Towards a Risk Management Based Approach for Protecting Internet Conversations

Dimitrios Michalopoulos[1], Ioannis Mavridis[1] and Vasileios Vitsas[2]
[1]University of Macedonia, Thessaloniki, Greece
[2]Alexander Technological Educational Institute of Thessaloniki, Greece
dimich@uom.gr
mavridis@uom.gr
vitsas@it.teithe.gr

Abstract: During recent years the number of online communication means between teenagers has been growing rapidly. However, the hazards that follow these new types of communication are growing as well. Predators use Internet conversations to attract minor users, usually resulting in catastrophic consequences. In this paper, a new risk management based approach is proposed, which aims to monitor internet-based conversations and identify possible attacks. In particular, a wide research on the area of children exploitation is first conducted, in order to identify the methods and the techniques that are used. Then, the implementation of a system capable of capturing and analyzing chat dialogs is proposed. The proposed system is under development and has not be implemented yet. It is based on three different sensors. The first one performs text analysis on captured traffic, as an attempt to look for known patterns that may indicate a possible attack. The Naive Bayes classifier method then follows, based upon the initial training set. In addition, this training set is enhanced and adopted to specific users' needs via the proposed "supervised learning technique". The second sensor captures files or web links that are sent through the chat conversation, indicating possible personal information leak or exposing unwanted material to minors. The third sensor counts how many times the same users talk with a particular child. As a result, a total risk factor is calculated as a weighted sum of the three risk factors, through applying the proper weight coefficients. In case the risk factor is above the predefined threshold, a warning signal is sent in order to warn on time that there is a possible grooming attack. The main challenge in the proposed system implementation is related to natural language processing, due to the fact that teenagers use their own acronyms and idioms when chatting, creating their own language. A deep research on these dialogs might result into different linguistic sets. Another important challenge is related with the privacy in internet related communications.

Keywords: Cybercrime, Grooming, Risk Assessment, Attack Recognition

1. Introduction

As the World Wide Web is expanding very rapidly, the communication between people is experiencing new forms. Children and teenagers are using more massively electronic communication media, such as IRC chat rooms and IM, or social networks such as Facebook, Myspace etc. For instance, in 2009 Facebook, which is the most popular social networking site, reports over 90 million users worldwide and MSN Messenger announces 27 million users (Nash 2009). Meanwhile, the hazards that follow these new types of communications are growing as well. Pedophiles and predators take advantage of these new methods in order to attract their victims. In a real world environment it is harder for them to attack, as they can easily be seen and attract attention. In an online environment they can hide themselves behind anonymity that internet provides. As a result, there are many incidents where minors are victims of similar attacks (Subrahmanyam *et al.,* 2006).

The consequences of online predation are catastrophic (Berson 2003). Children can be harmed for the rest of their lives and sometimes it is very difficult to be healed. Moreover, the role of parents is crucial. In most cases, parents are older people with lack of education about computers. As a result, they can not estimate the danger their child is exposed to while chatting many hours on the computer. The problem is getting worse considering that minors tend to talk about sexual matters with friends over the Internet instead of discussing them with their parents (Stanley 2001). In addition, not all predators are strangers. Indeed, only a 15% of the predators are strangers, a 40% are known to children, a 14% come from the family cycle and a 18% are friends or dates of the victims (Crosson-Tower 2004). Moreover, among 1500 young people from 10 to 17 years old who were interviewed about their experience in online chat rooms and social networking sites, sexual exploitation attempts were reported in addition with online harassment and frequent exposing to unwanted sexual material (Wolak *et al.,* 2008).

Hazards that children are exposed to vary through age and sex and can be divided into three main categories: (a) Cyberbullying, (b) sexual exploitation or grooming and (c) exposing to illegal material. Cyberbullying refers to all kind of attacks that terrify a minor user with threats for his/her life, parents, friends etc. Many times teenagers are exposed to anonymous threats through the internet (Bauman, 2007). In most of the cases these attacks come from their classmates, neighbors or friends who cover their identities behind the anonymity of internet. Moreover, teenagers do not talk to their parents about these threats and as a result they live in fear, they pay no attention to school and have an unsocial attitude (Berson, 2003). Sexual exploitation or grooming attacks are performed by people, who feel sexual attracted to children and use the modern communication methods to find their victims. Predators take advantage of the internet anonymity and build a profile suitable for their malicious purpose (Dean, 2007). At the other side, children are keen on discussing sexual maters with strangers on the internet which brings predators one step ahead on performing a grooming attack(Bakopoulos and Walker 2005).

