

MEASUREMENT AND ANALYSIS OF P2P ACTIVITY AGAINST PAEDOPHILE CONTENT

<http://antipaedo.lip6.fr>

Coordinator: Matthieu Latapy
Matthieu.Latapy@lip6.fr

Current Situation

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 [1], [2], [3], [4], [5], [6]. A report from the *United States General Accounting Office* in 2003 [1] concludes that "child pornography is easily accessed and downloaded from peer-to-peer networks".

A French working group (composed of administrations representatives, user associations and relevant economic actors) published in 2005 a study about paedophile contents available on the internet [2]. It reveals that French law-enforcement authorities typically observe 10 to 20 persons engaged in significant paedophile p2p exchanges per day *in France*. According to another report on child protection [3], written at the request of the french minister for family, the number of files with paedophile content available via p2p systems would be between 200 000 and one million.

This can easily be checked by any user, since a simple query on the keywords *porn* or *pedo* with a classical p2p client leads to hundreds, and up to several thousands, of answers.

The presence of such content, and its very easy access, make the current situation particularly worrying for p2p users, in particular children. Indeed, a significant number of children, in particular teenagers, nowadays use p2p systems [7], [3], [8], [1], [9], [5]. According to the 2005 Eurobarometer Survey on Safer Internet [7], 50% of the children of the European Union have access to the internet.

A study conducted in 2003 in France [3] established that 31% of children having access to the internet were using p2p systems. The presence of harmful contents in these systems, in particular paedophile ones, therefore constitute a worrying danger for a significant proportion of European children [8], [1], [9], [5]. Parents are in part aware of this situation: 69% of european parents believe their child has been exposed to harmful or illegal content on the internet [7].

This is even more alarming if one considers the fact that many fakes, *ie* files with contents that differ significantly from their names, are present in these systems. Because of this, all users, including children, face a high risk of downloading and visualising unwanted content¹. [8], [1], [9].

It is clear that viewing paedophile contents can be harmful for adults. Apart from the shock experienced by most users at the sight of such pictures, it is suspected that easy and/or unwanted access to paedophile content may increase or even create the user's interest for such contents². Also, a non negligible percentage of viewers of paedophile contents are paedophiles having already had sexual intercourse with children. The wide presence of paedophile content in p2p systems make these

people feel safe and unattainable in these systems, and leads to a trivialisation of such content³.

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. Actually, and despite some progress has been done thanks to the studies cited above, there is still an important lack of precise knowledge on this topic. It has been observed at many occasions that this has a deep impact on our ability to fight these exchanges [2], [10], [3]. For instance, the report written in 2005 at the request of the french minister for family [3] established the urgent need for studies of this phenomena, in order to understand better what is going on, help parents protect their children from unwanted content, and design appropriate tools for protecting children on the internet. This report emphasised the need for a watch, coordinated at the European level, to monitor not only the evolution of children's uses of the internet, but also the evolution of the risks they incur.

Objectives

The objective of this project is to tackle these issues by implementing key software, setting up reference databases and conducting leading studies, both to protect p2p users, in particular children, and help law enforcement authorities and other child protection organisations in their task. More precisely, we will focus on the following three areas, each with its own objectives.

Content rating and fake detection system

Our core objective is the design and implementation of a service able to give, for any file encountered in our measurements, a rating of its content as paedophile and/or pornographic, as well as an indication of the fact that it may be a fake or not. A confidence ratio will be associated to each of these indications. This service will be available on-demand to end-users through a web page form, but its use will be limited to avoid abuses (typically, we will limit the number of queries per user and per time unit in order to prevent users from searching paedophile content with it). A full unrestricted version will be provided to relevant institutions, with additional information like the date of first appearance of the content, the number of peers providing/downloading it during time, etc.

Such a tool would be a first step towards the possibility for ISP to filter p2p content, and for end-users to have indications on the content of a file they are interested in, before downloading it⁴. It may also be included in parental control systems and in p2p clients, which may send automatic queries to our system when needed. This would allow a significant reduction of exposure of p2p users, in particular children, to harmful content.

Paedophile keywords

One may identify three different kinds of paedophile keywords: the basic ones that anyone would think of to find paedophile content, more specific ones known mainly by people with experience in handling paedophile content (like paedophiles themselves and law enforcement personnel), and hidden, short-term keywords known only by small groups of people (who exchange these keywords in chat systems or other interpersonal communications). Identifying paedophile keywords therefore is a key issue for filtering, as well as law enforcement. It is also necessary to send appropriate queries to p2p systems for the measurement of paedophile activity. An objective of the project therefore is to use huge amount of recorded queries and file names to uncover such keywords, including hidden ones that serve only for short periods of time.

