Supervision des réseaux P2P structurés appliquée à la sécurité des contenus

Thibault CHOLEZ

Thèse pour l'obtention du Doctorat de l'Université Henri Poincaré - Nancy 1

réalisée sous la direction d'Isabelle CHRISMENT

au sein de l'équipe INRIA MADYNES

présentée et soutenue publiquement au LORIA, Vandœuvre-lès-Nancy, France

le 23 juin 2011

Nancy-Université
Université Henri Poincaré

Loria
Laboratoire lorrain de recherche en informatique et ses applications

INRIA

Les réseaux pair à pair

Principes :

- Systèmes distribués
- Échange direct de services : pairs à la fois clients et serveurs

Propriétés :

- Passage à l'échelle (agrégation des ressources)
- Tolérance aux pannes (réplication des informations)
- Distribution des coûts d'infrastructure (absence de centre)

Développement des réseaux pair à pair :

- Initialement : partage de fichiers, Napster (1999), Gnutella (2000)
- Aujourd'hui : nombreux services (systèmes de fichiers distribués, voix sur IP, streaming multimédia, etc.)
- Plus de la moitié du trafic global d'Internet (2008/2009)

Problèmatique

Sécurité des réseaux P2P :

- Absence de contrôle centralisé
- Autonomie des pairs

$\Rightarrow$ réseaux P2P victimes et supports d'activités malveillantes

- Problèmes de sécurité affectant les réseaux P2P :
  - Insertion de nœuds malveillants (attaque Sybil)
  - Pollution du mécanisme d'indexation (fichiers référencés par des noms trompeurs), suppression des contenus indexés (éclipse)
  - Diffusion de contenus illégaux (pédo-pornographie, malware)
- Problèmes liés :
  - Contenus malveillants diffusés à travers la pollution
  - Pollution dégrade la supervision (faux positifs)
  - Attaque Sybil support des applications malveillantes

Contributions de la thèse

Approche :

Développement de nouvelles méthodes de supervision appréhendant la sécurité des contenus

1. Supervision précise des accès aux contenus, évitant les faux positifs (pollution)
2. Recensement des attaques réelles (attaque Sybil, pollution) à l'échelle d'un réseau
3. Détection locale des attaques Sybil localisées et protection des pairs

Cas d'étude :

Réseaux P2P structurés à travers **KAD** : largement déployé, vulnérable, architecture distribuée, nombreux services

Plan

Plan

Supervision des réseaux P2P

Architectures P2P :

- Avec serveurs (centralisés, distribués, différenciés)
- Décentralisées non-structurées (pure, hybride [CRB+03])
- **Décentralisées structurées** : Tables de Hachage Distribuées (DHT, ex : Chord [SMK+01]) - Distribution complète, performances prouvées, localisation de données rares, nombreux déploiements

Architectures de supervision :

- Collecte du trafic P2P [SW04],
- Instrumentation de serveur [ALM09a], requêtes sur serveur
- Pots de miel [ALM09b], explorateurs [RIF02] [LHKM04] [SENB07b], sondes distribuées [MRGS09]

Supervision des réseaux P2P

Architectures P2P :

- Avec serveurs (centralisés, distribués, différenciés)
- Décentralisées non-structurées (pure, hybride [CRB+03])
- **Décentralisées structurées** : Tables de Hachage Distribuées (DHT, ex : Chord [SMK+01]) - Distribution complète, performances prouvées, localisation de données rares, nombreux déploiements

Architectures de supervision :

- Collecte du trafic P2P [SW04],
- Instrumentation de serveur [ALM09a], requêtes sur serveur
- Pots de miel [ALM09b], explorateurs [RIF02] [LHKM04] [SENB07b], sondes distribuées [MRGS09]

Limites :

- Supervision des contenus difficile pour les DHT
- Faux positifs non considérés

Sécurité des réseaux P2P

L'attaque Sybil [Dou02] :

- Découverte du réseau puis insertion de nombreux faux pairs
- Contrôle de la DHT possible (localement nb Sybils >> nb pairs)
- Applications : DDoS [NR06], éclipse [SENB07a], pollution etc.
- Nombreuses solutions proposées (contrôle centralisé [FMR+09], crypto-puzzle [CDG+02], contrainte sociale [YKGF06], etc.)

Contenus illégaux et pollution :

- Plusieurs types de pollution : fichiers corrompus [LKXR05], index poisoning [LNR06]
- Protections peu adaptées (vulnérables à l'attaque Sybil [CA07])

Sécurité des réseaux P2P

L'attaque Sybil [Dou02] :

- Découverte du réseau puis insertion de nombreux faux pairs
- Contrôle de la DHT possible (localement nb Sybils >> nb pairs)
- Applications : DDoS [NR06], éclipse [SENB07a], pollution etc.
- Nombreuses solutions proposées (contrôle centralisé [FMR+09], crypto-puzzle [CDG+02], contrainte sociale [YKGF06], etc.)

Contenus illégaux et pollution :

- Plusieurs types de pollution : fichiers corrompus [LKXR05], index poisoning [LNR06]
- Protections peu adaptées (vulnérables à l'attaque Sybil [CA07])

Limites :

- Aucune mesure récente de la pollution, peu de fichiers étudiés
- Vulnérabilités prouvées mais exploitation non observée
- Solutions inadaptées aux réseaux existants (rétrocompatibilité)

Le réseau P2P KAD

KAD est :

- Basé sur la DHT Kademlia [MM02] (localisation des pairs)
- Implanté par des clients open source (eMule et aMule)
- Largement utilisé (∼3 à 4 millions de pairs)

Identification des éléments du réseau :

- Éléments du réseau identifiés par KADID (128 bits)
- Pairs : KADID **aléatoire** ; Fichiers et Mots-clés : **hash** `MD4`
- Exemple : `MD4`(*avatar*) = `C0F70911A9C2E6F6960DDED0D4118244`

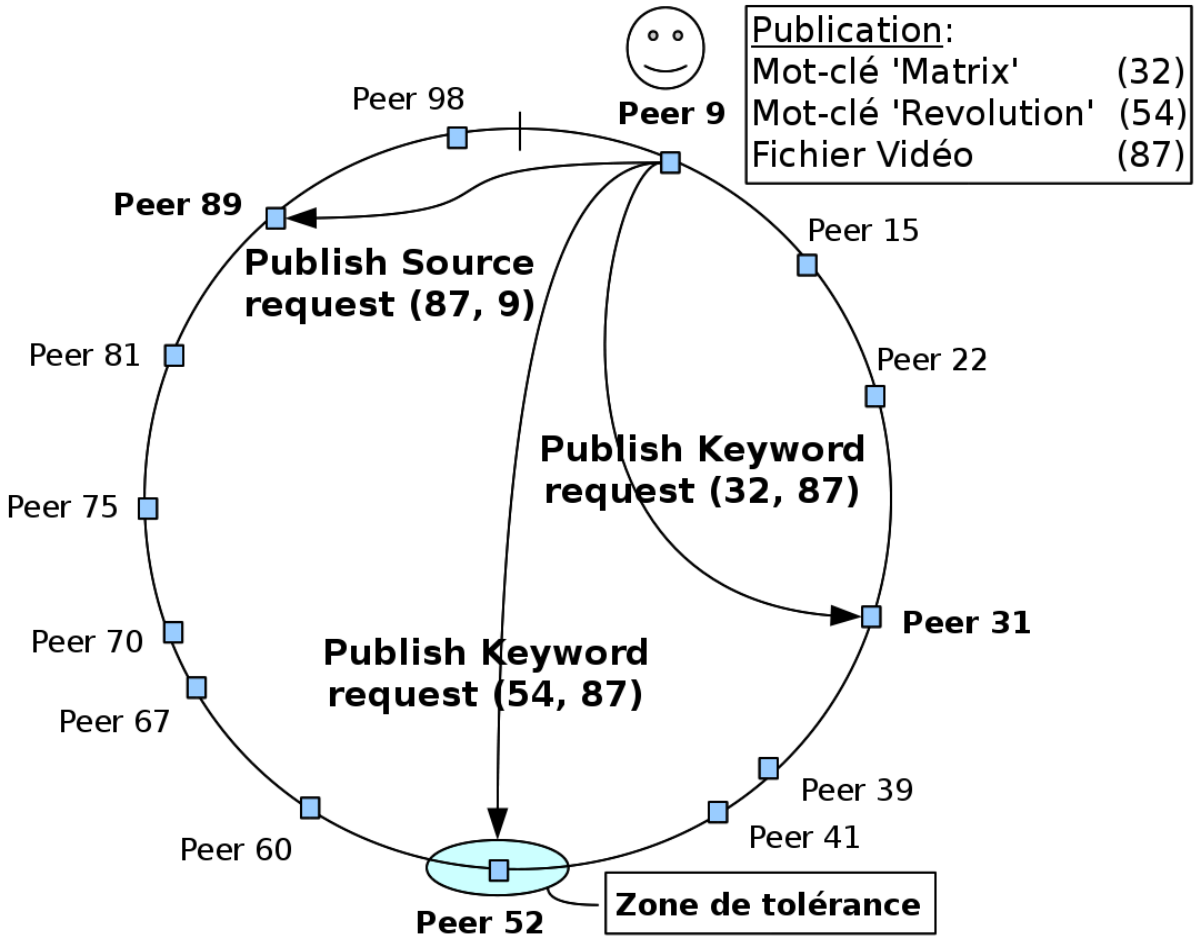
Distance entre identifiants : métrique XOR

- Préfixe : nombre de bits de poids fort consécutifs communs à 2 identifiants
- Exemple : préfixe de longueur 20 bits

`C0F70`911A9C2E6F6960DDED0D4118244
`C0F70`73FC6939D0FF3CE0E36B28E3644

Utilisation de la DHT de KAD

- Double indexation [Bru06] : mots-clés → fichiers ; fichiers → sources
- Indexation par les 10 pairs plus proches



Plan

Objectif

- Concevoir une architecture pouvant superviser l'accès aux contenus (DHT)
- Éviter les faux positifs (pollution)

Idée principale :