Indeed, there is little difference between cyberbullying and grooming. In many cases, a cyberbullying attack follows a grooming one, as predators terrify and threat their victims in order not to talk about what has been done or intended to be done. Moreover, some times cyberbullying comes first from grooming as predators intend to have their victims in fear for providing less resistance. Furthermore, Cyberbullying attacks occur more often in older teenagers (13+) where grooming techniques are followed more often by pedophiles whose target range victims are children bellow 13 years old.. Exposing to illegal material includes many types of images, video, music. Indeed, most of the published research is focused on violent media and pornography. Parents concern very much about the material on which minors are exposed to. In fact, many times children are not exposed without their will. As the internet is a huge source of information children can search for inappropriate material many times motivated by curiosity.

this issue of internet related hazards for minors is recent and as a result there is a lack of defense mechanisms. These mechanisms should be multidimensional (Subrahmanyam et al., 2006; Stanley 2001). Governments can play an important role by providing teachers at schools proper training for handling these situations. The lack of awareness can be issued by organizing seminars for parents and pupils in schools. In addition, psychologists can help victims to dispatch the consequences. Media might also have an important contribution by publishing similar cases, so minors get aware for the dangers underneath. Based on technological perspective we propose a system that can identify possible attacks. This system is under development and not implemented yet. It is called GARS (Grooming Attack Recognition System). GARS monitors chat and Instant messaging communications trying to identify dialogs that could harm the minor user. In this paper we address the security risks that exist for children in electronic communication media. GARS identifies the amount of risk involved in these means of communication and a method for quantitative calculation of risks is presented. Finally, a warning system is proposed for preventing on time a potential attack. In section 2 methods and strategies that are followed by predators are analyzed. In section 3 there is an analysis of the available technology for capturing IM and IRC chat. In section 4 the proposed system GARS is presented whereas in section 5 major project challenges are discussed.

2. Attacking Strategies

As it is mentioned, attacking on children through cyberspace is a new phenomenon mostly appeared with the internet integration at recent years. Unfortunately, little research has been published yet. Each case is unique of course, and the way that takes place is related with various characteristics such as age, location and the character of people in conversation. However, the research that has been published on this area has shown the there is a similar strategy followed in most of the cases by predators (Subrahmanyam *et al.,* 2006*;* Krone 2005; O'Connell *et al.,* 2003*;* Stanley 2001). The common attribute is our key for developing defense mechanisms against these kinds of attacks. The steps for performing such an attack are as follows (O'Connell *et al.,* 2003):

1) The friendship stage. The predator tries to know the victim, to collect as much information as possible for the child.

2) The relationship stage. Here, the attacker tries to establish some kind of relationship with the minor by giving the impression that he is the best friend. A cycle of trust is tried to be established.
3) The risk analysis stage. Now the attack takes a risk analysis plan. He tries to estimate the risk that has to be taken for such kind of attack.
4) The uniqueness stage. This stage is just one step before the final attack. Now, the predator estimates the trust that has been earned at previous conversations and tries to test it.
5) The attack stage. In most of the cases an offline meeting is trying to be arranged. Sometimes the predator forces the victim to have a conversation with him discussing sex-related topics whereas in other cases the victim is forced by blackmail of sending pictures or video with disgusting material. In general, this sensualist for pictures or video is very usual in this stage.

