Detection and analysis of paedophile activity in P2P networks

Raphaël Fournier-S'niehotta

Réseaux et individus, informatique et sciences sociales,
LIAFA

December, 3rd 2012

Outline

1. Paedophile queries
2. Paedophile users
3. Dynamicity
4. Conclusion

Large sets of queries

- Interaction between users and search engines
- Applications
  - traditional (system improvements)
  - original (Google Flu)
- set of queries : $q_i = (t, u, k_1, k_2, \ldots, k_n)$
  - t timestamp
  - u user information (IP address, connection port)
  - $(k_1, k_2, \ldots, k_n)$ sequence of keywords

Rationale

- Children victimization
- Danger for innocent users
- Societal problem

Very little is known

Goals

Increase knowledge
of paedophile activity in P2P systems

Detection

- Create an automatic tagging tool
- Elaborate a generic methodology

Analysis

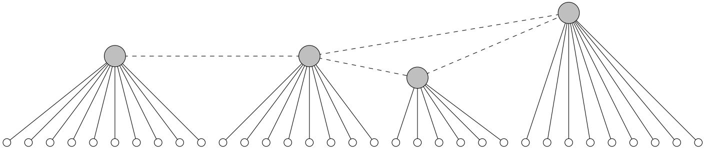
- Rigorous quantification of queries
- Study users

Challenges

- Appropriate data collection
  size, dynamicity, poorly documented protocols
- Automatic detection tool
  hidden activity, several languages
- Rigorous statistical inference
  low amount of paedophile queries

Datasets

- eDonkey (*eMule*, *MLDonkey*, *Shareaza*)



	Duration	Nb Queries	Nb IP
2007	10 weeks	107 226 021	23 892 531
09-12	147 weeks	1 290 377 956	82 264 897
2009	28 weeks	205 228 820	24 413 195

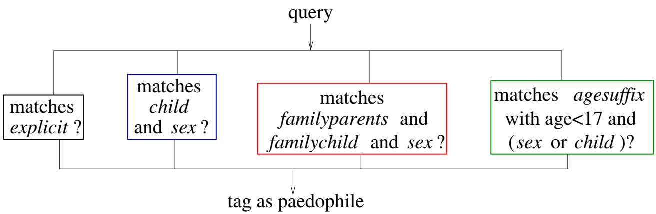
Duly anonymised

Outline

1. Paedophile queries
    - Tool design
    - Tool assessment
    - Fraction of paedophile queries

Tool design

- 4 categories of paedophile queries


query
matches *explicit* ?
matches *child* and *sex* ?
matches *familyparents* and *familychild* and *sex* ?
matches *agesuffix* with age<17 and (*sex* or *child*)?
tag as paedophile

raygold little girl

porno infantil

incest mom son video

12yo fuck video

Quality

False positive

“sexy daddy destinys child”
contains “sexy”, “daddy” and “child”
but most likely a music-related query

False negative

“pjk 12yo”
contains paedophile keywords that we don’t search for

How to estimate false positive and false negative rates?

Tool assessment – Survey

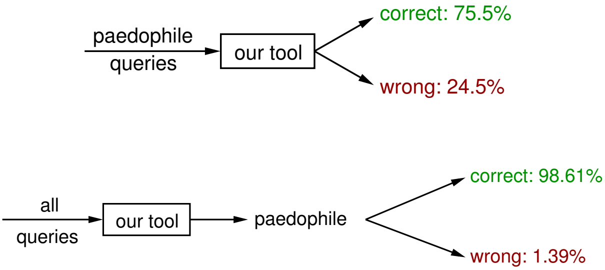
- set of 21 volunteering experts (Europol, national authorities, NGOs)
- set of 3,000 randomly selected queries:
  - paedophile
  - not paedophile
  - *neighbours* (submitted within the 2 previous or next hours of a paedophile query by the same user)
- tag queries as *paedophile*, *probably paedophile*, *probably not paedophile*, *not paedophile* or *I don't know*

	prob.	*je ne*	*prob.*	*pas*		
pédo	*pédo*	*sais pas*	*pas*	*pédo*	total	pertinence
...	...	...	...	...	...	...
1174	111	20	64	789	2158	99.1
...	...	...	...	...	...	...

