International Conference

# Advances in the Analysis of Online Paedophile Activity

Paris, France
2-3 June 2009

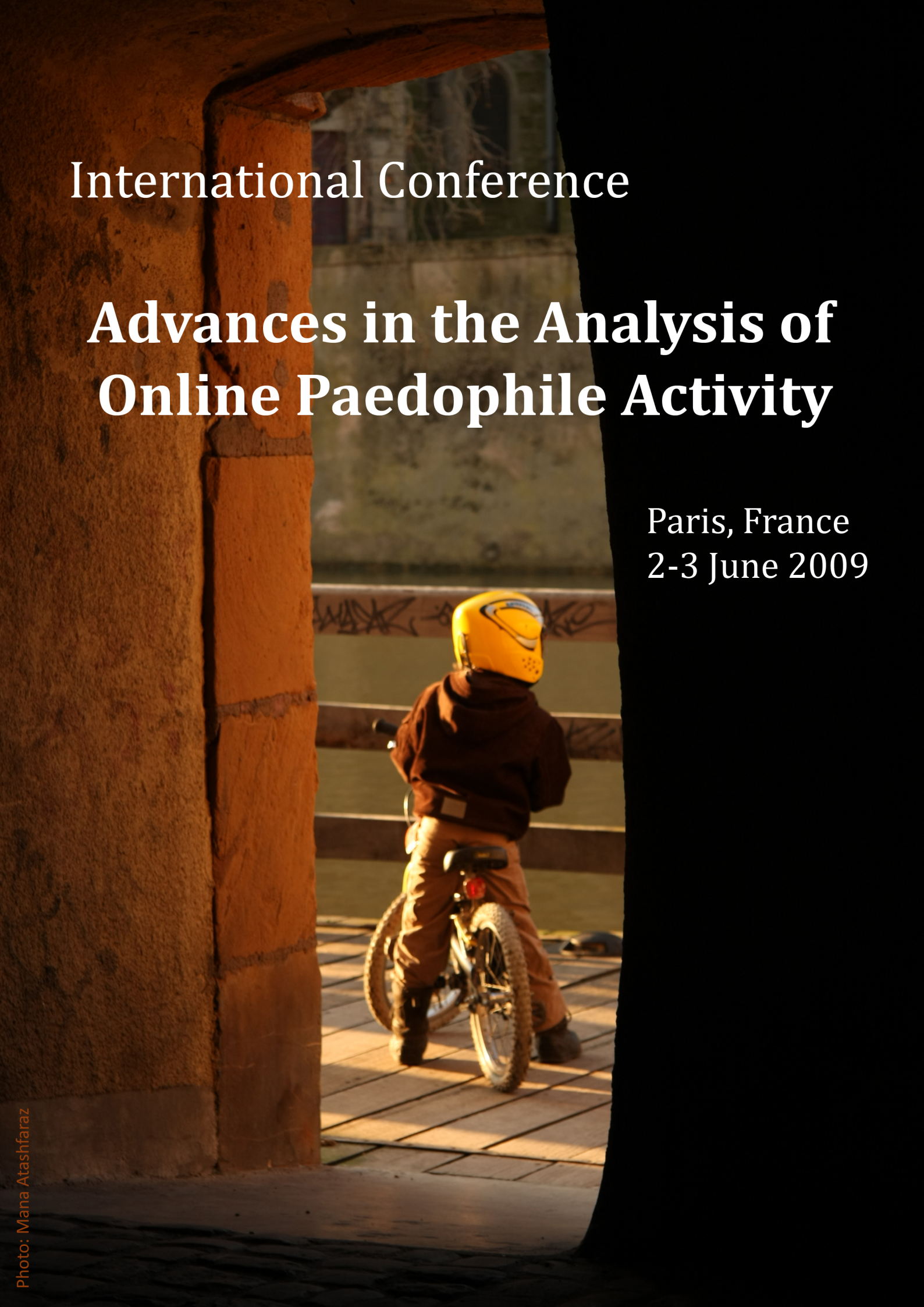


Photo: Mana Atashfaraz

International Conference

# Advances in the Analysis of Online Paedophile Activity

Paris, France; June 2-3, 2009

---

It is nowadays widely acknowledged that online pedophile activity constitutes a key issue with huge societal, ethical, legal, clinical and financial impacts.

However, measuring, observing and analyzing this activity is a challenging task. As a consequence, much effort is devoted to various kinds of works aimed at providing more knowledge and better control of this activity. This led to a wide set of significant results, but they remain relatively scattered.

The goal of this conference is to provide a clear and precise view of what is currently known on online paedophile activity as a whole. It will put together practitioners, analysts, and researchers, in an effort to build a unified vision of the current situation and key directions for future work.

The conference will rely on a set of key presentations given by leading experts of the topic who will present the state-of-the-art from their own perspective. In addition, participants are warmly invited to prepare posters to be presented during dedicated sessions at the conference, and discussion spaces will be provided to encourage the emergence of cooperation.


Co-funded by: the European Union and the CNRS

❖ *Printed in Paris; May 2009. Editor: Massoud Seifi*

# Program

## Articles



## Posters

# Measurement of Paedophile Activities in eDonkey

Guillaume Valadon, Clémence Magnien, Matthieu Latapy
UPMC Univ Paris 06 & CNRS, UMR 7606 LIP6 Paris, France
{firstname.lastname}@lip6.fr

## 1. CONTEXT

Nowadays, peer-to-peer (p2p) file exchange systems, such as eDonkey [1] are widely used on the Internet. Using keywords, users can search for files matching their interests, and retrieve them from several users at the same time. Many studies and independent contributions show that a huge amount of paedophile and harmful contents are distributed using p2p file exchange systems, and that the volume of such exchanges is increasing, see for instance [2]. The presence of such content, and its very easy access, make the current situation particularly worrying for p2p users, in particular children. Despite the fact that this situation is nowadays widely acknowledged, there is still no available filtering technique or content rating system to protect p2p users, in particular children, from harmful and paedophile content. Similarly, only few tools exist to help law enforcement authorities and other child protection organisations in fighting p2p paedophile exchanges. Furthermore, there is still an important lack of precise knowledge on this topic : the numbers of paedophile users, and files containing paedophile content are not precise.

Our objective is to improve the knowledge of the paedophile activity in p2p systems. We aim at giving an accurate and detailed view of what is going on concerning paedophile activity in currently running p2p systems. This includes the evaluation of the number of files/users involved, the identification of various kinds of files/users, and several other basic statistics, together with their evolution during time. We want to change the current situation into a situation in which we have a precise knowledge of paedophile activity in p2p systems. Moreover, we want to study how files are spread among clients, identify communities of common interests as well as provide a content detection system that identify paedophile or pornographic files without accessing their content.

## 2. THE EDONKEY PROTOCOL

The eDonkey protocol [1] is a file exchange p2p system that relies on servers and clients. In a nutshell, servers store lists of files shared by the clients, and act as search engines. On the other hand, clients send queries[1] to the servers and obtain a list of files that match the keywords contained in the query, as well as a list of clients that share these files. Using these retrieved lists, clients are able to download files from different clients at the same time.

From the technical perspective, communications in the eDonkey protocol can be narrowed down to three distinct types: (a) server to server, (b) client to server, and (c) client to client. The specification of the server to server communications being not publicly available, we were not able to observe them. However, they are not relevant to the study of paedophile content because they primarily contain statistics about clients and servers usage. The types of communications consequently impact the measurements that can be performed as well as the results that will be obtained concerning the study of paedophile activity. For example, the first provider of a paedophile picture cannot be observed the same way with client to server and client to client communications.

## 3. MEASUREMENTS ACTIVITIES

The goal of our measurements is to gather precise information about the eDonkey systems such as queries, filenames, or the first appearance of a client. Moreover, we aim at validating our measurement activities in order to assess that the observations are correct and that we gather as much information as possible. We performed three different measurements using: (1) servers, (2) honeypots, and (3) *regular* clients.

### 3.1 Server measurement

We conducted a *passive* measurement on an eDonkey server during ten consecutive weeks [3]. In this kind of measurement, it is possible to collect all of the queries sent to the server, as well as the files provided by the clients. Here, the measurement methodology is rather simple and consist in capturing all of the traffic sent and received by the server, anonymizing the data before

[1] sequences of keywords.

storage[2]. The resulting data set occupies 500 gigabytes in total. It contains 10 millions queries, 90 millions of IP addresses and 280 millions of distinct files.

This measurement as the benefit of capturing all of the data exchanged by the server, however it fails at collecting clients to clients communications. Hence, we are not able to study the file exchanges that occur between clients.

## 3.2 Client measurements

There are two different client based measurements. The first one, *Honeypot*, aims at collecting information about file exchanges between clients, whereas the second one performs queries using specific keywords to monitor files available in the eDonkey network. They are conjointly used to circumvent the limitations of the server measurements.

### *3.2.1 Honeypot*

We developed an eDonkey *honeypot* : a client that (1) informs the server that it is willing to exchange a predefined list of files, and (2) that allows connection from other clients, but will either send random content or no content at all. Moreover, the honeypot retrieves the list of files shared by clients contacting it. Consequently, it is now possible to measure the paedophile activity concerning client to client communication, and to precisely identify the interest of users concerning specific files.

Using this honeypot, we conducted an *active* measurement [4] during 32 days; 24 distributed honeypots were advertising 4 files. The resulting data set contains 110 049 IP addresses and 28 007 distinct files.

Two important facts were enlighten by this measurement : (1) long and distributed measurements are relevant and allow to discover more peers and files; (2) with the random content strategy, more peers contact the honeypots than with the no content one.

### *3.2.2 Client sending queries*

Here, we use an eDonkey client that exactly act as a regular client would. It periodically queries eDonkey servers using a list of predefined paedophile and non-paedophile keywords. The goal of this measurement is therefore to enumerate the files that are globally available in the eDonkey network, and to be able to detect when files appear.

This measurement was conducted during 140 days from October 2008 to February 2009. We observed 2 978 764 distinct IP addresses and 2 784 583 distinct files.

While simple, this experiment shows that it is really difficult to completely discover all of the files available in eDonkey : long measurements discover new files with an important growth rate.

[2]the data set is available at `http://content.lip6.fr/latapy/edonkey/weeks/`

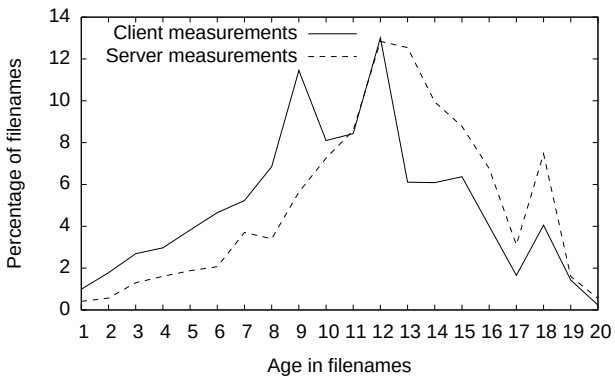



**Figure 1: Distribution of ages seen in filenames.** The y-axis indicates the percentage of filenames (containing ages) including the corresponding age on the x-axis.

## Example: Ages in filenames

A unique format was used to store the data from the three different measurements. It is therefore straightforward to analyse the resulting data sets using the same techniques. Figure 1 presents the distribution of filenames containing information about ages [3] for a given age.

Whereas we consider two different data sets, a striking result visually arises concerning ages : there is a clear interest for 18 years old, and another one around 12 to 13 years old. We are currently investigating the spike at 9 years old on the client measurement in order to check whether this is a measurement artifact : the server measurement contains much more filenames than the client one, and was performed two years ago.

## 4. REFERENCES

[1] Y. Kulbak and D. Bickson, "The eMule Protocol Specification," 2005, http://citeseer.ist.psu.edu/kulbak05emule.html.

[2] United States General Accounting Office, "File sharing programs : Child Pornography Is Readily Accessible over Peer-to-Peer Networks," 2003.

[3] F. Aidouni, M. Latapy, and C. Magnien, "Ten weeks in the life of an eDonkey server," in *Sixth International Workshop on Hot Topics in Peer-to-Peer Systems (Hot-P2P 2009), Rome, Italy*, May 2009.

[4] O. Allali, M. Latapy, and C. Magnien, "Measurement of eDonkey Activity with Distributed Honeypots," in *Sixth International Workshop on Hot Topics in Peer-to-Peer Systems (Hot-P2P 2009), Rome, Italy*, May 2009.

[3]such as 12 years old is encoded as 12yo or 12yr.

## Investigator's problems fighting against paedophilia on the Internet


- Philippe JARLOV
- Gendarmerie Nationale
- Section Recherches Bordeaux - FRANCE


1

## PLAN

- Introduction
- How do they use Internet ? (the problems)
- How will they use it in the future ?
- Conclusion

2

## Introduction

The Internet network became known to the general public in France in 1996 but has really started to grow after 1998, with a leap forward in 2000 and until now, thanks to the arrival of the broadband technologies.

3

## Introduction

The fight against paedophilia exists since the beginning of the internet, but the evolution of protocols and possibilities offered to the users make the problems always more and more complex to investigators :

- Many protocols

- easier to use

- increasingly easy to stay anonymous

4

## Introduction

To understand these difficulties, we are going to analyse the possibilities offered to paedophiles in order to share files, chat with victims and become predators or child abusers

5

## How they use Internet ? : IRC

What is IRC ?

**Internet Relay Chat**, commonly abbreviated **IRC**, is a chat protocol, a way to enable several people to talk to each other by entering text messages, but also transfer files as public or private servers.

6

## How they use Internet ? : IRC

IRC exists since the beginning of the internet and paedophile chats like « pedomon » and « preteengirlsexpics » have been well known on IRC Undernet or Dalnet for many years.

7

## How they use Internet ? : WWW

The presence of paedophilic files on Web sites is not to be demonstrated anymore. It's a fact.

8

## How they use Internet ? : WWW

Web 2.0 is a technology that can be used to develop more attractive websites, with new features like a private area, contact list, chat...

Web 2.0 is particularly valuable for paedophiles as they can use it to contact victims, create private groups, discuss and share files... in other word create and use a newsgroup, but in a much more simple way.

9

## W 2.0 Exemples : facebook

Users can join networks organized by city, workplace, school, and region to connect and interact with other people

10

## W 2.0 Exemples : aka-aki

A new social network where you can see on your mobile phone if one of your contact is near you (GPS technology).


11

## How do they use Internet ? : MSN

For instance, paedophiles can contact victims after obtaining they nickname from a blog. Then they can start a chat or exchange files, use a webcam and do even worse, like organise a real-life meeting.

12

## How do they use Internet ? : MSN Exemple

A MSN user was was arrested as paedophile on April 28th, 2009.
Profile : Education assistant in a technical high school of Savoy, aged 42. Found guilty of offenses of a paedophilic and pornographic nature. Apart from illegal videos and pictures discovered on his computer, the police have found proof of relations he had with teenagers using MSN chat.

13

## How do they use Internet ? : Virtual Worlds

Virtual world like SECOND LIFE can become hunting grounds for paedophiles.
There are more and more Virtual Worlds , as you can see on the next snapshot, and this is just an example.

14

## How do they use Internet ? : Virtual Worlds

www.virtualworldsreview.com


Active Worlds
Coke Studios
Cybertown
Disney's Toontown
Dreamville
Dubit
Habbo Hotel
The Manor
Mokitown
Moove
Muse
The Palace
Playdo
Second Life
The Sims Online
Sora City
There
TowerChat
Traveler
Virtual Ibiza
Virtual Magic Kingdom
Voodoo Chat
VPchat
VZones
whyrobbierocks
Whyville
Worlds.com
Yohoho! Puzzle Pirates

15

## How do they use Internet ? : Virtual Worlds

•Paedophilic Playground Discovered in 'Second Life' Virtual World

**•Every lifestyle group has its own place in the virtual world "Second Life" — including, apparently, paedophiles.**

•Britain's Sky News TV channel on Tuesday uncovered a virtual playground hidden away behind a strip mall in "Second Life" — a playground where little girls who looked about 10 years old offered the Sky reporter's avatar, or virtual representative, a variety of sex acts.

16

## How do they use Internet ? : Virtual Worlds

www.youtube.com/watch?v=dN_jr6xjs90

(The video)

17

## How do they use Internet ? : P2P Networks

MASSIVE FILE SHARING
EDONKEY/EMULE
GNUTELLA NETWORK

Thousands of paedophilic files are shared every day on these P2P networks, as controls are beeing performed on adaily basis and illegal users arrested. (the video)
Some of these arrests are very valuable to us as they help us identify child abuser's.

18

## How do they use Internet ? : BBS

A **Bulletin Board System**, or **BBS**, is a system running a software that allows post messages.

This protocol is old but predators use it, believing law enforcement does not operate anymore on this field, but only on P2P or chats...

19

## How do they use Internet ? : BBS Exemples

**485** Name: **Anonymous** : 2006-03-04 08:36 (sage)

>>484

Having lots of kids in my neighborhood, I do not need cp.

**489** Name: **Anonymous** : 2006-03-05 02:16

>>488

homosexual s&m pedo rape is best

**494** Name: **Anonymous** : 2006-03-06 17:31

i like little girls

20

## How do they use Internet ? : Newsgroups

A **usenet newsgroup** is usually within the system, for messages from many users in different locations.

Newsgroups are more easy to use than BBS

21

## How do they use Internet ? : Newsgroups

A list of newsgroups :

**alt.binaries.pictures.erotica enfant**
**alt.binaries.pictures.erotica.lolita**
**alt.binaries.pictures.lolita.fucking**
**alt.sex.children**
**alt.sex.pedophilia.girls**

22

## How will they use it in future ?

The future is now for some of them,

the solution is : be anonymous...but how ?

23

## Anonymous on IRC :

silcnet.org


SILC Pidgin is a full featured graphical SILC client included in the Pidgin distribution. Download your own SILC Pidgin immediately!

24

# Anonymous on Chats 1/2

scatterchat.en.softonic.com

ScatterChat represents a great way to make sure that your private conversations stay private, by allowing you to connect securely through Gaim encryption technology.

25

# Anonymous on Chats 2/2


26

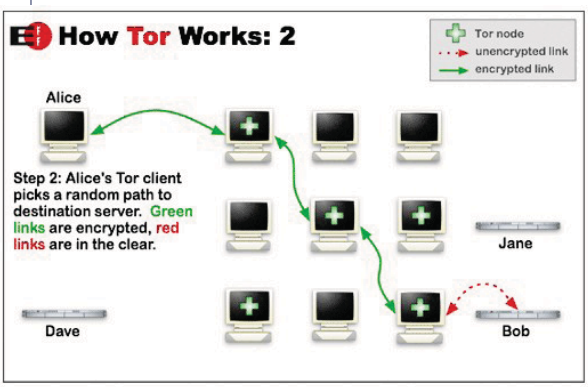
# Anonymous on the Web : Tor 1/2

www.torproject.org

Tor is free software and an open network that helps you defend against a form of network surveillance

27

# Anonymous on the Web : Tor 2/2




28

## Anonymous : Virtual Operating System

Anonym.OS project (based on bsd system)
is a bootable live cd based on OpenBSD that provides a hardened operating environment whereby all ingress traffic is denied and all egress traffic is automatically and transparently encrypted and/or anonymized.

29

## Anonymous : Virtual Operating System

Janus Virtual Machine
Internet Privacy Appliance : Encrypts your Internet traffic, hides your IP address, and is easy to setup.

STIA SYSTEM
Surf The Internet Anonymously, known as STIA for short, is an assemblage of various Free Software to provide a secure platform that improves your Anonymity

30

# Anonymous : a paedophilic message in a chat

----- Anonymous ----- 2008.02.07 - 19:30:17GMT -----

A lot of people say it, but I thought I'd mention "how" I did it in case it helps anyone else out there.

I have a 100+GB removal hard drive.

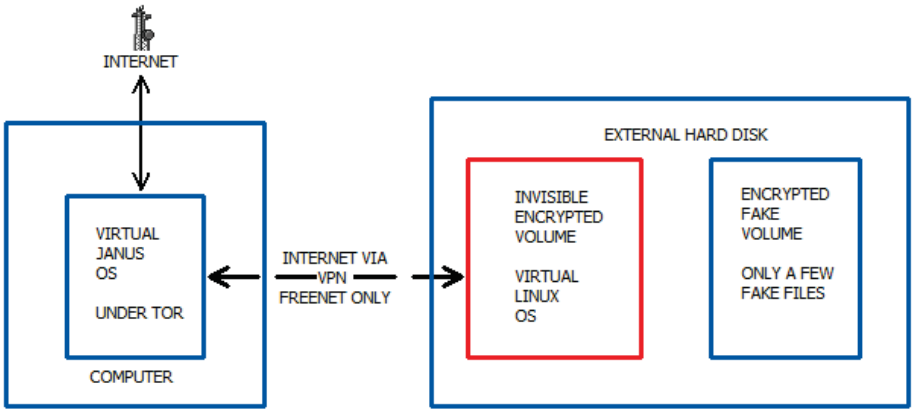
On that hard drive is sufficiently large Truecrypt container (configured with a fake visible container with some various private documents, and a hidden container for cp)

The hidden container holds a linux VM. This VM is where I have everything cp-related installed.
I also have a second VM running Janus, which runs Tor. I connect the linux VM to the Janus VM through openVPN, which eliminates any chance of Java or Javascript giving up my address. The only thing on the linux VM that is allowed to talk to the outside world without going through the VPN/Janus/TOR is Freenet.

Absolutely nothing leaks onto my "everyday" operating system.

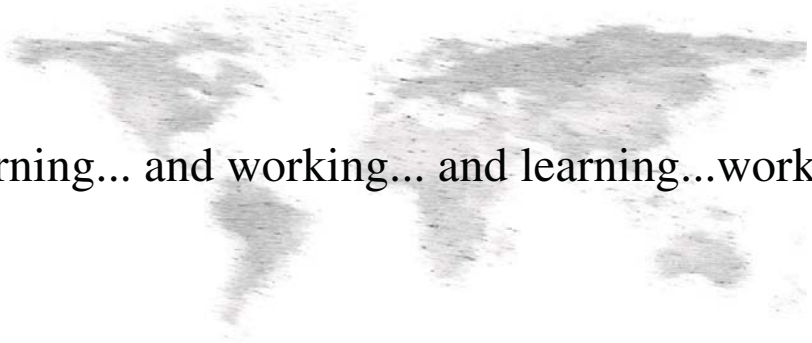
31

# Anonymous : a paedophilic message in a chat




32

# Conclusion



...learning... and working... and learning...working

# FIVES and P2P-based Intelligence Gathering

Johan Garcia
Dept of Computer Science, Karlstad University, Sweden

**Abstract**

This paper describes the FIVES project that has been initiated in order to develop tools that allows the police to efficiently evaluate large amounts of image and video material seized in connection with investigation of Child Sexual Abuse material (child-pornography). The FIVES project build on a modular forensic engine and a number of modules that extract different types of relevant information in an automated fashion. The modular approach of FIVES makes it possible to easily extend to also consider the issue of P2P-based CSA distribution and related investigations. The connection between P2P and FIVES can be done both in the sense that FIVES modules can use information derived form P2P monitoring, as well as using FIVES as a starting point for an automated support tool for P2P investigations.