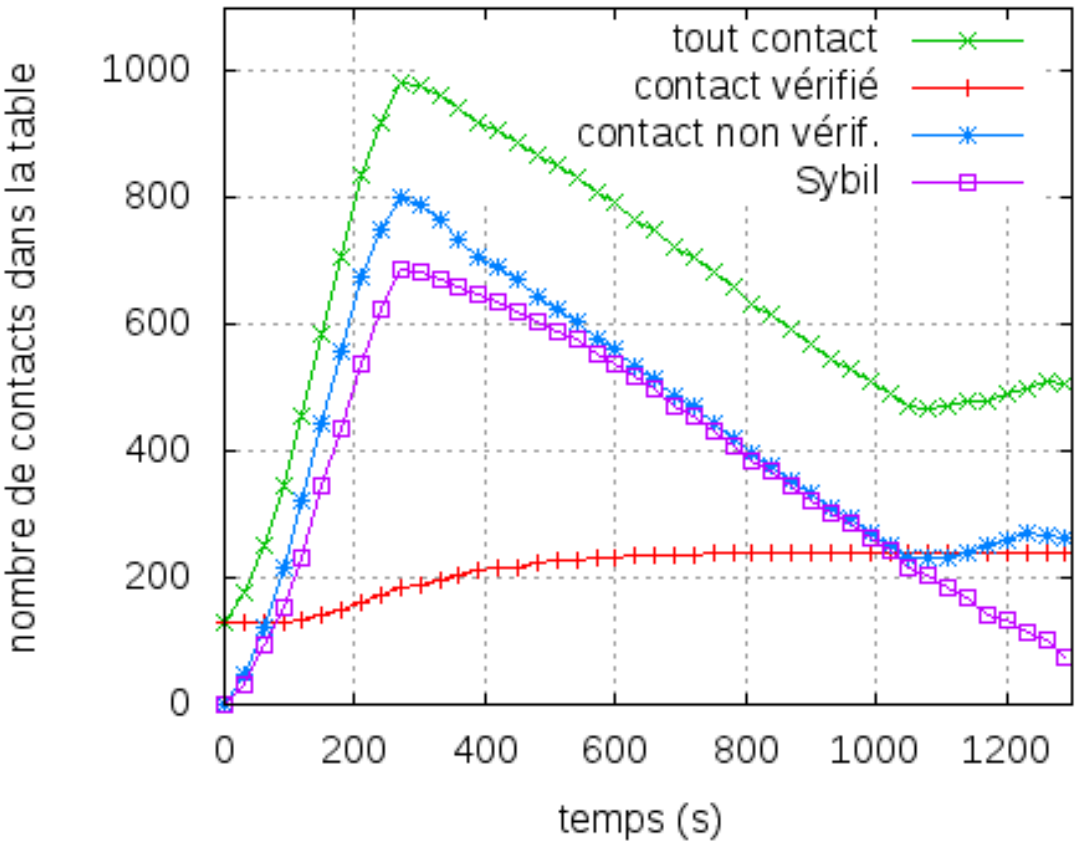
- Combiner les avantages des pots de miel et des sondes distribuées (informations complémentaires)

Étude préliminaire :

- Clients KAD récents : protections contre l'attaque Sybil
  1. Protection contre l'inondation
  2. Limitation par l'adresse IP (1 KADID par IP)
  3. Vérification d'identité (3-way handshake)
- Insertion de sondes toujours possible ?

Évaluation des protections de la table de routage

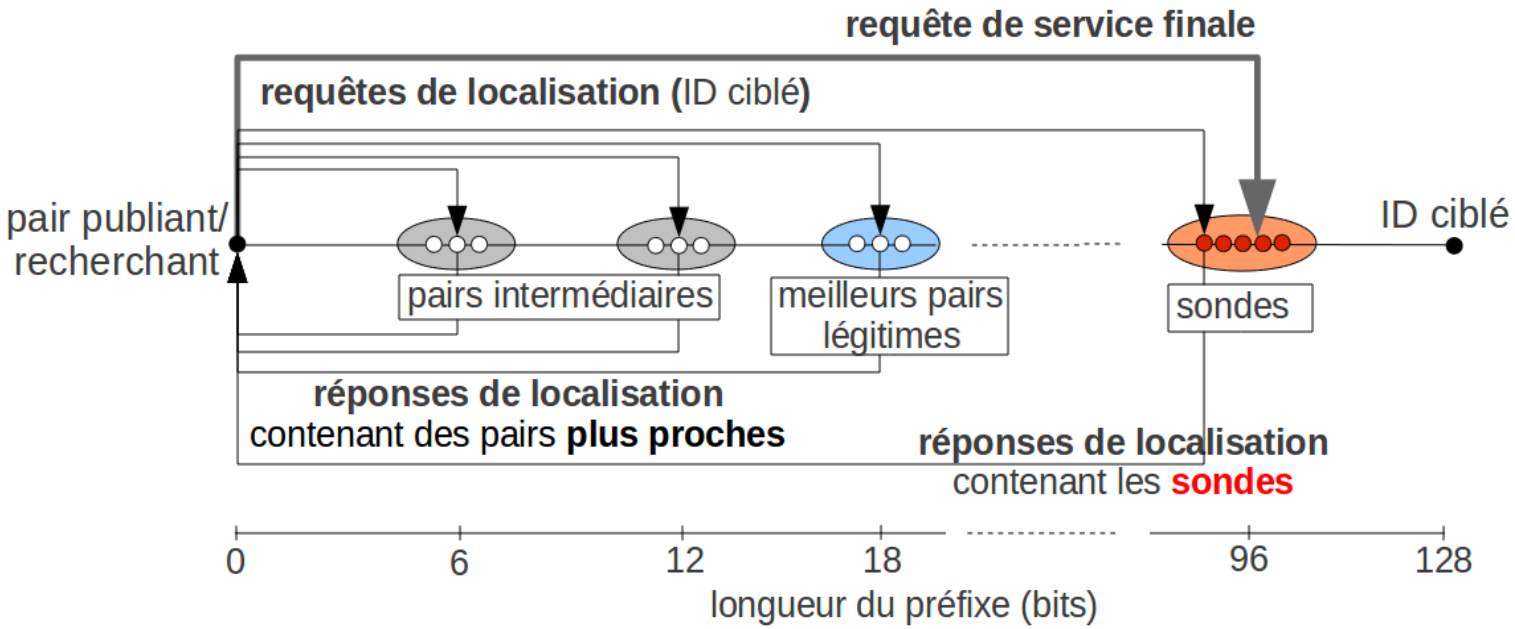
- Client récent instrumenté, injection de Sybils (adresse IP usurpée)
- Sybils marqués comme non vérifiés (inutilisables)
- Protection peu coûteuse, attaques massives limitées



Détournement du mécanisme de localisation

Nouvelle approche :

- Vulnérabilité : libre choix des IDs + localisation déterministe
- Placement de 20 pairs distribués proches de la cible (96 bits)
- Annonces régulières aux autres pairs, coopération entre sondes
- Capture des 10 requêtes de service répliquées après localisation



Stratégie de supervision

Supervision **passive** ou **active**

Intérêt supervision active :

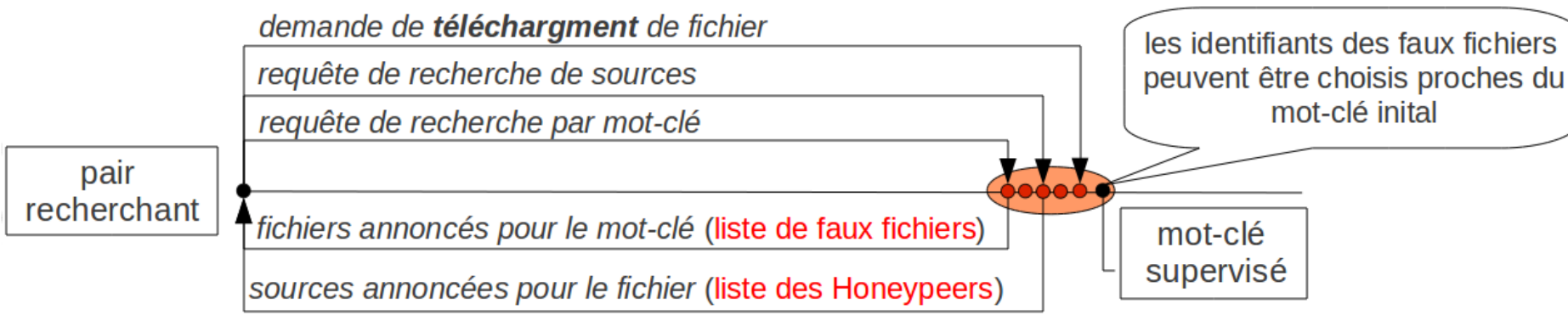
Contrôle possible de l'ensemble de la procédure de téléchargement pour éviter les faux positifs.

Annonce de faux fichiers :

- Appâts attractifs (contrôle du nombre de sources affiché)

Annonce des pots de miel :

- Capture des requêtes de téléchargement finales



Évaluation de l'architecture : modèle

Modèle probabiliste :

- Connaissance de la table de routage et du processus de localisation
- NP_n : nombre de pairs potentiels à l'étape n ; NH : nombre de sondes

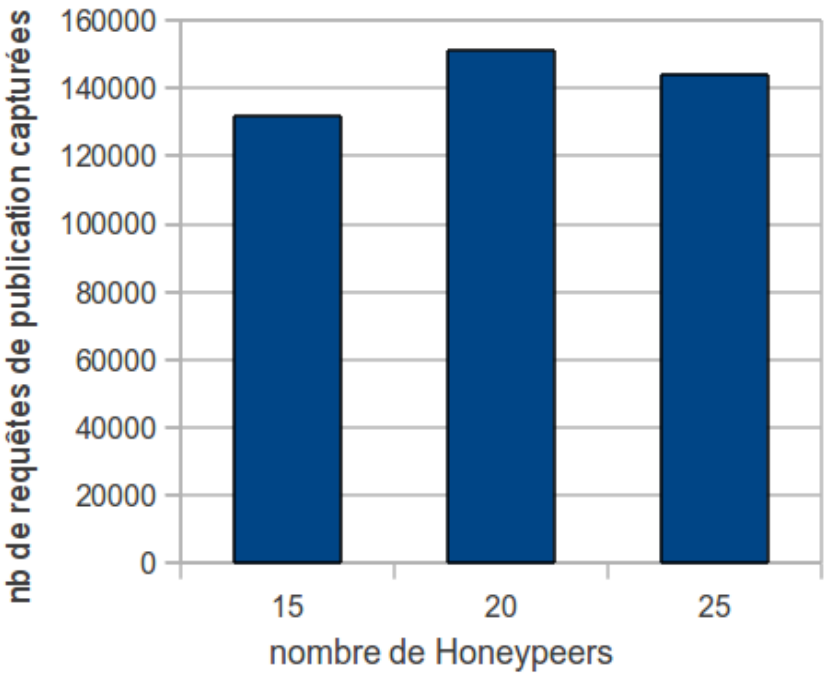
n	$NP_n + NH$	$P_n(H \geq 1)$
0	78145	0.00077
1	2461.5	0.0931
2	96.5	0.928
3	22.4	1

TABLE: Probabilité de trouver au moins un honeypeer à la n^{eme} étape de la localisation ($NP = 5 \times 10^6$, $NH = 20$)

Évaluation de l'architecture : mesures

Mesures réelles :

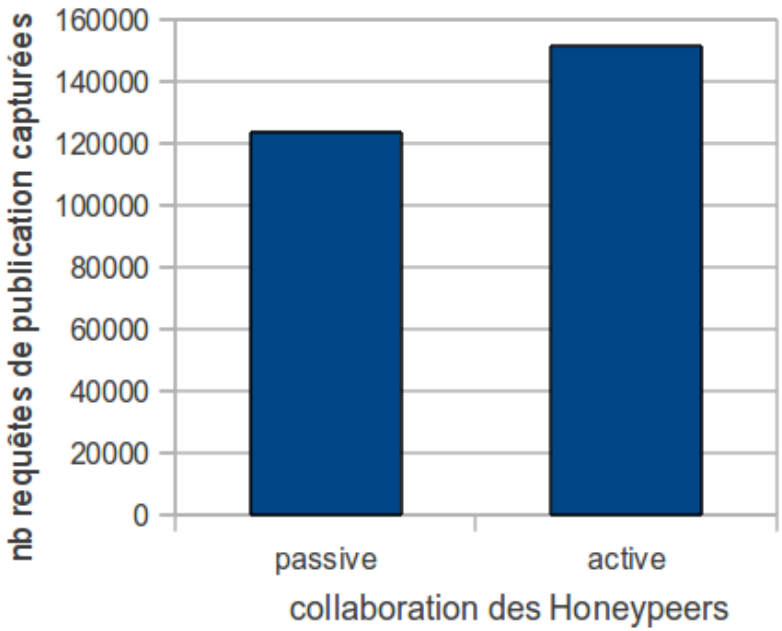
- Paramètres : nombre de sondes (20), collaboration, etc.