In the same direction, a similar published research based on the communication process shows that the steps that are followed by predators are the following (Olson *et al.,* 2009):

- Gaining access to the victim
- Entrapping the victim in a deceptive relationship
- Initiating and maintaining an abusive relationship

The second type of attacks is Cyberbullying (Bauman, 2007). What is important to be mentioned is fact that although it takes place very often in online communications, very few victims talk about that with their parents and even fewer actions are reported to police and other authorities. Indeed, only 10% of the minors who have experienced such an activity online have talked to their parents or police (Finkelhor and Ormrod 2000). Under these circumstances it is very difficult to investigate these cases and extract data that can be helpful. However, research has been published on this field analyzing the problem and the defenses (Bauman 2007).What is also important to be mentioned is the fact that in some cases children pretend themselves as cyberbullying victims. The main reason for such attitudes is the attention of their parents and friends that they want to attract. The most usual types of cyberbullying are:

- sycophantic defamation
- Assaulting and abusive messages
- Menace against life
- Social exclusion from online communication networks

Although cyberbullying is very frequent it is important to be mentioned that children are mostly unaware of the hazards that lie beneath online social networks. Besides, even if some children are informed about these dangers, they underestimate them (Berson, 2003). The last category of attacks is the exposing to problematic material. This includes violent media as well as adult or child pornography. In comparison with the previous categories, the problem is not so clear. First of all, exposing to adult pornography or violent media may not happen during an attack but during many other activities. It is hard to say, but children are exposed to problematic content during many other activities like watching TV or playing games etc. What comes next is the consideration that youths sometimes search the internet for such content. However, exposing minors to child pornography is the most problematic one. Most of the times minors are exposed to such media by predators as part of their attack methods as it is described previously (Wolak *et al.,* 2008).

3. Capturing IM and chat traffic

The published research on this field shows that there are some implemented systems which capture and analyze IM and chat dialogs. Indeed, it has to be mentioned that there are legal issues in terms of communication privacy that have to be overcome, which vary according to the country and its laws that the system is used for. A published server based system (Dewes *et al.,* 2003) captures all network traffic which comes through a router and then certain filters are applied to separate chat conversations from other network traffic. Similarly, a web chat monitor tool (Xiong *et al.,* 2005) can also block conversations that match specific blocking rules. One step further, a text mining tool with chat room analysis capability (Khan *et al.,* 2002) was also proposed and besides they proposed specific patterns for classifying captured dialogs. Furthermore, a corpus, which analyzes through natural language processing, the topic and emotions of internet based conversations, was published in (Forsythand and Martell, 2007). Although there is a lot of published research on analysis of chat room data, there are few attempts on creating algorithms for identifying predators. In such an effort, an implemented system (Pendar, 2007) uses automatic text categorization techniques and a k-NN classifier model can identify sexual predators. Indeed, its classifier reaches an f-measure of 0.943 on test data distinguishing. Similarly, an approach for detecting predators which is not based on a statistical approach but on the communicative theory was published in (Kontostathis *et al.,* 2009). What is also important to be mentioned is the difficulty of finding proceedings of known attacks. In most cases they are used for legal issues and consequently are not available. In parallel, police officers deny to give access to such files as they use them for investigation purposes. For additional guidance, the website www.perverted-justice.com provides in public some transcripts of grooming incidents in English.

4. Grooming Attack Recognition System

The proposed system is called GARS (Grooming Attack Recognition System). GARS is not implemented yet as it is under development. Regarding the Internet reference model, GARS will be located underneath the application layer.

4.1 System Architecture

The system architecture of GARS is depicted in Figure 1.

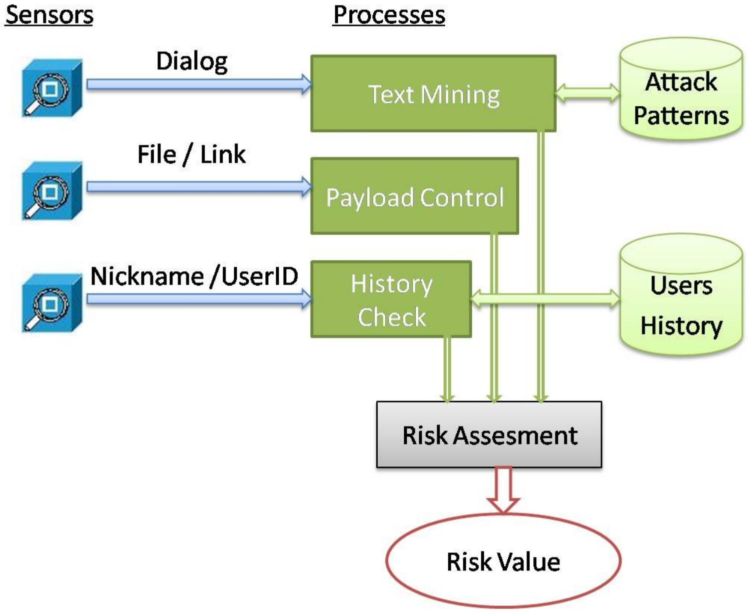

Figure1. GARS system architecture.