## 1 Introduction

The development of computer technology and computer communication during the last decades has lead to many benefits to society. However, it has also provided new means for the spreading of images and videos of child sexual abuse (CSA, also referred to as child-pornography). When the police seize material in this kind of investigations the examination of the seized material typically involves a considerable amount of manual labor, which consumes time and resources for the police. To improve the degree of automated work that can be performed the FIVES project [1] was initiated. The aim of the FIVES project is to develop a flexible tool-set that can be used to provide automated support to investigators working on cases related to CSA. Since the work-flow, investigative goals, available computing resources as well as the level of forensic computing sophistication varies considerably between different police forces, the system allows for a large degree of flexibility. In this paper, we provide an overview of FIVES and describe how the flexible architecture of FIVES could be leveraged to incorporate intelligence gathered from P2P-related monitoring activities, as well as sketch how FIVES could be extended to allow for more automation of monitoring and intelligence gathering in relation to P2P based exchange of CSA material.

## 2 FIVES overview

FIVES (Forensic Image and Video Examination Support) is a recently started project that develops a flexible software tool-set that allows law enforcement organizations to more efficiently handle large amounts of image and video material related to child sexual abuse. The tools target both image and video material while applying sound forensic procedures in a robust and scalable processing environment. The FIVES tool-set will have user interfaces with configurable complexity, catering for different categories of users. A basic interface can be configured for users such as vice detectives who are not necessarily specialized in computer forensics. This user interface will have a straight-forward design and contain a set of predefined processing chains targeted for child sexual abuse investigations. Specialized computer forensics investigators can work with an advanced user interface and get the full power and flexibility of the system. A simple boot version is also planned mainly for detecting known illegal material using fast file fragment matching.

### Technical Overview

The overall system architecture is illustrated in Figure 1. At the bottom there is the FCCU Linux forensic distribution [8] comprising of a large set of general forensic utilities. On top of this sits the OCFA forensic framework [9], which handles per-file processing in a distributed fashion, routing the files between different processing modules and recording relevant meta-data for each file. The various processing modules provide different specialized functionality. Extractor (E) modules can for example automatically decompress archive files for further processing. Image (I) modules perform specialized image processing such as image similarity matching. Video (V) modules are similarly specialized on video processing. Hashing (H) modules provide specialized hashing functionality. GUI modules offer a set of differently tailored interfaces to FIVES. The modular system design of FIVES provides considerable flexibility and adaptability, which will be helpful when considering the integration of P2P capabilities.

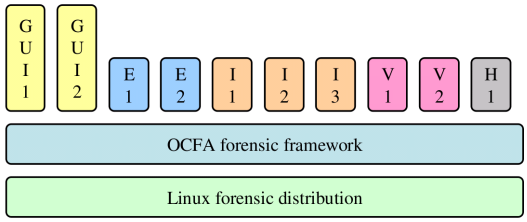



Figure 1: The FIVES overall architecture

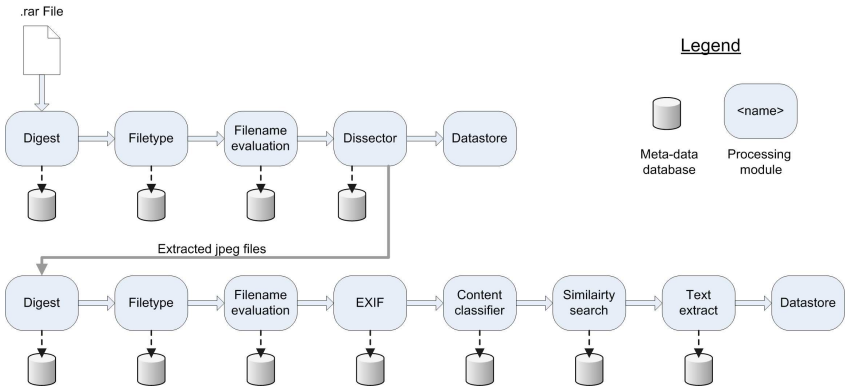



Figure 2: An example processing flow

### Example processing flow

This subsection illustrates how a chain of modules can be executed for different files found on a seized media. Typically, the system automatically scans all files on a storage medium, and may optionally also do a file fragment search. The actions taken on the different files are controlled by a rule list in the OCFA subsystem, which defines a number of possible module chains that can be executed depending on the meta-data extracted by previous modules. An example is provided in Figure 2, which shows how a possible flow can look. In this example the file is a .rar file, and involves the following modules.

1. The digest module calculates hash sums and automatically compares them against lists of known legal and known illegal files.

2. The filetype module identifies which type of file it is by examining file headers and contents. This ensures that files with modified suffixes, such as a .jpg file renamed to .exe file, does not trick the system.

3. The filename evaluation module examines the filename to find any potential clues of the content. This entails examining the filename for substrings known to signify illegal content such as “lolita”, “pthc”, and so on. This type of content-based naming is frequent in P2P systems. In addition, fuzzy matching can be used on a subset of the tokens by measuring the Levenshtein distance [2] of the filename parts to substrings that are believed to be varied in order to defeat exact matching schemes.

4. The archive dissector module produces a new set of files, the ones that were archived in the .rar. file. This module is the last processing module for the .rar file itself which is now sent to the datastore module for final storage.

5. The new files are resent into the system. In this example the files in the archive were a set of JPEG images. Each new file is handled as new evidence, and thus the digest module is the first module followed by the filetype module and the filename evaluation module.

6. In this example the rule-list is configured to send image files to a content classification module. The content classification examines a number of features of the image contents to classify the image into a category. If possible, it would clearly be useful to have an automated classification into the categories non-CSA / CSA. However, it is unlikely that this can be achieved with sufficient reliability using current technology. What more likely can be realized is an automated classification into non-porn / porn, where the porn class would include both adult pornography and CSA.

7. The similarity search module will look to find if the examined file is a visual copy, or close copy, to a known illegal file. While byte identical files will be detected by the digest module, this module can detect images which are visually similar but have been recoded or changed slightly.

8. The text extraction and OCR module will find and interpret text in images. This would for example be useful in order to automatically detect overlaid links advertising sites which are blacklisted, or which contain incriminating substrings. Like the filename evaluation module, also this module will apply fuzzy matching. However, for this module the main reason is to compensate for the errors likely to occur in the OCR step.

It should be noted that the example processing chain described above only shows one processing path with one set of example modules. Files with other content will obviously be processed by a different set of modules. The final set of module functionality set in FIVES is not definitive at this moment.

It is worth to note that the meta-data generated by the modules are stored in an SQL database. This allows powerful searches to be made based on arbitrary parts of all the recoded meta-data, to search among particular sub-sets of the data or to explore the generated meta-data for new investigative leads.

Finally, it should be noted that this is a simplified description and that there are a number of aspects which are not covered in this discussion, such as distribution and time efficiency trade-offs, processing robustness and traceability and so on.

## 3 The use of P2P derived data to improve analysis of seized material

While there are numerous projects that monitor, measure and examine various aspects of P2P data exchange, there are only a few that explicitly examine the issue from the perspective of CSA material exchange. One such project is ISIS [12], and another is MAPAP [14].

In addition to these project there also exists other projects such as INDECT [13], which is a large FP7 project that covers an array of security related issues including the constructing of agents assigned to continuous and automated monitoring of P2P networks, although not specifically targeted towards CSA.

Of the mentioned projects, MAPAP is relevant to FIVES since the data resulting from the collection and processing methodologies developed by that project could be incorporated in FIVES as discussed below. While the data generated by MAPAP is useful, it should be remembered that it provides only a snapshot of the conditions, and continuous monitoring would be needed to update the datasets if they are to be continuously used.

### Content classification

The MAPAP project has examined possibilities to perform automated content rating and fake detection. The approach suggested by MAPAP can likely be generalized to all non-anonymized peer-to-peer systems, but in their case the focus was on E2Donkey. The output data from MAPAP content rating is the E2Donkey hash-based file identifier and the likely content class the file belongs to. This information can easily be used in a FIVES module that computes the E2Donkey-specific hash and inserts the corresponding information as meta-data in the FIVES database. The meta-data can then be used in the later analysis step, where the MAPAP-derived classification meta-data can be fused with other classification meta-data generated by other modules.

### CSA keywords list

Another output from the MAPAP project is a list of likely keywords used by pedophiles to name material. This list is of obvious use in a file name examination module as well as in an image text extraction module. Fuzzy matching can be used to find a limited set of likely variations of CSA keywords. In this case, the MAPAP-generated keyword list is used as an input to these modules, and not as a base for a new module.

## 4 Aspects on automated P2P content analysis

The FIVES tool-kit is based on a forensic engine which is typically used to read forensic image copies of hard-disks. To employ FIVES modules for peer-to-peer related investigations would require a front-end capable of automatically participating in P2P monitoring, and generating the appropriate data. Additionally, the regular functionality of FIVES could also be useful for validation, and there are also a number of issues regarding P2P monitoring and anonymization technologies which are discussed below.

### Validating P2P derived information

The MAPAP project created an approach for large scale file classification. However, since no material was actually downloaded, it is not possible to exactly evaluate the precision of the classification. An evaluation of the classification accuracy could be done by downloading some subset of the classified material to validate the classification and get a measure of the accuracy. If such evaluations are performed, FIVES could be useful to automatically process the downloaded material to decrease the effort needed for such an evaluation.

### Infrastructure and module reuse

FIVES is based on a flexible and modular infrastructure and a range of image and video processing modules. The infrastructure provides the potential for robust distribution to several machines for expedited processing. These properties makes FIVES a potential candidate for integrating it with a P2P front-end to create an automated P2P monitoring and investigation platform. If FIVES is extended such a front-end, the back-end infrastructure can be re-used for the distributed processing functionality. Several of the image and video processing modules of FIVES will also be useful in a P2P monitoring scenarios. Some usage scenarios where such a P2P monitoring system could be useful include:

**New material detection** By utilizing various modules to identify suspected CSA material, and comparing it to previously known material it may be possible to provide automated support to the task of identifying when new, previously unknown CSA material become distributed in P2P networks. Such support could be helpful since it increases the chance of correctly identifying the the first distributor of new material, and thus provides a lead towards the creator of the material.

**Main distributor detection** Monitoring can also be used together with automated tools to expedite the classification of suspected CSA material so that the sources of actual CSA material can be detected, as opposed to the distributors of files that are named as CSA material but in fact contains advertising for non-CSA porn sites. Several of the FIVES modules may be useful to automatically prepare and prioritize which material is viewed when content is examined to find actual CSA material, and the corresponding main distributors in the P2P network are localized.

### P2P monitoring

Automated monitoring of P2P networks also creates a need to be observant to the technicalities involved. Large scale monitoring of P2P networks have been performed by various entities such as MediaDefender [10] in order to protect the rights of holders of immaterial rights, typically related to commercially produced music and movies. However, recent research has shown that it is fairly easy to implicate any arbitrary IP-address as performing illegal activities, tricking the naive monitoring approaches that are apparently used by some commercial media protection organizations and companies. This has been shown by Piatek et al [15] in their paper with the appropriate subtitle "Why my printer received a DMCA takedown notice". Obviously, to guarantee some kind of security against arbitrary accusations against private citizens it is of utmost importance that CSA monitoring do not use the severely flawed methodology used by some media monitoring companies. There are a number of aspects surrounding the legal aspects of monitoring that would need further examination and clarification, but these are outside the scope of this paper.

### P2P anonymization technologies

If one believes that free speech and free flow of information is a cornerstone in democratic governance it is also important to safeguard these rights, both for citizens in democratic countries, but especially for those in non-democratic countries. As a response to these concerns, and lately also as a response to flawed and aggressive monitoring and the resulting arbitrary accusations and legal procedures, various technologies have emerged to increase the privacy of users. There are a number of techniques that have been developed with the intent to increase the privacy both for general computer communications as well as specifically for P2P users. One such privacy-enhancing technology that is not specific to P2P is Tor [3]. Tor creates an overlay network using onion routing to create a number of layers such that an external observer cannot infer the sender and recipient of communication. While there exists various way to decrease the privacy provided by Tor, there are also continued development to improve the anonymity or usability aspects of this kind of anonymization services. One such example is provided in [4] which looks into how anonymity can be achieved without the reliance of a Public Key Infrastructure (PKI).

Privacy-enhancing technology can also be targeted specifically towards P2P as has been done with the Tarzan [11] approach, that uses address rewriting and a number of other techniques in a P2P context. This is an example of an approach that provides anonymity to the users of P2P networks, without placing

any restrictions on who will be allowed to share or access material. In addition to these approaches there are also approaches that use the concept of Friend-to-Friend (F2F) file-sharing. These approaches will not allow arbitrary users to share or access material as was the case for the previous techniques. Examples of such F2F sharing techniques are Herbivore [5], Turtle [6], and OneSwarm [16]. The idea of only sharing data with a set of known friends, or using reputation based-systems in general, is to increase privacy by decreasing the ability of monitoring nodes to listen in on the data exchange by posing as peers. Freenet [7] is something of a mix of the above approaches that can be configured to provide data forwarding through an overlay network of trusted friends, but where the data access is free for everyone thus providing an anonymous storage service.

An effort to automate the collection of CSA-related intelligence need to consider the emergence of these and other anonymization techniques, since they are likely to be more widely used as a response to the increased Internet surveillance that is put into place because of the perceived needs of the media industry, and the perceived need for intercepting communication for general security reasons. It is interesting to note that any monitoring for any reason is trivially circumvented by a determined adversary, such as a terrorist organization, by the use of one-time pad encryption which is proven uncrackable.

## 5 Conclusions

In this paper we have presented the FIVES project and various ways the functionality of FIVES can utilize or contribute to intelligence gathering relating to CSA exchanges in P2P networks. The current FIVES project is focused on processing seized material, and the discussion in this paper has shown that there are opportunities to integrate P2P derived information into FIVES modules to improve the processing. Furthermore, the possibility to utilize aspects of the FIVES tool-set to support powerful and automated P2P intelligence collection has been discussed, along with some reflections on P2P monitoring and anonymization.

## References

[1] FIVES: Foresic Image and Video Examination Support, http://fives.kau.se

[2] V. I. Levenshtein, Binary codes capable of correcting deletions, insertions, and reversals. Soviet Physics Doklady 10 (1966):707–710.

[3] R. Dingledine, N. Mathewson, and P. Syverson. Tor: the second-generation onion router. In Proceedings of the 13th conference on USENIX Security Symposium, San Diego, CA, 2004.

[4] S. Katti, D. Katabi, and K. Puchala. Slicing the onion: Anonymous routing without PKI. in Proceedings of the 4th ACM Workshop on Hot Topics in Networks (HotNets), 2005.

[5] E. G. Sirer, S. Goel, M. Robson, and D. Engin. Eluding carnivores: file sharing with strong anonymity. In Proceedings of the 11th workshop on ACM SIGOPS European workshop, 2004.

[6] B. C. Popescu, B. Crispo, and A. S. Tanenbaum. Safe and private data sharing with turtle: Friends team-up and beat the system. In Proc. of the 12th Cambridge Intl.Workshop on Security Protocols, 2004.

[7] I. Clarke, O. Sandberg, B. Wiley, and T. W. Hong. Freenet: a distributed anonymous information storage and retrieval system. In International workshop on Designing privacy enhancing technologies: design issues in anonymity and unobservability, Berkeley, California, 2001.

[8] The FCCU Forensic CD, http://www.lnx4n6.be/

[9] The Open Computer Forensics Architecture, http://ocfa.sourceforge.net/

[10] MediaDefender, http://www.mediadefender.com/

[11] M. J. Freedman and R. Morris. Tarzan: a peer-to-peer anonymizing network layer. In Proceedings of the 9th ACM conference on Computer and communications security, 2002

[12] ISIS: Protecting children in Online Social Networks, http://www.comp.lacs.ac.uk/isis/index.html

[13] INDECT. http://www.indect-project.eu

[14] MAPAP: Measurement and Analysis of P2P Activity Against Paedophile Content. http://antipaedo.lip6.fr/

[15] M. Piatek, T. Kohno, A. Krishnamurthy. Challenges and Directions for Monitoring P2P File Sharing Networks –or– Why My Printer Received a DMCA Takedown Notice, 3rd USENIX Workshop on Hot Topics in Security (HotSec '08), 2008

[16] OneSwarm: Privacy preserving peer-to-peer datasharing, http://oneswarm.cs.washington.edu/index.html

# Isis[1]: Protecting Children in Online Social Networks

Awais Rashid, Paul Rayson, Phil Greenwood, James Walkerdine
*Computing Department, Lancaster University, UK*
*{awais | greenwop | paul | walkerdi} @comp.lancs.ac.uk*
Penny Duquenoy, Patrick Watson
*Middlesex University, UK*
*P.Duquenoy@mdx.ac.uk, P.Watson@mdx.ac.uk*
Margaret Brennan
*Child Exploitation and Online Protection Centre, UK*
*maggie.brennan@ceop.gov.uk*
Matt Jones
*Swansea University, UK*
*mattjonez@gmail.com*

**Abstract**

The aim of the Isis project is to develop an ethics-centred monitoring framework and tools for supporting law enforcement agencies in policing online social networks for the purpose of protecting children. The project is developing natural language analysis techniques to help identify child sex offenders from chat logs and monitoring mechanisms that can be non-invasively attached to file sharing systems for identifying the distributors of child abuse media. The ethical issues associated with such monitoring activities are studied through consultations with representatives from stakeholder communities and fed back into the development of the framework and tools. The project results are to be used and evaluated by specialist international law enforcement agencies as part of their own policing activities.

## 1. Introduction and Overview

Recent years have seen a rapid rise in the number and use of online social networks. Such social networks vary in nature from chat systems, for example, MSN, Skype and IRC, to online communities, such as, MySpace and YouTube, through to file sharing systems, for instance, peer-to-peer networks: Gnutella, BitTorrent, FastTrack, etc. Amongst the many types of 'risk' on the internet as mentioned in the Byron review in the UK [1] and Internet Safety Technical Task Force in the US [2], these social networks pose two significant risks in terms of child exploitation. The first major type of risk is *paedophiles and other child sex offenders predating on children*. Children actively participate in social interactions using forums such as chat rooms and web-based communities. Offenders can use such forums to predate on children, or even to plan the commission of sexual offences against children. These concerns are reflected by the formation of the Virtual Global Taskforce and specialist UK enforcement agencies and Scottish legislation to criminalise the 'grooming' of children in chat rooms in October 2004. The second risk is the offence of *distributing and sharing child abuse media*. Child sex offenders can formulate their own social networks using mechanisms, such as file-sharing systems, in order to distribute and share child abuse media. The scale of distribution of illegal media (including child abuse media) on such file-sharing systems was highlighted by a recent study at Lancaster University [3], which found that 1.6% of searches and 2.4% of responses on the Gnutella peer-to-peer network relate to illegal sexual content. Given the system's scale, these results suggest that, on the Gnutella network alone, hundreds of searches for illegal images occur each second. The study also found that, of those users sharing illegal sexual content, 57% were solely devoted to such distribution while half of the material shared by another 17% involved such content.

[1] This research is supported by a UK research grant from EPSRC/ESRC (reference EP/F035438/1) involving the Universities of Lancaster, Middlesex and Swansea. For further details, see the project website (http://www.comp.lancs.ac.uk/isis/)

Given the vast amount of information that is communicated within online social networks, new monitoring and analysis technologies need to be developed in order to tackle the growing problem of child grooming and the distribution of child abuse media. The development of such technologies faces three significant research challenges:

1. *How to identify active child sex offenders across online communities?*
   Paedophiles and other child sex offenders often masquerade as children in order to establish contact with potential victims and gain their trust. Distinguishing the "innocent" interaction amongst children or amongst children and adults from such predatory advances is a non-trivial task yet effective, early and accurate identification of sexual offenders is vital for the protection of children. At the same time such offenders may use multiple online identities and known child sex offenders may move to other online social networks upon detection in one network. It is, therefore, vital that once a suspected child sex offender is detected in one network, s/he can be successfully detected in other networks which s/he may attempt to employ for grooming children.
2. *How to identify the core distributors of child abuse media?*
   The key research challenge is to accurately identify child abuse media from the plethora of perfectly legal material that exists within file sharing systems. The problem is compounded by the fact that offenders often use specialised vocabulary to describe their shared media—a vocabulary that evolves and changes over time—and operate over different file sharing networks. Any monitoring framework must be non-invasively attachable to existing file sharing systems given the wealth of such systems and clients available today. In addition to identifying child abuse media within such systems, any monitoring tools must be able to distinguish core distributors of such media from mere users. This is essential for child protection as this would help law enforcement agencies in tackling the problem at its roots.
3. *How to ensure that such developments maintain ethical practices?*
   The development of such monitoring and analysis techniques raises a number of ethical challenges pertaining, on the one hand, to utilising the framework and tools in a beneficial way for child protection and, on the other hand, the need to protect innocent users of online social networks from the potential of falsely being identified as child sex offenders and safeguarding their privacy.

Isis is aiming to tackle the above three challenges by developing novel chat log analysis and non-invasive file sharing monitoring techniques based on natural language processing and aspect-oriented programming [4] practices respectively. The resulting framework and tools will assist law enforcement agencies while ensuring that they fall within current ethical bounds—note our goal is not automation but to provide support for detecting potential sexual offences through analysis of large amounts of data which cannot manually be analysed in an efficient manner.

## 2. Challenges Tackled by the Isis Approach

Existing work on policing online social networks has focused primarily on the monitoring of chat and file sharing systems. Chat policing software for home use such as Spector Pro[2], Crisp[3] and SpyAgent[4] allow the logging of online conversations, but are restricted in that they need to be installed on the actual PC that is partaking in the activity. Less obtrusive chat policing systems, as used by policing organisations, typically use a network-level tracing methodology [5] to identify and log chat traffic at the network-level for later analysis.

[2] www.spectorsoft.com
[3] www.protectingeachother.com
[4] www.spytech-web.com

However, police surveillance tactics deployed at network-level present real challenges to law enforcement in terms of detecting edge-based criminal activity and achieving effective online guardianship [6]. Three significant shortcomings can be observed. Firstly, too much data is produced to make pro-active analysis practical. Secondly, child sex offenders often masquerade as children in order to make contact, making detection difficult. Thirdly, systems tend to be developed to monitor a predator stereotype (adult male) which does not reflect patterns of internet based sexual predation of children and young people [7]. For example, Finkelhor [8] found that young people themselves make aggressive sexual solicitations in almost half of all cases and that of those known to be adults (25%), the majority are aged between 18-25. In 27% of cases in this study (conducted in the US) the age of predators was unknown and could well include adults masquerading as young people. A key question yet to be addressed is how to distinguish both between adults predating as young people and between 'normal' youth sexual behaviour on the internet and youth predation. Due to these challenges, policing organisations focus primarily upon reactive policing, wherein known culprits are identified and tracked and children are provided with mechanisms to report suspicious behaviour. Unfortunately, this approach is incapable of tackling many cases where children do not report incidents (in [8] only 3% reported) and where offenders may be unknown to the authorities. Moreover, these policing tactics do little to advance a preventive approach to the problem of online grooming and predation within social networks by enabling effective guardianship and the potential for law enforcement intervention in pre-criminal situations, e.g., at the point of an early "friendly" online encounter between a prospective offender and a child (see [6] for a discussion of the significance of offender search, pre-criminal situations, opportunity and other contextual factors in the prevention of Internet crimes against children).