Évaluation de l'architecture : mesures

Mesures réelles :

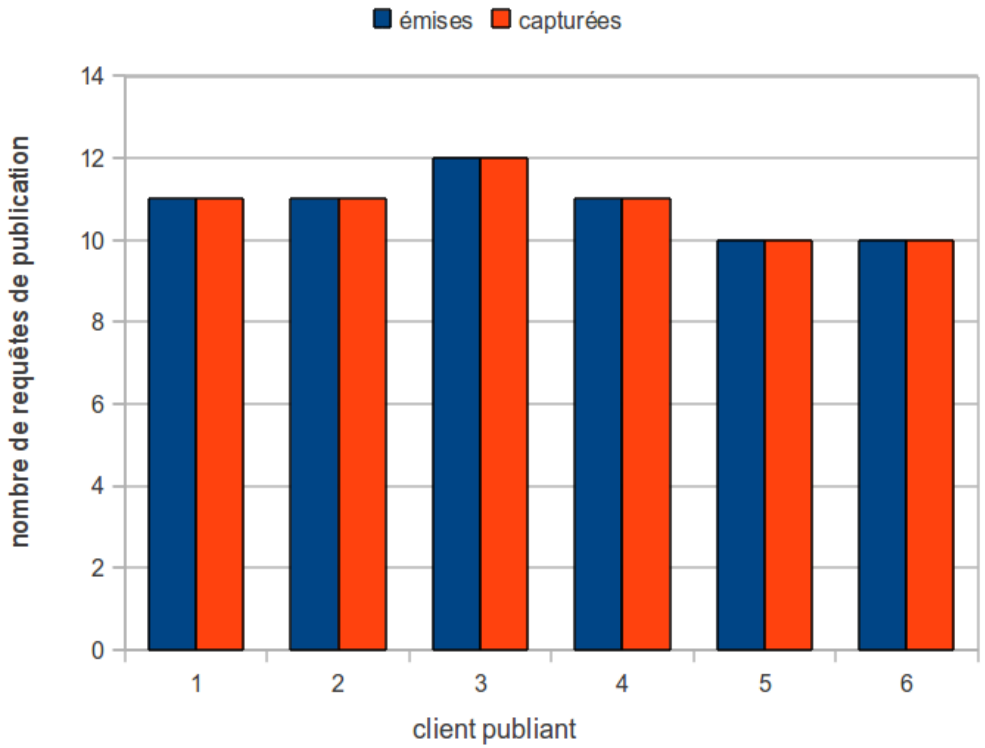
- Paramètres : nombre de sondes (20), collaboration, etc.



Évaluation de l'architecture : mesures

Mesures réelles :

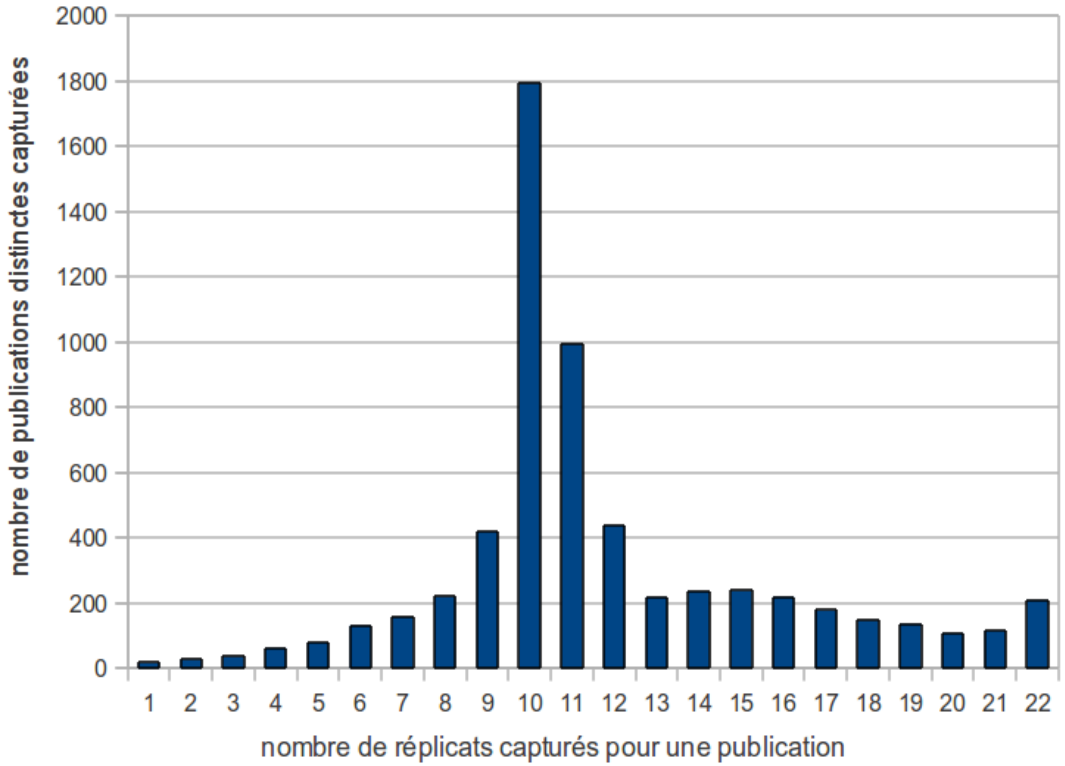
- Paramètres : nombre de sondes (20), collaboration, etc.
- Efficacité : publications par clients instrumentés, nombre de réplicats



Évaluation de l'architecture : mesures

Mesures réelles :

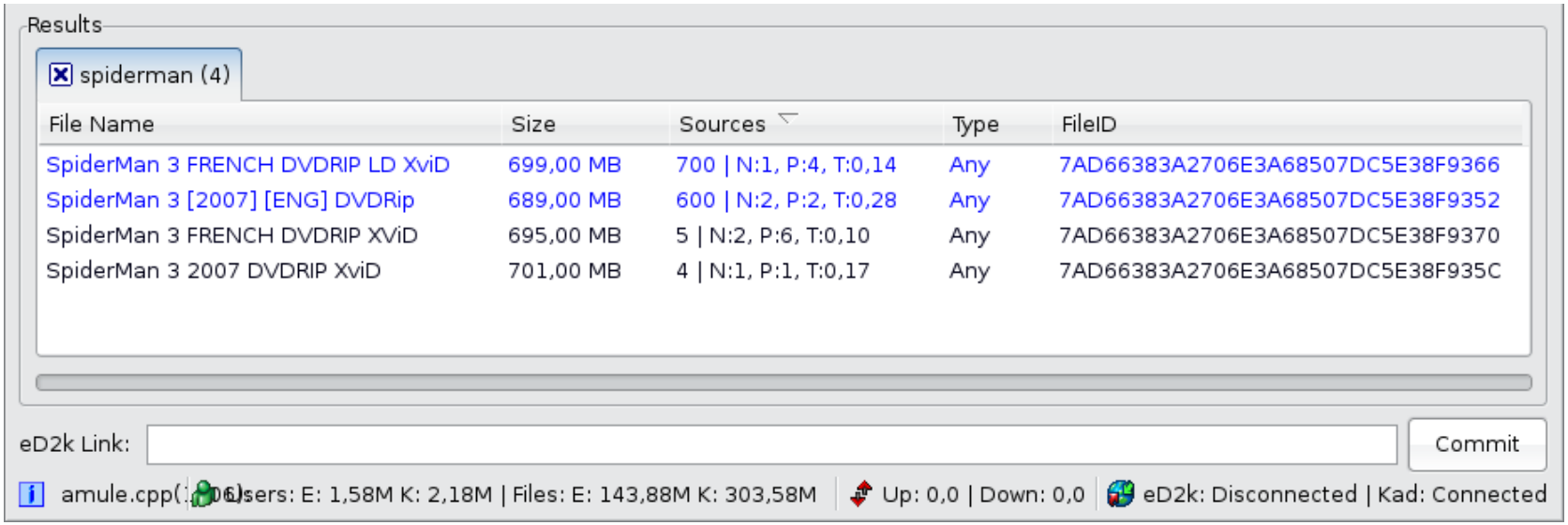
- Paramètres : nombre de sondes (20), collaboration, etc.
- Efficacité : publications par clients instrumentés, nombre de réplicats



Évaluation de l'architecture : mesures

Mesures réelles :

