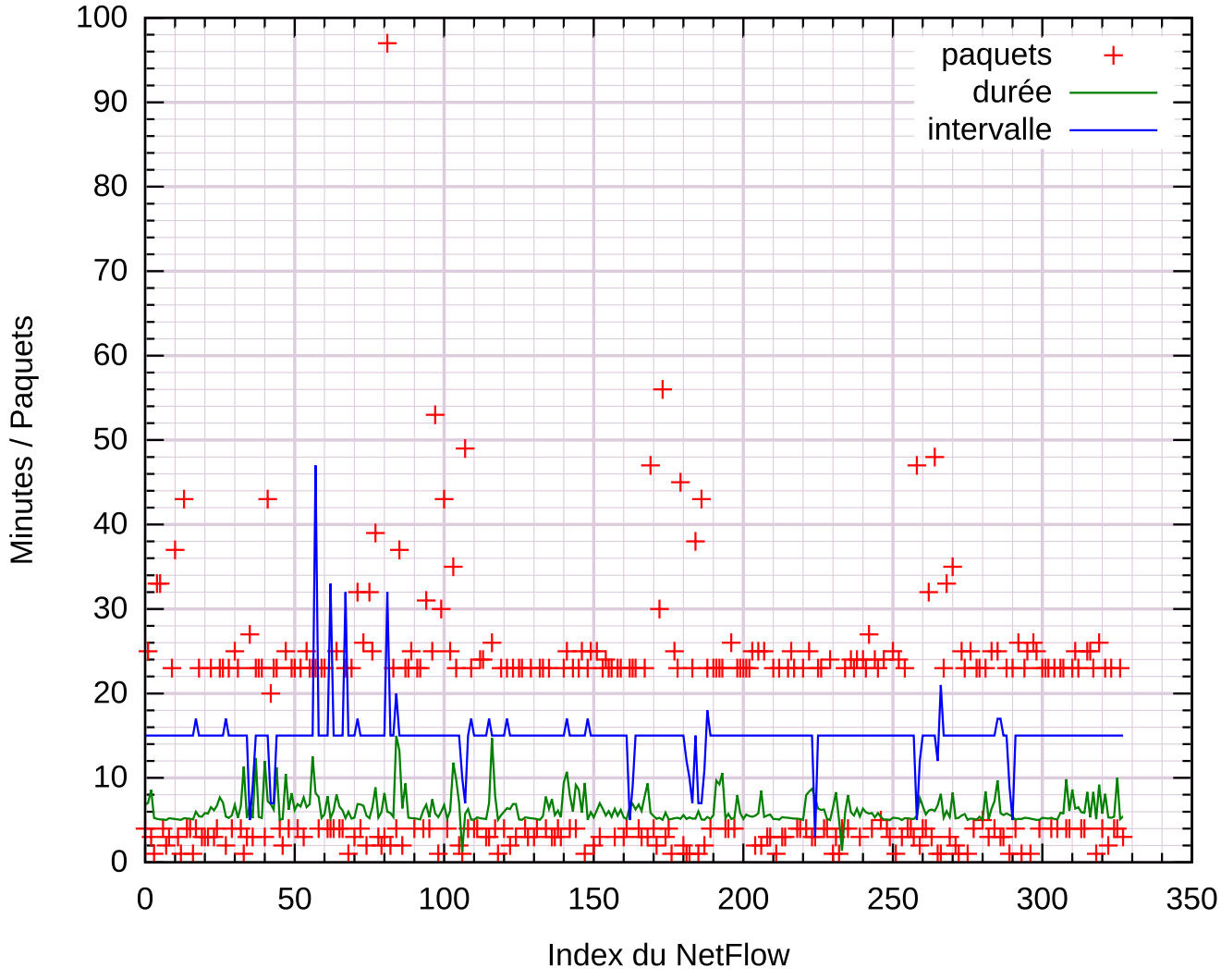
Résultats

- Changements supposés de lieu à des heures étranges
- Patterns pollués par des IP récurrentes
- Bruit de fond (routines) trop important

Distinction des routines



Distinction des routines



Distinction des routines

- 59% courriels (vérification régulière)

Distinction des routines

- 59% courriels (vérification régulière)
- 25% Facebook (synchronisation)

