A Pseudonymous Peer-2-Peer Review System for Child Protection On-line [*]

T. Martin, C. Durbin, M. Pawlewski, and **D. Parish**
Loughborough University,
e-mail: {thomas.2.martin, chris.durbin, mark.pawlewski}@bt.com
e-mail: d.j.parish@lboro.ac.uk

Abstract. Children are using the internet more and more, and from a younger age. This is despite the commonly known dangers of predators. There is no policing of the internet, nor would it be possible to instigate. Parents are in the difficult position of trying to monitor and control their children's internet usage, when more often than not the children know the technology better than they do. This can lead to either ineffective measures, or measures that the children will deliberately circumvent for their own privacy. There are also technical issues that are far from trivial. The problem of distinguishing the dialogue of a child from a predator pretending to be a child is extremely difficult. This paper presents a solution which can accurately identify threats, while satisfying the apparently conflicting needs for safety of, and privacy for, the children.

1. Introduction

This paper looks at the problem of protecting children from on-line stalkers/predators. A recent survey of 1,500 children (aged 10-17) in the United States found that approximately 1 in 7 (13%) received unwanted sexual solicitations, and 34% communicated online with people they did not know in person (Wolak, Mitchell & Finkelhor, 2006). This often took the form of crude or vulgar comments in chat rooms - the victims were not bothered and handled the situation well. However some victims felt traumatised and some are targets of aggressive online solicitations (Mitchell, K., Finkelhor & Wolak. 2007). There is also a growing gap between what children do online, and what their parents think they are doing (Lemish, 2008). With the increased importance of the internet in all of our lives, there is more and more pressure on children to be active on-line, and from a younger age. The dangers permeate almost the entire internet, and change rapidly as the technology evolves. Parents are ill-equipped to protect their children through no fault of their own, but (partly) because while they did grow up in a society where these threats existed, they did not exist in this new form. Today's reductions in barriers to communication have made the problem of protecting children much more complex. Children are often taught not to talk to strangers but with the variety of social interactions available today, teaching a child to block all communications from unknown parties would be challenging to even the most technically minded parent. This is probably undesirable too (Wolak, Finkelhor, Mitchell, & Ybarra. 2008)

This area has understandably received a lot of attention. There is a wide variety of content-control software available to prevent children from accessing illicit material. This mainly works by blocking known URLs, but also by dynamically analysing the content. While by no means trivial, this problem is limited in that it is only the content being sent to the child that needs analysis. These approaches do not apply to two way interactions. Firstly, blocking entire sites/protocols is not necessarily desirable as some safe use may be allowed (or else the child would be motivated to try work around the blocks). Secondly, the danger a predator poses is not merely displaying unwanted material to the child, but in arranging a meeting outside the parents' control. This can (and may) be done without mentioning anything overtly sexual. Predators are a danger because they can effectively mimic normal child-to-child conversations. If nothing else, one half could be simply copied and pasted from other conversations between actual children. The only difference may be in attempting to meet in person.

This paper discusses the existing approaches to online child protection and the conflicting requirements of the parent and the child in a moderated approach to chat. It presents an idea for a system of anonymous review with various options for added functionality, along with a justification of the system. The penultimate section expands on the two key technical components – pseudonymous messaging and intelligent text analysis.

[*] This paper was originally published in Kierkegaard, S. (2009) Legal Discourse in Cyberlaw and Trade. IAITL.

2. Previous Work

Existing approaches to online child protection typically fall into the following broad categories Block, Review, Filter and Moderate. To enable the reader to better understand the problem domain these approaches and the weaknesses associated with them are examined.

Blocking restricts access to protocols and applications deemed "unsuitable", for example peer-to-peer (P2P) networks or Instant Messaging (IM). Operating in a simple deny/permit fashion can make blocking something of a blunt and unwieldy tool. This lack of flexibility restricts its usefulness only to situations where something must be prohibited.