In terms of language monitoring capabilities the existing chat policing software tools rely on human monitoring of logs or simple-minded keyword or phrase detection based on user-defined lists. Such techniques do not scale. Nor do they enable identification of adults masquerading as children or support pro-active policing. Techniques do exist which make use of statistical methods from computational linguistics and corpus-based natural language processing to explore differences in language vocabulary and style related to age of the speaker or writer. The existing methodologies, such as key word profiling [9], draw on large bodies of naturally occurring language data known as *corpora* (sing. *corpus*). These techniques already have high accuracy and are robust across a number of domains (topics) and registers (spoken and written language) but have not been applied until now to uncover deliberate deception. The second relevant set of techniques is that of authorship attribution. The current methods [10] would allow a narrowing in focus from the text to the individual writer in order to generate a stylistic fingerprint for authors.

For policing file sharing systems two significant tools exist, Peer Precision[5] and LogP2P[6]. Both systems also use a network-level tracing methodology in conjunction with a 'honey-pot' approach, wherein the policing peer offers an illegal file to the network and when an offender attempts to download this file, client-side software captures the offender's IP address at the packet-level. This approach suffers from two significant shortcomings. Firstly, it is unable to differentiate between those who download and share a single file, and those who are the 'core' distributors of child abuse media (e.g. distributing many thousands of files, producing and distributing child abuse imagery or uploading newly-produced child abuse material for the first time). This is a significant problem for frequently backlogged child protection agencies with limited resources. Secondly, as these systems work at the network-level, they can potentially be thwarted by encryption at the application-level. This is of particular significance as recent research has shown that users are migrating to more anonymous and secure file sharing systems [3]. Finally, and perhaps most critically, the honey-pot approach

[5] www.icactraining.org/P2P.htm
[6] aidounix.com/?LogP2P

relies upon the use of well-known files. Hence, it is incapable of identifying those offenders who may be sharing recently-produced material. The incorporation of monitoring functionality in file sharing systems requires significantly altering multiple components to ensure that monitoring takes place at the right points in the system. However, such invasive changes are expensive and hard to maintain and evolve across various releases of a system. The recent rise of aspect-oriented software development techniques [4] has facilitated non-invasive composition of such systemic concerns as monitoring, which makes in-step evolution of such functionality with changes in the rest of the system more modular and manageable. Though aspect-oriented techniques have been used in individual systems (e.g., the widely used mysql database system) for logging purposes, to date, they have neither been applied for monitoring online social networks nor on a scale spanning multiple systems and various releases of such systems. A particular issue in file sharing systems is that filenames reflect specialised vocabulary which changes over time [11].

Taking an ethical perspective on this research is not only important in respect of the abuse of children [12], and the protection of the researchers who conduct this type of research, but also because of the issues surrounding the use of monitoring technologies that have an impact on user privacy [13]. Researchers in the field of computer ethics have noted that values are embedded within technology design, e.g., [14, 15], and, as a result, there have been numerous calls for the integration of ethical assessment, evaluation and stakeholder impact analysis within the design and development of computer systems to mitigate adverse effects, e.g., [16, 17]. In advocating this approach there is a recognition that not only are the potential risks associated with the software development reduced [18], but also that the awareness of the development team to the ethical aspects inherent in these systems is raised – thus creating a body of 'ethically aware' information professionals [6]. To date there has been a lack of suitable case studies in the computer ethics literature and appropriate guidance for technology developers to incorporate ethical considerations within the development cycle. This project is developing new understandings of user-centred methods for highly sensitive systems and of effective designs of privacy/awareness interfaces that will benefit other developments and mitigate the effects of adverse outcomes that impact on public acceptability.

## References

[1] T. Byron, "Safer children in a digital world: the report of the Byron review", http://www.dcsf.gov.uk/byronreview/, 2008.

[2] J. Palfrey, "Enhancing child safety & online technologies: final report of the Internet safety technical task force", Berkman Center, Havard University, 2008.

[3] D. Hughes, S. Gibson, J. Walkerdine, G. Coulson, "Is deviant behaviour the norm on P2P file sharing networks?" *IEEE Distributed Systems Online* 7(2), 2006.

[4] R. Filman, T. Elrad, S. Clarke, M. Aksit (eds.), "Aspect-Oriented Software Development", Addison-Wesley, 2001

[5] D. Hughes, J. Walkerdine, K. Lee, "Monitoring challenges and approaches for P2P file sharing systems", Proc. 1st International Conference on Internet Surveillance and Protection (ICISP'06), 2006.

[6] M. Taylor, E. Quayle. "The Internet and Abuse Images of Children: Search, Pre-criminal Situations and Opportunity" *Situational Prevention of Child Sexual Abuse*, R. Wortley, S. Smallbone (eds.), Criminal Justice Press, 2006. pp. 169-195.

[7] S. Dombrowski, K. Gischlar, T. Durst, "Safeguarding young people from cyber pornography and cyber sexual predation: a major dilemma of the Internet". *Child Abuse Review* 16, 2007, pp. 153-170.

[8] D. Finkelhor, K. Mitchell, J. Wolak, "Online Victimization: A Report on the Nation's Youth", National Center for Missing and Exploited Children, Alexandria, VA, 2000.

[9] P. Rayson (2008). From key words to key semantic domains. *International Journal of Corpus Linguistics*. 13:4 pp. 519-549.

[10] P. Juola, J. Sofko, P. Brennan, "A prototype for authorship attribution studies", *Literary and Linguistic Computing* 21, 2006, pp. 169-178.
[11] D. Hughes, P. Rayson, J. Walkerdine, K. Lee, P. Greenwood, A. Rashid, C. May-Chahal, C., M. Brennan (2008) Supporting law enforcement in digital communities through natural language analysis. In *proceedings of the 2nd International Workshop on Computational Forensics (IWCF 2008)*, Washington DC, USA, August 7-8, 2008. Lecture Notes in Computer Science 5158, pp. 122-134.
[12] M. Eneman, "The new face of child pornography", in *Human Rights in the Digital Age*, Cavendish Publishing, 2005.
[13] D. J. Cook, S. K. Das, "How smart are our environments? An updated look at the state of the art", *Pervasive and Mobile Computing* 3(2), 2007, pp.53-73.
[14] H. Nissenbaum, "Values in the design of computer systems", *Computers and Society*, March, 1998.
[15] J. van den Hoven, "ICT and value sensitive design", in *The Information Society: Innovation, Legitimacy, Ethics and Democracy*, P. Duquenoy, P. Goujon, K. Kimppa, S. Lavelle (eds.), Springer, 2007.
[16] P. Duquenoy, O. Burmeister. "Exploring ethical aspects of Pervasive Computing" in *Risk Assessment and Management in Pervasive Computing: Operational, Legal, Ethical and Financial Perspectives*, Varuna Godara (Ed.), IGI Global. 2008. pp.264-284.
[17] D. H. Gleason, "A software development solution", *Proc. Ethicomp: Systems of the Information Society*, Poland, 2001.
[18] D. Gotterbarn, "Reducing software failures: Addressing the ethical risks of the software development lifecycle", Australian Journal of Information Systems, 1999.
[19] P. Duquenoy, D. Whitehouse, "A 21st century ethical debate: Pursuing perspectives on Ambient Intelligence", *Proc. Landscapes of ICT and Social Accountability*, Finland, 2005.


Centre for Abuse & Trauma Studies

# Internet Child Abuse: Understanding Offender Online Grooming Behaviour

**Professor Julia Davidson, PhD(ECON)**
**Director of Research in Criminology,**
**Co-Director Centre for Abuse & Trauma Studies**
**(Kingston University & Royal Holloway, University of London)**

**Paper prepared for : *'Advances in the Analysis of Online Paedophile Activity'* Paris, France 2009, June, 2 - 3**

## Young People's use of the Internet

Internet use has grown considerably in the last decade. Information technology now forms a core part of the formal education system in many countries, ensuring that each new generation of Internet users is more adept than the last. Research studies in the UK suggest that the majority of young people aged 9-19 accessed the Internet at least once a day. The Internet provides the opportunity to interact with friends on social networking sites such as MySpace, Facebook and Bebo and enables young people to access information in a way that previous generations would not have thought possible. The medium also allows users to post detailed personal information, which may be accessed by any site visitor and provides a platform for peer communication hitherto unknown. The majority of children (65%) in Davidson and Martellozzo's (2005) study had access to at least one computer at home, 49% had computers in their bedrooms. Other children did not have a computer at home, but had access to a computer at relatives' or friends' houses, 15% used Internet cafés on a regular basis (more than once a week). The findings suggest that almost all of the children had access to the Internet outside school. 60% accessed the Internet more than four times per week, this was particularly the case for the 12 to 14 age group. Of those children accessing the Internet, 76% were largely unsupervised and spent long

periods of time on their computer particularly during school holidays and at weekends. Generally, the children had a great deal of knowledge about computing, and the majority of 12-14 year olds were extremely confident Internet users.

Individuals join virtual communities, where they meet other persons who have the same interest. A virtual community provides an online meeting place where people with similar interests can communicate and find useful information. Communication between members may be via e-mail, bulletin boards, online chat, web based conferencing, or other computer-based media. As a business model, a virtual community can make money from membership fees, direct sales of goods and services, advertising, click-through, and sales commissions (Gottschalk, 2006).

Vidnes and Jacobsen (2008) surveyed 772 persons from 16 to 29 years of age. One of the key findings was that especially the youngest ones are active users of web cameras, and that young people who use web cameras have certain characteristics. They are more socially active on the Net than others, and more interested to get to know new people. 53 percent of youngsters between 16 and 19 years that use web cameras have come in contact with people on the Net that they later on have met outside the Net. Only 25 percent of those who do not use web camera report the same. This shows that web camera users to a greater extent than others expand their social network on and outside the Net. At the same time it is evident that web camera users to a far greater extent are involved in activities that are perceived as potentially dangerous.

However, most young web camera users communicate mainly with persons that they know already. When getting to know a new person, they seldom start by using their web cameras. Rather, they start using Facebook or Nettby (Net village), which are characterized by openness in the sense that individuals present themselves by real pictures, name, interests and friends. These net societies function as a virtual, social gathering place, where one can be introduced to each other and move on, get contacts and expand networks (Vidnes and Jacobsen, 2008).

The study by Vidnes and Jacobsen emphasized the following findings (Thorgrimsen, 2008):

- 7 out of 10 youngsters between 16 and 19 years old have access to web camera on their computer, 48 percent of them use it at least once a month.
- Almost all youngsters are using MSN on a daily basis, but very few use their camera daily.
- First and foremost camera is used with persons that one seldom meets.
- Web camera is far less used than other communication tools such as net community and MSN.
- Young people are careful in providing identifiable information before feeling safe in new net relationships.
- The web camera also works as a control mechanism, as one cannot be sure who is on the line before the camera is turned on. Resistance to turn it on is interpreted as the other person has something to hide.

If young people want to get to know each other better, then they may move into more private tools such as MSN, which intensifies the communication. If the relationship is developed further, then the private arena of web cameras emerges. By choosing different tools for different relationships and for different phases in a relationship, Vidnes and Jacobsen (2008) thus found that young people are able to regulate the degree of intimacy in the relationship.

Myspace and other social networking sites like it offer thriving communities where young people engage in countless hours of photo sharing. In addition to Myspace, other social networking and blogging sites such as Friendster.com, Facebook.com and MyYearbook.com allow users to post pictures, videos, and blogs, and they support email and instant messaging. Myspace and Facebook differ in that Myspace is open to anyone, and has loose age restrictions, while Facebook users are encouraged and often required to register using their real name (Kierkegaard, 2008).

## Sex Offender Online Behaviour

There is, however, increasing evidence that the Internet is used by some adults to access children and young people in order to groom them for the purposes of sexual abuse. MySpace have recently expelled 29,000 suspected sex offenders and are being sued in the United States by parents who claim that their children were contacted by sex offenders on the site and consequently abused. The Internet also plays a role in facilitating the production and distribution of indecent illegal images of children, which may encourage and complement online grooming.

Davidson and Martellozzo (2008: 277) suggest that Internet sex offender behaviour can include: "the construction of sites to be used for the exchange of information, experiences, and indecent images of children; the organization of criminal activities that seek to use children for prostitution purposes and that produce indecent images of children at a professional level; the organization of criminal activities that promote sexual tourism". Child grooming is a process that commences with sexual predators choosing a target area that is likely to attract children. In the physical world, this could be venues visited by children such as schools, shopping malls or playgrounds. A process of grooming then commences when offenders take a particular interest in the child and make them feel special with the intention of forming a bond. The Internet has greatly facilitated this process in the virtual world. Offenders can now seek out their victims via online games and social networking sites. According to Wolak et al. (2008), most Internet-initiated sex crimes involve adult men who use the Internet to meet and seduce underage adolescents into sexual encounters. The offenders use Internet communications such as instant messages, e-mail, and chat rooms to organize meetings and develop intimate relationships with victims.

Child sex offenders are forming online communities and bonds using the Internet. They are openly uniting against legal authorities and discussing ways to influence public thinking and legislation on child exploitation. While paedophile web sites are being tracked down and removed from Internet servers in countries all over the world, they are popping up again at a higher pace in most parts of the world.

An example of a web site representing a virtual community for paedophiles is "Boylove". On the web site, The Boylove Manifesto could be found, which argued the case for intergenerational relationships (www.prevent-abuse-now.com):

> As boylovers we distance ourselves from the current discussion about "child sexual abuse". Human sexuality plays the same part in a boylove relationship as it undoubtedly does in any relationship between human beings. A boylover desires a friendly and close relationship with a boy.

One of the key problems in policing the global crime is the variation in legislation between countries. The concept of grooming is now recognized in some legislation. The Protection of Children and Prevention of Sexual Offences (Scotland) Act 2005 includes "meeting a child following certain preliminary contact". Where a person arranges to meet a child who is under 18, having communicated with the child on at least one previous occasion (in person, via the Internet, or via other communication technologies), with the intention of performing sexual activity on the child, is an act of grooming.This new offence category was also included in the Sexual Offences Act 2003 in England and Wales (this section of the Act also applies to Northern Ireland). Section 15 makes "meeting a child following sexual grooming" an offence. This applies to the Internet, other technologies such as mobile phones and to other communication forms. Grooming involves here a process of socialization during which an offender seeks to interact with a child (young person under 18 in Scotland, England and Wales), possibly sharing their hobbies and interests in an attempt to gain trust in order to prepare them for sexual abuse. The process may also involve an attempt to normalize sexual relations between adults and children.

In the United States it is an offence to electronically transmit information about a child aged 16 or under, for the purpose of committing a sexual offence (US Code Title 18, Part 1, Chapter 117, AS 2425). The Australian Criminal Code (s218A) makes similar restrictions, as does the Canadian Criminal Code (s172.1). The legislation in Scotland, England and Wales differs in that the sexual grooming offence applies both to the Internet and to the "real world". Legislation in other countries addresses only electronic grooming via the

Internet and mobile phones. In reality it would be extremely difficult to police and evidence grooming behaviour in the "real world". It is therefore unsurprising that few cases have been brought to court on this basis under the Protection of Children and Prevention of Sexual Offences (Scotland) Act 2005 and the Sexual Offences (England and Wales) Act 2003. Several other countries are beginning to follow the UK lead in legislating against grooming behaviour. For example, sexual grooming was added to the legislation in Norway in 2007.

Research suggests that sex offenders are likely to experience some form of abuse in childhood and have diffciulty in building and maintaining adult relationships (Smallbone and Dadds,1998; Ward and Keenan1999). It would seem that new information technologies have afforded some sex offenders the opportunity to maintain anonymity and it is probable that online relationships are more managable and easily maintained than offline relationships. Beech (2009) suggests that online sexual relationships can become particularly significant for individuals who have difficulty in building actual sexual relationships: *'Along with its ability to provide a level of perceived pseudo-sexual intimacy with children, the Internet also would appear to provide a social outlet for individuals who have difficulties initiating, and maintaining, relationships with other adults'* (Beech, 2009, p8).

What do we know about online grooming offenders use of the Internet and the role the Internet plays in their fantasy and offending behaviour? Laulik et al suggest that online sex offenders lack assertiveness and empathy in relationships, and demonstrate low levels of self-esteem (Laulik, Allam, & Sheridan, 2007), whilst Middleton et al (2006) note that such offenders are inadequate, have low self-esteem, and show little victim empathy (Middleton et al., 2006). . Recent research conducted by Hernandez (2009) with 150 convicted, incarcerated online offenders (indecent image collectors) in an attempt to explore the link between image collection and contact abuse,suggests that by the end of the treatment programme 131 (85%) of this group admitted that they had perpetrated a contact offence, the number of victims involved in this offending was 1,777 an average of 13.56 victims per

offender (SD= 30.11), polygraph testing was employed to validate the offenders self reports, but the validity of this method of validation has been questioned. Seto (2009) suggests however on the basis of his research with over 300 child pornography offenders that this group is unlikely to be convicted for contact sexual offences. 4% of 301 child pornography offenders were convicted for sexual abuse contact offences over a 3.9 year period. This study is however based upon reconviction data and not offender self report.

Preliminary findings from ongoing doctoral research employing documantary analaysis of online grooming offender case files and conducted with a Police High Technology Crime Unit in the UK (Martellozzo, 2009) suggests that:

1. The majority of sex offenders in the study began to satisfy their fantasies of abusing children by simply exploring cyberspace. In doing so they quickly discovered sites that would satisfy their imagination to the point of actively engaging in erotic dialogues with children and other offenders and downloading pornography.
2. The majority of groomers sent undercover police officers some type of adult pornography or child abuse images during the grooming process. Almost two-thirds of the subjects exposed themselves to the undercover officer via photograph or web-cam. However, some suspects used images as a grooming tool, most stated that they used them principally for sexual gratification.
3. Existing literature demonstrates that sex offenders have a tendency to deny responsibility for their offending and to attribute blame to either the victim or to circumstances (Gudjonnssen, 1988). Martellozzo's (2009) research with online offenders demonstrates that they tended to minimise their intention to sexually abuse a child. Several subjects claimed the communication was just 'fantasy' and that they thought they were talking to an adult. However a significant proportion of subjects who met the undercover police officers thinking they were meeting a 12 year old girl. Of these subjects only one declared his intention to abuse the child. The others were in denial even when found in possession of condoms and other child abuse items.Respondents

had a tendency to deny their interest in children and blame either their personal circumstance, boredom or the easy access provided by the Internet.

4. Finally Martellozzo's research demonstrated that the majority of the sex offenders viewed the Internet as a tool that can offer security and anonymity to the point that the risk of committing offences is minimised. Offenders did not blame the Internet but saw it as a safe environment where all forms of behaviour are acceptable, including child abuse.

(*Martellozzo, 2009)*

## Offender Case Studies

Three cases are described to illustrate online behaviour, these may not be typical but police officers acting undercover suggest that they are (provided by the Metropolitan Police).

**Case One:**

This case involved the 'grooming' of a 14 year-old girl from Canada by a 34 year old businessman from South West England. The offender and victim spent 6 months chatting online before the offender began sending money to the girl for the purchase phone cards so they could communicate also via text messages.

The girl's parents became suspicious and contacted the local police, who in turn contacted the London Metropolitan police, when:

- He offered to go to see her in Canada
- He was prepared to pay $500 to have sex with the girl
- He offered to pay more money in exchange for sex with the girl's 10-year-old friend
- He asked if she could post her underwear
- He claimed to be an 'abuser of children'

The suspect's behaviour was sufficient to constitute 'grooming', as defined in the Sexual Offences Act 2003, and as such justified formal police intervention. The investigation revealed that the offender had other previous criminal convictions ('kerb-crawling', soliciting a prostitute and for possession of cocaine). The suspect's computer was seized and analysed and it was found that previously he had met and sexually abused another child who he had groomed online.

Also, around 6000 images of children under the age of 16 were found in the suspect's computer, and many of these were indecent. Of the total collection of images, 4500 depicted children dressed in underwear or swimwear, and 1117 portrayed children being physically abused. It was further discovered that the offender regularly discussed the sexual abuse of children with other like-minded individuals by logging on to dedicated websites, including a number of different child love forums.

Furthermore, it emerged from interviews with the Canadian girl that the suspect had been sending her indecent images of other children. This represents another 'grooming' tactic whereby child abuse images are shown to the child to lower that child's inhibitions concerning sexual activity (Krone, T., 2005) and to make the child believe that it can be regarded as 'normal' to have sex with adults or to pose naked in front of a camera. The suspect then asked the girl to send him indecent pictures of herself. He also offered to pay the girl money to produce and send indecent images of one of her younger friends.

**Case Two**

The same 34 year-old businessman discussed in Case Study One arranged to meet one of the many victims he groomed online. The victim was a 15-year-old girl from the USA. She met the offender when she was in England on holiday with her mother and grandmother. The mother knew about the relationship between her daughter and the man, but did not at first suspect

any indecent or criminal intent. She admitted speaking many times on the phone with the man, and they even agreed to meet during their visit in England.

After having spent some time together and gaining their trust, the man persuaded the mother to let him look after her daughter and promised he would bring her back to the hotel. The mother accepted on the condition he would have her back by 11 p.m. This process of ongoing interaction and negotiation further evidences the offender's manipulative capacity to groom not only the child, but also the parents (Finkelhor et al, 2000).

It is alleged that he then invited the girl back to his house and took her to his bedroom, where he performed oral sex on her. Much later, the mother noticed love bites on the girl's body and took her immediately to the doctor, where it was discovered that she had contracted a sexual transmitted infection.

The suspect was then arrested and interviewed. During the interview with the police, the suspect claimed that he and the girl were very close and that, at the time of the investigation, he still felt close to her. The suspect admitted his 'unusual' close relationship with the girl, but attempted to transfer blame to the mother. The officer in charge of the case claimed that victim appeared to be naive and easily led (Interview with Officer, Metropolitan Police HTCU, 2nd May 2005).

The forensic analyst in charge of the case described the suspect as leading a double life. On the one hand, he was involved with charities and participated in awareness campaigns for child amputees. On the other hand, they served as a means of increasing his access to disabled and/or vulnerable children for the purposes of abuse.

**CONVICTION**

The offender was subsequently charged with grooming, producing and distributing indecent images of children under the Sex Offences Act (2003), and was sentenced to two years in prison. He pleaded not guilty to sexually

assaulting the young girl, and there was insufficient evidence to secure a conviction for this alleged offence.

**Case Three:**

This case relates to a 55-year-old man who committed a series of serious sexual assaults against boys between the ages of 12 and 16 years. The offences were committed both in the UK and Ghana.

Briefly, the offender befriended a 12-year-old boy in the UK while working in a toy model shop. After a period of grooming he began sexually abusing the boy, having convinced him that this sexual activity was normal.

The offender travelled also to Ghana where his eldest son had been working for the Voluntary Services Overseas as a teacher. Due to the extreme poverty in this country the offender was able to groom young boys with gifts and money in order to fulfil his sexual fantasies.