Distinction des routines

- 59% courriels (vérification régulière)
- 25% Facebook (synchronisation)
- 9% Yahoo (météo)

Distinction des routines

- 59% courriels (vérification régulière)
- 25% Facebook (synchronisation)
- 9% Yahoo (météo)
- 7% Google (cadre de travail Android)

Distinction des routines

- 59% courriels (vérification régulière)
- 25% Facebook (synchronisation)
- 9% Yahoo (météo)
- 7% Google (cadre de travail Android)

Potentiellement 100% de routines

Un résultat peut en cacher un autre

Résultats inattendus

- **Il n'y a pas plus régulier qu'une machine**

Résultats inattendus

- **Il n'y a pas plus régulier qu'une machine**
- Modèle condamné à **détecter les routines**

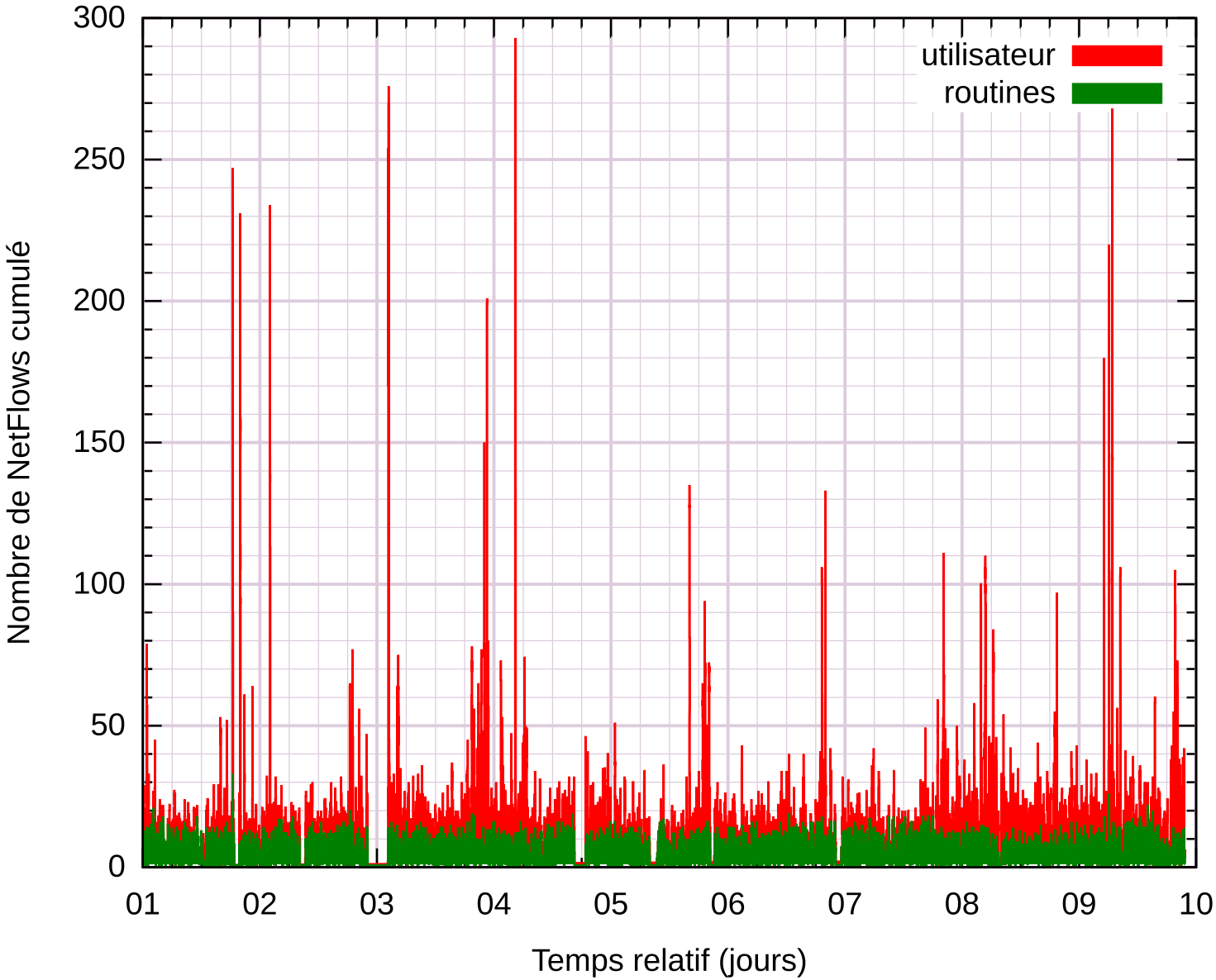
Résultats inattendus

- **Il n'y a pas plus régulier qu'une machine**
- Modèle condamné à **détecter les routines**
- Reprise du jeu de données