Reviewing technologies vary in type and application but the core ethos is to allow the parent to monitor the child's activity. Website histories, messaging logs, emails, even full replay of video conference sessions maybe recorded. This may be impractical if the child is an avid internet user or in families with multiple children. Reviewing also suffers from problems of privacy (older children are particularly sensitive about their privacy and may be tempted to circumvent the system) and the generation gap - parents may not be able to penetrate youth lingo and slang.

Filtering may be considered a subset of blocking, usually applied to restrict access to websites considered unwholesome in content. Filters generally consist of blacklisted (or whitelisted) URLs, or dynamic blocking of websites based on content - typically examining sites for a list of proscribed keywords and phrases. Each of these methods suffers drawbacks - blacklisting often involves content labelling, sites labelled as containing certain content are blocked. Labelling is performed by the website operator (who may not be aware of the labelling scheme or may neglect to use it). Some providers of filtering software manually review sites but this is an unscalable approach and the quality of this filtering has been brought into question as has its subjective nature (National Research Council. 2002).

Many internet forums use moderation to enforce rules, edit posts, and ban disruptive users, trolls and spammers. Some child oriented forums, including those of the BBC[1], operate a process of pre-moderation - each message is examined before it is posted. Moderators are trained to screen messages for signs of bullying, harassment, or anything that may result in a child being in exposed to harm. Moderation suffers two key drawbacks - scalability, and the human bias (subjectivity).

The system proposed here addresses the issues highlighted above without sacrificing safety. The child can feel their privacy is being maintained, although messages of concern are being reviewed, the contents of the message will not seen be their parents (thus shielding them from any embarrassment). In this regard the system may be considered similar in fashion to traditional moderation - their messages may be reviewed but not by their parents - but able to overcome its limitations.

3. Conflicting requirements

On-line child protection cannot be solved with technology alone. This paper therefore proposes a system that uses a combination of automation and human judgement to recognise threats. There are many potential pitfalls in trying to solve this problem. One solution might be to give parents comprehensive logs of their child's internet usage. This would be giving them too much information to manage effectively, and would be a tempting target for identity theft. If the parent has the power to control exactly what their child does on-line, it is possible they can better protect them, but the controls may be overwhelming. Also, children do not want to have their privacy violated, and will circumvent the system one way or another if it is too invasive. Even if they do not have the level of skill necessary to circumvent the system, they could always spend the majority of their internet time away from the home (at school, library, friends, etc). So some level of privacy for the child is needed. Similarly, all access that can be given and kept "safe" needs to be allowed. It would also be naive to expect children to suddenly migrate onto a new "safe" social network, IM network, etc. Any solution must cater for what they already use.

4. Description

The proposed system uses existing technology as a pre-filtering stage to create a prioritised list of 'suspect' chat conversations. This is subsequently analysed using human judgement via a pseudonymous volunteer who sees an appropriately sanitised version of the data which does not divulge the identity of the child, thus protecting their privacy.

[1] http://www.bbc.co.uk/chatguide/glossary/moderation.shtml

The system works as a software client that can be downloaded and run on any PC. The primary user (presumably a parent of at least one child), installs and sets up the client. There are two stages to the setup. First, the parent must record any sensitive/personal data unique to themselves and their child. This could include names, addresses, email/contact info, credit card number, phone numbers, etc. These will be used to determine when the child may be giving inappropriate information to a stranger, but also when they are inadvertently identifying themselves. The data can be stored as hashed values, or at least encrypted. The second part of the setup is performed in conjunction with the child. The parent reviews or "vets" all contacts the child has (in all applications / platforms /networks). The parent determines which contacts can be considered "safe". This should only include contacts the parent has/can meet in person, or know through some trusted organisation/third-party. At the very least, contacts who neither the child nor the parent has met should be considered unsafe. All contacts are labelled either "safe", "uncertain" or "blocked".