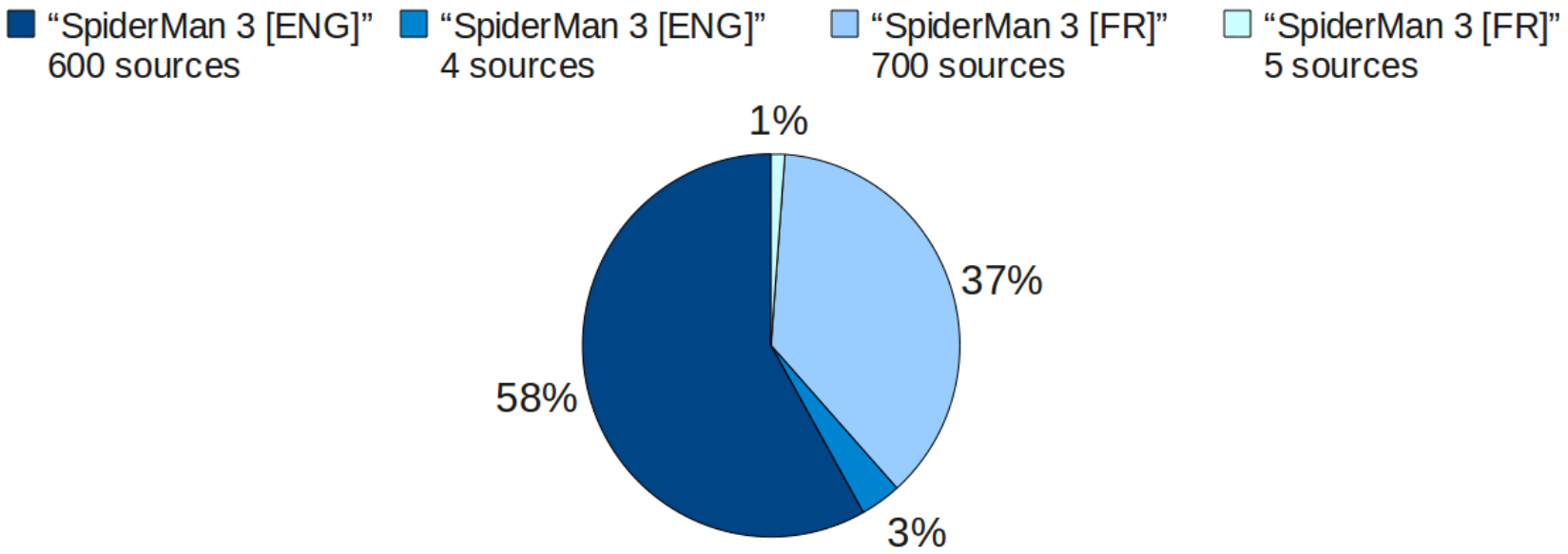
- Paramètres : nombre de sondes (20), collaboration, etc.
- Efficacité : publications par clients instrumentés, nombre de réplicats
- Supervision active de « spiderman » : quatre faux fichiers, 96% des téléchargements pour les deux fichiers populaires ;



Évaluation de l'architecture : mesures

Mesures réelles :

- Paramètres : nombre de sondes (20), collaboration, etc.
- Efficacité : publications par clients instrumentés, nombre de réplicats
- Supervision active de « spiderman » : quatre faux fichiers, 96% des téléchargements pour les deux fichiers populaires ;



Application : supervision des contenus pédophiles

Objectif : comprendre et caractériser les activités pédophiles, application à grande échelle

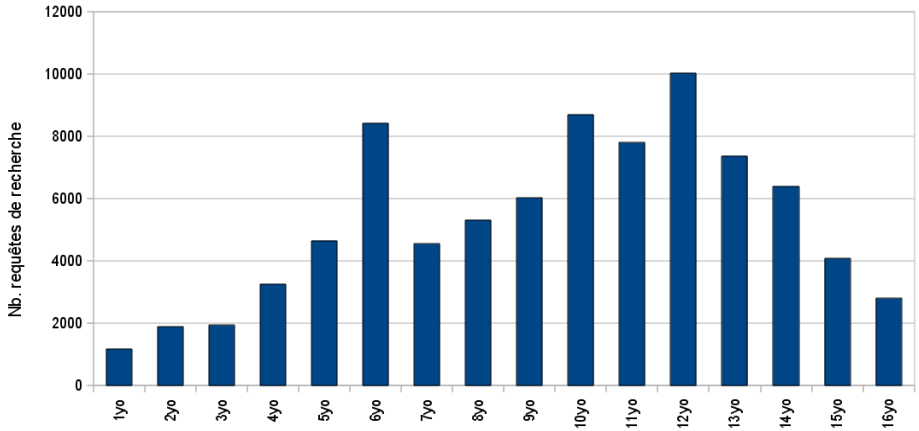
Contexte : projet ANR MAPE (Measurement and Analysis of Peer-to-peer Exchanges for pedocriminality fighting and traffic profiling), collaboration avec le LIP6

Expérience :

- Supervision (passive) de mots-clés
- Deux semaines de capture, 72 mots-clés supervisés
- 28GB collectés au LHS (158M fichiers, 36M publications, 12.8M pairs, etc.)
- Cinq sondes par mot-clé, 360 sondes déployées, 35 machines PlanetLab

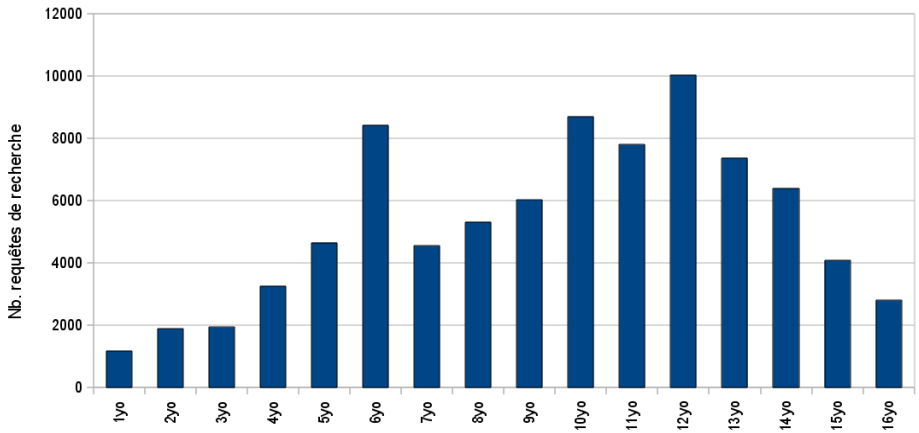
Analyse des données collectées

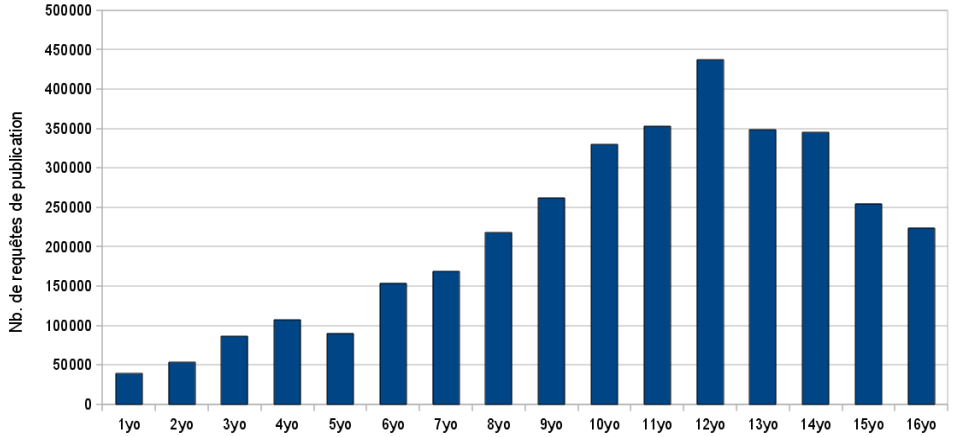
- Analyse quantitative des activités pédophiles : recherches, publications, activité horaire des pairs, corrélation des mots-clés, etc. (collaboration avec Andreea Orosanu)



Analyse des données collectées

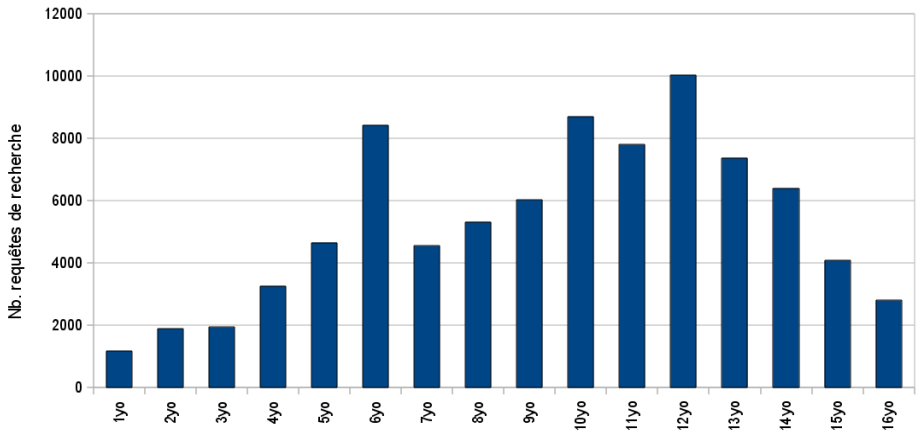
- Analyse quantitative des activités pédophiles : recherches, publications, activité horaire des pairs, corrélation des mots-clés, etc. (collaboration avec Andreea Orosanu)

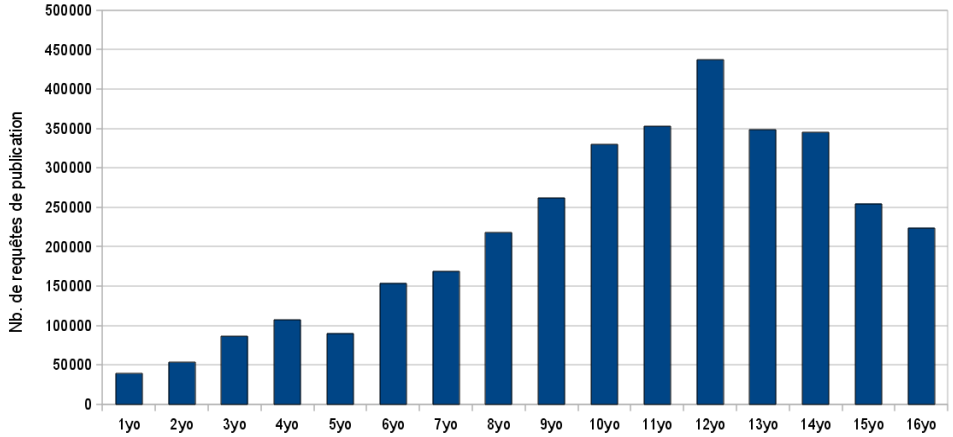


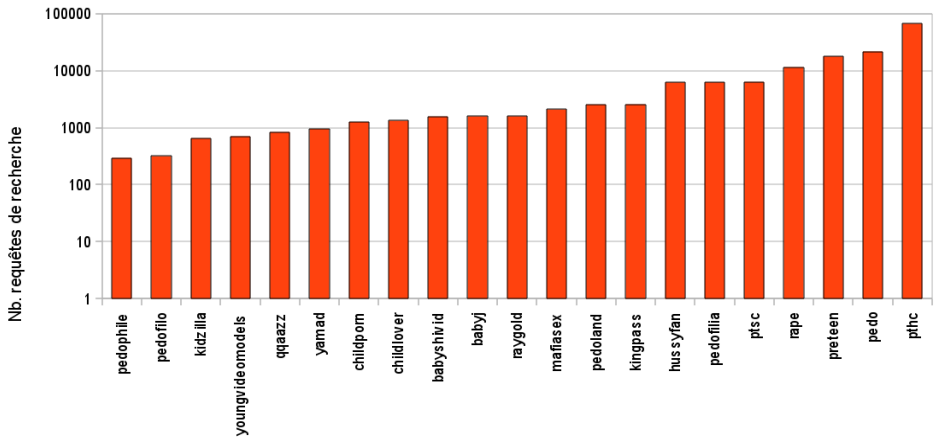


Analyse des données collectées

- Analyse quantitative des activités pédophiles : recherches, publications, activité horaire des pairs, corrélation des mots-clés, etc. (collaboration avec Andreea Orosanu)

