Fighting against paedophile activities in the KAD P2P network

Thibault CHOLEZ Isabelle CHRISMENT Olivier FESTOR

INRIA

Nancy-Université
Université Henri Poincaré

P2P networks challenges

Advantages

- Decentralized systems: no infrastructure cost, good scalability and robustness
- Allows millions of users to share files

Limits

- No central control & autonomous users
- P2P networks are a support to spread paedophile files
- Normal users can have access to malicious contents unintentionally

Objectives

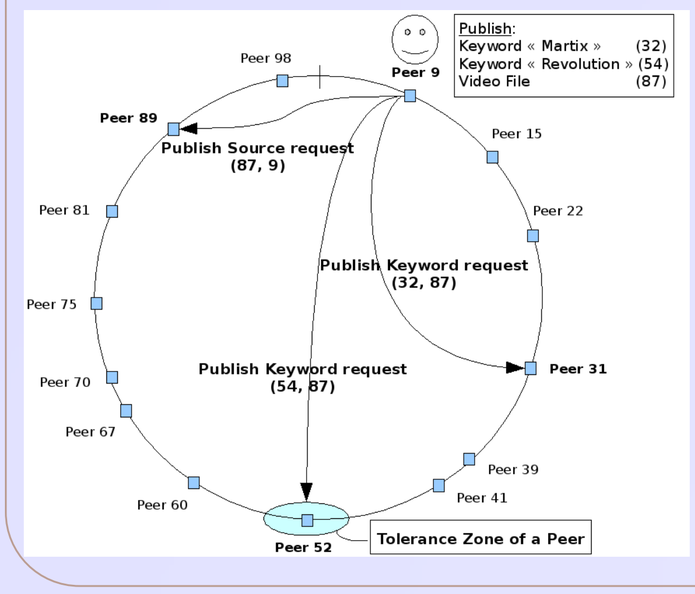
- **Monitor paedophile activities**
- **Monitor and act on paedophile contents**

The KAD network

KAD is a part of **eMule** and one of the **major** P2P networks ($\sim$ 3 millions of simultaneous users).

KAD is used to index and **retrieve the files shared** by the users. Unlike eDonkey or Bittorrent, it is **fully distributed**: no central component knows "who is sharing what".

KAD uses a specific architecture called **Distributed Hash Table** and a **double indexation mechanism**. Each participant is responsible of a part of the overall indexation of contents.



- Peers, Files and Keywords share the same **address space** (2^{128}). The **tolerance zone** defines which peers index what contents, regarding their **addresses**.

- Each file shared by a peer is **published** in two steps:
  - Each **Keyword** composing the filename is linked to the **File** *(Publish Keyword request)*
  - Each **File** is linked to the **Peer** sharing it *(Publish Source request)*

- Searching for a file involves similar **Search requests**.

Technical difficulties

Observing users and **controlling contents** in a P2P network are very difficult tasks:

- To keep the information available, each file and keyword is published on **dozens** of peers.

- Monitoring only files can lead to **false positive** (normal users considered as paedophiles).

- Attracting paedophiles with **Honeypots** (fake files) is **resource consuming**: popular files need to show a **high number of sources**.

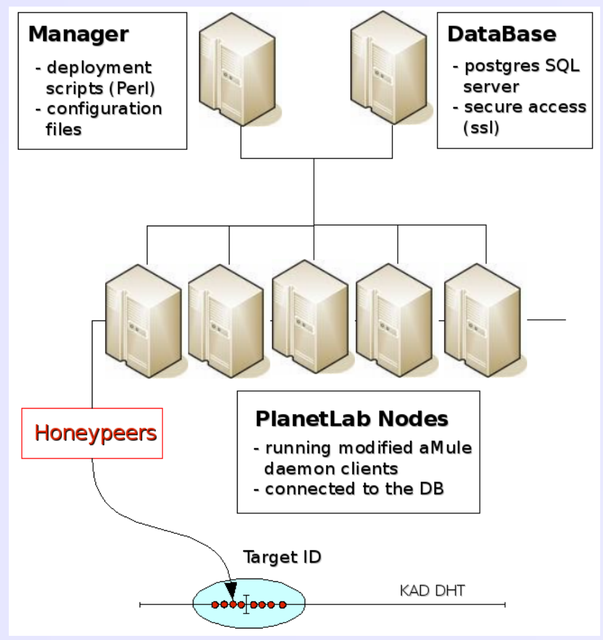
- Recent **protection mechanisms** inserted in KAD mitigate the **Sybil attack** (insertion of many fake peers from a single computer to disturb the network).

Our solution: a specific Honeynet architecture

Our solution, **HAMACK** (Honeynet Architecture Against MAlicious Contents in KAD), relies on **3 KAD properties**:

- The weakness of KAD allowing to **freely** choose the **place of a peer** in the network
- The fact that content is always tried to be **published** on the **closest peers** possible
- The **very efficient lookup algorithm** used to find the peers responsible for a specific content

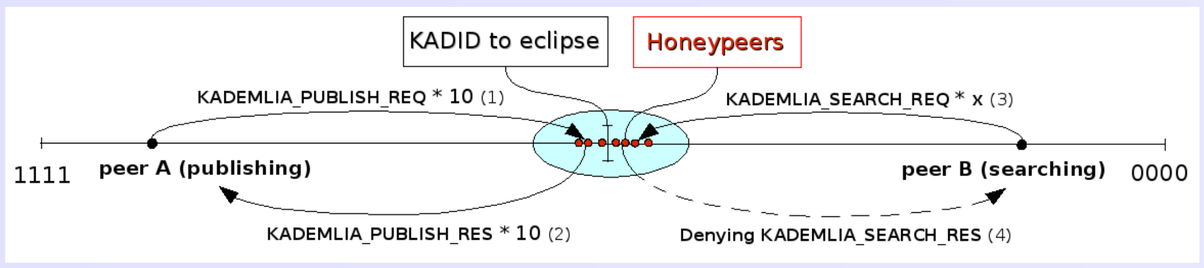
We proved that placing **20 Honeypeers** closer that any other peer to a given file or keyword **allows to control it**.