Once setup, the client runs in the background when the child logs onto the computer. The client intercepts all text-based communication protocols before they are presented to the user (the idea is text-based, but see Section 4.8 for voice/video extension). All communications between the child and contacts that have been explicitly labelled safe by the parent continue unimpeded and un-monitored. Any communications to/and from a contact that has been labelled uncertain, or from a new contact get processed. The processing works as follows:

1. All received text from the contact and keystrokes from the child are stored.
2. The text is checked against a list of known problem words/phrases ("sex", "drugs", "would you like to meet", etc.)
3. Other probabilistic analysis is performed (Bayesian network analysis, Gaussian modelling, etc.) looking for indicators of unwanted behaviour.
4. The text is searched for any of the sensitive/personal data entered by the adult. If any is found it is removed and noted.
5. The results from all these tests are combined in a single weighted score.
6. The identifier for the contact (email address, *Skype* name, etc.) is stored as a keyed hash using the child's password (could be their login password).

The processing is done on a section of text of a limited size (a page), and only on communications between the child and one contact. The client stores the processed logs in a list prioritised by the results of the analysis in the above list (top of the list will be the highest match with expected pattern of a predator). As noted in the fourth step, all sensitive/personal data will be removed from the log (can be replaced with a generic placeholder). Rather than just using the data the parent entered in the initial setup, a dictionary of names/local places could also be used to sanitise the logs. Periodically the client will send the logs at the top of the list to another client. Even with the sanitisation of personal data and replacing contact details with hashed values, in order to protect the privacy of the child, the logs will be sent over a pseudonymous network (Kinateder, Terdic & Rothermel, 2005). This allows messages to be passed from client to client, without either being able to discover who the other party is. This works through repeated layers of encryption and routing through various different nodes on the network. The idea is to create a community of effort where parents are reviewing each others children's logs, but in an anonymous manner. The technology used in anonymous communications is described in Section 6.

The receiving client (of the logs) will be administered by another parent. They will be presented with the sanitised logs and asked "Is this something the parent should be concerned with?" The reply can either be a yes/no button, or a scale e.g. from 1 to 10. The second parent can also highlight the text of the logs that is objectionable and a specific reason, e.g. "contact trying to meet child", "inappropriate sexual content of messages", "child is revealing personal information/identifying him/herself or home location". The results are returned to the first parent, again through the pseudonymous network. The first parent will be presented with only those logs (still sanitised) where the reviewer thought there was a problem. The particular application and time/date can be made known to the parent, but the identity of the other contact is still protected by the child's password. The parent can then decide whether to list the contact as "blocked" preventing any further communication, discuss the matter with their child to determine if the contact can be added to the safe list, or leave the contact "unknown" and continue monitoring.

Figure 1 shows a typical example. In the example, there is a central server called the Match Maker that maintains a list of pseudonymous clients that are on-line. An adult can request the details of one or more clients from the Match Maker. An alternative method that distributes this information throughout the network is also possible, and discussed in Section 6. The numbered steps in Figure 1 are explained below:

1. Adult1 downloads software client.
2. Adult1 installs client and configures settings (records sensitive data).
3. Adult1 and Child1 agree all safe contacts.

4. Child1 interacts with contacts. Communication with safe contacts continues as normal. Communication with uncertain contacts monitored.
5. Monitoring consists of analysis as described in previous list.
6. Suspect logs are stored in priority list based on results of analysis.
7. Client1 queries *Match Maker* server over pseudonymous network for list of online clients.
8. Client1 selects from list and sends sanitised logs for review.
9. Adult2 checks sanitised logs for undesirable communication.
10. Logs returned to Client1 with rating results.
11. Adult1 takes any necessary action based on feedback.

The most important part of the system is step 9. In order to encourage good behaviour in the clients, each adult will only receive results from the pseudonymous network about their children when they have finished a certain amount of reviews for other clients. This promotes good behaviour. This can be taken a step further by enabling a reputation system at the Match Maker server. This would record feedback from adults who found the results they received helpful. The reputation would be tied to the pseudonym and the person behind it would remain unknown, but they could be rewarded by being given a higher priority when requesting (clients for) reviews of their own logs, both in terms of speed and the quality of the rated reviewer.

The following section presents several variations that may improve the overall system, but which should be considered optional as they may have downsides, depending on the implementation details and user requirements.