This will result in a dynamic list of paedophile keywords, that will evolve during time, which we plan to send to law enforcement authorities and a restricted set of other relevant institutions⁵. This list will contain detailed information on the keywords, like their frequency during time, the other keywords with which they appear, their date of first appearance, etc.

Improved knowledge of paedophile activity

Our objective here is to give 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 also seek more subtle information, like studies of how users develop an interest in paedophile content, global maps of paedophile contents, including their nested *community* structures, and methods to make the difference between people that probably download paedophile content accidentally and people that focus on such contents.

The objective here therefore is to obtain rigorous and deep enlightenment on p2p paedophile activity, which will lead to the publication of detailed reports on each aspect, as well as both technical and general public synthesis reports at the end of the project. We want to change the current situation into a situation in which we have a precise knowledge of paedophile activity in p2p systems.

References

- [1] United States General Accounting Office. *File sharing programs - peer-to-peer networks provide ready access to child pornography*, 2003.
- [2] *Recommandation du forum des droits sur l'internet : Les Enfants du Net (2) Pedo-pornographie et pedophilie sur l'internet*, 2005.
<http://www.foruminternet.org/telechargement/documents/reco-enfance2-20050125.htm>
- [3] *Protection de l'enfant et usages de l'internet*, 2005. Rapport remis au ministre en charge de la famille.
- [4] Le Bouclier. *Study: Peer-to-peer: Sharing pedophilia*. <http://www.bouclier.org/dossier/318.html>
- [5] Jerry Ropelato. *Peer-to-peer pornography - kids know, do mom and dad?*
<http://www.familysafemedia.com/peer-to-peer-pornography-ki.html>
- [6] Parent2parent organisation. <http://wiki.morpheus.com/~p2punited/parents.php>
- [7] *Eurobarometer survey on safer internet*, 2005. Survey commissioned by the Directorate-General Information Society and Media.
- [8] *Recommandation du forum des droits sur l'internet: Les Enfants du Net (1) Les mineurs et les contenus prejudiciables sur l'internet*, 2004.
http://www.foruminternet.org/telechargement/documents/reco-enfance1-20040211_2.htm
- [9] *Fbi cyber education letter*. <http://www.fbi.gov/cyberinvest/cyberedletter.htm>
- [10] Thierry Breton. *Chantier sur la lutte contre la cybercriminalite*, 2005. Rapport remis à Monsieur le Ministre de l'Interieur, de la Securite Interieure et des Libertes Locales.
- [11] M. Taylor, E. Quayle, and G. Holland. Child pornography, the internet and offending. ISUMA, *The Canadian Journal of Policy Research*, 2:94-100, 2001.
- [11] E. Quayle and M. Taylor. Child pornography and the internet: Perpetuating a cycle of abuse. *Deviant Behavior*, 23 (4):331-362, 2002.
- [11] The fifth estate: Landslide: *PORNOGRAPHY IS LINKED TO ABUSE*.
<http://www.cbc.ca/fifth/landslide/abuse.html>.

- 1 On this subject, the report from the *U.S. General Accounting Office* [1] warns that "when searching and downloading images in peer-to-peer networks, juvenile users face a significant risk of inadvertent exposure to pornography, including child pornography"
- 2 In some cases it may even encourage people to try and have sexual intercourse with children [2], [11], [12].
- 3 In some cases, paedophile pictures are used by paedophiles to lure children into thinking that sexual intercourse between adults and children is normal [2], [13], [12].
- 4 The use of this system may be outlined as follows: when a user finds with his/her P2P client a file that seems to fit his/her interests, the P2P client automatically provides him/her with the hash code for this file (this is a common feature of P2P clients); the user then enters this code into our system (through its web page), that tells him/her if this content seems safe, or on the contrary probably is harmful (pornographic or paedophile), depending on the case.
- 5 The dissemination of this list will be carefully controlled, since it would also be a valuable resource for paedophiles themselves

Partners and Support

This work is conducted by the following institutions:

CNRS and University Pierre and Marie Curie

Contact: Clémence Magnien
(clemence.magnien@lip6.fr)

University College Cork

Contact: Ethel Quayle
(E.Quayle@ucc.ie)

Nobody's Children Foundation

Contact: Agnieszka Wrzesie
(awrzesien@fdn.pl)

University of Ljubljana

Contact: Vasja Vehovar
(vasja@ris.org)

INRIA Lorraine

Contact: Isabelle Chrisment
(ichris@loria.fr)

In addition to these institutions, it is co-funded by:

European Union, Safer Internet Plus Programme

Agence Nationale de la Recherche (ANR)