Résultats inattendus

- **Il n'y a pas plus régulier qu'une machine**
- Modèle condamné à **détecter les routines**
- Reprise du jeu de données
- Différenciation des flux détectés par les modèles

Résultats inattendus



Conclusion

- **Exportateur de NetFlows pour Android** fonctionnel

Ingénierie

- **Exportateur de NetFlows pour Android** fonctionnel
- **Modulaire et scalable**

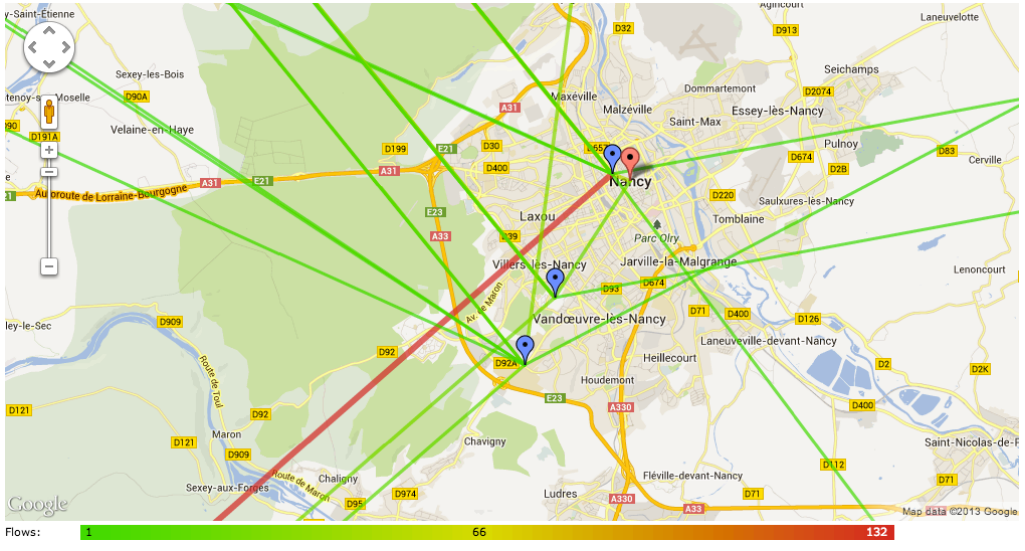
Ingénierie

- **Exportateur de NetFlows pour Android** fonctionnel
- **Modulaire et scalable**
- Entièrement **libre**

Ingénierie

- **Exportateur de NetFlows pour Android** fonctionnel
- **Modulaire et scalable**
- Entièrement **libre**
- Bibliothèque disponible pour **encourager la diversité**

Ingénierie



Recherche

- **Première expérience** très instructive

Recherche

- **Première expérience** très instructive
- Écriture d'un **papier de recherche**

Recherche

- **Première expérience** très instructive
- Écriture d'un **papier de recherche**
- **Résultats inattendus mais prometteurs**

Recherche

- **Première expérience** très instructive
- Écriture d'un **papier de recherche**
- **Résultats inattendus mais prometteurs**
- Et plus intéressants en terme de recherche

Perspectives

- **Confirmer** les résultats (expérience)

Perspectives

- **Confirmer** les résultats (expérience)
- **Affiner** le modèle en vue des nouveaux objectifs

Perspectives

- **Confirmer** les résultats (expérience)
- **Affiner** le modèle en vue des nouveaux objectifs
- **Adapter** le papier de recherche

Perspectives

- **Confirmer** les résultats (expérience)
- **Affiner** le modèle en vue des nouveaux objectifs
- **Adapter** le papier de recherche

Doctorat dans l'équipe MAIA
(en collaboration avec EDF R&D et Madynes)

Questions