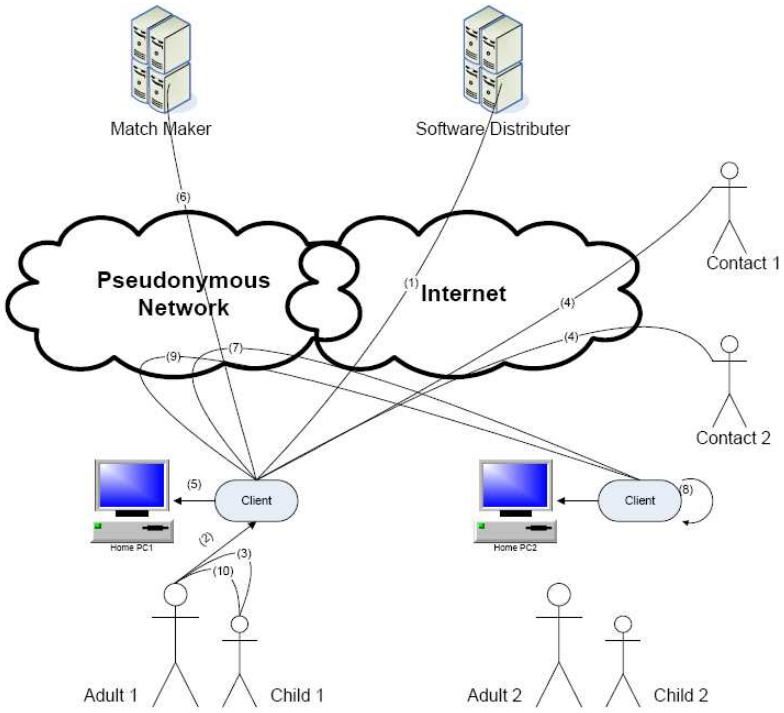

Fig. 1 Communication flow for Pseudonymous Peer Review

4.1. Monitoring all communications

The client could be configured to monitor all communications, including those between the child and safe contacts, but with a higher threshold needed for logs to be stored and sent for review. This negates more of the privacy the child has, but is still a great deal better than traditional monitoring: the contact can still be protected with the child's password, and only logs that have been considered problematic get passed on to the parent. It may be appropriate for younger children.

4.2. Multiple reviewers

The client can send the logs to several other clients for review. This will likely reduce the time it takes for a response, and the redundancy will give the parent more confidence in the results. The downside is the increased

burden on reviewers. But given that there may only be a few logs sent, and they can probably be reviewed very quickly, it is likely a good balance can be achieved. For example, the client sends all of the top 3 logs to 3 different clients (sending out 9 in total), but can only review results once they have reviewed 9 other logs. This approach has the benefit of normalising the reviewing process – the effect of a wildly liberal or conservative response would be brought closer to the prevailing societal attitude.

4.3. Instant reaction

Because of the human component of the review process, there will be a delay in the response time. One option that could be considered is that if a given log has a particularly high score, then the client can immediately block communication with that contact and notify the adult, bypassing the anonymous reviewer process. Careful configuration would be needed to avoid too many false alarms. Likewise, the sanitisation could be applied as a filter to all out going communications, not just the stored logs.

This may not be necessary – research indicates (Walsh & Wolak. 2005) that where a victim and sex offender met, the online relationship typically formed over a month or more, from multiple conversations. These would likely have already being flagged, reviewed and reported to a parent.

4.4. Learning behaviour

If the feedback from the reviewers is sufficiently detailed (selection of problem words/phrases, reasons for choice), then it would be possible for the clients to improve their scoring/prioritisation algorithms. This could either be done locally, or centrally at the *Match Maker* server. Learning at the central server could be very efficient, but would introduce problems with keeping submissions anonymous and being able to trust submissions. Learning locally at the client would not have any privacy problems, but would have a much smaller selection of results to draw from and hence a slower rate of learning.

4.5. Use of dictionary

The client could use a dictionary of common names to sanitise the child's logs. Certain patterns could be identified, such as phone numbers, credit card numbers, postcodes, etc. When the adult enters the home address, a central service might be able to provide a list of all nearby street names, local monuments, and locations that could be used by the child to give directions. This could also be incorporated into the scoring mechanism as well as the sanitisation.