Following complaints by local people, authorities in Ghana interviewed the suspect about his activities. However, they did not have the technological facilities to analyse his computer and cameras, and he was released. British authorities were notified of his imminent return to the UK. Having been stopped at Heathrow Airport, the suspect was found to be in possession of hundreds of indecent images of children. He was arrested immediately by Customs officers, interviewed, charged with the importation of indecent images, and remanded in custody. Police were then notified that the material seized was likely to contain evidence of hands-on abuse. Officers from the London Metropolitan Police Paedophile Unit commenced an investigation, which revealed that the man had:

- Sexually abused a number of boys under the age of 16 in Ghana
- Sexually abused a boy under the age of 13 in the UK
- Abused boys in the UK both in a converted van and an office
- Recorded his sexual abuse of the victims
- Downloaded a quantity of indecent images of children from the Internet
- Made indecent pseudo images of children

In June 2005 the offender was arrested and charged with:

- Rape of a child under 13 years
- The sexual penetration of a child
- Distribution of Indecent Images of Children

When he was interviewed by the police he gave a full admission of having sex with other boys in Africa. He also admitted downloading indecent images of children from the Internet and also made a number of pseudo-images.

**Conviction**

The offender was successfully convicted of rape and sexual assault of young boys both in the United Kingdom and in Africa. He was sentenced to an indeterminate Prison Sentence in January 2006 under the new Sex Offences Act 2003 (Section 72).

*Adapted from Davidson & Martellozzo, 2008, p10-15*

Analysis of recent offender case files suggests that offending behaviour is fluid and that where offenders are using the Internet anyway, their offending is varied and not necessarily focused on one form of behaviour eg. exclusively online grooming.

## Moves to Protect Children Online

Despite moves on the part of law enforcement agencies, governments, the IT industry and organizations such as VGT and IWF to control online abuse, John Carr of the NCH in the UK suggests in a recent report (2006) that such efforts are largely failing as the number of indecent images of children on the Internet continues to increase and the images become ever more disturbing, involving a greater degree of violence and increasingly younger children. It is suggested that governments are failing to make the growing trade in indecent images of children a high enough political priority and that, as the title of the report suggests '*Out of Sight, Out of Mind*' (2006) the hidden nature of the offending and lack of public awareness makes this possible. Indeed other recent research conducted in the UK suggests that child victimization and protection issues are not a high priority for the criminal justice agencies involved in the investigative process (Davidson, Bifulco, Thomas & Ramsay, 2006, in print). Carr advocates a global initiative, and key areas are identified where action should be taken. Carr is correct in suggesting that the key issue is one of effective leadership, and that a '*global leadership mechanism'* (p1) should be developed. This mechanism, it is suggested, should take the form of a new NGO or a network that draws upon existing NGOs. This central body would act to scrutinize and advice governments, law enforcement agencies and the industry. It would also provide a hitherto absent degree of IT industry public accountability. This is undeniably an essential move as at present attempts to protect children online are ad-hoc and some international police forces have only just begun to recognize the scale of the problem (International Police Executive Symposium Conference, 2006). The difficulty will be in setting up a central mechanism that is really able to scrutinize international approaches to the problem and that will have the power to intervene effectively where there is inaction or indifference.

## Paper and Other Useful References

Ashenden, S. (2004) *Governing Child Sexual Abuse. Negotiating the Boundaries of Public and Private, Law and Scienc.* London & New York Routledge

Beckett, C. (2003) *Child Protection. An Introduction.* Sage Publications

Cobley, C (2005) '*The legislative framework'* in Matravers, A (Ed) (2005) ' Sex Offenders in the community: managing and reducing the risks' Cambridge: Willan Publishing

Carr, J. (2006) '*Out of sight, out of mind: Tackling child sexual abuse images on the Internet- a global challenge*' NCH The Children's Charity

Castells, M. (1996) The Network Society, ????

Castells, M. (2004) The Power of Identity, Oxford: Oxford University Press.

Cobley, C (2005) '*The legislative framework'* in Matravers, A (Ed) (2005) ' Sex Offenders in the community: managing and reducing the risks' Cambridge: Willan Publishing

Corby, B. (2006) *Child Abuse. Towards a Knowledge Base.* Open University Press.

Craig, L. , Browne, K. & Beech, A (2004) *'Identifying Sexual and Violent re-Offenders'* British Psychological Society Conference, Division of Forensic Psychology, 22nd March 2004, Leicester University.

Davidson, J. (2004) ' *Child Sexual Abuse Prevention Programmes: The Role Of Schools'* In '*Sex Offending Is Everybodies Business*' (Giotakos, O., Eher, R & Pfafflin, F (Eds). 8th International Conference of the International Association for the Treatment of Sexual Offenders, 6-9 October 2004: Pabst: Lengerich

Davidson, J. & Martellozzo, E. (2005) *'The Internet And Protecting Children From Sex Offenders Online: When Strangers Become 'Virtual Friends'* http://www.oii.ox.ac.uk/research/cybersafety/extensions/pdfs/papers

Davidson, J. and Martellozzo, E (2004) '*Educating children about sexual abuse and evaluating the Metropolitan Police Safer Surfing Programme'* *http://www.saferschoolpartnerships.org/ssp-topics/evaluations/documents/ssfindingsreport.pdf*

Davidson, J. (2006) *'Victims Speak: Comparing Child Sexual Abusers and Child Victims Accounts, Perceptions and Interpretations of Sexual Abuse'* Victims and Offenders. Vol 1, No 2, 159-174

Davidson J. Bifulco, A., Thomas G., and Ramsay, M. (2006) '*Child Victims Of Sexual Abuse: Children's Experience Of The investigative Process In The Criminal Justice System*' Practice Journal, in print. Taylor Francis

Davidson, J. (3/2008) '*Child Sexual Abuse, Media Representation and Government Reactions'* (Series Eds D. Downes and P. Rock, London School of Economics) Routledge. ISBN: 978-1-90438569-1 (4th book in social policy series)

Davidson, J and Gottschalk, P '*Online Groomers Profiling, Policing and Prevention'*
Russell House (forthcoming 12/2009)

Davidson, J and Gottschalk, P (Eds) '*Internet Child Abuse: Current Research, Policy & Police Practice*' Routledge (forthcoming 10/2010)

Finkelhor, D. (1984) *'Child sexual abuse: New theory and research'* New York: Free Press

Finkelhor, D (1984) 'Four conditions: *A model" Child Sexual Abuse: New Theories and Research.*'New York: The Free Press

Finkelhor, D.; Kimberly, J. ; Wolak, J. (2000) *On Line Victimisation: a report on the Nation's Youth.* Alexandria, Virginia: National Cemtre for Missing & Exploited Children.

Gallagher, B., Fraser, C., Christmann, K., Hodgson, B., (2006) *'International and Internet Child Sexual Abuse and Exploitation'* Centre of Applied Childhood Studies. University of Huddersfield

Gillan, A. (2003)*'Race to save new victims of child pornography'* Guardian Newspaper November 4, 2003

Goldson, B et al. (eds) (2002) *Children, Welfare and the State.* Sage Publications.

Gottschalk, P & Davidson, J (2009) *'Digital forensics in law enforcement: The case of online victimization of children'* Electronic Government: An International Journal

Greer, C., Jewkes, Y. (2005) *Extremes of Otherness: Media Images of Social Exclusion.* Social Justice (Special Edition on Emerging Imaginaries of regulation, Control and Oppression), 2005, 32, 1:20-31

Home Office Task Force on child protection on the Internet (2003) '*Good practice Models and Guidance for the Internet Industry on: Chat services; Instant Messages; Web Based Services*'

ICAC (National Centre for Missing & Exploited children and Boys and Girls). NetSmart Presentation. 2004 University of New Hampshire.

Jewkes, Y. (ed.) (2003a) '*Dot.cons.: Crime, Deviance and Identity on the Internet'*. Cullompton: Willian.

Jewkes, Y. (2003b) `*Policing the net: crime, regulation and surveillance in cyberspace'* in Y. Jewkes (ed.)Dot.cons.: Crime, Deviance and Identity on the Internet. Cullompton: Willian, 15-35

Krone, T. (2004) '*A typology of online child pornography offending'* Trends and Issues in Crime and Criminal Justice, No 279. Canberra: Australian Institute Of Criminology

Krone, T. (2005) *Combating Online Child Pornography in Australia* in Quayle, E. & Taylor, M. *Viewing Child Pornography on the Internet. Understanding the Offence, managing the Offender, Helping the Victims.* Russell House Publishing.

Livingstone, S. & Bober, M. (2004) '*UK Children Go Online: Surveying the experiences of young people and their parents*'. (LSE, 2004)

Livingstone, S & Bober, M. (2005) '*Internet Literacy Among Children and Young People* '(LSE, 2005)

Martellozzo, E. (2004) *'Child Pornography on the Internet: Police Strategies'* In 'Sex Offending Is Everybodies Business' (Giotakos, O., Eher, R & Pfafflin, F (Eds). 8th International Conference of the International Association for the Treatment of Sexual Offenders, 6-9 October 2004: Pabst: Lengerich

National Offender Management and the Scottish Executive (2005) *Consultation Possession Of Extreme Pornographic Material.*

Middleton, D., Elliott, I.A., Manderville-Norden, R., Beech, A.R. (2006) *'An Investigation into the Applicability of the Ward and Siegert Pathways Model of Child Sexual Abuse with Internet Sex Offenders'* Psychology, Crime and Law, in press.

O'Connell, R (2003) '*A Typology of Child Cybersexpolitation and Online Grooming Practices'*

Pritchard, C. (2004) *The Child Abusers. Research and Controversy.* Open University Press.

Quayle , E & Taylor, M. (2001) *'Child Seduction and Self-Representation on the Internet'* Cyberpsychology and behaviour, 4, 5: 597-607

Quayle , E & Taylor, M (2002) *'Paedophiles, Pornography and the Internet: Assessment Issues'* British Journal of Social Work, 32: 863-75

Quayle, E. & Taylor, M. (2003) *'Model of Problematic Internet Use in People with a Sexual Interest in Children'* Cyberpsychology and behaviour, 6, 1. 93-106

Robbins, P. & Darlington R. (2003) *"The Role of the Industry and the Internet Watch Foundation"* In 'Policing Paedophiles on the Internet', MacVean & Spindler (Eds.). New police Bookshop.

Thomas, T. (2000) *Sex Crime: Sex Offending and Society.*Willan Publishing

Wyre, R (2003) '*No excuse for child porn'* Community Care 1489: 38-40

# Child pornography: the exploitation and abuse of children.

**Ethel Quayle and Terry Jones**

## Abstract

While there has been considerable investment in the detection of online sexual offending against children there has been little interest in the children portrayed in the images. This is reflected in the paucity of research in this area and the fact that so few children have ever been identified. This presentation examines the existing literature and presents new data to explore both the nature of the images and what they might tell us about the children who are both abused and exploited.

In recent years there has been a considerable investment in the detection of online sexual offending against children and the provision of specialized treatment facilities. In part this was in response to a dramatic increase in the number of such people entering the criminal justice system. For example, in the UK Middleton (2009, p 7) reports that, 'For England and Wales in 1999, there were 238 convictions for publication, possession or distribution of obscene matter and indecent photographs of children. By 2005 they had reached 1,296 (Hansard, 2008)*an almost 500% increase in convictions. The total sexual offence convictions (all sexual offences) in 2005 was 4,800 (Home Office, 2006). Therefore, in this year convictions for internet-related sexual offences accounted for almost one-third of all sexual offence convictions. This level of convictions had a significant effect on the proportion of sex offenders entering or waiting to commence treatment programmes, leading to questions of suitability of the treatment programme content, appropriate treatment dosage and possible ''contamination effects'' of exposure to contact child sex offenders'. The emergence of this population has also resulted in an increased body of research that explores the underlying aetiology and contexts associated with online sexual offences against children (for example, Elliott and Beech, 2009). There has, however, been little published literature that relates to the children within these images and interest has focused largely on populations that predate the Internet. This short paper seeks to address this by turning our focus back to child pornography as it relates to victimisation.

Online sexual offences against children are often associated with downloading, trading or the production of child pornography (Quayle, 2008) and while interest in child pornography is not new (Lanning, 2000), it is the case that with each technological advance we have seen an increase in the availability of such materials, and this has been most noticeable in relation to the advent of the Internet. There are historical accounts of child pornography and its distribution, which appeared to be facilitated by the popular use of photography (Taylor and Quayle, 2003). The criminalisation of such material, however, made access both difficult and dangerous, although there was a period of approximately ten years when in some European countries all pornographic materials were decriminalised. In part this appears to reflect an absence of concern with those involved in the production of the material (both children and adults) and more a concern with the consumer. In Denmark the anti-pornography laws were repealed in 1969 and Sweden then followed in 1971, and this resulted in a booming trade in child pornography, with the appearance of material in the media and in ordinary shops (Schuijer and Rossen, 1992). Such decriminalisation of child pornography and its subsequent ease of both production and distribution arguably resulted in the depathologising of some interests and behaviors, or at least a decoupling of some of our assumptions between the nature of sexual interest in children and the notion of sexual harm (Quayle, 2008). However, it is of interest that much of the commercial production of child pornography during this period involved images of children that were produced in a domestic context and sent on to magazine editors in exchange for

money.

Most of early studies, and virtually all legal documents, use the term child pornography, but more recently questions have been raised as to whether this term both reflects the content of what is produced, and whether it implicitly implies consensual activity (Taylor and Quayle, 2003). The term 'abusive images' is now widely used by those who advocate for children's rights in relation to sexual abuse through photography (Jones and Skogrand, 2005), but this change is not straightforward. The term child pornography is consistently used in the majority of laws and policy documents internationally (Akdeniz, 2008), and attempts to change terminology are thought by some to be both confusing and to not adequately capture the complex nature of the material (Lanning, 2008). This is worth further consideration, as concerns about the language used are not simply a question of semantics and this is reflected in the language used in this paper.

Cassell and Cramer (2008) argue that throughout history there has been a recurring moral panic about the potential danger of communication technologies (particularly for young women) but that when investigated it is less the technology that appears to be to blame, but rather the potential sexual agency of young women, parental loss of control, and the 'specter of women who manifest technological prowess'. In a similar vein, the Internet Safety Technical Taskforce (2008) argue that although they are frequently reported in the media, US Internet sex crimes against minors have not overtaken the number of unmediated sex crimes against minors, nor have they contributed to a rise in such crimes. The report states that the increased popularity of the Internet in the United States has not been correlated with an overall increase in reported sexual offenses. Evidence is cited from the US that overall, sexual offenses against children have declined in the last 18 years (National Center for Missing and Exploited Children, 2006), with research indicating a dramatic reduction in reports of sexual offenses against children from 1992 to 2006 (Calpin 2006; Finkelhor and Jones, 2008). However, seemingly at odds with this, data from the FBI (2006) indicated that between 1996-2006 there was a 1,789% increase in the number of open cases, a 2,174% increase in arrests and summons, and a 1,397% increase in convictions for sex related crimes on the Internet. While the percentages are alarming, the figures from the FBI are largely meaningless without us knowing what they represent, but they do illustrate where the FBI's perception of the changing incidence of these crimes.

In contrast to the decline of reported contact sexual offences against children, public behaviour in relation to illegal or problematic Internet content has led to a substantial number of reports of child pornography. In 2006, CyberTipline (a US congressionally mandated system for reporting child crimes) received 62,365 reports of child pornography (National Center for Missing and Exploited Children, 2006). The 2007 Global Internet Trend Report of INHOPE (the International Association of Internet Hotlines) indicated that during the last quarter of 2006 the hotline network processed an average of 91,000 reports per month. Approximately 35,000 of these reports were received from the public and 19,000 were determined to refer to either illegal or harmful content. INHOPE determined that 9,600 reports related to child pornography and that this number was increasing at an average of 120 reports per month. It is unclear whether these figures represent an increasing volume of child pornography, or rather the actions of an increasingly concerned public. The reality is that we have no idea of the numbers of people who commit sexual offences related to child pornography on the Internet. We can examine conviction rates, but these reflect only the countries where possession and distribution of child pornography is both illegal and where there are either the resources or inclination to act upon detection (Quayle, 2008).

In the US, Wolak, Mitchell and Finkelhor (2003) reported that law enforcement made an estimated 2577 arrests during twelve months (starting July 1st 2000) for Internet sex crimes against minors. Two-thirds of offenders who committed any of these crimes possessed child pornography. Finkelhor and Ormrod (2004) examined child pornography patterns from the FBI's National Incident-Based Reporting System (NIBRS). The data from 1997-2000 on 2469 crime incidents involving pornography revealed that over these three years pornography offences increased by 68% and juvenile victim/child exploitation pornography offences increased 200%. But at the time of this report, only a small minority of all pornography offences known to the police was coded as involving a computer.

However, these statistics reflect only those who are caught. Other data, such as that provided

by one leading UK Internet Service Provider suggested that in July 2004 they blocked more than 20,000 attempts per day to access child pornography on the Internet. More recent data from the Swedish and Norwegian blocking of access to known sites carrying child abusive images reveal as many as 15,000 – 18,000 daily attempts in Norway (Quayle, Lööf and Palmer, 2008). Such use of the WWW as a possible means of accessing child pornography was investigated by Demetriou and Silke (2003) who established a website to examine whether people, who visited for the purposes of gaining access to legal material, would also attempt to access illegal or pornographic material if it was offered. Over an 88-day period, 803 visitors entered the site and it was found that the majority of visitors accessed those sections purporting to offer illegal or deviant material. However, material that is produced legally can also be used in a problematic way. This was demonstrated by Lehmann, Cohen and Kim (2006) in relation to the detection and management of pornography seeking in an online clinical dermatology atlas. During the study period, one third of the search queries related to anatomical sites and over half specified children.

So are the anxieties expressed about the new technologies, and in particular about the use of abusive images of children, simply part of a moral panic? Mears et al., (2008) in a survey examining US views towards sex crimes indicated that the public supports incarceration as an appropriate response to prosecuting child pornography offenders. The suggestion made by these authors was that Americans may support incarceration because of a belief that behaviour such as downloading abuse images will lead to a contact offence against a child. This might also be the case why convictions sometimes attract long sentences within the US, with Greenhouse (2007) giving details of a US Supreme Court decision declining to review a case in which an Arizona man was given a 200-year sentence for possessing 20 'pornographic' images of children. Yet Jenkins (2009) has argued that we see nothing constituting moral panic. It may be that while we have seen increasing anxiety about children's agency online, which result in displays of sexual behaviour or the establishment of sexual relationships, we show much greater ambivalence towards child pornography. We are concerned with the relationship between the use of images and harm against 'real' children in the offline world (largely *our* children), but we seem to demonstrate a much more complex level of interest in the photographic depiction of abusive and exploitative practices towards children.

Jenkins (2009) attributes what he calls this 'failure to launch' to several factors. He argues that one main reason is technical in that law enforcement agencies work at a technological level that is too low to comprehend trade in images as it actually is. But perhaps one factor, mentioned by Jenkins, to explain some of our ambivalence towards abusive and exploitative images of children lies in our lack of knowledge on the one hand of what these images are, and an overexposure on the other to sexualised visual materials. It is not simply that we lack concern about the illegality of images, or, as demonstrated by the 200-year sentence, the potential of such offenders to pose a further threat. Adler (2008, p3) reflects that, 'Claims about the changing nature of child pornography are difficult to verify for a number of reasons: above all, it is extremely hard, if not impossible, to measure accurately the online environment; in addition, no one outside of government can fully assess these claims because child pornography law prohibits researchers, academics, or anyone outside of law enforcement from looking at child pornography'. Yet sexualised images of children, clearly not defined in law as child pornography, are often found in contexts such as advertising, (described by Rush and La Nauze, 2006, as examples of corporate paedophilia). Adler's (2001) argument is that as we legislate more and more to control abusive images of children we potentially create a 'a vast realm of discourse' in which the image of the child as sexual is not only preserved but multiplied. Child pornography law socially constructs the child as sexual and one result of this construction may be that more people feel sexual desire for children.

If we look at the material found in the collections of offenders, the kinds of pictures that can be identified range from pictures of clothed children, through nakedness and explicit erotic posing to pictures of a sexual assault of the child or children in the photograph. We can make some objective sense of this by thinking of them in terms of a continuum of increased deliberate sexual victimisation (Taylor et al., 2001). This continuum ranges from everyday and perhaps accidental pictures involving either no overt erotic content, or minimal content (such as showing a child's underwear) at one extreme, to pictures showing actual rape and penetration of a child, or other gross acts of obscenity at

the other. Taking this perspective focuses attention not on just illegality as a significant quality of pictures, but on the preferred type of pictures selected by the collector, and the value and meaning pictures have to collectors (Taylor and Quayle, 2003). In trying to understand the ways in which children are victimised within the images, Taylor et al., (2001) generated a typology based on an analysis of publicly available images obtained from Newsgroups and Websites (made possible under Irish Law). This 'COPINE Scale' had ten levels ranging from indicative images to ones depicting sadism or bestiality. In 2002, in England and Wales, the Sentencing Advisory Panel (SAP) published their advice to the Court of Appeal on offences involving child pornography. The SAP believed that the nature of the material should be the key factor in deciding the level of sentence, and adapted the COPINE scale to five levels. They dropped levels 1 to 3 completely, arguing that nakedness alone was not indicative of indecency. The proposed structure was therefore that COPINE levels 5 to 6 constitute sentencing level 1 and COPINE levels 7 onwards each constitute an individual sentencing stage (Gillespie, 2003). One consequence of using such this measure has been that it provides a means of communication about the images without, for most people, the images ever having been seen. It is interesting that this way of talking about child sexual abuse has also entered into ordinary discourse.

Much earlier work by Lanning (1992) introduced an important distinction between child pornography (the sexually explicit reproduction of a child's image) and child erotica (any material, relating to children, that serves a sexual purpose for a given individual). In a similar fashion, Tate (1990 pp 203-217), commented on how the material ranged from, "posed pictures of naked and semi naked children, through more explicit shots of their genitalia thumbed apart to still, film and video recordings of oral, vaginal and anal sex". While legal definitions of child pornography have to be objective and expressed in terms that allow for the proper application of due process, it becomes apparent that not all of the material that is currently circulating on the Internet would meet any legal definition of child pornography, and the definition of such images as 'abusive' is a largely subjective one. Svedin and Back (1996, p9) defined child pornography as, "a text or an image – i.e. photo, slide, film, video or computer program – that is intended to evoke a sexual feeling, fantasy or response in adults". However, expressing criteria in terms of a capacity to generate fantasy may be problematic when objective definitions are required, as the range of materials that might evoke fantasy includes photographs that can be found in any family album or clothes catalogue.

The challenge posed by such a debate is, in the context of the huge volume of legal, but sexualised material relating to children on the Internet, as to how we might define these images, and whether we should be attempting to control their distribution. Clearly we cannot legislate against fantasy, but King (2008, p 332) has argued that, 'It is not clear… that the consumer (or the rest of society) can always (or ever) be sure what category a particular image falls into, how much harm to the subject it represents, for however happy and carefree the child seems to be, we cannot know what later effects she suffered (or, indeed, what she was subjected to after or as a result of that photograph). In fact it's clear that some degree of harm is almost always done to the subject in the production and distribution of child pornography of all kinds…'. King (2008) goes on to suggest that child pornography not only harms its immediate victims, the children whose abuse is at its centre, but also harms other children through the actions and attitudes of its consumers.