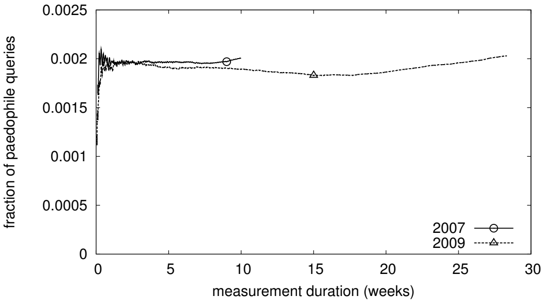
Assessment results

Limited filter precision

- False negatives
- False positives



Fraction of paedophile queries



Result

- detected queries: slighlty above 0.19% for both datasets
- after correction: 2,5 queries out of 1,000 are paedophiles
- 1 paedophile query every 33 seconds

Outline

2 Paedophile users

- Distinguishing different users
- Fraction of paedophile users

Distinguishing users

Classical hypothesis:
user $\sim$ IP address

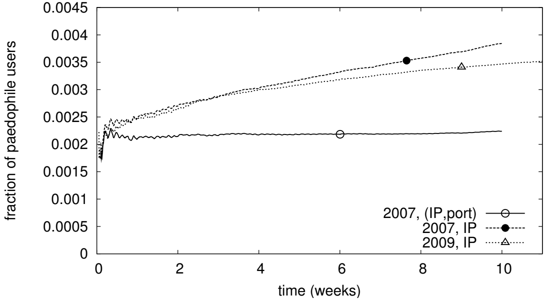
Problems

- gateway/firewall (NAT) IP addresses
- dynamic addresses allocation
- several users per computer
- several computers per user

Improvements

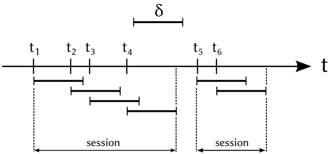
- user $\sim$ IP adress + connection port
- measurement duration
- sessions

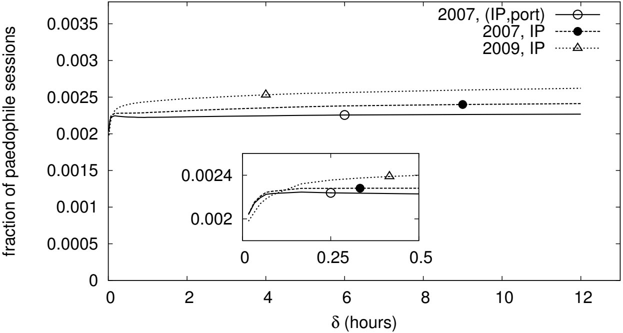
User: IP *vs* (IP,port)



- hypothesis: user tagged as paedophile after one such query
- pollution: all dynamic/public IP addresses may be considered as paedophile *after some time*
- convergence when considering (IP, port)

User: sessions





Fraction of paedophile users

- False positive/negative rate on users
- $p(u \in U^{+} \mid u \in V(n, 0)) = 1 - (1 - f'^{-})^{n}$
- $p(u \in U^{-} \mid u \in V(n, k)) = (f'^{+})^{k}(1 - f'^{-})^{n-k}$