4.6. Re-vetting

A good precaution would be for the adult to periodically re-vet the child's safe contacts. Strangers are not the only sources of potential threats to young children, and asking the child about their relationships with their peers could bring to light early warning signs.

4.7. Impersonation prevention

The times and duration the child has been on-line can be safely stored without fear of privacy invasion. When the adult logs in, a simple calendar with the child's usage can be graphically presented. This prevents the most obvious circumvention method. If there are no usage hours recorded for large amounts of time the parent knows the child was using the computer, then the parent knows the child has circumvented the system (most likely by using the parent's or another un-monitored account).

***4**.8. Voice/Video chats*

As previously described the system is text-based. This can be extended to voice/video, such as commonly found in IM products such as *Skype/MSN*, etc. This requires the use of Speech Recognition Techniques. This technology is currently of limited maturity, but could potentially be employed for keyword spotting. This will slot easily into the proposed system converting the video/audio stream into a text document. This also adds further protection of privacy than would be had by direct monitoring of the video/audio.

4.9 Age of Children

The definition of "a child" covers a large range of ages and maturities. Ultimately the decision to allow a child to use the internet for chat resides with the parent; however, a total prohibition would likely lead to using the technology anyway and without parental oversight. The best result is achieved when child and parent have a strong relationship and agree on what is, and what is not permitted. Internet initiated grooming of pre-pubescent children is extremely rare (Wolak, et al. 2008) (Lanning, K. 2001), partly because they are more closely supervised and also because it is difficult to engage them in sexual/romantic conversation because of their immaturity.

This system is best suited to children approaching puberty. They can be informed that (like adults) they are being rewarded with privacy in return for abiding by the rules. As they enter puberty, become sexually aware, and start to desire privacy they will already be familiar with the system. The child would continue to use the system until they reached a suitable level of maturity so that it was no longer needed or they reached adulthood.

Conversations of concern that are sent for moderation would include an indication of the child's age. This would have a bearing on what would be deemed appropriate.

5. Justification

In order for this system to work, active participation of all parties is required. The system has deliberately been designed to encourage good behaviour. The child is motivated to convince the parent that his/her contacts are trustworthy in order to have confidential communications. The parent is encouraged to provide meaningful reviews, in order to get a better reputation, which will result in speedier replies. The process of reviewing can also be of direct benefit to the reviewing adults themselves. It is educational, in that they are made aware of the kind of dangers that exist on-line and what their child may be exposed to. This better places them to discuss the problems with their child and agree what is safe/acceptable internet usage.

The shortcomings of existing approaches to this problem have been highlighted, and the approach described here overcomes these issues. Conversations, emails and message board postings will be monitored by the emerging generation of intelligent text analysis tools to spot conversations of concern. Only these conversations will be submitted for manual/human review. This process overcomes the scalability issues associated with traditional moderation, vastly reducing the reviewer's workload. The community nature of the approach also negates the cost associated with the traditional moderation model.

Any single moderator reviewing a conversation of concern will be subject to the human condition - bringing their own bias and prejudice into the review process. Using multiple moderators for each conversation of concern will reduce the impact of any one error in judgement. For example say a conversation of concern is reviewed by a conservative individual and someone of a more liberal mindset, an average of their result/feedback will tend towards the centre ground. Here "centre ground" is meant as the general view of society at large.

The moderator will be faced with the hurdle of slang employed by children and may need assistance interpreting the contents of some message (conversations of concern). It is hoped that such information will be sought from the moderator's own children. The benefits of this will be twofold; the parent will gain a better understanding of trends, challenges and experiences of young people in the modern world helping them to better understand their own children. From reviewing the contents of the messages (conversations of concern) both parent (moderator) and child can learn first hand of the dangers that exist. This will be informative for the parent and can serve as a warning for the child.

6. Components

This section gives a very brief description of how the technical aspects of the system might be implemented. This system relies heavily on the use of pseudonymous communications, along with the dependent technology of Distributed Hash Tables. The first subsection describes the work by Kinateder, et al (2005), the second deals with the work of Stoica, Morris, Karger, Kaashoek & Balakrishnan, (2006) and the third describes how text classifications are best applied to this system.