One further challenge relates to pseudo (digitally altered) images and virtual child pornography. Gillespie (2003) has raised important issues about how different an image has to be for it to constitute a pseudo–image, possession of which in England and Wales is likely to attract a lower sentence. In the US, the constitutionality of virtual child pornography remains a critical issue. In Ashcroft v. Free Speech Coalition (2002) a majority of the Supreme Court struck down portions of the Child Pornography Prevention Act of 1996, stating that virtual child pornography created without real or identifiable minors was unconstitutionally overbroad (Quayle, 2008). It might be thought that these 'pseudo-photographs' complicate our understanding of the problem and challenge our understanding of harm. Harm, however, need not always be harm towards a specific child. Most legislation against the distribution and possession of child abuse images builds on the fact that even unaware victims somehow come to harm, much in the way described by King (2008), and the increased number of abusive images in circulation may add to the likelihood that children are seen as possible objects of real abuse.

In 2003, Taylor and Quayle wrote that, "Pseudo-photographs are constructed photographs, often very cleverly done with great technical sophistication, using digital reconstruction techniques to create an image that is not a photograph of a real person, or of real events. Thus the head of a child might be placed onto the body of a woman, where the body features are manipulated to make it appear to be that of a child (breast reduced in size or eliminated, and pubic hair eliminated)…". However, while the production of such material a few years ago might have been a technological challenge, this would not be the case today. With the advent of software packages such as Adobe Photoshop, the majority of us would be able to create quite complex digitally altered images. The prediction was made that easier and more accessible computer aided animation and 3D computer graphics would lead to a growth in animated child pornography. This has now happened and we see evidence of wholly constructed computer images, although it is unclear as yet what impact this might have on the availability of such image distribution. One of the primary producers of such imagery is Japan where there is a huge market in manga, and other forms of animation, that many believe are sexually exploitative. A report in the UK's *Guardian* newspaper (*Guardian*, 2008) suggested that sexually explicit comics account for a large proportion of Japan's Yen 500 bn *manga* market, with many featuring schoolgirls or childlike adults being raped or engaged in sadomasochism. However, the article suggested that *manga* belonging to the popular '*lolicon*' – Japanese slang for Lolita complex – genre are likely to escape the proposed ban in Japan on the possession of child pornography, "as MPs are concerned that outlawing them could infringe on freedom of expression and drive men who use them as an outlet for their sexual urges to commit more serious sexual offences". However, outside of offender accounts there is little empirical research to support this while there is evidence to suggest that such manga is often found as part of the collections of seized images from offenders (Seto, 2009).

In countries outside of Japan there has been a bid to criminalise the possession of non-photographic visual depictions of child sexual abuse. In the UK a formal period of consultation began in relation to this in April 2007 and concluded in June of that year. Prior to this, the Criminal Law Sub Group of the Home Secretary's Task Force on Child Protection on the Internet had been considering the issues raised by computer generated images (CGIs) drawings and cartoons which show graphic depictions of sexual abuse of children or child-like characters. The Consultation document recognised that these images do not involve harm to real children in their creation, but that the possession of such material was a cause for concern, particularly as technological advances have increased the availability of such material. In the summary of the responses to the Consultation, it was noted that many people viewed the definition of what would constitute 'pornographic' as both troublesome and opaque. There was also concern that, 'stylisations of animations freely mix aspects typifying different ages', which would make the allocation of age subjective and therefore an impossible assessment of legality. Opponents of these measures, such as the American Civil Liberties Union, have argued that people's thoughts are their private thoughts, and that prohibition of pseudo-child pornography is a violation of free speech rights (Taylor and Quayle, 2003). However, Oswell (2006) has presented an important argument against this stating that, although the evidential value of the virtual image is different from an actual image (and hence the forms of police investigation and legal prosecution are different), until an image can be said to correspond to an actual case of child sexual abuse, all Internet child pornography can be viewed as real. In this sense, the primary concern is not one of the effects of the image on others or one of the relations of power encoded in the image, but one of the virtual evidentiality of the image (i.e. on the image's capacity to refer to an objective reality that is both internal and external to the image).

Within the last few years we have witnessed the development of supranational and international policy documents which set out to define 'child pornography' and four policy documents that are central to this issue. The European Union's Framework Decision on combating the sexual exploitation of children and child pornography entered into force in 2004 and required member states to take steps to ensure compliance by 20th January, 2006. Here child pornography is defined as pornographic material that visually depicts or represents:

i. a real child involved or engaged in sexually explicit conduct, including lascivious exhibition of the genitals or the pubic area of a child; or
ii. a real person appearing to be a child involved or engaged in the conduct mentioned in (i); or
iii. realistic images of a non-existent child involved or engaged in the conduct mentioned in (i).

As we can see, the definition in the EU Framework Decision talks about a 'real' child, 'real' person and 'realistic' images, which may prove unlikely to cover virtual images or cartoons. The Council of Europe's Cybercrime Convention (2001) came into force in July 2004, and Article 9 defines child pornography as pornographic material that visually depicts: a minor engaged in sexually explicit conduct; a person appearing to be a minor engaged in sexually explicit conduct; or realistic images representing a minor engaged in sexually explicit conduct. This relates to all people under the age of 18, but it is possible for a lower age limit of 16 to be set. The third document is the United Nation's Optional Protocol to the Convention on the Rights of the Child on the Sale of Children, Child Prostitution and Child Pornography which came into force in January 2002 and defines child pornography as, 'any representation, by whatever means, of a child engaged in real or simulated explicit sexual activities or any representation of the sexual parts of a child for primarily sexual purposes'. In all three a child is defined as someone under the age of 18 years and includes both photographs of actual children as well as representations of children, which would appear to include computer generated images. However the issue of age is subject to several reservations and complicated by the age of sexual consent established under national law. Akdeniz (2008) draws our attention to the fact that the UN definition is broad and, as it refers to 'any representation', would also include textual material, cartoons and drawings.

The most recent relevant instrument establishing a definition of child pornography is the Council of Europe Convention on the Protection of children against sexual exploitation and sexual abuse. While this definition is restricted to visual materials it does not require that a real child be used in their production (as is the case in the US). However member states may opt not to criminalise the production and possession of virtual child pornography. Importantly the Convention has chosen not to criminalise the consensual production and possession of materials created by children who have reached the age of consent. However, most instruments do not directly address the issue of adolescents who make or access indecent images of children, and this in itself may prove to be problematic. Piper (2001) has argued that one of the landmark changes in terms of criminal justice policy in recent times has been the approach to juvenile crime, which in the UK led in the 1990s to the effective reduction in the age of criminal responsibility to 10 years accompanied by a series of measures that were designed to tackle youth crime. She convincingly argued that adolescents involved in criminality became less victims of social failings in need of protection but rather criminals who require the intervention of the criminal justice system (Quayle et al., 2008). Gillespie (2008) has argued that in the UK the criminal justice system is increasingly adopting a harsher approach to adolescents who break the law, with the law adopting very different approaches to adolescents involved in indecent images of children and those who have direct sexual contact with another adult.

It is important to note that while child pornography on the Internet is generated in a number of ways, these photographs (and non-photographic depictions) are a permanent product of abusive practices, some of which involve a direct sexual assault on a child. As we will see, their permanence may itself be an issue for the children photographed. We can think of child pornography production coming from many sources, some of which predate the Internet. This may include: hard copy images relating to the period of decriminalisation, that have been scanned; those produced during the sexual abuse of children in domestic settings, where the child *may* know about the photography taking place; hidden, or stolen images, made, for example, by placing cameras in shower heads, or surreptitiously photographing children in swimming pools; commercial images, where the photographer may be involved in abusive practices towards the child, and self-generated material, produced by children in response to sexual demands by others, as well as through activities initiated by young people themselves. Clearly one issue here relates to confidence in the ability of professionals to identify images that are of children (as opposed to adults positioned as children). This hinges on the definition of child, which in many, but not all, jurisdictions is set at 18 years. A study by Cattaneo et al., (2008) examined some of the difficulties relating to this and noted that there is great variability in physical

maturity due to biological, pathological and environmental factors. In this study photographs of eleven adult females were taken and two groups, experts and lay people, were asked to establish if each were less than 18 years and the basis for making that judgment. Their results indicated that all assessors performed poorly and the authors conclude that a more reliable categorisation might be between pre-pubertal and pubertal children, but this would exclude all the images of children (under the age of 18) who are sexually mature.

Schuijer and Rossen (1992) analysed the content of 'child pornography magazines and videos' that had been in circulation prior to the change of law across a number of European countries. They suggested that on the basis of the estimate of 1,065 published magazines, a conservative estimate would be the involvement of 6,000 children. Their analysis indicated that 42% of the pictures were of boys, with more girls appearing in extreme forms of images. However Australian data from Baartz (2008) describing the gender, ethnicity and age of the victims portrayed in the images examined by investigators would suggest that the children were mostly white, westernised females, aged between 8 and 12 years. Asian children were the next most common ethnic group, and there was a comparative absence of indigenous Australian children. An analysis of the COPINE archive in 2003 indicated that the majority of images available were of white Caucasian and Asian children, with very few African or African-American children (Taylor & Quayle, 2005). Indeed, in 2003, Websites started to appear advertising specialist sites that included interracial pictures. Similarly, Carr's (2004) study, which was one of the few to analyse the images used by offenders, indicated that the vast majority of offenders selected material portraying Caucasian and Asian children. These findings are supported by a recent analysis of images from one of the existing police data bases and will be discussed in greater detail in the Conference presentation.

Our lack of knowledge about children being abused through photography is reflected in the relatively small numbers who are ever identified. Where identification does take place, there is little consistent empirical data, although the National Center for Missing and Exploited Children (NCMEC) suggested that as of September 2008, 1,660 children had been identified through distributed and non-distributed images (73% female and 27% male). At the time of writing this chapter, this was the most complete data set of identified children available, but is based only on what has been reported to NCMEC by law enforcement (Lee, 2008). The numbers for 'Gender' represent actual individual *children*, whereas the numbers for Age Category (Infant/Toddler, 6%; Prepubescent, 49%; Pubescent, 45%) represents the percentage of identified *series.* There can be more than one child within a series so, for example, a series that has four prepubescent boys will be counted once, as the percentage represents the series, not children. The statistics for Ethnicity also represent *series* and include all the identified series in the NCMEC system, as well as some other known child sexual abuse series that are currently being investigated (Asian, 16, Biracial, 0, Black, 23; Hispanic, 19; Other, 5; Unknown, 0, and White, 1186). It is unclear whether this data on identified children reflects the actual distribution of images currently circulating on the Internet and is certainly different in its gender distribution than is suggested by Shuijer and Rossen (1992) in relation to earlier hard copy images.

A report from Baines (2008, p34) suggests that while our knowledge of these children remains imprecise, 'In terms of content, the number of non-commercial images showing babies or toddlers is on the increase: victims in commercial images also are increasingly young, with 80 per cent estimated to be less than 10 years old. Moreover, a number of investigations by UK and overseas law enforcement agencies have highlighted the fact that there are many series of images in which the victims appear to have been abused a number of years earlier but where the images have only just come to light. This is particularly true for images of boys and where the material has been seized from a contact sexual abuser – in turn suggesting that offenders who have previously been content to keep a record of the abuse for their own personal gratification may have been detected after succumbing to the urge to share this material on the Internet. In recent years law enforcement has also seen the emergence of images – albeit so far a relatively small number – containing victims of non-white origin, including those of South American and Southeast Asian origin. This proliferation of images from a variety of source countries points to the role of the Internet in facilitating truly global communications and networking across obvious language and cultural barriers'.

Clearly children within the images do not necessarily come from countries where Internet access is widely available. There have been reports of children exploited through the production of child pornography in Mexico (Azaola, 2000), South Asia (Huda, 2006) and India (Kacker, Varadan and Kumar, 2007). What is of concern about the production of such images is that we have little knowledge of how they have become part of abusive practices against children, and no knowledge as to how they were used. We do not know if these photographs were sold on, whether they became part of commercial sexual exploitation (through the sale of DVDs) or whether their images will ever find their way onto the Internet (Quayle et al., 2008). There has also been an increase in images that are commercially produced and distributed through the Internet that involve the exploitation of children in modelling sites. The Regional Overview on Child Sexual Abuse Images through the Use of Information and Communication Technologies in Belarus, Moldova, Russia and Ukraine (2008) indicated that a significant challenge identified in the region through their analysis is, 'the use of children for the production of sexual abuse images by organised criminal networks, especially under the guise of the modeling business. The involvement of a wide range of actors … particularly structured networks of people who have organised themselves around the production and distribution of sexual abuse images of children… it appears that boys are as much at risk of sexual exploitation in the production of abusive images as girls'. While many of these images would not meet a legal definition of child pornography they are clearly exploitative of such children and may be part of a range of abusive and exploitative practices.

# An Examination of Problematic Paraphilic use of Peer to Peer Facilities.

**Authors**

Sean Hammond,
Ethel Quayle,
Jurek Kirakowski,
Elaine O'Halloran,
Freda Wynne

**Abstract**

This paper describes a methodology for investigating the paraphilic use of Peer to Peer facilities. The focus is upon problematic paraphilias, by which we mean those that imply illegal and/or non-consensual activity. The methods applied involve a new technique for evaluating the co-occurrence of paraphilic themes in order to inform a psychological profiling of P2P users. A typal analysis derived from Configural Frequency Analysis is reported. This shows in particular, that hebephilic and paedophilic behaviour are interrelated in a more complex manner than is expected by pure legal classification.

## Background

Decentralised and anonymous P2P systems offer scope for the pursuit of socially dubious sexual interest in a relatively safe and secure environment. The ease with which pornographic materials can be accessed through P2P networks has raised serious concerns, particularly for the protection of children who may be recipients, or indeed the subjects, of such material (Congressional Committee on Government Reform, 2001; Greenfield, 2004). Nevertheless, the small amount of empirical research in the area suggests that pornographic exchange forms only a small part of the total P2P traffic. Thus, Hughes, Walkerdine, Coulson and Gibson (2006) found in a study of the Gnutella P2P network that pornography constituted only 1.6% of searches and 2.4 % or responses. This is in stark contrast to the warnings from US government agencies such as the US Federal Exchange Commission (2004), the US General Accounting Office (2003) and the US Congressional Committee on Government Reform (2001) on the pervasiveness of pornography on P2P networks. The CCGR (2001) demonstrates that the most popular Gnutella search terms in 2001 contained a number that were unequivocally sexual.

It should be said that Hughes et al (2006) were only concerned with 'illegal' sexual material and they used a very strict filter in order to reduce the number of false positives. This might suggest that their findings offer a conservative estimate of the sexual use of P2P networks. However, Hughes et al (2006) point out that those people using P2P networks to exchange pornographic materials, while representing a small sub-community of users, were

particularly active. It must also be born in mind that even if a small proportion of exchanges are sexual in nature, the vast number of exchanges in general still suggests a very large number of sexually motivated P2P users. Hughes et al. do not profile the pornography users in their study but they do cite the importance of group-specific norms as a basis for the notion of a tightly identified sub-community of sexually motivated users. This leads to the intriguing possibility of a more specific analysis in which more tightly defined paraphilia based sub-communities may be uncovered. The most troubling and contentious of these, of course, is paedophilia.

Much of the polemic against the P2P exchange of pornographic material is fuelled by the fact that such exchanges include material depicting the sexual abuse of children. The report from the General Accounting Office estimates that 42% of pornography exchanges on P2P networks involve children. Even allowing for fairly lax criteria for making such judgements, the anonymity and ease of access of P2P networks clearly facilitates the exchange of depictions of child abuse.

This paper describes a methodology for profiling paraphilic use of P2P networks. We use a specifically psychological-psychiatric focus to the profiling problem. The focus here is upon deviant patterns of sexual interest rather than the legal framework that defines sexual offending and abuse. For this reason we discriminate two paraphilic groupings associated with child molestation, these are the paedophile and the hebephile (Blanchard et al., 2008). While the paedophile manifests a sexual interest in prepubescent children, the hebephile is sexually interested in pubescent or recently post-pubescent children. Legally, the difference is largely unclear but psychologically the arousal pattern of the hebephile to secondary sexual characteristics is more in keeping with those of a normal adult. This does not suggest that hebephilia is less socially problematic than paedophilia but merely that there may be different underlying pathologies at work, so grouping them together may ultimately prove unhelpful.

### Aims

The aims of this paper are to report on developments following an earlier study in which we explored the associations between sexual interest themes using exploratory and heuristic multidimensional scaling methods (Quayle, Hammond and Wynne, 2007). The aim of this report is to examine taxonomies of paraphilic sexual interest informed by the search terms individuals employ in their P2P interactions and it takes a more model-based approach to the examination of sexual interest themes.

## Method

### The Data

The present report is based upon two data sets made available through the MAPAP project team. The first of these data sets comprises a list of the 119,869 most commonly used terms in P2P submissions over the period of one week. This data set we will call the Search Terms List (STL). The second data set comprises all eDonkey transactions over this

period and numbers over 3,000,000 records. This data set we will term the P2P Submissions List (PSL).

### *Identification of Thematic Categories*

Our first priority was to try to impose some order upon the rich variety of the data collected. In the first instance, this involved an analysis of the STL in which the list was trawled for words with a sexual connotation. An exhaustive search of the list was carried out to identify terms that indicated sexually related material. In addition, a computer program was written for our Windows system in order to isolate words or part-words according to a given theme.

The result of this process was the identification of a number of specific themes defined by their sexual and fetishistic content. Specific terms and words associated with these categories were identified from the STL data. It should be born in mind that these are not a final categorization an on-going refinement continues among the partners in MAPAP project. It should be clear that the motive behind using such search terms may vary and the assumption that these terms were being used to seek out sexually stimulating material is likely to be erroneous in some cases. However, given the large number of cases considered it was felt that this would constitute a manageable degree of error and would not invalidate the current methodology.

### *Identification of Individual's Sexual Interest Profiles*

In order to provide a sexual interest profile for each record we next turned to the PSL data set. This contains over 3,000,000 records of submissions to eDonkey. A computer program was written for our Windows platform to scan these records in a serial fashion to find instances of the words identified as representing the thematic categories.

For each case, a record containing variables representing the 25 themes was created. Each variable was initially set to zero. If a sought after word occured in the PSL data set for that case, then the variable representing the theme in which it is placed is incremented by one.

If, after scanning, a case has no occurrence of the critical words it is jettisoned and the program moves onto the next case. If, on the other hand, the case does contain critical words the record of 25 themes is written to a second data file. In this way the program identifies those cases where a user has made one or more sexually related submissions as defined by the terms recorded in appendix 1.

Thus a second data file was generated containing only these cases that manifest at least one of the 25 categories and this contained 62940 cases. Each of the 25 variables contains the frequency with which terms are submitted within each of the 25 thematic categories. In order to control for the fact that each theme is built of differing numbers of terms we chose to represent the data in binary form thus:-

If $y_i>0$ then $x_i=1$
If $y_i=0$ then $x_i=0$

Where $y_i$ is the observed frequency of words in thematic category i.
$x_i$ is the binary value

Each case, then, is recorded as a profile of 25 binary variables in which at least one variable is recorded as 1. Recording the data in this way provides a tractable data set to address the exploration of the relationships between sexual themes and a typal analysis of deviant sexual interest.

## Results

The focus for this study is on problematic or 'deviant' sexual interests and to this end, a subset of the 25 themes was selected. These themes are presented in table 1 along with the percentage of total sexually motivated searches they each account for. A number of unexpected findings emerge. In stark contrast to the GAO claim that 42% of sexual exchanges on P2P networks involve children, we have found only 0.82% of searches to be explicitly paedophile orientated. It is also surprising to note that zoophilic or bestial interest appears to outstrip paedophilic searching.

The hebephilic theme had the largest incidence of these seven themes at 2.29% and the fact that this theme outnumbers the paedophile theme is not unexpected (Studer, 2004). All in all the problematic themes selected here account for a relatively small amount of the 79427 sexually motivated searches observed in our study epoch (8.27%).

**Table 1:**
**Frequencies of Thematic Categories in the PSL Data Set**

| Theme | n | % |
|---|---|---|
| Gerontophilic | 124 | 0.15% |
| Incest | 455 | 0.57% |
| Paedophilic | 657 | 0.82% |
| Bestial | 907 | 1.14% |
| Sadistic | 1046 | 1.31% |
| Rape | 1561 | 1.96% |
| Hebephilic | 1819 | 2.29% |
| Total Sexual | 79427 | |

### *Configural Frequency Analysis*

In order to examine the presence of distinct types in the data the search profiles using the seven paraphilic search themes were first subjected to a confirmatory zero-order Configural Frequency Analysis (CFA). The zero-order model was applied rather than the more typical first-order because the aim was to test the simple main effect of each paraphilic theme. In the zero-order model observed frequencies are tested against expected frequencies

generated on the assumption of a uniform frequency distribution (von Eye 1990). This confirmatory analysis serves two purposes. First, it tests the hypothesis that each of the paraphilic themes represents a 'pure' and discrete sub-community in the P2P network space and second it serves as a tentative validation for the coding scheme used. The results are reported in table 2.

**Table 2**
**Zero-Order Confirmatory CFA: Typal Identification of Paraphilic Interests**

| Theme | f | Lehmacher z | Adjusted p | Profile P H G B S R I |
|---|---|---|---|---|
| Incest | 425 | -3.02 | p<0.0071 | 0 0 0 0 0 0 1 |
| Rape | 1377 | 40.07 | p<0.0071 | 0 0 0 0 0 1 0 |
| Sadistic | 919 | 19.34 | p<0.0071 | 0 0 0 0 1 0 0 |
| Bestial | 872 | 17.21 | p<0.0071 | 0 0 0 1 0 0 0 |
| Gerontophilic | 123 | -16.69 | p<0.0071 | 0 0 1 0 0 0 0 |
| Hebephilic | 1763 | 57.55 | p<0.0071 | 0 1 0 0 0 0 0 |
| Paedophilic | 615 | 5.58 | p<0.0071 | 1 0 0 0 0 0 0 |

Bonferroni adjustment for p at 0.05 = 0.0071

These findings support the contention that each of the themes describes a specific and independent sub-community. For example, of the 657 searches betrayiong paedophilic interest (table 1), 615 or 93.61% manifested a 'pure' profile with no other associated interest. It should also be noted that Incest and Gerontophilia are found to be significantly under-represented in the P2P network space, as indicated by the negative z ratio in column 3. These profiles, where the observed frequency is less than the expected uniform frequency, are sometimes named anti-types to describe their scarcity in the sample under scrutiny (Krauth 1985).

As it stands this analysis provides limited information except that P2P users seeking paraphilic materials appear to be pretty specific in their searching behaviour. This does support Hughes at al.'s (2006) contention that the P2P network space is made up of distinct sub-communities providing the basis for targeted strategies for policing and managing problematic users. However, the question still remains as to whether there exists other 'comorbid' or multi-paraphilic 'communities'. A first order analysis will be required to test this hypothesis as the interaction between themes will need to be examined. For this analysis the expected frequencies are generated by conditioning out the main effects between themes. Such an analysis is summarised in table 3.

In this analysis all possible profiles or combinations of the 7 themes are included, making this an exploratory analysis. The number of possible profiles is $2^7$ or 128, and it is

instructive to observe that only 28 combinations are to be found in practice, suggesting a high degree of specificity in paraphilic searching behaviour.

