Monitoring and Controlling Content Access in KAD

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
- Allow millions of users to share files

Limits

- No central control & autonomous users: P2P networks are a support to **spread malicious files** (paedophilia, malware...)
- Normal users can access to malicious contents unintentionally (**pollution**)

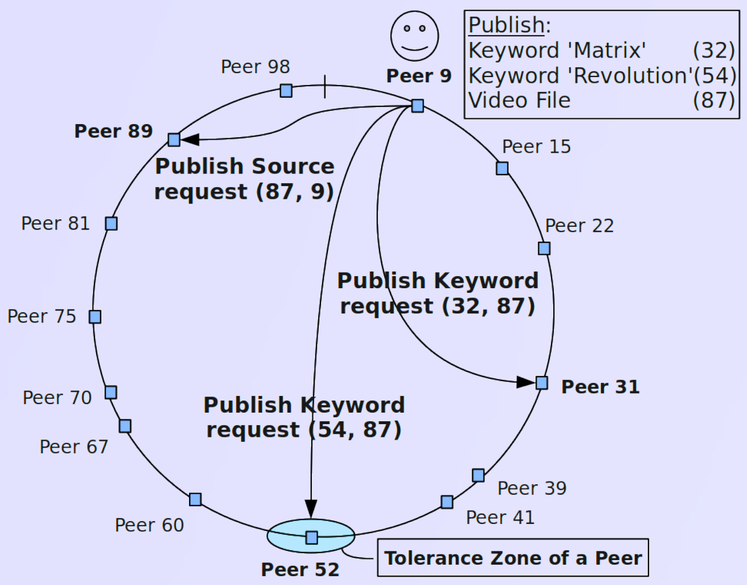
Objectives

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

The KAD network

KAD is used by **eMule** to index and **retrieve the files shared** by the users ($\sim$ 3 millions). 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 **KADID**.

- 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 uses similar **Search requests**.

Technical difficulties

Monotoring users activity or **controlling contents** in a P2P network are 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).

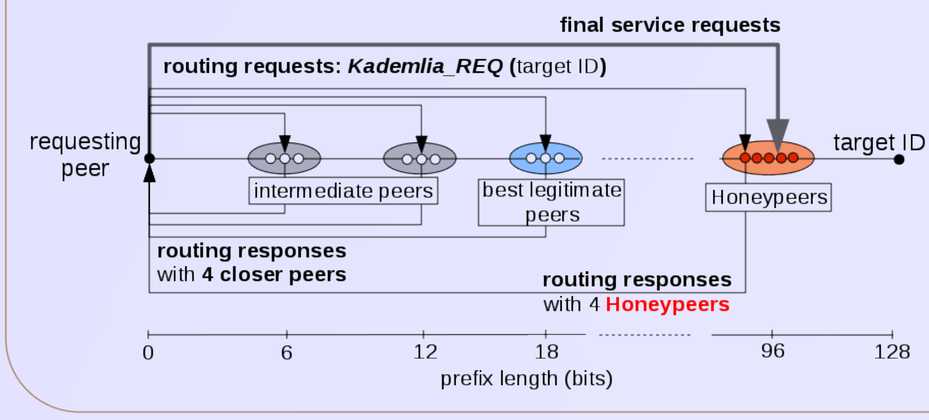
- Attracting users with **Honeypots** (fake files) is **resource consuming**: popular files need to show a **high number of sources**.

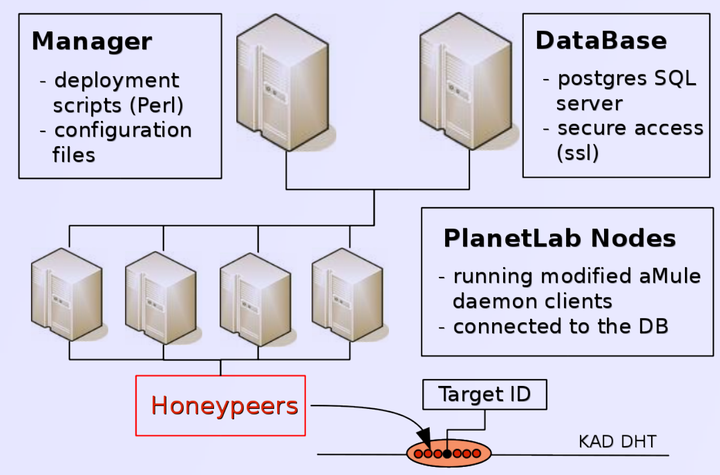
- Recent **protection mechanisms** inserted in KAD mitigate the **Sybil attack** (insertion of many fake peers from a single computer to disturb the network).