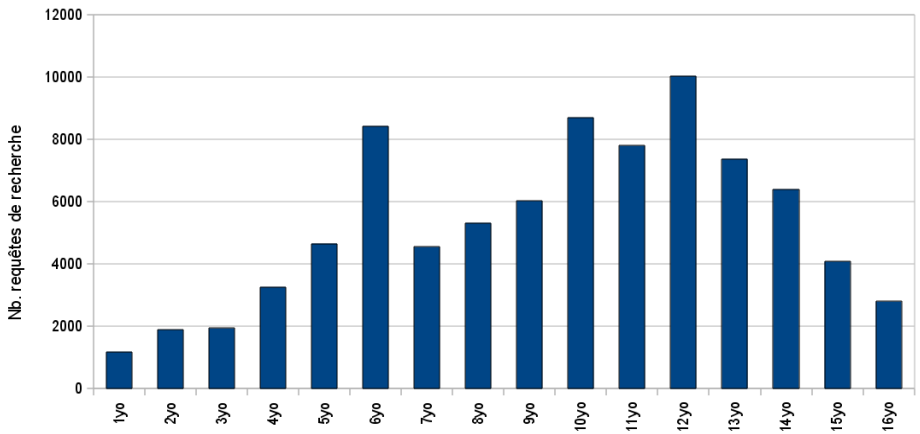


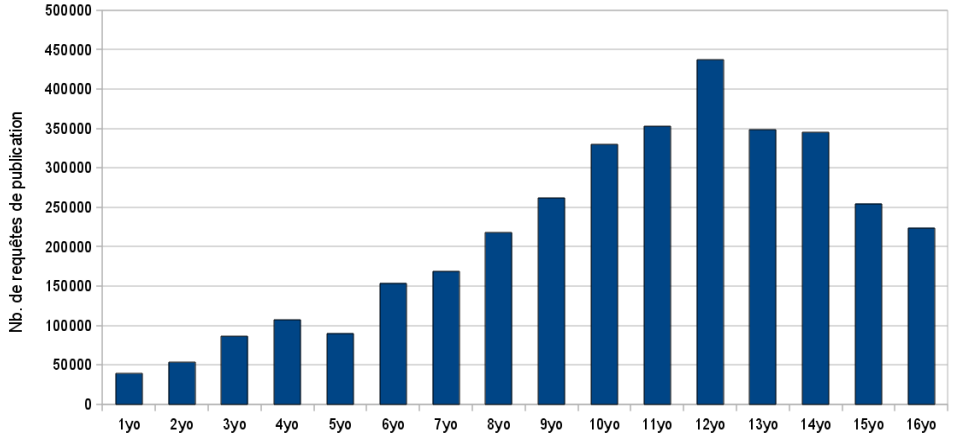


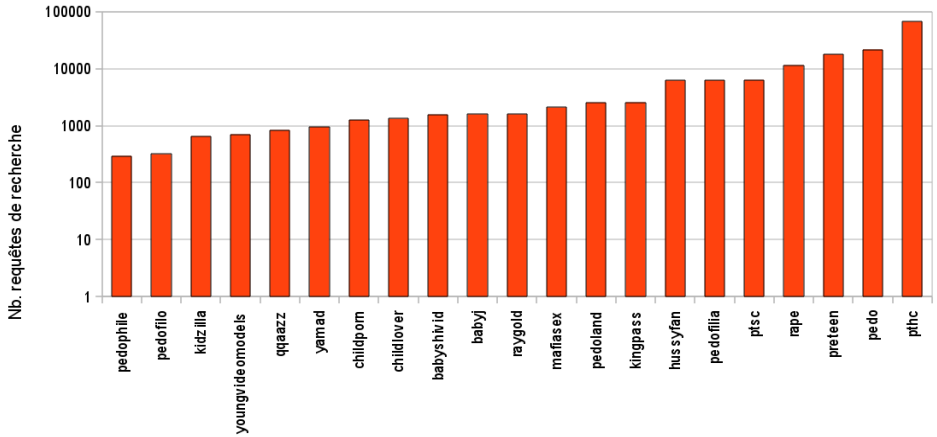


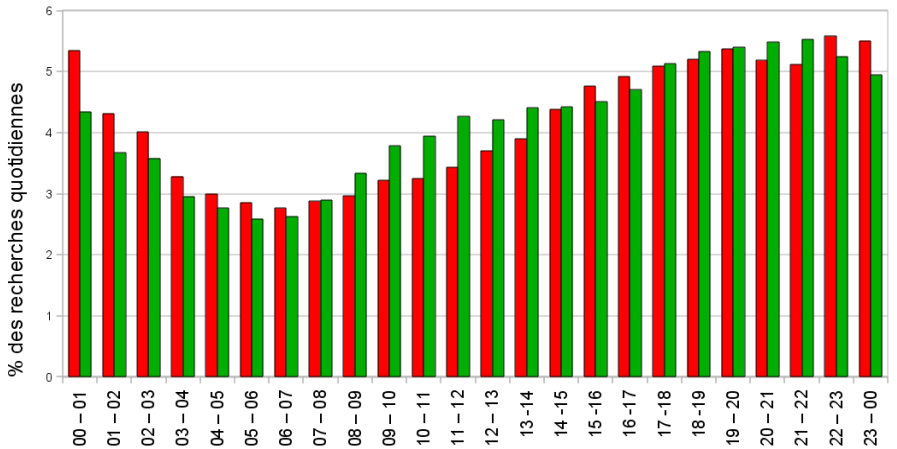
Analyse des données collectées

- Analyse quantitative des activités pédophiles : recherches, publications, activité horaire des pairs, corrélation des mots-clés, etc. (collaboration avec Andreea Orosanu)









Validation et Bilan

Géo-localisation et quantification :

- Application d'algorithmes d'étiquetage aux requêtes, puis aux pairs
- Géo-localisation par adresse IP (GeoLite City)
- Validation : quantification des pairs pédophiles conforme à eDonkey (LIP6)

Résumé :

- Évaluation des protections existantes [AIMS 2009]
- Conception d'une nouvelle stratégie de supervision (sondes + pots de miel) limitant les faux positifs [NOTERE 2009]
- Évaluation de l'architecture (modélisation et mesures) [ICC 2010], Implantation (collaboration avec F.Beck et JP. Timpanaro)
- Application à grande échelle : supervision de contenus pédophiles

Plan

La pollution dans KAD

- Contexte : projet GIS 3SGS ACDAP2P, collaboration avec l'UTT

Falsification d'indexation :

- Indexation d'un fichier par de (très) nombreux noms différents
- Pollution dangereuse (virus, contenus pédophiles, etc.)

Détection :

- Impossible par la DHT (informations inutiles ou indisponibles)
- Obtention des noms de fichiers depuis chaque source

Collecte de données :

- Top 100 des contenus les plus téléchargés en 2010
- Pour chaque contenu : découverte des 20 fichiers les plus populaires (2000 fichiers étudiés)

Métrique de similarité pour détecter la pollution

Coefficient de Similarité de Tversky

- X et Y : 2 ensembles de mots-clés
- X : mots-clés du nom de fichier donné par la DHT
- Y : mots-clés d'un nom de fichier donné par une source

$$S(X,Y) = \frac{|X \cap Y|}{|X \cap Y| + \alpha * |X - Y| + \beta * |Y - X|} \in [0;1] \quad (1)$$

Coefficient de pollution

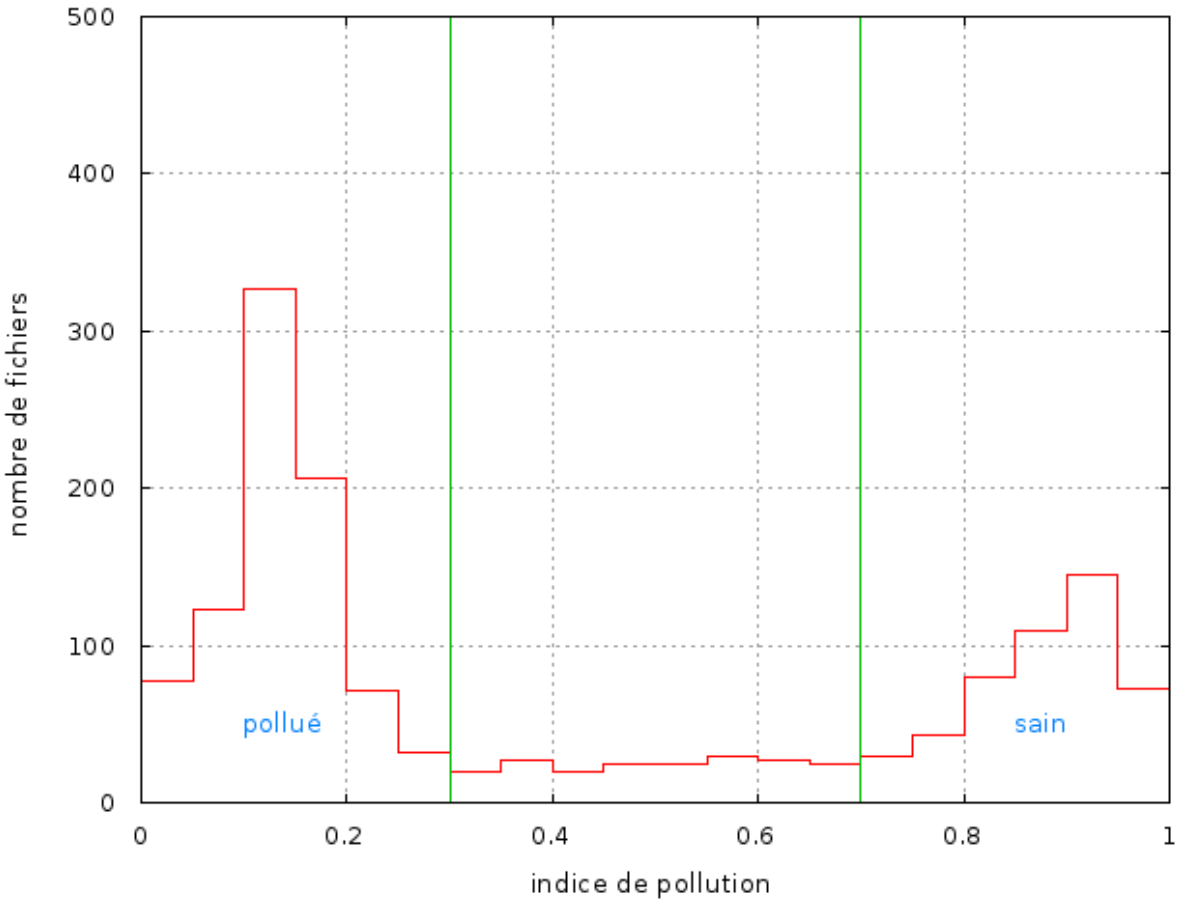
- Coefficient de pollution P calculé pour chaque fichier X
- Étant donné Y_i noms obtenus des sources

$$P(X) = 1 - \frac{\sum_i S(X,Y_i)}{i} \quad (2)$$

Application de la métrique

Expertise et définition des seuils de détection :