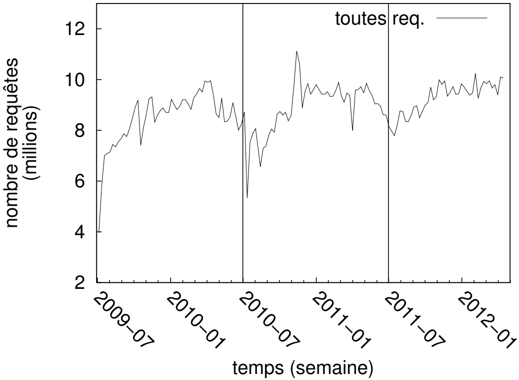
Result

- Fraction of paedophile users close to 0,22% for both datasets

Outline

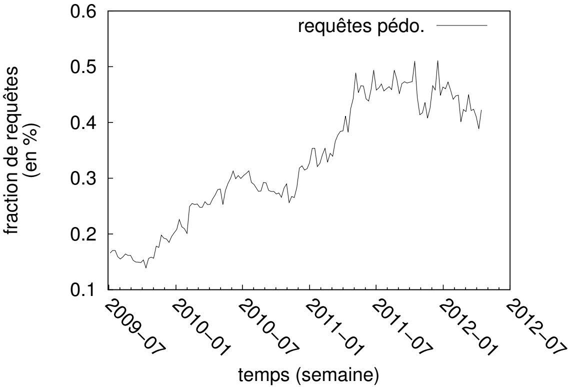
3 Dynamicity
- Long-term evolution of paedophile activity
- Daily evolution of paedophile activity

Long-term evolution



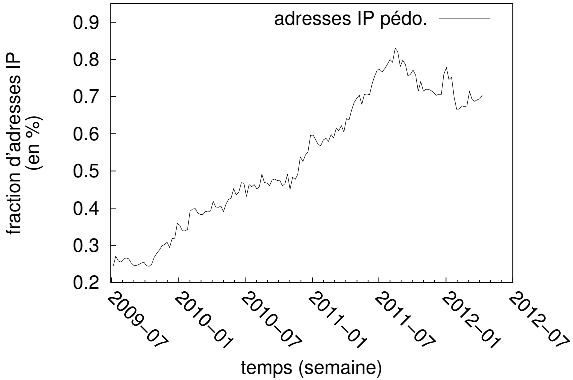
- stability of global traffic over 3 years
- fraction of paedophile queries strongly increasing
- fraction of paedophile users also increasing

Long-term evolution



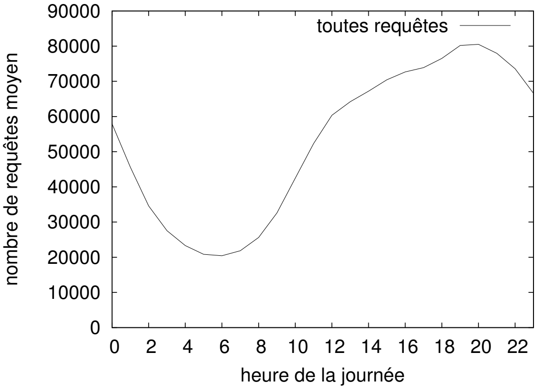
- stability of global traffic over 3 years
- fraction of paedophile queries strongly increasing
- fraction of paedophile users also increasing

Long-term evolution



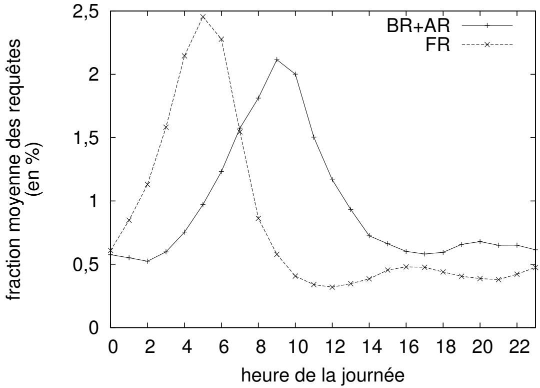
- stability of global traffic over 3 years
- fraction of paedophile queries strongly increasing
- fraction of paedophile users also increasing

Daily evolution



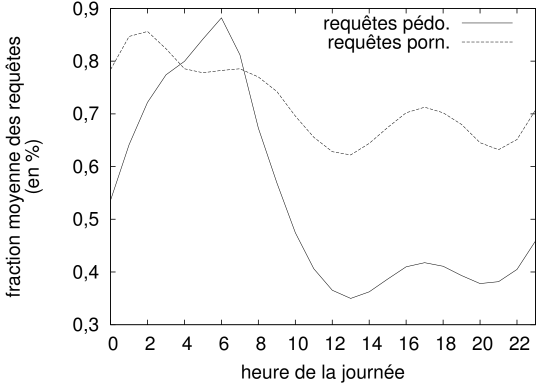
- circadian cycle (day/night effect)
- fraction of paedophile queries peaks at 6 AM
- paedopornagraphy and traditional pornography differ