In table 3 column 5 indicates whether each profile may be statistically identified as a 'Type' or an 'Anti-Type'. For present purposes we consider 'Types' to represent potential sub-communities in P2P network space. Note that the statistical criterion has changed because we now use a Bonferroni adjustment for the 128 potential profiles, allowing a more conservative estimate of significance.
In addition, a number of statistically significant profiles must be treated with caution because the expected frequencies are very small. This rules out profiles 36, 68, 98 and 100.

**Table 3**
**First Order Exploratory CFA: Paraphilia Profiling based on Thematic Interactions**

| Profile | f | Expected | Lehmacher z | | Profile P | H | G | B | S | R | I |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 56617 | 56642.76 | -1.668 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 425 | 412.45 | 2.028 | | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 3 | 1377 | 1440.52 | -6.098 | A | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 4 | 3 | 10.49 | -2.347 | | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| 5 | 919 | 957.24 | -4.277 | A | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 7 | 119 | 24.34 | 19.565 | T | 0 | 0 | 0 | 0 | 1 | 1 | 0 |
| 9 | 872 | 828.17 | 5.202 | T | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 11 | 24 | 21.06 | 0.652 | | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| 13 | 1 | 14.00 | -3.525 | A | 0 | 0 | 0 | 1 | 1 | 0 | 0 |
| 17 | 123 | 111.81 | 3.378 | T | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 19 | 1 | 2.84 | -1.107 | | 0 | 0 | 1 | 0 | 0 | 1 | 0 |
| 33 | 1763 | 1685.70 | 7.055 | T | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 34 | 6 | 12.27 | -1.822 | | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| 35 | 14 | 42.87 | -4.525 | A | 0 | 1 | 0 | 0 | 0 | 1 | 0 |
| 36 | 7 | 0.31 | 11.976 | T | 0 | 1 | 0 | 0 | 0 | 1 | 1 |
| 37 | 3 | 28.49 | -4.880 | A | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| 39 | 1 | 0.72 | 0.324 | | 0 | 1 | 0 | 0 | 1 | 1 | 0 |
| 41 | 9 | 24.65 | -3.217 | | 0 | 1 | 0 | 1 | 0 | 0 | 0 |
| 65 | 615 | 597.49 | 2.392 | | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 66 | 9 | 4.35 | 2.247 | | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 67 | 11 | 15.20 | -1.094 | | 1 | 0 | 0 | 0 | 0 | 1 | 0 |
| 68 | 2 | 0.11 | 5.681 | T | 1 | 0 | 0 | 0 | 0 | 1 | 1 |
| 69 | 3 | 10.10 | -2.262 | | 1 | 0 | 0 | 0 | 1 | 0 | 0 |
| 73 | 1 | 8.74 | -2.648 | | 1 | 0 | 0 | 1 | 0 | 0 | 0 |
| 97 | 12 | 17.78 | -1.397 | | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| 98 | 2 | 0.13 | 5.200 | T | 1 | 1 | 0 | 0 | 0 | 0 | 1 |
| 99 | 1 | 0.45 | 0.815 | | 1 | 1 | 0 | 0 | 0 | 1 | 0 |
| 100 | 1 | 0.00 | 17.369 | T | 1 | 1 | 0 | 0 | 0 | 1 | 1 |

Bonferroni adjustment for p at 0.05 = 0.00039

Only one combination type emerges and this is represented by profile 7 which includes rape and sadistic themes. It is perhaps not unexpected that such a combination would arise and it would seem to represent a sub-community of sexually sadistic individuals with an interest in rape. Of particular interest to us is the finding that combining paedophilia and hebephilia (profiles beginning 1,1,…) do not emerge in any significant manner. This suggests that the Child Molester label really does describe two distinct sub-groups in terms of sexual interest and may further suggest the development of discrete strategies for tackling both offender groups.
The anti-types are also revealing as they show the unlikely combinations. Thus hebephilc interest is particularly unlikely to covary with an interest in rape and sadism.

To conclude the typal analysis of this data a model-based approach was taken utilising a Latent Class Analysis (Lazarsfeld, 1950; Goodman, 1974; Magidson and Vermunt 2004). This is a probabilistic approach as opposed to the more deterministic CFA. The data, as described above, was truncated by removing all null profiles (0,0,0,0,0,0,0) and was then fitted to a number of unrestricted latent class models ranging from 2 to 8 underlying classes. The 7-class model was found to be the best fitting using the log-likelihood statistic and the Bayesian Information Criterion (BIC). This solution is summarised in table 4.

Given the relatively large sample size, the statistical indices can be inflated so a more heuristic index of fit may be useful. The disimilarty index is a descriptive measure indicating what proportion of the sample should be moved to another cell to get a perfect fit. On that basis the latent class solution reported suggests an excellent fit to the data.

**Table 4**
**Latent Class Analysis of Paraphilia Themes: The 7-Class Solution**

| Themes | Classes….. 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| Gerontophilic | 0.000 | 0.998 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Bestiality | 0.008 | 0.028 | 0.026 | 0.000 | 1.000 | 0.019 | 0.054 |
| Paedophilic | 0.000 | 0.000 | 0.001 | 0.000 | 0.046 | 0.004 | 1.000 |
| Hebephilic | 0.000 | 0.035 | 0.008 | 1.000 | 0.010 | 0.020 | 0.003 |
| Sadistic | 0.000 | 0.031 | 0.001 | 0.000 | 0.000 | 1.000 | 0.000 |
| Rape | 0.000 | 0.000 | 1.000 | 0.001 | 0.000 | 0.000 | 0.001 |
| Incest | 0.997 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Class Probabilities | 0.019 | 0.072 | 0.143 | 0.279 | 0.230 | 0.101 | 0.154 |

| Diagnostic Statistics | |
|---|---|
| Log Likelihood | 59.12 |
| 2LL | 118.23 |
| Pearson $\chi^2$ | 194.60 |
| Pearson $\chi^2$ under independence | 8281.38 |
| Dissimilarity Index | 0.004 |
| BIC | 93.90 |

It is clear from these results that the best way to describe the underlying latent structure is as 7 discrete sub-communities each defined by one specific sexual theme. This is unsurprising given the CFA results above but it serves to further emphasise the contention of extant paraphilic sub-communities in the P2P network space.

## Discussion

The research programme that supported this work is entitled '*Measurement and analysis of peer to peer activity against paedophile content*'. The study reported here builds upon an earlier study exploring the paraphilic space indicated by P2P searches (Quayle et al 2007). In this study we take a further step in trying to derive a method for examining the latent groups of paraphilic users of P2P networks. As the title of the programme states our primary interest is in paedophile interest although this psychological/psychiatric term is often confounded with legal attempts to define sexual offenses against children. We have attempted to distinquish between offenders who may have a paedophilic orientation from those who may have a hebephilic interest. Not because one is of lesser concern than the other, but because if, as Hughes et al, (2006) suggest, P2P networks are constituted of specific sub-communities, an effective targeting strategy may prove to be a viable alternative to scattergun policing of the networks.

The findings presented here, are certainly suggestive of the existence of group specific social-norms, and while that cannot be directly ascertained with this data, there is ample evidence for specificity in the sexual interests expressed by people exhibiting paraphilic motivation for P2P use.

## References

Blanchard, R., Lykins, A. D., Wherrett, D., Kuban, M. E., Cantor, J. M., Blak, T., Dickey, R., & Klassen, P. E. (2008). Pedophilia, hebephilia, and the DSM–V. *Archives of Sexual Behavior.* 38, 335–350

Congressional Committee on Government Reform (2001) Children's Access to Pornography Through Internet File-Sharing Programs. Special Investigations Division Committee on Government Reform. U.S. House of Representatives

General Accounting Office (2003) File-sharing programs: Peer-to-Peer networks provide ready access to child pornography. GAO-03-351. Washington: US General Accounting Office

Goodman, L.A. (1974) Explanatory latent structure analysis using both identi8able and unidenti8able models. *Biometrika* 61, 215–231.

Greenfield, P.M. (2004) Inadvertent exposure to pornography on the Internet: Implications of peer-to-peer file-sharing networks for child development and families. *Applied Developmental Psychology* 25, 741–750

Huges, D., Walkerdine, J., Coulson, G. and Gibson, S. (2006) Peer-to-Peer: Is deviant behaviour the norm on P2P file-sharing networks? *IEE Distributed Systems Online.* 7 (2), 1-11.

Krauth, J. (1985) Typological personality research by Configural Frequency Analysis. *Personality and Individual Differences.* 6, 161-168.

Lazarsfeld, P.F. (1950) The logical and mathematical foundation of latent structure analysis. In: Stouffer, S., et al. (Ed.), *Measurement and Prediction.* Wiley, New York.

Lienert, G.A. (1988) *Angewandte Konfigurationsfrequenzanalyse.* Frankfurt: Athenaum.

Magidson, J., and Vermunt, J.K, ( 2004) Latent class analysis. D. Kaplan (ed.), *The Sage Handbook of Quantitative Methodology for the Social Sciences,* Chapter 10, 175-198. Thousand Oakes: Sage Publications.

Quayle, E., Hammond, S., Wynne, F. (2008) An Empirical Investigation into the Sexual Interest Profiles Manifest in P2P Activity: A Preliminary Report. MAPAP Project. SIP-2006-PP-221005.

Studer, L. H., Aylwin, A. S., Clelland, S. R., Reddon, J. R., & Frenzel, R. R. (2002). Primary erotic preference in a group of child molesters. *International Journal of Law and Psychiatry, 25,* 173–180.

US Federal Exchange Commission (2004) Peer-to-Peer File-Sharing Technology:Consumer Protectionand Competition Issues: Staff Report. http://www.ftc.gov/reports/p2p05/050623p2prpt.pdf

von Eye, A. (1990) *Introduction to Configural Frequency Analysis.* Cambridge: Cambridge University Press.


# INHOPE – The International approach to combating the proliferation of Images of Child Sexual Abuse on the Internet.

The mission of the INHOPE Association is to support and enhance the performance of Internet Hotlines around the World; ensuring swift action is taken in responding to reports of illegal content to make the internet a safer place.

## Who are INHOPE?

INHOPE, the International Association of Internet Hotlines, was founded in 1999 under the European Commission's Safer Internet Action Plan to combat these growing concerns.

INHOPE represents and co-ordinates a global network of Internet Hotlines, supporting them in their aim to respond to reports of illegal content to make the Internet safer.

## Goals:

- To maintain a worldwide network of national Hotlines.*
- To ensure rapid and effective response to illegal content reports by developing consistent, effective and secure mechanisms for exchanging reports between Hotlines internationally and ensuring a coordinated approach is taken.
- To implement INHOPE policies and best practice standards for hotlines and encourage exchange of expertise.
- To expand the network of INHOPE members and provide consultation and training to meet INHOPE best practice standards.
- To educate and inform policymakers and stakeholders at an international level, including government, law enforcement, child welfare groups, Industry and other related bodies, with the aim of achieving better co-operation internationally.
- To raise awareness of INHOPE and member hotlines as a "one stop shop" for global reports of illegal content.
- To identify new trends in cyber crimes and develop solutions.

***"INHOPE is working towards a safe environment for Internet users which will protect our children and respect the privacy and dignity of our citizens."***

## Illegal use of the Internet

Over the last decade the Internet has changed the way we communicate, the way we do business and ultimately the way we live. Unfortunately, as with many new forms of technology, there is also a downside to the Internet and in the mid 1990's concerns were raised over the new types of illegal content being found online, primarily Online Child Sexual Abuse Images.

Hotlines have a key role to play in tackling the problems mentioned above and therefore making the Internet a safer place.

## What is a Hotline?

Internet Hotlines have proven to be an effective first line of defence against illegal activity online. Through a Hotline, Internet users can make a report of something they suspect to be illegal on the Internet – mainly via email or web-interface. The Hotline will investigate these reports to determine if they are illegal, and if so, trace the origin of the content.
If the content is illegal and hosted in their own country the Hotline will refer this onto law enforcement agencies and also the Internet Service Provider for removal. If the content is hosted overseas the will refer it onto the corresponding hotline. Hotlines have the support of their national government, Internet industry, law enforcement, and Internet users in the countries of operation and offer effective transparent procedures for dealing with complaints. INHOPE was established to support and co-ordinate Internet Hotlines around the world in dealing with illegal content online.

## International Cooperation

Illegal activity on the Internet is a cross border problem that no one organisation can effectively tackle alone. This is where the INHOPE network demonstrates its effectiveness. Often material reported to Hotlines is hosted beyond the borders of their own country or the perpetrator is located abroad. International coordination between Hotlines is essential.

## Membership of INHOPE

Any Hotline can apply for INHOPE membership and be subject to the INHOPE membership process. All INHOPE members enjoy the benefits of the association including access to best practice papers, training programmes, knowledge management systems as well as members' meetings.

To qualify for membership, Hotlines must meet the following criteria:

- Provide a mechanism for receiving complaints from the public about alleged illegal content.
- Have effective transparent procedures for dealing with complaints
- Have the support of government, industry, law enforcement, child welfare groups and Internet users in the countries of operation
- Co- operate with other members in exchanging information about illegal content and use and share their expertise
- Make a commitment to maintain confidentiality
- Respect the procedures of other members.

Contact details:

**Adrian Dwyer**
Membership Coordinator

## INHOPE MEMBERS

| | | |
|---|---|---|
| Australia | ACMA | www.au.inhope.org |
| Austria | Stopline | www.at.inhope.org |
| Belgium | Child Focus | www.be.inhope.org |
| Bulgaria | ARC FUND | www.bg.inhope.org |
| Canada | cybertip.ca | www.ca.inhope.org |
| Chinese Taipei | ECPAT Taiwan | www.tw.inhope.org |
| Cyprus | Safeweb | www.cy.inhope.org |
| Denmark | Red Barnet | www.dk.inhope.org |
| Finland | STC Finland | www.fi.inhope.org |
| France | AFA | www.fr.inhope.org |
| Germany | ECO | www.de.inhope.org |
| Germany | FSM | www.de.inhope.org |
| Germany | jugendschutz.net | www.de.inhope.org |
| Greece | SafeNet | www.gr.inhope.org |
| Hungary | MATISZ | www.hu.inhope.org |
| Iceland | Barnaheill | www.is.inhope.org |
| Ireland | ISPAI | www.ie.inhope.org |
| Italy | HOT 114 | www.it.inhope.org |
| Italy | STC Italy | www.it.inhope.org |
| Japan | Internet Association Japan | www.jp.inhope.org |
| Luxembourg | LISA Stopline | www.lu.inhope.org |
| Netherlands | Meldpunt | www.nl.inhope.org |
| Poland | NASK | www.pl.inhope.org |
| Russia | Friendly RUNet Foundation | www.ru.inhope.org |
| Russia | National Internet-Safety Node | www.ru.inhope.org |
| Slovenia | SPLETNO OKO | www.si.inhope.org |
| South Africa | Film Publication Board | www.za.inhope.org |
| South Korea | KISCOM | www.kr.inhope.org |
| Spain | Protegeles | www.es.inhope.org |
| United Kingdom | Internet Watch Foundation | www.uk.inhope.org |
| United States | Cybertipline | www.us.inhope.org |

# INHOPE

## The International Association of Internet Hotlines

Advances in the Analysis of Online Paedophile Activity Conference

Paris 2nd – 3rd June 2009

Adrian Dwyer
Membership Coordinator

"The project is co-funded by the European Union, through the Safer Internet plus programme".
Visit: http://ec.europa.eu/saferinternet

## What is INHOPE?

- INHOPE is an umbrella organization of the national Hotlines providing a possibility for the internet users to report about illegal content
- Founded in 1999 under the European Commission's Safer Internet Action Plan to combat growing concerns related to the illegal content
- INHOPE represents and co-ordinates the global network of Internet Hotlines and supports them in their fight against illegal content
- The global network currently consists of 35 Hotlines in 31 different countries all over the world

# Mission of INHOPE

To support and enhance the performance of Internet Hotlines around the World, ensuring swift action is taken in responding to reports of illegal content to make the internet a safer place.

co-funded by the European Union

# INHOPE Members

| | |
|---|---|
| **Australia** | ACMA 1999 |
| **Austria** | Stopline 1999 |
| **Belgium** | Child Focus 2001 |
| **Bulgaria** | ARC Fund 2006 |
| **Canada** | cybertip.ca 2005 |
| **Chinese Taipei** | ECPAT Taiwan 2005 |
| **Cyprus** | CNTI 2008 |
| **Czech Republic** | Our Child Foundation 2007 |
| **Denmark** | Red Barnet 2002 |
| **Finland** | STC Finland 2002 |
| **France** | AFA 1999 |
| **Germany** | ECO 1999 |
| **Germany** | FSM 1999 |
| **Germany** | jugendschutz.net 1999 |
| **Greece** | SafeNet 2004 |
| **Hungary** | MATISZ 2005 |
| **Iceland** | Barnaheill 2001 |
| **Ireland** | ISPAI 1999 |
| **Italy** | HOT 114 2006 |
| **Italy** | STC Italy 2003 |
| **Japan** | Internet Association Japan 2007 |
| **Latvia** | Secretarial of Special Assignments Minister for Electronic Government Affairs 2008 |
| **Lithuania** | Communications Regulatory Authority of the Republic of Lithuania 2008 |
| **Luxembourg** | LISA Stopline 2008 |
| **Netherlands** | Meldpunt 1999 |
| **Poland** | NASK 2006 |
| **Portugal** | FCCN 2007 |
| **Russia** | Friendly RuNET Foundation 2009 |
| **Russia** | ROCIT 2009 |
| **Slovenia** | Spletno Oko 2007 |
| **South Africa** | Film Publication Board 2009 |
| **South Korea** | KISCOM 2003 |
| **Spain** | Protegeles 2002 |
| **United Kingdom** | Internet Watch Foundation 1999 |
| **United States of America** | Cybertipline 1999 |

co-funded by the European Union

# National Hotlines

- National contact point for the public to report illegal content on the Internet
- Illegality of the reported sites is investigated
- Origin of the illegal sites is traced
- Information concerning the illegal site is passed to the member Hotline in the country of origin
- Hotline in the country of origin informs the national law enforcement agency
- Law enforcement procedures in national and international level
- Internet Service Provider contacted by the Law Enforcement or the Hotline – depending on the national agreement
- INHOPE best practice standards

inhope

co-funded by the European Union

# Illegal content

- The definition of illegal content relies on national legislations
- Illegal content might concern
  - Child Sexual Abuse Images
  - Extreme violence
  - Racism and xenophobia
  - Bestiality
  - Grooming
  - Online hate & xenophobia websites
  - Adult pornography

inhope

co-funded by the European Union


# Global Trend Report 9/2004-12/2006

- INHOPE network received 900,000 reports from the general public
- In total 1.9 million reports processed
- 160,000 reports forwarded to law enforcement agencies for action (5,800 per month)

- 21% of all processed reports were about illegal or harmful content (20,000 per month)

- 50% of these Images of Child Sexual Abuse
- 19% of these other child-related content
- Images of Child Sexual Abuse grew by 15% per year

co-funded by the European Union

# 2007 Trends

- All together 900,000 reports processed by the member hotlines
- 6,000 reports per month assessed as potentially illegal – passed to Law Enforcement
- Note – the figures describe the work of the Hotlines – not the actual amount of reported or existing illegal sites



# INHOPE

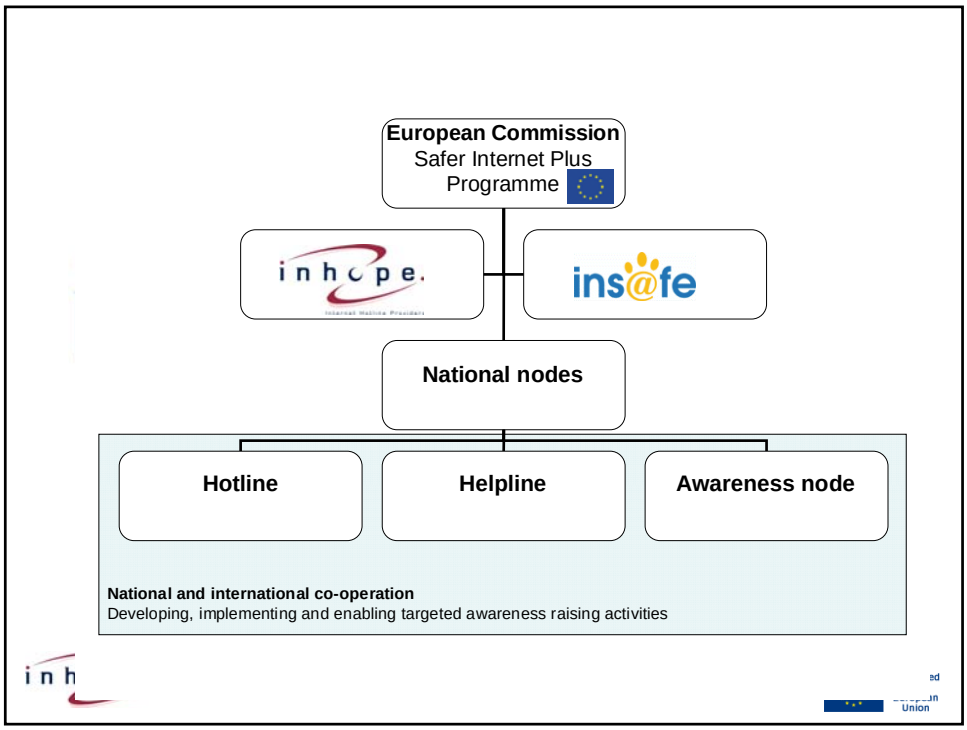
- Provides a single point of contact for global reports of illegal content to initiate global activities
- Offers an easy way for the public to report anonymously suspected illegal sites and content they find on the Internet
- Provides the fastest and most effective way to get the information concerning illegal content to the party that has the best possibilities to take action
- Established procedures and support from various national and international stakeholders

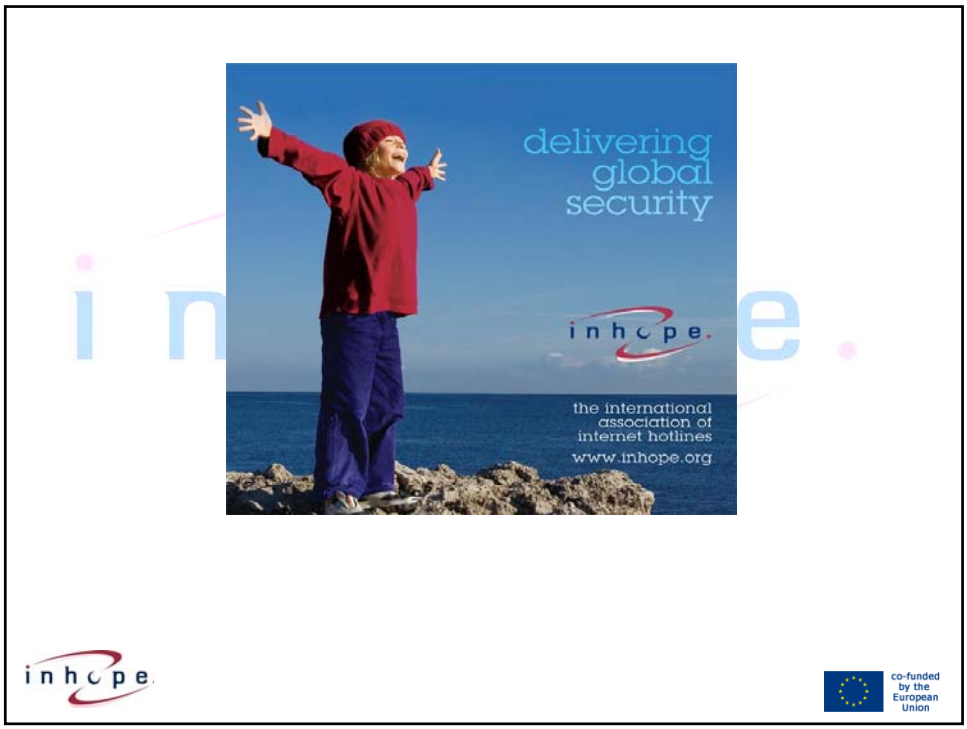
# INHOPE URL Database

- In order to gain more relevant information and to intensify the work a shared URL database will be created
- Reduce duplication of reports passed to Law Enforcement
- Provides a global view of the problem related to the images of child sexual abuse
- More relevant information for developing strategies to tackle the problem

# INHOPE & INSAFE cooperation

- Participating and supporting the Safer Internet Day celebration
- Links and logos
- Exchange of quality information
- Participation in meetings and events
- Joint events and trainings
- Invitations to attend and present at the network meetings
- Strategic meetings
- Regional meetings
- Reciprocal sharing of knowledge


European Commission
Safer Internet Plus
Programme
inhope
ins@fe
National nodes
Hotline
Helpline
Awareness node
National and international co-operation
Developing, implementing and enabling targeted awareness raising activities


delivering
global
security
inhope
the international
association of
internet hotlines
www.inhope.org
co-funded
by the
European
Union

# Contact Details

Adrian Dwyer

Membership Coordinator

Email: adrian.dwyer@inhope.org
Tel: +353 12938860
Mobile: +353 8797 04587

# Youth Protection Roundtable

Jutta Croll and Katharina Kunze

The Internet should be a tool for gathering knowledge and information as well as for entertainment for users of all age groups. But nowadays more and more people are afraid of encountering unwanted and harmful content instead of useful information while being online.