- Expertise humaine : étiquetage de 20% des fichiers (pollué, sain)
- Faux positifs : 3.78% ; Faux négatifs : 0.88%



Quantification de la pollution dans KAD

Résultats de la quantification :

- falsification d'indexation : 41.1%, index poisoning : 20.5%, fichiers sains : 28.6%
- 8.8% de noms à caractère pédophile, 55.7% pornographique

Résumé : [P2P 2011] (short 19%)

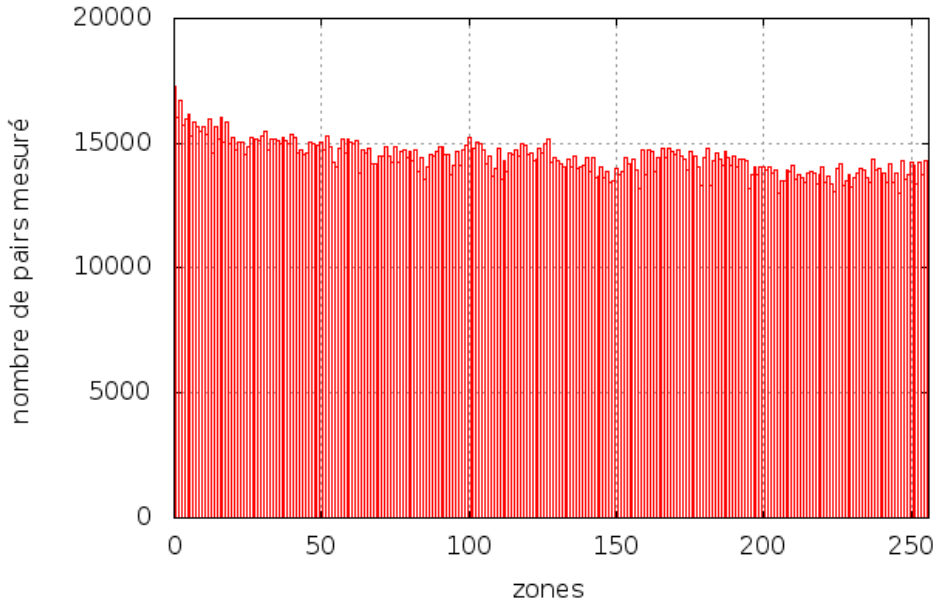
- Définition d'une métrique de détection, validation par expertise humaine
- Collecte de fichiers populaires, quantification de la pollution (collaboration avec Guillaume Montassier, UTT)

Amélioration possible :

- Détection de la pollution par la DHT
- Publication d'une source : ajout du nom de fichier
- Limite : attaques Sybil localisées

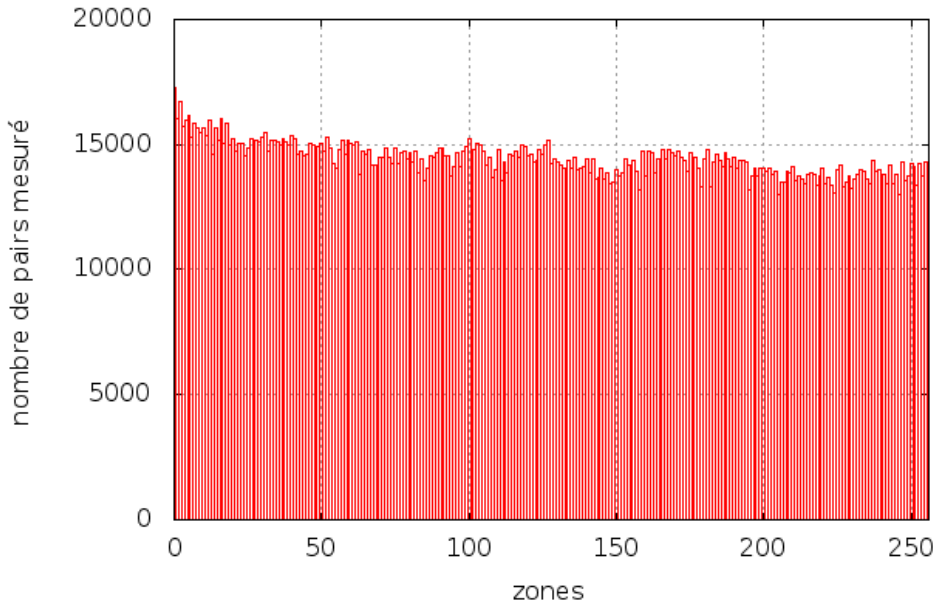
Exploration du réseau

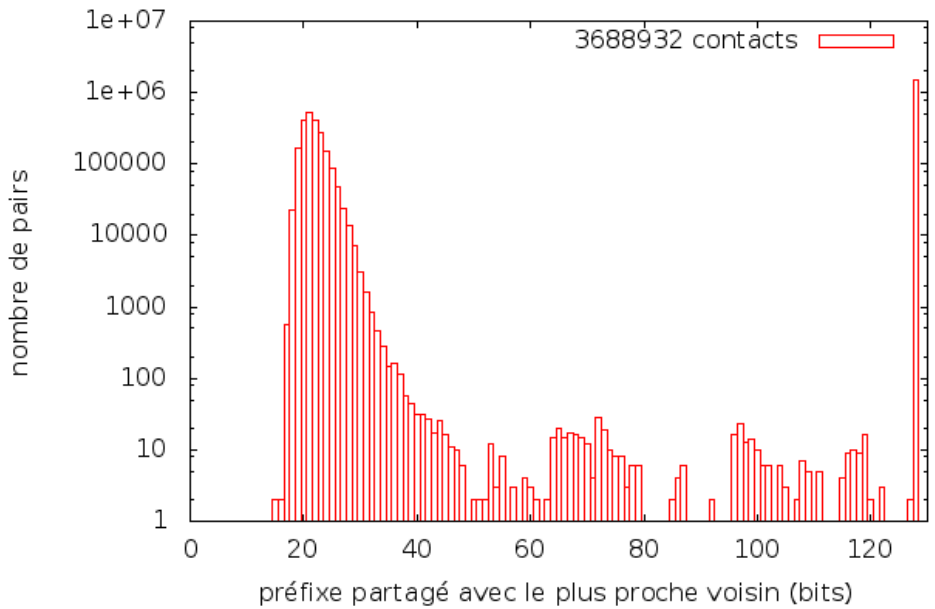
- Explorateur précis et peu intrusif (stage de Christopher Hénard)
- Cartographie des pairs de KAD (KAD ID, IP, ports, etc.)
- Analyse des données, détection des placements anormaux
- Ex : exploration (8 Juillet 2010) : 3 688 932 pairs



Exploration du réseau

- Explorateur précis et peu intrusif (stage de Christopher Hénard)
- Cartographie des pairs de KAD (KAD ID, IP, ports, etc.)
- Analyse des données, détection des placements anormaux
- Ex : exploration (8 Juillet 2010) : 3 688 932 pairs

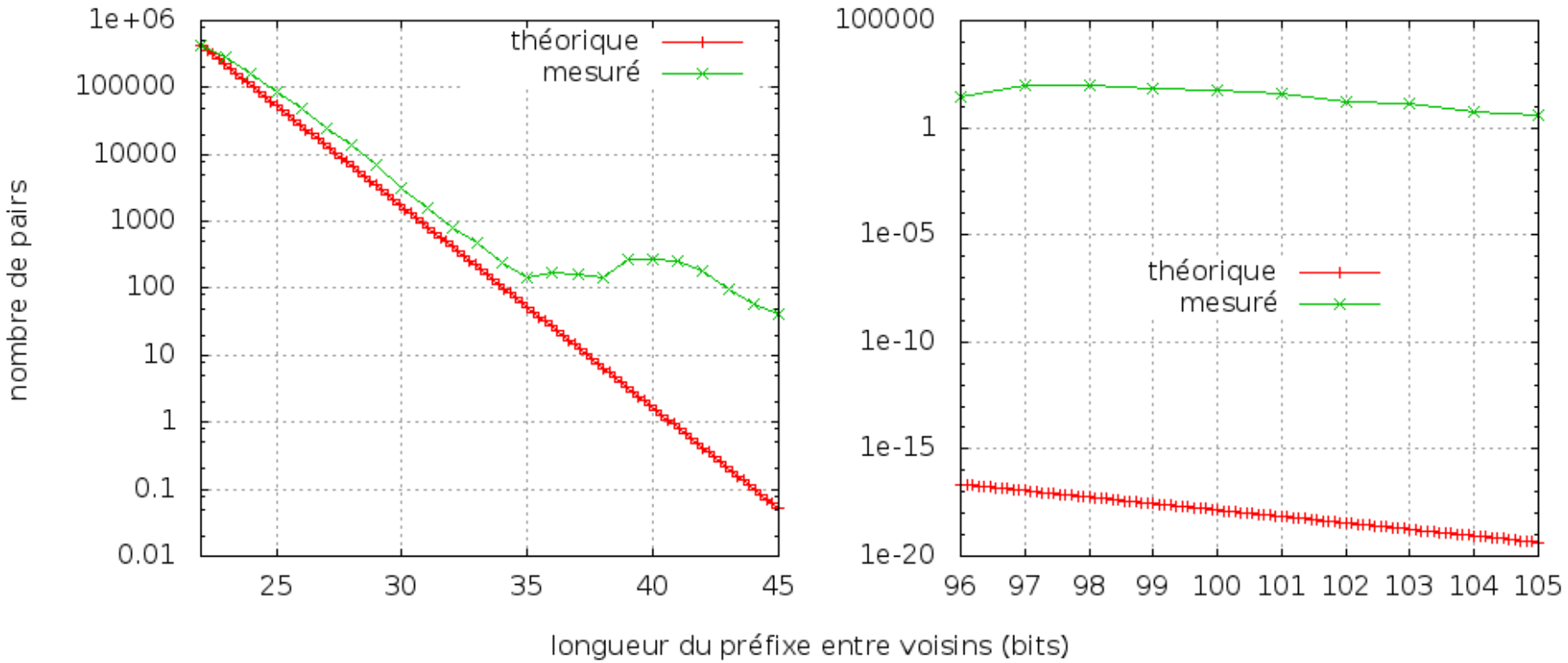




Distribution théorique des distances

Nombre moyen de pairs potentiels partageant x bits avec un identifiant cible étant donné N pairs dans le réseau :

$$F(x) = \frac{N}{2^x} \text{ avec } N = 4 \cdot 10^6 \text{ et } x \in [1; 128] \qquad (3)$$