6.1 Pseudonymous Communications

In order to protect the anonymity of the child, the parent and the reviewer, a communication system with the following properties is required:

- the sender has some ephemeral/indirect knowledge of the receiver, but not their actual address
- the receiver cannot know the address of the sender
- the receiver can reply to the sender
- no observation of the network gives any information about who is communicating with whom

These are all achieved with the use of public key encryption and intermediaries. It is assumed that there is a network of nodes, called mixes, capable of processing messages (encryption/decryption) and passing them on. These nodes are not trusted, and while they can disrupt the communications by not participating as expected, it is imperative that they learn nothing by analysing messages as they pass through. It is also assumed that public keys for all parties can be readily obtained (as well as the addresses of the mixes).

However they are stored, the public keys are used to ensure that only the recipient can decrypt the message, but does not help in anonymously delivering the message. Along with the public key, one or more pseudonym is stored. These are created by those who want to receive anonymous communications (the recipient) and will be used by a sender. The pseudonym is a sequence of addresses of mixes, but nested in layers of encryption that ensure that only the next mix can decrypt the current layer and they only get the address of the next mix. The further layers of encryption mean that a mix cannot know any further destinations than the next mix, and they do not know any of the previous mixes in the chain since these addresses have been stripped off. Since knowledge of the path of the route is effectively split amongst all these independent entities, it would be impossible to determine the ultimate sender and receiver for a given message without compromising the majority of the network.

There are other ways to increase anonymity, sending messages in batches to confound traffic analysis, sender adding more mixes to the sequence to protect against dishonest recipients, etc. For more details see (Kinateder, et al. 2005). It is also preferable to store the public keys/pseudonyms in a Distributed Hash Table, rather than a single server. This is described in (Stoica, et al. 2006).

6.2. Text Classification

Traditional keyword and regular expression filters are inadequate for analysing IM and chat forum conversations. The new generation of text mining and text analysis tools offer far superior classification abilities. Word frequency, word distance, word pairs (bi-grams), Latent Semantic Analysis, term strength, term frequency-inverse document frequency (tf-idf), and term by document matrices have all proven successful at extracting features from textual sources. These features are subsequently applied to statistical modelling techniques including Bayesian analysis, k-Nearest Neighbour and Support Vector Machines (Tretyakov, 2004) (Conrad & Hunter, 1994). These techniques have proven results in spam detection, document categorisation, authorship attribution and information mining (Aas & Eikvil, 1999).

Classification techniques such as these typically compare new information against known values and categorise it accordingly. For example, a new email will be analysed and the result compared against the result for known spam and ham (legitimate email). This process poses a problem in the child protection domain.

In order to classify conversations as safe or uncertain they need to be compared against known paedophile, and normal chat models. Constructing normal chat models is a trivial task, constructing paedophile (chat) models is not. To construct an accurate model access to paedophile chat logs is required. Law enforcement does not typically share this information freely. Honey trap organisations such as Perverted Justice[2] publish chat logs from sting operations (against paedophiles) on their website. Research (Wolak, J., Finkelhor, D., Mitchell, K., & Ybarra, M. 2008) indicates the in the majority of online sexual predation the offender quickly reveals his true age and intentions – this tallies with the chat logs of Perverted Justice. Pendar's (2007) work shows that Bayesian trigram analysis is effective at detecting grooming (based on Perverted Justice chat logs). The operating practices of Perverted Justice have been called in to question and allegations of entrapment and poor evidentiary quality have been made (Stokley, 2008) (Salkin, 2006). As a consequence of this, the reliability of these chat logs to represent "real world" logs must be considered uncertain. The following proposes another method that may also be suitable - a system of thresholding.

Rather than classifying based on the results of messages matching against two categories/models (normal chat and suspicious), a match against a single model with a threshold value could be used instead. Conversations would be compared against the normal chat model; those matching closely (with a probability above a certain threshold) would be considered normal and not be affected. Conversations poorly matching the model (below the threshold) would be marked as "uncertain" and passed on to the system for evaluation by a moderator. In addition, those conversations falling well below the threshold might be marked for priority moderation/inspection.