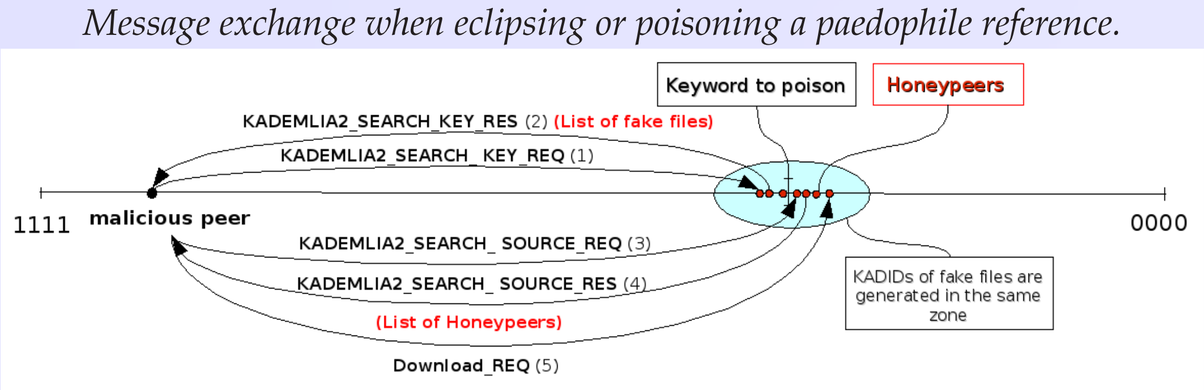
By attracting all the **publications** and **searchs** of **paedophile contents**, HAMACK can **assess and control** the **paedophile behavior** from the **initial search** of keyword to the **final download**.

HAMACK features to fight against paedophile activities

- **Passive monitoring:** attract all Publish & Search requests, store them in database, answer normally.
- **Eclipsing content:** attract all Publish & Search requests, deny Search responses.
- **Index poisoning:** attract all Publish & Search Keyword requests, answer with fake paedophile files.
- **Promoting Honeypots:** attract all Publish & Search Source requests, answer with Honeypeers.
- **Discover** the **new** published **paedophile files** for a given keyword & the **peers** sharing a file.
- **Remove the paedophile content** from the network: prevent users from accessing it.
- Announce very **attractive fake** paedophile files showing a high number of sources.
- Attract the **final download** requests for our fake files.

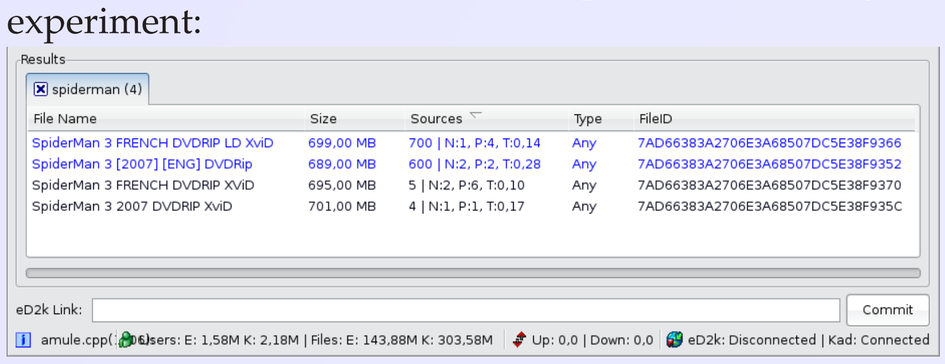


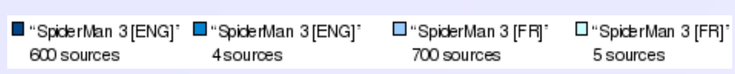
Message exchange when eclipsing or poisoning a paedophile reference.

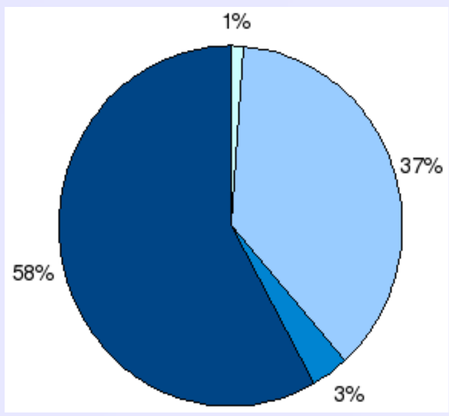


Experiments on the real network

We **eclipsed** the good references for the keyword "**spiderman**" and **poisoned** them with **4 fake files**.
Results returned for a search for "spiderman" during the experiment:







The 2 **fake files** announced with a **high** number of sources **received** much more **download requests** from users. It shows the importance to **control the DHT** to build an **efficient Honeypot** to attract paedophiles.

Acknowledgment: This work is funded by the French ANR Research Project MAPE(Measurement and Analysis of Peer-to-peer Exchanges for pedocriminality fighting and traffic profiling), under contract ANR-07-TLCOM-24 and by the EC IST-EMANICS Network of Excellence (#26854).