Analyse des distances inter-pairs

- Juillet 2010 : 426 groupes suspects (préfixe commun >35bits)
- Avril 2011 : 2074 groupes suspects, motifs évidents

```
Prefix "E0E4286C280000000000000000000000", length 37, shared by 4 contacts:
<E0E4286C288DB7E593E7C38343A3B7FE, 58.17.X.X, 11571, 11571, 8, R>
<E0E4286C2A96962BA6F79FE873B91BAF, 123.144.X.X, 11571, 11571, 8, R>
<E0E4286C2B680D27AD5B192EC0A026C1, 123.145.X.X, 11571, 11571, 8, R>
<E0E4286C2C5DB68A18390EE08906C5F7, 123.144.X.X, 11166, 11166, 8, R>

Prefix "E0F6D84E000000000000000000000000", length 37, shared by 6 contacts:
<E0F6D84E000E6F7036C539D8609DC76D, 123.144.X.X, 10476, 10476, 0, T>
<E0F6D84E0115F18099354BF758F5A722, 123.145.X.X, 10476, 10476, 0, R>
<E0F6D84E011AD304907B4FC1244FE124, 123.145.X.X, 10867, 10867, 8, T>
<E0F6D84E034B693D9413A6750EEBBD4F, 123.144.X.X, 0, 58977, 8, T>
<E0F6D84E0597D765274E4610D4C524FE, 123.144.X.X, 10121, 10121, 0, T>
<E0F6D84E075BBF7498EDCB9B853CCA67, 123.144.X.X, 10476, 10476, 8, T>
```

- Limite : au moins deux pairs insérés pour permettre la détection
- Méconnaissance des contenus ciblés

Analyse des distances pairs-contenus

- Contenus connus a priori : collecte de 888 mots-clés populaires
- Recensement des pairs suspects (préfixe commun >35bits)
- Nombreuses attaques (216/888), attaques isolées
- 44 Adresses IP suspectes, 2120 pairs

```
twilight 4D62D26BB2A686195DA7078D3720F60A
<4D62D26BB2A686195DA7078D3720F632, X.Y.#.#, 7290, 7294, 8, R> [prefix = 122]
soundtrack AC213377BB53F608390BD94A6AE6DD35
<AC213377BB53F608390BD94A82582F42, #.#.#.#, 5003, 5002, 8, R> [prefix = 96]
harry 770CF5279AB34348C8FECF9672747B94
<770CF5279AB34348C8FECF96524D8CDE, #.#.#.#, 5003, 5002, 8, P> [prefix = 98]
```

Résumé : [SAR-SSI 2011]

- Exploration complète du réseau (DHT), détection par analyse des distances
- Recensement de nombreuses attaques Sybil localisées

Plan

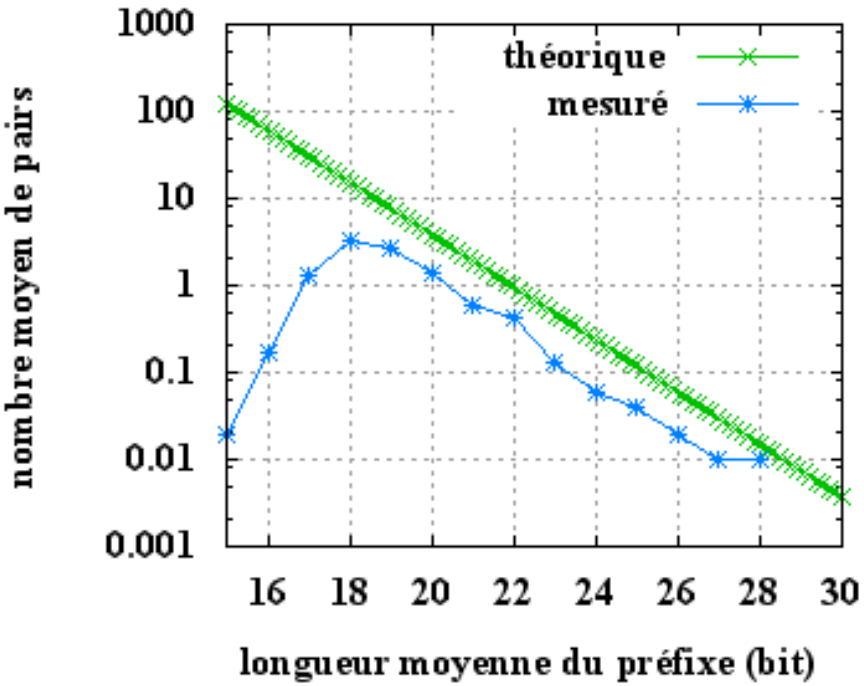
Distribution régulière des identifiants

Approche :

- Laisser les pairs choisir aléatoirement l'ID
- Vérifier que la distribution des pairs est aléatoire

Mesures sur la DHT :

- 1800 procédures de localisation de KADID aléatoires, enregistrement des 10 meilleurs pairs



Métrique pour détecter les attaques localisées

Comparaison de distributions

- Difficulté : échantillon très petit (10 pairs)
- Majorité des outils statistiques peu efficaces
- KL-divergence efficace mais nécessite interprétation

Kullback-Leibler divergence (G-test) utilisée pour détecter les attaques :

$$D_{KL}(M \mid T) = \sum_i M(i) \log \frac{M(i)}{T(i)} \tag{4}$$

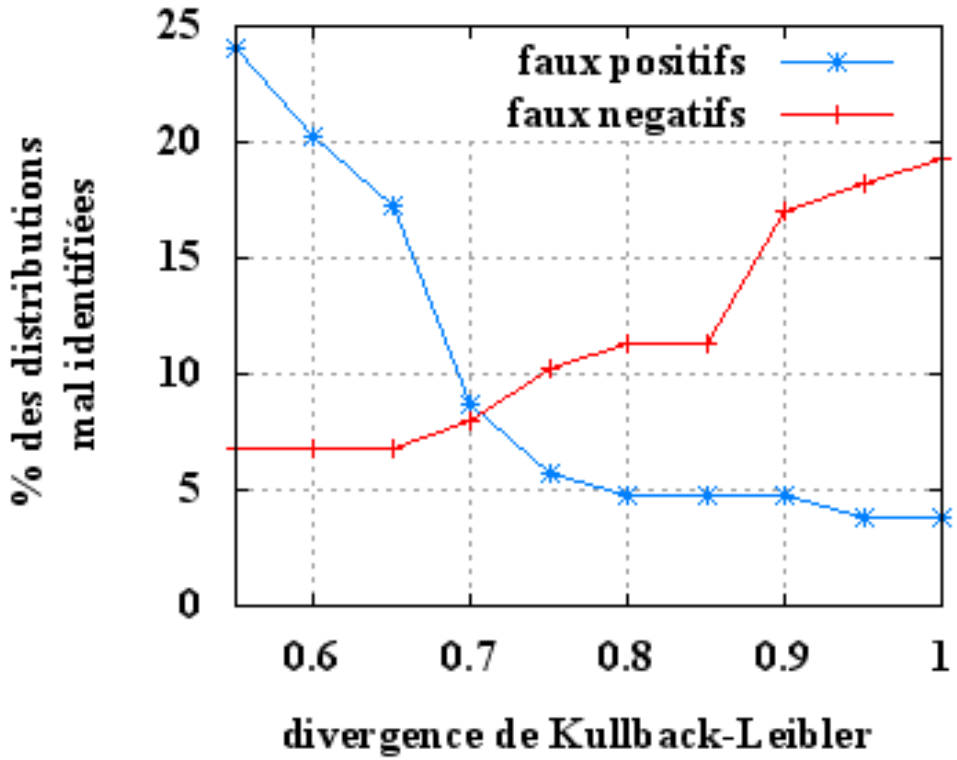
Préfixe	18	19	20	21	22	23	24	25	26	27	28
M (attaque)	0	0	0	0	0	0	0	0	0.5	0.5	0
M (saine)	0.6	0.2	0.1	0.1	0	0	0	0	0	0	0
T	$1/2$	$1/2^2$	$1/2^3$	$1/2^4$	$1/2^5$	$1/2^6$	$1/2^7$	$1/2^8$	$1/2^9$	$1/2^{10}$	$1/2^{11}$

TABLE: Exemples de distribution de préfixes comparée par la divergence K-L pour détecter les attaques

Application de la métrique

Évaluation de la métrique de détection et seuillage :

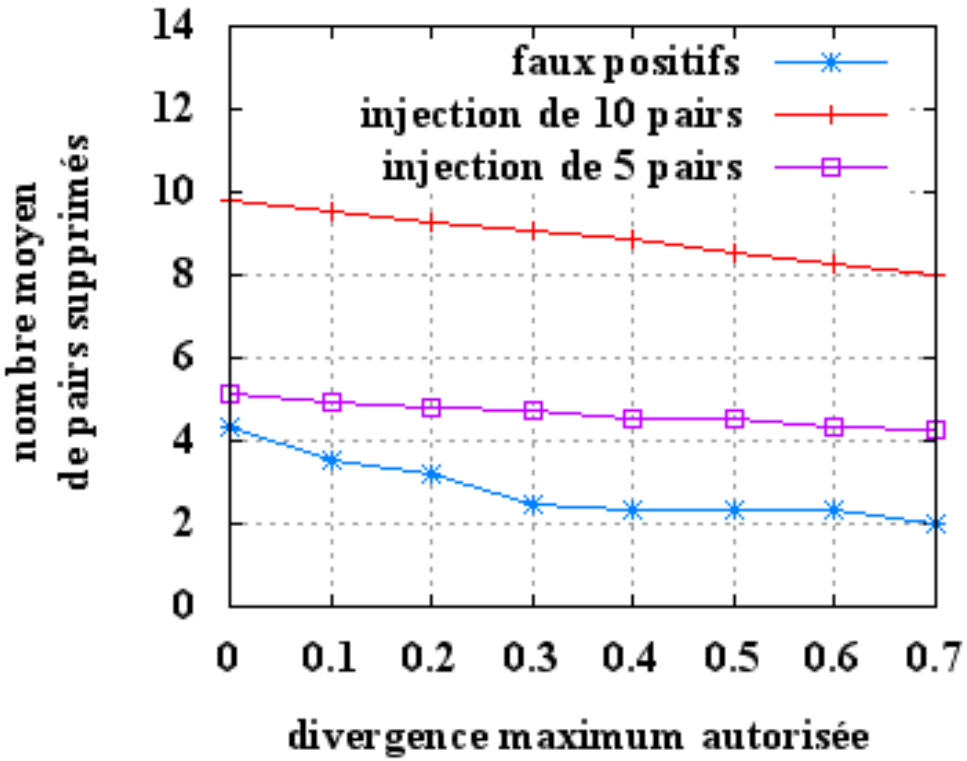
- Deux ensembles de données : distributions d'attaques simulées et distributions régulières mesurées
- Faux négatifs : peu de pairs insérés (<5) avec des préfixes faibles (18-19 bits)