A combination of technical tools with increased effectiveness and approved educational measures seems to be the solution. To develop the ideal mix of both there is a need for collaboration between specialists from the technical side and experts from the pedagogical side. Therefore at the Youth Protection Roundtable relevant players worked together at five bi-annual international meetings within the project's duration of 30 months, from November 2006 till April 2009. Considering the various cultural backgrounds of European countries, the activities emphasised the following elements:

- Facilitate and coordinate the exchange of views between technical experts and children's welfare specialists
- Find a common language
- Conducting a survey on matters of youth protection online and safer Internet
- Enable technicians to take account of the potential effects of newly developed technologies on their safe use by children
- Improve the usability of filtering technologies
- Provide European parents and educators with the information necessary to decide on appropriate content in accordance with their cultural values
- Motivate children's welfare experts to include consultation on supportive technologies into their portfolio
- Identify good practise approaches

At the Youth Protection Roundtable

- youth protection online was defined
- different viewpoints of technicians and welfare experts were identified,
- a common view on risks and measures was developed,
- technical tools for youth protection were identified and reviewed with regard to their effectiveness,
- a new approach to digital literacy was elaborated with input from young people themselves, and
- an European approach on shared responsibility for youth protection online was developed and eventually adopted.

While many fruitful and distinct ideas have blossomed at the Youth Protection Roundtable, one overarching goal was to develop a common strategy embedded in the cultural situation to prevent children and youths from encountering unwanted and harmful content while using the Internet.

This goal was achieved by delivering the

- **YPRT Principles** for the improvement of youth protection online: eight principles to which the YPRT members commit themselves and declare their intent to co-operate on their implementation and dissemination, and the
- **YPRT Toolkit**, a detailed catalogue of references for the improvement of youth protection online.

# Fighting against paedophile activities in the KAD P2P network

Thibault Cholez, Isabelle Chrisment and Olivier Festor

MADYNES - INRIA Nancy Grand Est, France
{thibault.cholez,isabelle.chrisment,olivier.festor}@loria.fr

**Abstract.** In this poster, we present a solution to fight against paedophile activities in KAD. Our distributed architecture can monitor and act on paedophile contents in a very efficient way by controlling keywords and files. Early results on the real network demonstrate the applicability of our approach.

**Keywords:** Honeypot, KAD, DHT

## 1 Motivation

KAD is a part of the **eMule** software and one of the major file sharing P2P networks ($\sim$ 3 millions of simultaneous users). KAD uses a structured architecture called Distributed Hash Table (DHT) to allow users to retrieve a specific file from keywords and the possible sources for a file. Observing users and controlling contents in KAD are real technical challenges:

- Each file and keyword is published on dozens of peers on the DHT, in order to keep the information available.
- As paedophile contents can be referenced through normal keywords, monitoring only files can lead to false positive (normal users considered as paedophiles).
- Attracting users with Honeypots (fake files) is resource consuming because 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).

## 2 Our features to fight against paedophile activities

Thanks to our specific distributed architecture exploiting some KAD weaknesses, we can provide several features helping to study and fight against paedophile activities on that network. Being given the hash of specific contents, we can do:

- **Passive monitoring:** we transparently monitor all the requests sent to the targeted contents in the network. We can discover all the new published files for a given keyword and all the peers sharing a file.

– **Eclipsing content:** we eclipse entries of the DHT to remove the targeted contents from the network and prevent users from accessing it.
– **Index poisoning:** we poison the DHT references with very attractive fake files showing a high number of sources.
– **Promoting Honeypots:** we attract the final download requests for the controlled files towards our Honeypots.

By attracting all the publications and searchs of paedophile contents (keywords and files), our architecture can assess and control the paedophile behavior from the initial search of keyword to the final download.

## 3 Experiments

To test our solution, we eclipsed the good references for the keyword "spiderman" and poisoned them with 4 fake files for one day. The results in figure 1 and 2 show that our architecture is effective and the importance to control the number of sources to build an efficient Honeypot. Our upcoming work consists to deploy our architecture to specifically study and fight against paedophile activities.

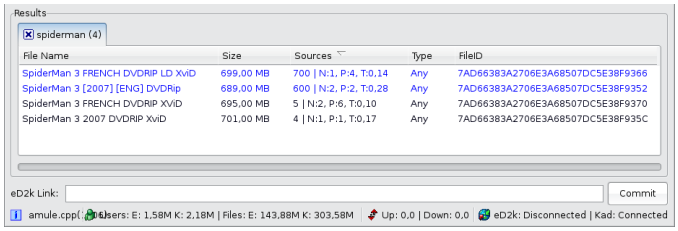



**Fig. 1.** Results of a search for "spiderman" under eclipse and poison (4 fake files)

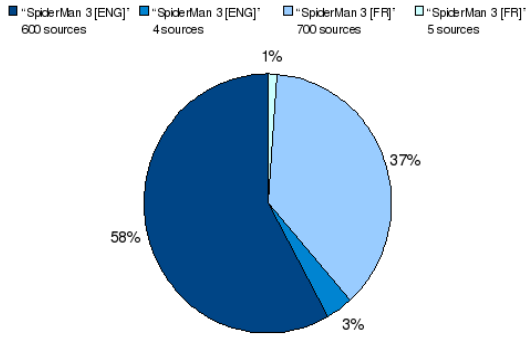



**Fig. 2.** Proportion of download requests received for each fake file

**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.

# Measurement of paedophile activity in eDonkey using a client sending queries

Firas Bessadok, Karim Bessaoud, Matthieu Latapy and Clémence Magnien
LIP6 - CNRS and University Pierre & Marie Curie
firas.bess@gmail.com, karim.bessaoud@complexnetworks.lip6.fr, matthieu.latapy@lip6.fr,
clemence.magnien@lip6.fr

## 1. INTRODUCTION

The observation of peer-to-peer (P2P) file exchange systems is a hot topic covered by several works [1][2]. In our proposal, we mainly focus on observing and analyzing paedophile activity in P2P networks. Our goal is to provide different indicators related to paedophile files found in the eDonkey network. To this end, we used a modified client which automatically sends queries to eDonkey servers and collects information about files sent back by the servers (file-id, name of file, size of file, or users who possess it). The obtained data is then analyzed to gain insight on paedophile activity in the system.

## 2. MEASUREMENT

Our system is composed of a unique client which connects itself to several servers belonging to the eDonkey P2P network. The client restarts after each session of 12 hours. A session is defined by the execution of the client during 12 hours. In each session, the client sends 15 different queries (8 of them are well-known paedophile keywords) one by one every 6 minutes. Once the answers (file-id, filenames or file sizes) are received, the client formats them into XML files. We run this measurement during 140 days from October 2008 to February 2009.

## 3. OBSERVATIONS

We collected 2 784 583 distinct files during this experiment. Later on, we computed many statistics on these files. For instance, the number of files found in each session, the distribution of ages in filenames, and also the number of paedophile files and other non-paedophile. We present some of the obtained results below.

### 3.1 Number of file-id

Figure 1 shows the evolution of the number of distinct file-id (vertical axis) observed during our measurement as function of the number of sessions executed : similar to the time elapsed since the beginning of this measurement (horizontal axis).

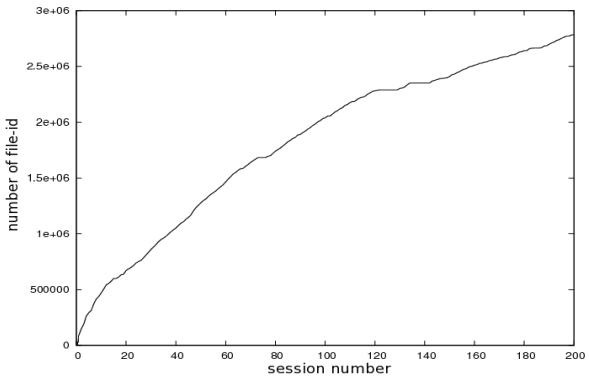



**Figure 1: Evolution of the number of file-id observed during the measurement.**

We note that the plot is growing, as expected, but in a non-linear way. In the first twenty sessions, the curve has a large slope because most file-id have not yet been seen. But after that, the slope decreases gradually due to the fact that most file-id have already been seen. On the other hand, this slope remains significant. We therefore conclude that the number of file-id present in the eDonkey network is so great that we can't see all of them with our measurements, even if conducted for long period of time (140 days here).

### 3.2 Number of paedophile files

Figure 2 presents, the evolution of the number of distincts paedophile file-id (vertical axis) observed during our measurements as a function of time (horizontal axis) represented by the number of sessions. We assumed that a file is a paedophile one if its name contains at least one paedophile keyword.

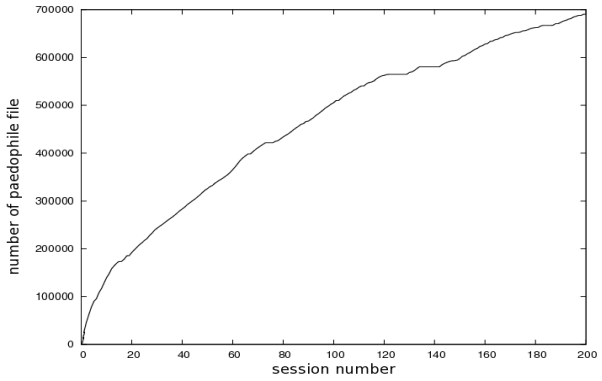



**Figure 2: Evolution of the number of paedophile files found during the measurements.**

Here we found that this curve is growing like Figure 1. Similary, we can't discover all paedophile files during our measurement. Here, we detected 701 857 paedophile files.

### 3.3 Ages contained in filenames

Figure 3 represents the distribution of ages found in the filenames obtained during our measurement. It describes the percentage of the occurences (vertical axis) for each age (horizontal axis) in the filenames viewed during our measurement. There are 77 030 filenames that contains an age.

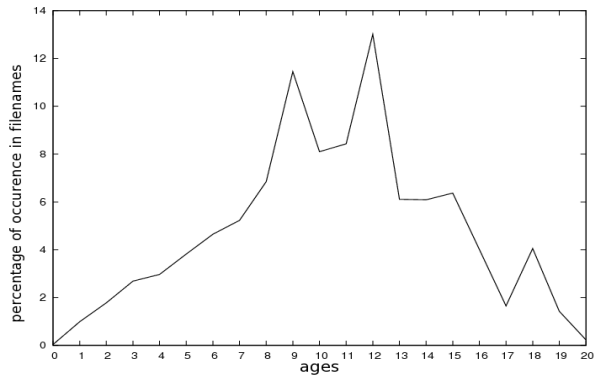



**Figure 3: Ages distribution in filenames.**

We observe that the interval of ages ranges from 0 years old up to 20, with an important focus between 8 and 15 years old and two peaks, at 9 and 12 years old.

## 4. ONGOING AND FUTURE WORK

We are planning to perform a new experiment using multiple distributed clients using the PlatnetLab platform and with one session. These measures will give partial views of the network. We plan to use them to estimate the real value of several parameters such as : the number of paedophile files and the number of copies of a particular paedophile file. We will use the multiple-recapture model [3] for these estimations. The multiple-recapture model is used to estimate the unknown size of a population using multiple samples. Until now, we have installed our client on 94 PlanetLab machines. We have collected measures simultaneously from all these machines. These collected measures are under study. The next step is to analyse the results in order to determine if there is a positive or negative dependency between the results obtained by each client and to determine if the results are homogeneous.

## 5. ACKNOWLEDGEMENT

This work is supported by the European MAPAP (SIP-2006-PP-221003) and the French ANR/MAPE projects.

## 6. REFERENCES

[1] F. Aidouni, M. Latapy, and C. Magnien. Ten weeks in the life of an edonkey server. *Proceedings of HotP2P'09*, 2009.

[2] E. Adar and B.A. Huberman. Free riding on gnutella. *First Monday, vol. 5*, 2000.

[3] Z. Schnabel. The estimation of the total fish population of a lake. *Am Math Monthly*, 1938.

# Detecting keywords used by paedophiles

Christian Belbeze
Université de Toulouse, Institut de Recherche en Informatique de Toulouse. Email: christian@belbeze.com
Matthieu Latapy
LIP6 – CNRS and UPMC. Email: matthieu.latapy@lip6.fr

## Abstract

We propose here a method to compute sets of strongly related keywords from logs of user queries. This method relies on the construction and analysis of a huge (weighted) (directed) graph, from which we extract aggregates of words. The challenge which we address is to obtain relevant aggregates with reasonable computational costs.

Keywords : Keyword, log file, search engine, cluster, aggregate, graph.

## 1. From a log file to a directed weighted graph

Given a large amount of queries (170 millions) entered by users searching for files in a P2P network (log file and experimentation conditions to get it are described in the article “Ten weeks in the life of an eDonkey server”*) , we create a graph. In this graph nodes are keywords and links are co-occurrence links: two words are linked together if they appear in the same query. Therefore only queries including more than one keyword are considered. The obtained graph has 2.8 million nodes (keywords) and 68 million links.

In the graph each element, node and link, has an associated **weight**. This weight is the number of times that the node or link has been found in the log file (excluding one keyword queries). We denote by $\mathbf{W_A}$ the weight of node A and by $\mathbf{W_{AB}}$ the weight of the link between A and B.

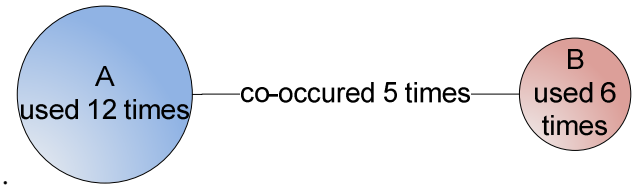



**Figure 1:** From log file to a graph: weighted nodes and link.

We now turn the obtained weighted undirected graph into a directed version as follows. Given two nodes A and B we define the Coefficient of Reliability (CR) from A to B, denoted by

$$CR_{A\longrightarrow B} = \frac{w_{AB}}{w_A}$$

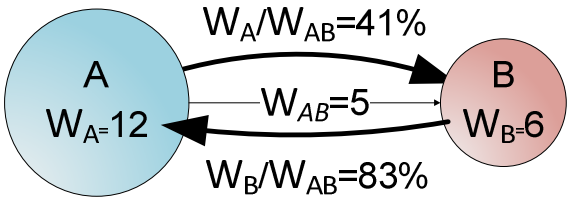
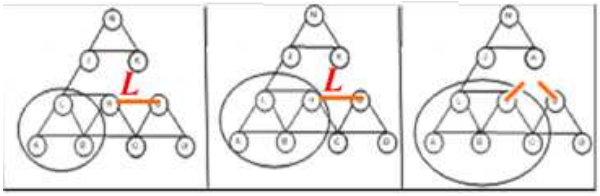



**Figure 2:** Directed version of the graph.

## 2. First method

The graph defined above encodes much information on the relations between words entered by users in their queries. Our goal now is to use this encoding to exhibit groups of words, called aggregates, with particularly strong relations. To achieve this, we will proceed as follows. We first compute the triangles in the graph (i.e. sets of three words with all these possible links between then) which constitute our initial aggregate. We then grow aggregates by merging triangles following these rules:




*L* is **removed** if $CR_{A\rightarrow B}$ **And** $CR_{B\rightarrow A}$ < **Vdw** [Too weak to keep]
But *L* is **kept** if $CR_{A\rightarrow B}$ **Or** $CR_{A\rightarrow B}$ > **Av** [Too big to ignore]

**Figure 3:** aggregate triangles and removing links.

- **Aggregates must remain biconnex :** A biconnex graph is a graph where each node is connected by at least two paths to any other node of the graph.
- **Links with a CR lower than a predefined minimal Value (Vdw) for the two nodes are removed except if one of its CR is higher than an Activation value (Av), see figure above.**

### Experimentation and results.

Applying this method produces a massive aggregate containing typically more than 2 000 000 of words. This aggregate is too large to have any sense. The reason of that is that many links concern very rare words. Indeed than 50% of keywords are used once or twice. These links are always kept because their **CR is at least 50%.**

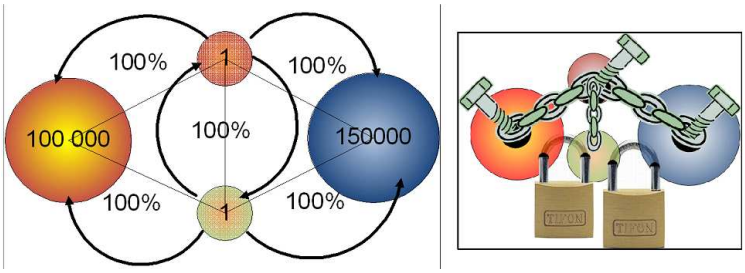



**Figure 4:** Rare words are too strongly linked with others words and especially with frequent words.

## 3. Improved method

Removing rare words could be the basic solution to improve the method above. But this would not really make sense in pursuing our goal. We want to keep the opportunity to get rare and very used words in aggregates. To make this possible, we propose an algorithm that limits the size of aggregates. It relies on adapting parameters when aggregates reach a maximal defined size.

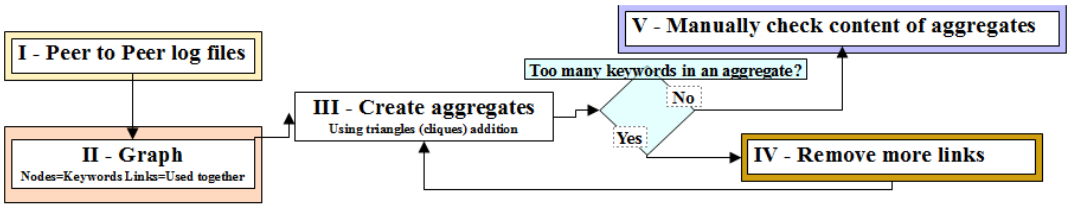



**Figure 5:** Aggregation including size limitation principle

Using the same algorithm as above, we first define a maximum aggregate size (**MAS**). Then we adapt 4 criteria used in the algorithm to keep aggregates size under **MAS** value. We defined for each of the four criteria, a start value and a final value. We defined too a number of step (**NbSteps**) to reach final criteria values. Each time aggregate reaches the maximum predefined **MAS** size we will increase Value Double Way (**Vdw**) and Activation Value (**Av**) to remove more links. At this step, we also modify the minimum (**Min valid weight**) and maximum (**Max valid weight)** weight of a node. We increase minimum weight and decrease maximum weight of node using specific formulas. By this way we will get limited number of keywords in each aggregate.

| Parameters | Start value | Final value | New incremented values **IStep**=**IStep**+1 |
|---|---|---|---|
| **Vdw** | 3 | 10 | 3 + 7 * (**IStep** / **NbSteps**) |
| **Av** | 10 | 51 | 10 + 41 * (**IStep**/ **NbSteps**) |
| **Min valid weight** | **Min(G.weight)** 1 | **Avg(G.weight)** 70 | **Avg( G.weight )** ^ (**IStep**/**NbSteps**) |
| **Max valid weight** | **Max(G.weight)** 328000 | **Avg(G.weight)** +1 71 | Max(**G.weight** ) ^ ((**NbStep-IStep**)/**NbStep**) + **Avg(G.weight)** |

**IStep:** Number of times aggregate reached the maximum size (**MAS**); **Max(G.weight):** Maximum number of use for a word in the graph; **Min(G.weight):** Minimum number of use for a word in the graph; **Avg(G.weight):** Average weight of keywords in the graph.

**Figure 6:** Limits and step modification. Numeric values are the ones use in the experimentation.

## Experimentation and results.

We chose to use 50 steps and a size limit of 80 keywords by aggregate. Starting with a set of well known keywords the improved method creates aggregates including these keywords. More the keywords' weight decreases and more the number of aggregates, including the keyword, decreases too (figure 7).

| Keywords | Weight | Number of aggregates | Max Size | Avg Size | Min Size |
|---|---|---|---|---|---|
| pthc | 45737 | 96 | 78 | 9 | 3 |
| incest | 13609 | 70 | 52 | 11 | 3 |
| ygold | 9183 | 19 | 61 | 15 | 3 |
| ptsc | 3189 | 14 | 11 | 6 | 3 |
| incesti | 1277 | 2 | 4 | 3.5 | 3 |
| inceste | 1220 | 3 | 17 | 12 | 7 |
| 4yo | 1042 | 4 | 14 | 9 | 4 |
| 3yo | 832 | 3 | 12 | 10 | 8 |

**Figure 7**: Aggregates including 8 well known words.

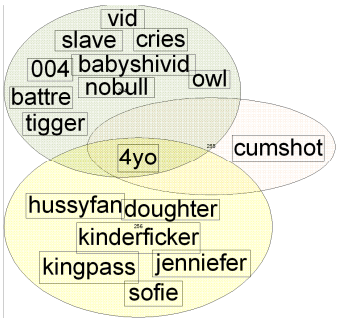



**Figure 8**: Sample of aggregates.

Figure 8 presents parts of aggregates including the word "4yo".

# CONCLUSION

We presented a method to create aggregate of keywords from user queries encoded in a large graph. The method presents the advantage to keep a reasonable and a predefined maximum size for each aggregate. It produces seemingly relevant results, while keeping its computational cost reasonable. Assessing and refining the results is the next step to investigate.

# REFERENCES

*Aidouni, F. and Latapy, M. and Magnien, C. Ten weeks in the life of an eDonkey server, Sixth International Workshop on Hot Topics in Peer-to-Peer Systems (Hot-P2P 2009), May 29, 2009, Rome, Italy.