GARS includes three sensor subsystems for capturing dialogs, files / links, and nicknames / userID. Data gathered from these sensors are then used as input to three corresponding processes. Text mining analyzes captured dialogs based on predefined attack patterns. Files and links are used for Payload Control. Nicknames and userID are utilized to retrieve users history and then perform history checks. For this purpose, GARS contains two data bases for recording attack patterns and users history, respectively. The final risk value is calculated from risk assessment, as a summary of the three risk factors which come from processes and are summarized with the proper weight.

In the same way a spam filter, based on text analysis, can identify spam messages, the first sensor captures dialogs and the corresponding Text Mining process calculates a probability of attack. The algorithm used for probability estimation is the Naive Bayes classifier (Moore 2009*;* Murphy 2009*;* McCallum and Kamal 2009*;* Rish 200; Thorsten 1998). Initially, GARS is fed with suspicious dialogs (indicating attacks) that can be extracted from websites like www.perverted-justice.com. Indeed, there are basic keywords and dialogs that indicate the presence of a predator (O'Connell *et al.,* 2003). This set of keywords and dialogs is included in the initial training set. Afterwards, a supervised learning technique is followed that allows minor users to mark conversations as annoying (Kontostathis *et al.,* 2009). These conversations are added to the set of attacks patterns. Further detailed analysis and an example are analyzed in the followed subsection.

The second sensor captures sending files or web links through conversations. The Payload Control process that follows calculates the risk factor that remote users attitude generate. With the term attitude we refer to possible files or http links exchanged between users. The sensor identifies web links as "http://.." , "http://www..." or even E-mail addresses. In addition, it is capable of identifying file transferring especially photos or videos, for instance .jpg, .tif, .bmp, .avi, .wmv and similar. This kind of attitude may be harmless, for example the minor could exchange information about his/hers favorite rock star. However, it could be very suspicious as well. Finally, the sensor would capture the number of question marks "?". The more of them are located in a conversation may indicate that the remote user asks for information. It should be indicated that all the captured files or links will not be stored permanently in terms of respecting communication privacy.

The third sensor captures remote user's Nickname and userID, if possible corresponding to the time of conversation. The process that follows, History Check, checks if the captured Nickname or ID is recognized in recent users database or not. This database is informed every time there is conversation activity. From one perspective it is not possible to monitor one user's past as he/she can change many times the nick name and hide the traces. Indeed, during a grooming attack, the attacker tries to earn minors trust (Olson *et al.,* 2009; O'Connell *et al.,* 2003). This indicates that he might use the same nickname many times in order to be a friend with the child and earn the desired trust. The risk factor that this process counts indicates whether the minor communicates with the same user again and again. Of course, by its own it can not generate a warning as the minor may talk with a friend - classmate many times. However, the presence of communication with the same user several times for lot of time in a combination with the previous calculated factors increases the total risk factor.

The above processes calculate the specific risk factors and send them to the final process where the total risk factor would be calculated based on the indications and the weight of each factor signal. Therefore, the total risk factor is calculated as:

$$R = w_{da}R_{da} + w_{sf}R_{sf} + w_{uh}R_{uh} \quad (1)$$

where the R represents the total risk factor, the w_{da} and R_{da} the weight and the risk of the dialog analysis part, the w_{sf} and R_{sf} the weight and the risk of the sending file-link part and the w_{uh} with R_{uh} the weight and the risk of the user history part. It can be indicated that $w_{da} + w_{sf} + w_{uh} = 1.$

4.2 Probabilistic Risk Assessment

The Text Mining process classifies the captured dialogs into two categories: potential attack and normal dialog. Internet based dialogs consist of sentences or phrases from both users (only conversations between two users are examined). For simplicity sake, each word is considered to be randomly located in the phrase and possible articles, conjunctions, prepositions and pronouns are excluded.