Contre-mesure : évitement des attaques

En cas de détection :

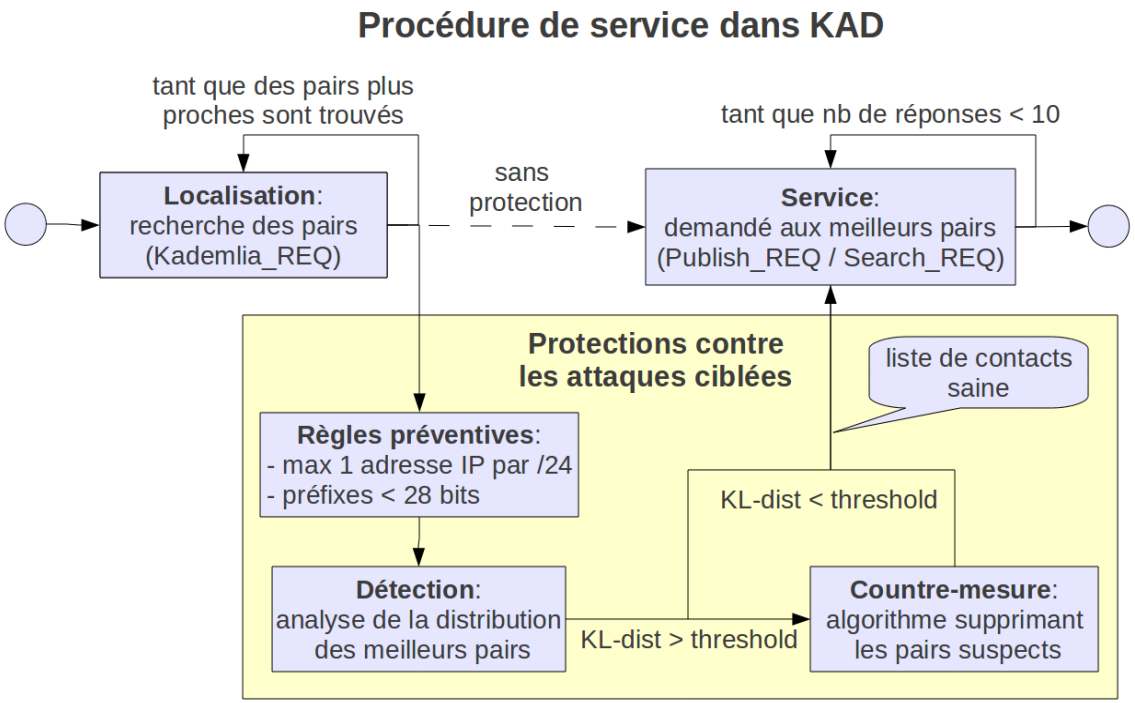
- Filtrage progressif des pairs suspects
- Tant que divergence $>$ seuil, suppression des pairs du préfixe le plus suspect, mise à jour de la distribution



Bilan de la protection contre les attaques localisées

Résumé : [HotP2P 2010]

- Détection locale des attaques par comparaison de distributions
- Évaluation par simulation d'attaques
- Implantations et optimisations de la protection



Plan

Bilan

Supervision des accès aux contenus dans une DHT

- Évaluation des protections de KAD
- Conception, évaluation et implantation d'une architecture efficace et précise
- Validation à grande échelle : supervision de contenus pédophiles

Métrologie des attaques

- Détection d'une nouvelle pollution, métrique validée par expertise humaine
- Quantification à grande échelle (2000 fichiers, 41% pollués)
- Détection des attaques par analyse des distances inter-pairs ou pairs-contenus
- Cartographie du réseau et recensement des attaques

Bilan

Protection des accès à la DHT

- Détection locale des attaques par analyse des distributions (Kullback-Leibler)
- Évaluation métrique et contre-mesure par simulation
- Implantation et expérimentation dans KAD et gtk-gnutella (Raphaël Manfredi)
- Solution efficace : aucun surcoût, rétro-compatible, implantation simple

Perspectives

Collecte et traitement des données :

- Analyse manuelle de grandes bases de données limitée : utilisation de techniques de fouille de données
- Approche non-supervisée (apprentissage de nouveaux mots) puis supervisée (caractérisation des pédophiles)

Caractérisation des pairs malveillants

- Organisation : étude des motifs des attaques (IP, ports, distances),
- Comportement (surveillance, pollution, DDoS, etc.), par communications directes

Protection et amélioration des services distribués :

- Détection directe de la pollution par la DHT
- Conception d'un mécanisme de réputation global

Publications

Travaux présentés :

- IEEE P2P 2011 (short-19%) : Content Pollution Quantification in Large P2P networks : a Measurement Study on KAD
- SAR-SSI 2011 : Détection de pairs suspects dans le réseau pair à pair KAD
- HotP2P 2010 : Efficient DHT attack mitigation through peers' ID distribution
- ICC 2010 : Monitoring and Controlling Content Access in KAD
- NOTERE 2009 : Une architecture de honeypots distribués pour superviser le réseau P2P KAD
- AIMS 2009 : Evaluation of Sybil Attacks Protection Schemes in KAD

Travaux connexes :

- HotP2P 2011 : When KAD meets BitTorrent - Building a Stronger P2P Network
- NTMS 2011 : BitTorrent's Mainline DHT Security Assessment
- ICN 2008 et JDIR 2008 (best paper) : A Distributed and Adaptive Revocation Mechanism for P2P networks

Frederic Aidouni, Matthieu Latapy, and Clémence Magnien.
Ten weeks in the life of an edonkey server.
In *IPDPS* [IEE09], pages 1–5.

Oussama Allali, Matthieu Latapy, and Clémence Magnien.
Measurement of edonkey activity with distributed honeypots.
In *IPDPS* [IEE09], pages 1–8.

Rene Brunner.
A performance evaluation of the Kad-protocol.
Master's thesis, University of Mannheim and Institut Eurecom, 2006.

Cristiano Costa and Jussara Almeida.
Reputation systems for fighting pollution in peer-to-peer file sharing systems.

In *P2P '07 : Proceedings of the Seventh IEEE International Conference on Peer-to-Peer Computing*, pages 53–60, Washington, DC, USA, 2007. IEEE Computer Society.

Miguel Castro, Peter Druschel, Ayalvadi Ganesh, Antony Rowstron, and Dan S. Wallach.
Secure routing for structured peer-to-peer overlay networks.
SIGOPS Oper. Syst. Rev., 36(SI) :299–314, 2002.

Yatin Chawathe, Sylvia Ratnasamy, Lee Breslau, Nick Lanham, and Scott Shenker.
Making gnutella-like p2p systems scalable.
In *SIGCOMM '03 : Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 407–418, New York, NY, USA, 2003. ACM.

INFOCOM 2006. 25th IEEE International Conference on Computer Communications, Joint Conference of the

IEEE Computer and Communications Societies, 23-29 April 2006, Barcelona, Catalunya, Spain. IEEE, 2006.

John R. Douceur.
The sybil attack.
In *IPTPS '01 : Revised Papers from the First International Workshop on Peer-to-Peer Systems*, pages 251–260, London, UK, 2002. Springer-Verlag.

Romano Fantacci, Leonardo Maccari, Matteo Rosi, Luigi Chisci, Luca Maria Aiello, and Marco Milanesio.
Avoiding eclipse attacks on kad/kademlia : an identity based approach.
In *Proceedings of the 2009 IEEE international conference on Communications*, ICC'09, pages 983–987, Piscataway, NJ, USA, 2009. IEEE Press.

IEEE Computer Society.

23rd IEEE International Symposium on Parallel and Distributed Processing, IPDPS 2009, Rome, Italy, May 23-29, 2009. IEEE, 2009.

Fabrice Le Fessant, Sidath B. Handurukande, Anne-Marie Kermarrec, and Laurent Massoulié.
Clustering in peer-to-peer file sharing workloads.
In *Peer-to-Peer Systems III, Third International Workshop (IPTPS'04), La Jolla, CA, USA, February 26-27, 2004, Revised Selected Papers*, Feb 2004.

Jian Liang, Rakesh Kumar, Yonjian Xi, and Keith W Ross.
Pollution in p2p file sharing systems.
In *IN IEEE INFOCOM*, pages 1174–1185, 2005.

Jian Liang, Naoum Naoumov, and Keith W. Ross.
The index poisoning attack in p2p file sharing systems.
In *INFOCOM* [DBL06].

Petar Maymounkov and David Mazières.
Kademlia : A peer-to-peer information system based on the xor metric.
In *IPTPS '01 : Revised Papers from the First International Workshop on Peer-to-Peer Systems*, pages 53–65, London, UK, 2002. Springer-Verlag.

Ghulam Memon, Reza Rejaie, Yang Guo, and Daniel Stutzbach.
Large-scale monitoring of DHT traffic.
In *International Workshop on Peer-to-Peer Systems (IPTPS)*, Boston, MA, April 2009.

Naoum Naoumov and Keith Ross.
Exploiting p2p systems for ddos attacks.
In *InfoScale '06 : Proceedings of the 1st international conference on Scalable information systems*, page 47, New York, NY, USA, 2006. ACM.

Matei Ripeanu, Adriana Iamnitchi, and Ian Foster.

Mapping the gnutella network.
IEEE Internet Computing, 6 :50–57, 2002.

Moritz Steiner, Taoufik En-Najjary, and Ernst W. Biersack.
Exploiting kad : possible uses and misuses.
SIGCOMM Comput. Commun. Rev., 37(5) :65–70, 2007.

Moritz Steiner, Taoufik En-Najjary, and Ernst W Biersack.
A global view of kad.
In *IMC 2007, ACM SIGCOMM Internet Measurement Conference, October 23-26, 2007, San Diego, USA*, 10 2007.

Ion Stoica, Robert Morris, David Karger, M. Frans Kaashoek, and Hari Balakrishnan.
Chord : A scalable peer-to-peer lookup service for internet applications.
In *SIGCOMM '01 : Proceedings of the 2001 conference on Applications, technologies, architectures, and protocols for*

computer communications, pages 149–160, New York, NY, USA, 2001. ACM.

Daniel Stutzbach and Reza Rejaie.
Improving lookup performance over a widely-deployed dht.
In *INFOCOM* [DBL06].

S. Sen and Jia Wang.
Analyzing peer-to-peer traffic across large networks.
Networking, IEEE/ACM Transactions on, 12(2) :219–232, April 2004.

Haifeng Yu, Michael Kaminsky, Phillip B. Gibbons, and Abraham Flaxman.
Sybilguard : defending against sybil attacks via social networks.
In *SIGCOMM '06 : Proceedings of the 2006 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 267–278, New York, NY, USA, 2006. ACM.