Cui, H., Wen et al 2002, Probabilistic query expansion using query logs. Proceedings of the eleventh international conference on World Wide Web, pp. 325 – 332.

Fu, L. et al , 2003, Collaborative querying through a hybrid query clustering approach, 2003, Digital libraries: Technology and management of indigenous knowledge for global access, ICADL, pp. 111-122..

Gangnet, M. and Rosenberg, B., 1993, Constraint programming and graph algorithms, Annals of Mathematics and Artificial Intelligence 8(3–4): 271–284.

Hoffmann, C. et al, 2000, Decomposition plans for geometric constraint systems, Proc. J. Symbolic Computation 2000.

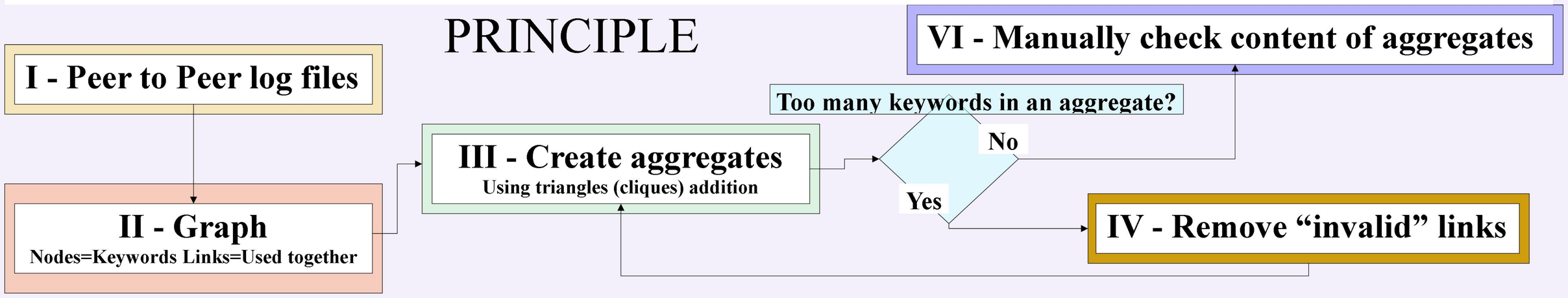
Latapy, M., 2007. Grands graphes de terrain – mesure et métrologie, analyse, modélisation, algorithmique. Habilitation `a diriger des recherches, Université Pierre et Marie Curie, Paris, France.

# DETECTING KEYWORDS USED BY PAEDOPHILES

Christian Belbeze Université de Toulouse, Institut de Recherche en Informatique de Toulouse. Email: christian@belbeze.com
Matthieu Latapy LIP6 – CNRS and UPMC. Email: matthieu.latapy@lip6.fr

We propose here a method to compute sets of strongly related keywords from logs of user queries. This method relies on the construction and analysis of a huge (weighted) (directed) graph, from which we extract aggregates of words. The challenge which we address is to obtain relevant aggregates with reasonable computational costs.

## PRINCIPLE

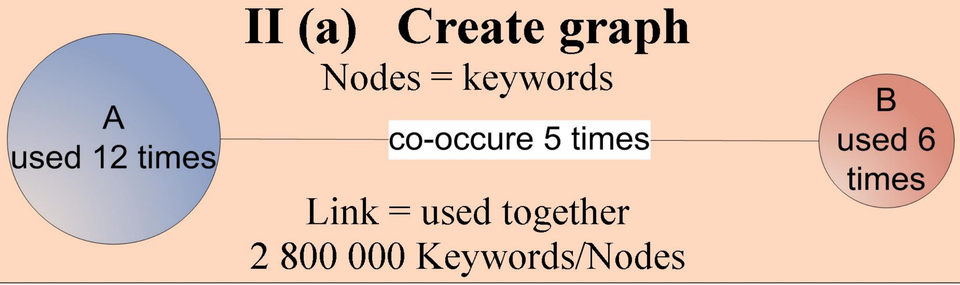



## I Peer to Peer log file

| UserID | Date | Time | Mots |
|---|---|---|---|
| 15 | 12/05/2008 | 08:25:12 | 2125 2255 589 1351 |
| 16 | 12/05/2008 | 08:26:45 | 545 54545 54588 565 |
| 17 | 12/05/2008 | 08:27:00 | 65328 6548 5684 |
| 18 | 12/05/2008 | 08:28:01 | 86 3213 |
| 19 | 12/05/2008 | 08:28:02 | 2669 6933412 3226 |

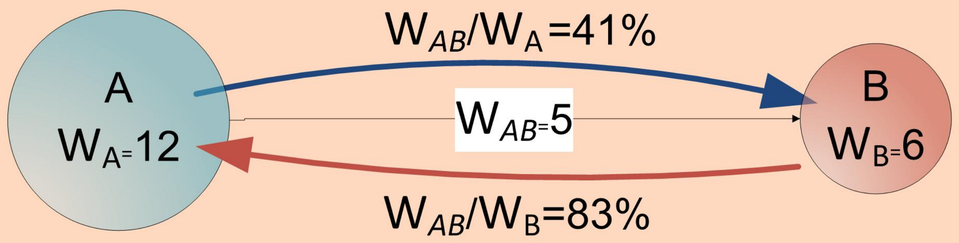
## II (a) Create graph

Nodes = keywords




Link = used together
2 800 000 Keywords/Nodes

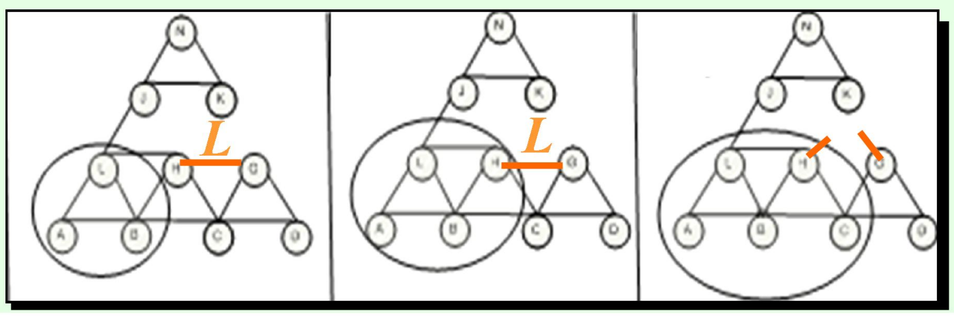
## II (b) Create an *oriented* graph




Coefficient of Reliability of Bond (CRB) of the word A towards the word B noted $\mathbf{CRB_{A\to B}}$ is calculated as

$$\mathbf{CRB_{A\to B}} = \frac{w_{AB}}{w_A}$$

## III (1) Create aggregates by connected triangles and remove weak links

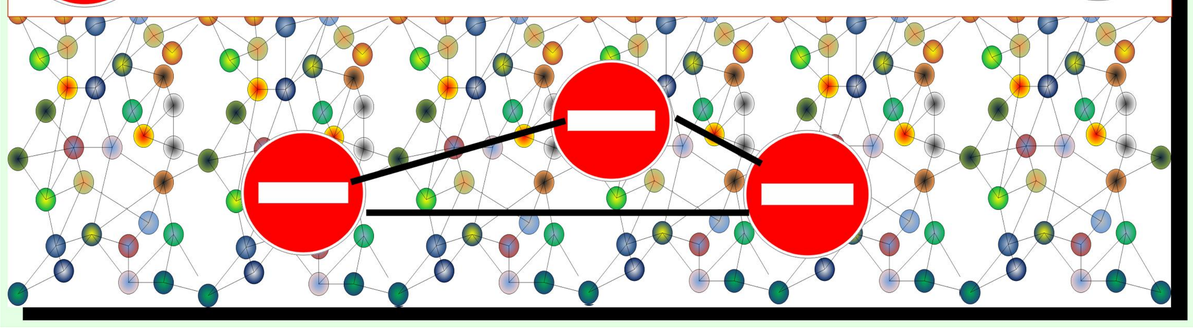



*L* is ***removed*** if
$\mathbf{CRB_{A\to B}}$ **And** $\mathbf{CRB_{B\to A}}$ < **Vdw** [Too weak to keep]

But *L* is ***kept*** if
$\mathbf{CRB_{A\to B}}$ **Or** $\mathbf{CRB_{B\to A}}$ > **Av** [Too big to ignore]

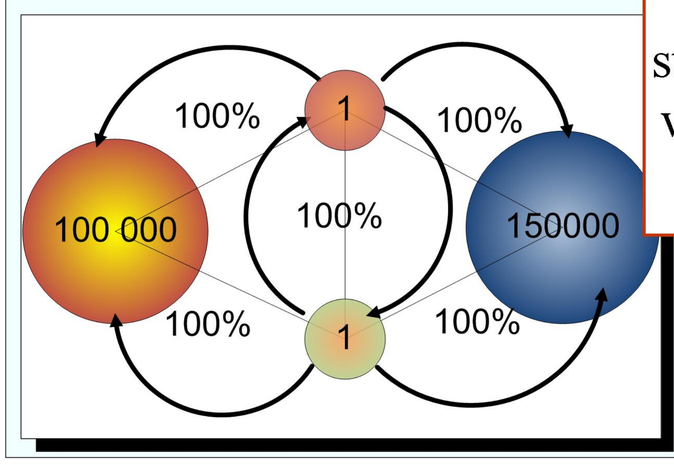
### Result

One aggregate of 2 000 000 keywords

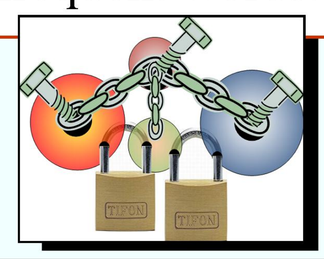


## Too many keywords in an aggregate ?

**why upgrading Av and Vdw is not efficient?**




Rare words are too strongly linked with others words and especially with frequent words.



## III (2) Create aggregates with a limited number of keywords by changing parameters of aggregation

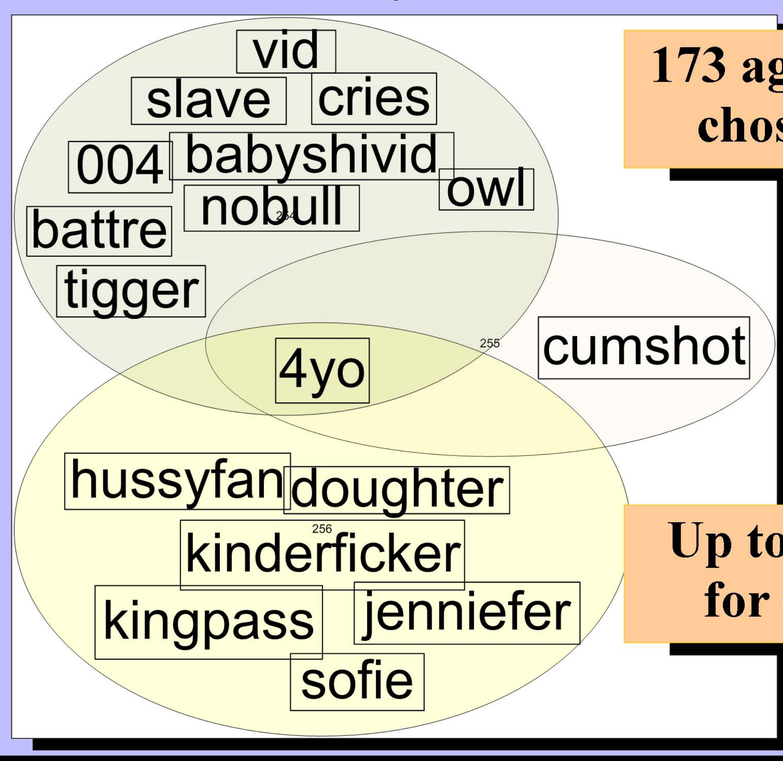
### IV - Remove "invalid" links

| | |
|---|---|
| Av | Activation value of CRB to valid the link in anyway |
| Avg(G.weight) | Average of weight of keywords in the graph |
| IStep | Number of times aggregate reached the maximum size |
| Max(G.weight) | Maximum number of use for a word in the graph |
| Min(G.weight) | Minimum number of use for a word in the graph |
| NbSteps | Number of increments to reach parameters' final value |
| Vdw | Minimal CRB value for both nodes to valid the link |

| Parameter | Start value | Final value | New incremented values IStep=IStep+1 |
|---|---|---|---|
| **Av** | 10 | 51 | 10 + 41 * (IStep/ NbSteps) |
| **Max valid weight*** | Max (G.weight) | Avg (G.weight) +1 | Max(G.weight) ^ ((NbSteps-IStep)/NbStep) + Avg(G.weight) |
| **Min valid Weight*** | Min (G.weight) | Avg (G.weight) | Avg(G.weight) ^ (IStep/NbSteps) |
| **Vdw** | 3 | 10 | 3 + 7 * (IStep / NbSteps) |

*Word with a number of use lower than « Min valid Weight » or higher than «Max valid Weight» are ignored.

## VI - Manually check content of aggregates




**173 aggregates for 18 chosen keywords**

**Up to 96 aggregates for one keyword**

# Tracing paedophile eDonkey users through keyword-based queries

Raphaël Fournier, Guillaume Valadon, Clémence Magnien, Matthieu Latapy
LIP6 – CNRS and University Pierre & Marie Curie
104, avenue du Président Kennedy, 75016 Paris, France

## 1. MOTIVATION

Recent studies showed that Internet bandwith is now massively dedicated to *peer-to-peer* (P2P) exchanges. Thus, it is important to precisely examine this new way of transmitting data and the kind of content that is exchanged. In particular, authorities are willing to obtain reliable information on paedophile activity on these networks, in order to fight against cybercriminality[1].

One of the most prominent P2P networks is eDonkey, a semi-centralized system. Users connect to servers and submit content queries, servers return lists of files, then users exchange files directly. A study [1] was designed to record, all the exchanges between an eDonkey server and the connected peers. The exchanges collected include:

- keyword-based search queries submitted to the server and the server's answer consisting in lists of files;
- specific file requests and server's answers (lists of users sharing the file).

The experiment lasted ten weeks without interruption. There were 127 million queries submitted to the server, by 28 million IPs. As a first and rather rough definition, we consider that a user uses only one IP and that an IP is not shared by several users. The measures on paedophile activity were based on a set of 21 keywords. Their "paedophile" nature was previously assessed by co-occurence studies [2]. We consider that a query is paedophile if it contains at least one of these words – we call it model "PQ". An IP is considered as paedophile as soon as it submits such a paedophile query.

## 2. GOALS

This study aims at establishing some accurate facts about paedophile users on the eDonkey server. Above all, counting paedophile users is our priority. Thus, the study will first require to clearly define what a paedophile user is and what makes a query a paedophile one.

[1]This work is supported by the European MAPAP (SIP-2006-PP-221003) and the French ANR/MAPE projects.

## 3. RESULTS

### Number of queries by IP (Fig. 1)

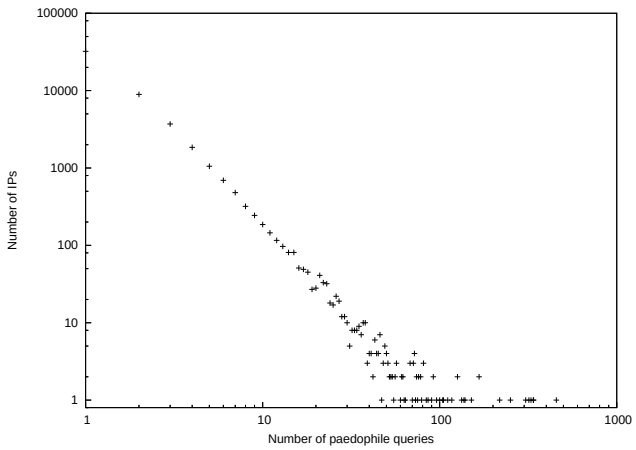



Figure 1: **Distribution of the number of paedophile queries by paedophile IP**, *i.e.* for each encountered number of paedophile queries, the number of IPs which submitted this number of queries.

The distribution is highly heterogeneous: more than 66% of the total paedophile IPs submit only one query – and 94% less than 5 –, while some IPs submit up to 456 queries within ten weeks. This figure raises the question of the characterization of a paedophile IP: is a single query within 10 weeks enough to consider the IP as paedophile ? 456 queries in the overall experimentation means more than 6 paedophile queries a day, is it the behavior of a very active human user or of a robot ? Our underlying assumptions on the way eDonkey clients work are also called into question here: are queries sometimes automatically re-submitted to the search engine ?

This distribution is crucial to have a good model to count paedophile users with a low false-positive rate.

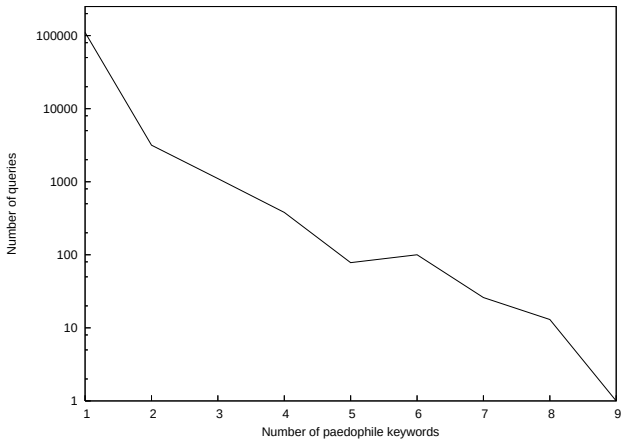



Figure 2: **Distribution of the number of paedophile keywords in queries**, *i.e.* for each number of paedophile keywords from our set contained in queries, the number of queries (vertical axis), with logarithmic scale.

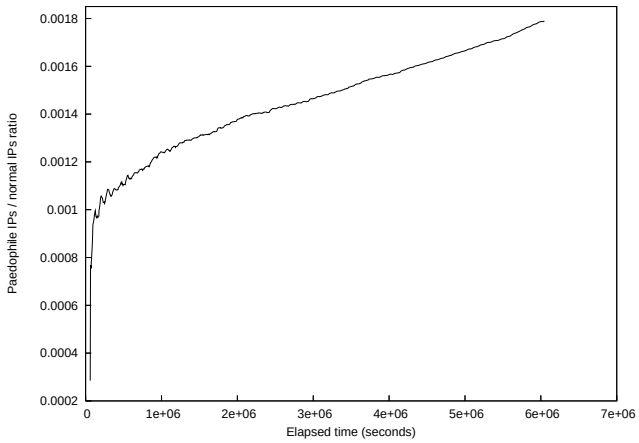



Figure 3: **Time-evolution of the ratio of paedophile IPs over normal IPs.**

## Number of keywords in queries (Fig. 2)

The distribution shows that, among queries containing at least one paedophile keyword, most of them contains a single keyword. These results raise the question of the correct definition for a paedophile query. Our first model (PQ) might be improved to eliminate queries with more than $x$ paedophile keywords.

The table below shows the number of paedophile queries, for two definitions of a paedophile query. In row A, the number of queries containing exactly x paedophile keywords. In row B, the number of queries containing at least x paedophile keywords.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| A | 110614 | 3166 | 1105 | 380 | 78 | 100 | 26 | 13 | 1 |
| B | 115483 | 4869 | 1703 | 598 | 218 | 140 | 40 | 14 | 1 |

## Paedophile IPs / normal IPs ratio (Fig. 3)

If we now consider NAT or dynamic IP allocation, an IP may remain paedophile until the end of the ten weeks, even if it is not the same user anymore. This kind of contamination may lead to eventually consider every IP as paedophile. The growing ratio (see Fig. 3) shows that, even by the end of the experiment, we still discover many new paedophile IPs in queries.

## Total number of queries by keyword (Fig. 4)

One may notice several groups of keywords, sorted by frequency of appearance. But this plot also shows that a specific keyword (the right most one) is used in slightly less than 60% of all the paedophile queries. The gap between this paedophile word and others suggests that it may be a not-so-specific keyword. Further investigation, such as combination of this keyword with others,

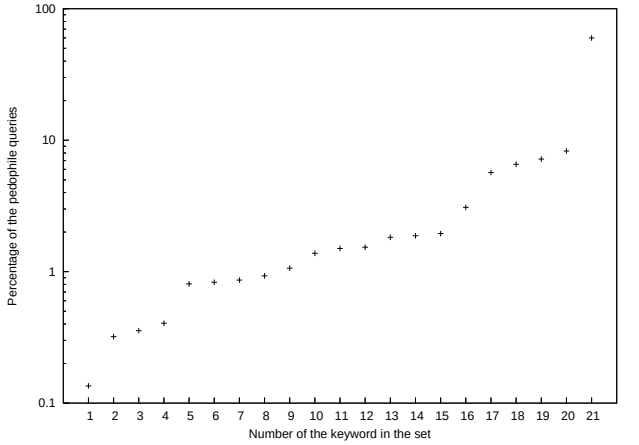



Figure 4: **Number of queries by paedophile keyword.**

must be carried out to decide whether all queries containing this keyword should be considered paedophile.

# 4. FUTURE WORK

Most of the future work will consist in studying combination of keywords, so as to corroborate our hypothesis on the model (specific paedophile keywords). A deeper work on ages will also be carried out. We will progressively refine our definition of paedophile queries and IP, before establishing accurate and reliable results.

# 5. REFERENCES

[1] F. Aidouni, M. Latapy, and C. Magnien, "Ten weeks in the life of an edonkey server," *Proceedings of HotP2P'09*, 2009.

[2] M. Latapy, C. Magnien, and G. Valadon, "First report on database specification and access including content rating and fake detection system," *http://antipedo.lip6.fr*, 2008.

# Poster Presentation

## Adverse Effects of Cyber Laws on Online Child Porn Detection and Prevention

Philippe Langlois

**Abstract:**

Policy makers are pressured into passing new cyber law for various agenda. We will present in this talk how the current trend in cyber security and copyrights law affecting the internet might have a detrimental impact on the fight against Online Paedophile Activity. We will have a specific look on laws such as the proposed HADOPI law (Internet et Création) in France and other legislation and try to see the organic effects of such law. Particular attention will be given to the importance of cryptography as a privacy tool and the needle-and-stack situation that may arise if massive "gray" content is mixed with "black" content.

International Conference

# Advances in the Analysis of Online Paedophile Activity

Paris, France; June 2-3, 2009

---

## Organization Committee

- Sean Hammond, University College Cork (Ireland)
- Matthieu Latapy, CNRS - UPMC (France)
- Clémence Magnien, CNRS - UPMC (France)
- Raphaëlle Nollez-Goldbach, CNRS - UPMC (France)
- Massoud Seifi, CNRS - UPMC (France)
- Vasja Vehovar, University of Ljubljana (Slovenia)
- Agnieszka Wrzesie, Nobody's Children Foundation (Poland)

**An International Conference organized by the Antipaedo Project**

### Co-funded by:

- Laboratoire d'Informatique de Paris 6 (LIP6), Université Pierre et Marie Curie (UPMC) et Centre National de la Recherche Scientifique (CNRS), France
- COPINE Project, University College Cork, Ireland
- Nobody's Children Foundation, Poland
- University of Ljubljana, Slovenia
- Safer Internet Plus programme, European Union

**The European Union**