Our solution: a specific Honeynet architecture

HAMACK (Honeynet Architecture for Monitoring content ACcess in KAD), relies on **2 properties**:

- The weakness of KAD allowing to **freely** choose the **place of a peer** in the network
- The **lookup algorithm** used to find the peers responsible for a content: KAD always **publishes** the content on the **closest peers** possible.

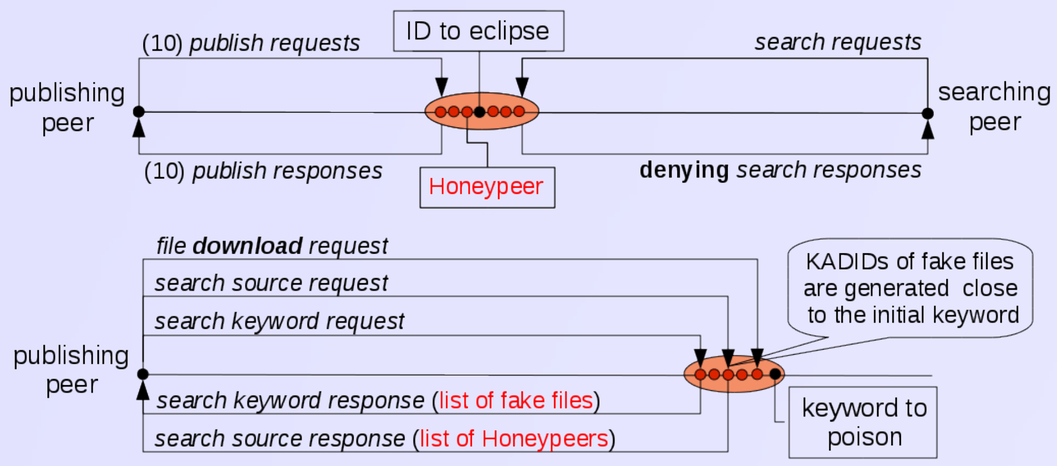




We proved that placing **20 Honeypeers** closer that any other peer to a given file or keyword **allows to control it**.

HAMACK features against malicious contents

- **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 generated files.
- **Promoting Honeypots:** attract all Publish & Search Source requests, answer with Honeypeers.
- **Discover** the **new** published **malicious files** for a given keyword & the **peers** sharing a file.
- **Remove the malicious content** from the network: prevent users from accessing it.
- Announce very **attractive fake** files showing a high number of sources.
- Attract the **final download** requests for our generated files.

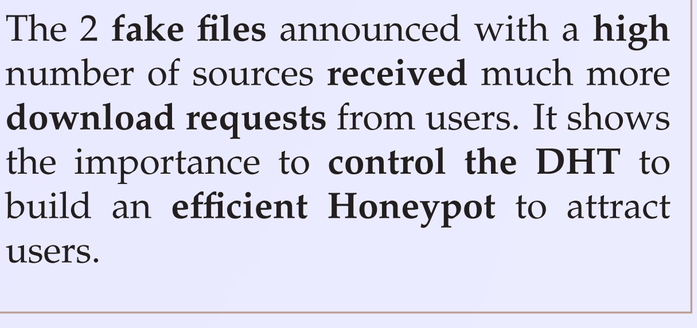


By attracting all the **publications** and **searchs** of a particular **malicious contents**, HAMACK can **assess and control users behavior** from the **initial search** of keyword to the **final download**.

Experiments on the real network

Performance evaluation: (see paper)

- Number of probes, probes configuration
- Load distribution
- Number of replicated requests captured...



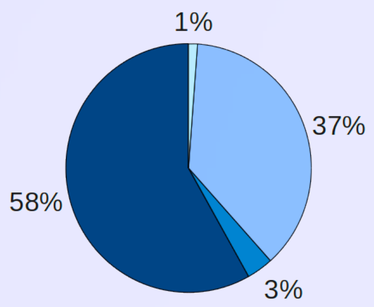
Controlling access to real contents:
We **eclipsed** the good references for the keyword "**spiderman**" and **poisoned** them with **4 fake files**.

Results

spiderman (4)

File Name	Size	Sources	FileID
SpiderMan 3 FRENCH DVDRIP LD XviD	699,00 MB	700	7AD66383A2706E3A68507DC5E38F9366
SpiderMan 3 [2007] [ENG] DVDRip	689,00 MB	600	7AD66383A2706E3A68507DC5E38F9352
SpiderMan 3 FRENCH DVDRIP XViD	695,00 MB	5	7AD66383A2706E3A68507DC5E38F9370
SpiderMan 3 2007 DVDRIP XviD	701,00 MB	4	7AD66383A2706E3A68507DC5E38F935C

"SpiderMan 3 [ENG]" 600 sources — "SpiderMan 3 [ENG]" 4 sources — "SpiderMan 3 [FR]" 700 sources — "SpiderMan 3 [FR]" 5 sources



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 users.