Based on the document classification algorithm (McCallum and Kamal 2009*;* Kotsiantis and Pintelas 2005), two classes are assumed: attack class A and not-attack class A'. Class A consists of words that indicate attacks. During an online conversation, the sensor captures a phrase (say F) which consists of i words. The probability that the i-th word appears in class A is denoted by $p(w_i \mid A)$. The conditional probability of a particular phrase F, given class A, is $p(F \mid A) = \prod_i p(w_i \mid A).$ From Bayes' theorem it holds (Murphy 2009):

$$p(A|F)=\frac{P(A)}{P(F)}\prod_i p(w_i|A) \tag{2}$$

We are focused on are the probabilities $p(A|F)$ and $p(A'|F)$. From (2) we have

$$p(A|F)=\frac{p(A)}{p(F)}\prod_i p(w_i|A) \tag{3}$$

And

$$p(A|F)=\frac{p(A)}{p(F)}\prod_i p(w_i|A) \tag{4}$$

Dividing the left-hand side, as well as the right-hand side of (3) and (4) yields

$$\frac{p(A|F)}{p(A'|F)}=\frac{p(A)}{p(A')}\prod_i \frac{p(w_i|A)}{p(w_i|A')} \tag{5}$$

And as long as $p(A|F)+p(A'|F)=1$ we obtain (Cramer 2003)

$$\log(\frac{p(A|F)}{p(A'|F)})=\log(\frac{p(A)}{p(A')})+\sum_i \log(\frac{p(w_i|A)}{1-p(w_i|A)}). \tag{6}$$

Given the above process, the text mining process should decide now whether the captured words from the dialog indicate an attack or not. For this purpose, it calculates the second part of (6) and compares it with zero. In case

$$\log(\frac{p(A)}{p(A')})+\sum_i \log(\frac{p(w_i|A)}{1-p(w_i|A)})>0 \tag{7}$$

it holds $p(A|F)>p(A'|F)$ and the factor indicates a possible attack. The fraction $\frac{p(A)}{p(A')}$ can be calculated through stochastic simulation from previous research. What is more, this parameter can take feedback from other factors, for example a user that writes suspicious words without sending files-links or without previous history record lowers the probability of being an attacker.

As an example of the above theoretical analysis we can assume that the dialog-capturing sensor captures a phrase F “Do you want to take a photo of yourself?” which consists of i=9 words. Four of these words are recognized in attack patterns ‘want’, ‘take’, ‘photo’, ‘yourself’. We can extract that $p('want'|A)=0.3$ as ‘want’ appears 3 out of 10 times in phrases in attack patterns. Likewise $p('take'|A)=0.6$, $p('photo'|A)=0.8$ and $p('yourself'|A)=0.5$. Based on previous research on the area, it can be extracted that a 60% of children, for example, have experienced an online attack. Therefore, the fraction $\frac{p(A)}{p(A')}$ can be adjusted to 0.6. From (5) it holds

$$\log(\frac{p(A|F)}{p(A'|F)})=\log(0.6)+\log(\frac{0.3}{1-0.3})+\log(\frac{0.6}{1-0.6})+\log(\frac{0.8}{1-0.8})+\log(\frac{0.5}{1-0.5})=0.012$$

Indicating a possible attack.

The derivation (1) denotes that the total risk factor is calculated by the weighted summary of the three above factors. However, how these weights can be adjusted for best accuracy? They can be regulated by system adoption to the specific needs of each case and experimental research. Actually, as it is mentioned before, there are approximately four categories of minors: boys or girls above 13 year old or less. Each of these is exposed to slightly different hazards (Subrahmanyam *et al.,* 2006). For example, boys under 13 are more vulnerable to grooming attacks whereas these over 13 are more vulnerable to cyberbullying ones. This fact gives us the opportunity to adjust our system to the specific needs of the user. In addition, adjusting can be achieved with the fraction $\frac{p(A)}{p(A')}$ of (6). In the previous example, for a boy under 13 years old the fraction $\frac{p(A)}{p(A')}$ is higher than for a boy over 13 years old.

In parallel, an experimental confirmation of the above parameters is needed. In order to achieve the best results for this project we need to implement it on an experimental basis and test not only the adjustable variables but also its effectiveness. A warning signal could be irritating in case of a false positive alarm or the lack of a warning signal may lead to catastrophic result. Consequently, the sensitivity of the implemented sensors should be adjusted properly.