[2] http://perverted-justice.com

In order for the process to work a normal chat model would first need to be trained. The child's online chat activities over a given time would be used to train the initial model. This model will serve as baseline of normal behaviour. Once the training process is complete, online conversations are compared against the model. During the initial training of the model it will not be able to flag problems. As the text classification is just one of several tools used to identify potentially suspicious behaviour, other mechanisms are still available. Key word matches with known general problem phrases and scrubbing of specific personal details is still performed. And to further protect the child (and ensure a "clean" model) the parent can require that no communication with new contacts is allowed for the period of the training. After the initial training period there are likely to be a number of false positive, especially if the training period is brief. If, for example, the child only converses with their peer group during the training of the model, a conversation with a parent or grandparent would be flagged as anomalous. This should be considered a period of normalisation. As the review process records the conversation as a false positive, the text of the conversation can be used to tune the initial model.

7. Conclusions

This paper presents a concept, whereby parents can have "conversations of concern" reviewed anonymously by other parents in return for their own actions as reviewers. The limits of existing technical measures to protect children have been highlighted and it is proposed that the system described here could help bridge the gap through a community approach.

The merits of this idea do not come solely from the technology, but rather from several deliberate reward mechanisms for the users. Children are encouraged to play-by-the-rules and are rewarded with privacy. Parents who are conscientious reviewers will get a better view of the dangers their children are exposed to. The technology provides the means for what would be sensitive information to be shared in a safe way. Great lengths have been made to avoid anything that could be considered censoring by the end-users. All of this is combined to strike the best balance between the child's safety and their freedom on-line.

References

1. Aas, K., & Eikvil, L. (1999). Text categorisation: A survey. Norwegian Computing Center, Oslo, Norway, http://citeseer.ist.psu.edu/aas99text.html
2. Conrad, J. G., & Hunter, M. U. (1994). A system for discovering relationship by feature extraction from text databases. Annual ACM Conference on Research and Development in Information Retrieval, 1994
3. Kinateder, M., Terdic, R., & Rothermel, K. (2005). Strong Pseudonymous Communication for Peer-to-Peer Reputation Systems. ACM Symposium on Applied Computing, 2005. http://portal.acm.org/citation.cfm?id=1067033
4. Lanning, K. (2001). *Child molesters: A behavioral analysis* No. Fourth Edition)National Centre for Missing & Exploited Children.
5. Lemish, D. (2008). Generation Gap? 'Online Gap' Widens Divide Between Parents and Children. Science Daily. http://www.sciencedaily.com
6. Salkin, A. (2008). As perverted-justice.com battles web pedophiles, some raise concerns over its tactics. International Herald Tribune. http://www.iht.com/articles/2006/12/13/news/justice.php?page=1
7. Stoica, I., Morris, R., Karger, D., Kaashoek, M.F., & Balakrishnan, H. (2001). Chord: A Scalable Peer-to-peer Lookup Service for Internet Applications. Proceedings of the 2001 ACM SIGCOMM Conference. http://pdos.csail.mit.edu/papers/chord:sigcomm01/chord sigcomm.pdf
8. Stokley, S. (2008). 'To catch a predator' sex stings net mixed results. The Press Enterprise. http://www.pe.com/localnews/inland/stories/PE News Local R dateline28.6b3814.html
9. Tretyakov, K. (2004) Machine Learning Techniques in Spam Filtering. Institute of Computer Science, University of Tartu.
10. Walsh, W., & Wolak, J. (2005). Nonforcible internet-related sex crimes with adolescent victims: Prosecution issues and outcomes. *Child Maltreatment, 10*(3), 260-271.
11. Wolak, J., Finkelhor, D., Mitchell, K., & Ybarra, M. (2008). Online "predators" and their victims myth, realities and implications for prevention and treatment. *American Psychologist, 63*(2), 111-128.
12. Wolak, J., Mitchell, K., & Finkelhor, D. (2006). Online Victimization of Youth: Five Years Later. http://www.missingkids.com/en US/publications/NC167.pdf