Daily evolution



- circadian cycle (day/night effect)
- fraction of paedophile queries peaks at 6 AM
- paedopornagraphy and traditional pornography differ

Daily evolution



- circadian cycle (day/night effect)
- fraction of paedophile queries peaks at 6 AM
- paedopornagraphy and traditional pornography differ

Outline

4 Conclusion

Conclusion

General approach for detecting rare contents

Contributions

- Automatic detection tool
- Large set of paedophile queries
- Rigorous quantification
  2.5 queries out of 1,000 are paedophile
- User identification
- Quantification of paedophile users

Other contributions

- Comparing eDonkey and KAD

Perspectives

Tool improvement

- previous/next queries
- languages, word order, categories
- machine learning

Analysis

- different threshold for being considered paedophile
- geolocation
- community detection (graph topology)
- detailed study of sequences of queries

Contact

Thank you for your attention.

raphael.fournier@lip6.fr

KAD network

- Completely distributed protocol of clients
- No server for file indexing
- Some peers are in charge of some files and keywords

Principle:

- Precise and targeted injection of peers into the network to control files or keywords
- Peers catch queries and control replies

Applications:

- Which files are published for a given keyword? Which peers share them ?
- Eclipse : prevent peers from accessing content

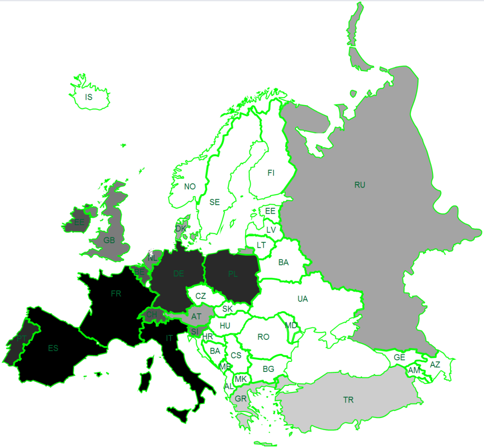
Geo-location: statistics

country	# queries	# paedo	ratio
IT	19569361	15426	0.08 %
ES	8881405	5177	0.06 %
FR	7583815	8059	0.11 %
BR	2795090	4849	0.17 %
IL	2139697	2618	0.12 %
DE	2093106	11238	0.54 %
KR	1386799	336	0.02 %
US	1053183	6184	0.59 %
PL	975170	1178	0.12 %
AR	810466	1465	0.18 %
CN	635392	337	0.05 %
PT	513327	434	0.08 %
IE	511185	54	0.01 %
TW	417893	138	0.03 %
BE	402565	646	0.16 %
CH	320054	1710	0.53 %
GB	319386	1698	0.53 %
NL	243646	1131	0.46 %
CA	241460	1233	0.51 %
SI	239572	167	0.07 %
MX	210504	1098	0.52 %
RU	200958	2712	1.35 %
AT	184248	977	0.53 %

Biased by:

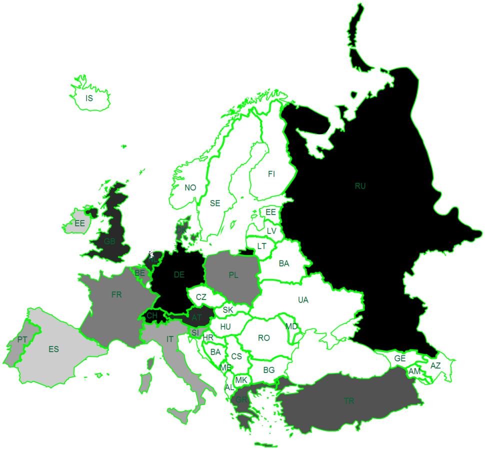
- language knowledge
- decoding problems

Geo-location: maps



# queries

Geo-location: maps



ratio # paedophile queries / # queries