5. Discussion

The development of GARS brings up important challenges. First of all, the implementation of such a system that recognizes known attack patterns through captured dialogs requires deep knowledge on how these attacks are performed. This attack analysis is required not only on methods and strategies that predators follow, but also on the language that is used by minors while they are talking online. The natural language processing is crucial. Minors are keen on working out common words with replacing letters with symbols or changing the letter order. Indeed, sometimes adults are unable of understanding the conversation (Forsythand and Martell 2007*;* Khan *et al.,* 2002). This fact in addition with the grammatical errors that are often in a chat conversation will increase the difficulty of developing GARS for recognizing the written words. A potential solution seems to be the supervised learning technique that will be followed. Attacking patterns will be enriched with dialogs including all these idioms and linguistic characteristics that minors use.

In addition, another important challenge references to communication privacy. The implementation of GARS includes a sensor that captures dialogs. However, legal legislations around the world deny such an employment as it violets the privacy of online communications. Indeed, GARS will correspond to this challenge with analyzing without storing conversations. In addition, in terms and conditions during the installation process will be mentioned out that GARS will capture and analyze dialogs without storing them so, the system administrator (the parent) is aware of. In general, the aim of GARS is not monitoring or loging of communications. The main purpose is the protection of minor users focused on real time detection and warning. This instant warning could be crucial. Grooming predators follow a strategic procedure through which gain access to minors' feelings. The sooner such an attack is detected and the parent is warned the less catastrophic affects will have on minors' future life.

6. Conclusion - future work

Fighting child sexual exploitation through internet communication channels is definitely a multidimensional approach. From government's perspective, there is a need for law enforcement focused on social networking websites, their regulations and the removal of known offenders. Besides, there is a demand for awareness and training for children, parents and teachers as well. From technological perspective, it would be very useful the implementation of a system capable of identifying these kinds of attacks and warn parents on time. Authors propose GARS, a system that is on the early stages of development, to capture internet based dialogs and compare them with known patterns. The warning signal should be diacritical with the form of an e-mail or even better a sms for more instant alarm. This paper, in addition with describing the problem, presents a mathematical approach for automatic attack detection as it is described in equations [2]-[7].

However, the implementation of such a system may encounter certain challenges. In order to overcome these challenges and develop a system for proper protection there is a need for deep analysis on grooming attack methods and effective natural language processing. In addition, lots of attention has to be paid during the initial training set building. A deep research on real dialogs might provide different possible linguistic sets and therefore GARS will be more effective.

References

Bakopoulos, B. and Walker, R., (2005) Conversations in the Dark: How young people manage chat room relationships, [online], *First Monday*, Vol. 10, No 4, http://firstmonday.org/issues/issue10_4/walker/index.html

Bauman, S., (2007) CyberBullying: a Virtual Menace, National Coalition Against Bullying, Melbourne, Australia, [online], www.ncab.org.au/Assets/Files/Bauman,%20S.%20Cyberbullying.pdf

Berson, I., (2003) Grooming Cybervictims: The Psychosocial Effects of Online Exploitation for Youth, [online], *Journal of School Violence*, Vol. 2, No. 1, www.cs.auckland.ac.nz/~john/NetSafe/I.Berson.pdf

Cramer, J.S, (2003) *Logit models from economics and other fields,* Cambridge University Press.

Crosson-Tower, C., (2004) *Understanding child abuse and neglet*, Allyn & Bacon.

Dean, S., (2007) *Sexual Predators: How to recognize them on the internet and on the street - how to keep your kids away*, Silver Lake Publishing.

Dewes, C., Wichmann A. and Feldmann A., (2003) "An analysis of internet chat systems", Paper read at the proceedings of the 3rd ACM SIGCOMM conference on Internet measurement [online] http://portal.acm.org/ft_gateway.cfm?id=948214&type=pdf&coll=GUIDE&dl=GUIDE&CFID=75253887 &CFTOKEN=36979828

Finkelhor, D., and Ormrod, R., (2000) *Characteristics of Crimes against Juveniles*, Juvenile Justice Bulletin, pp. 1-11.

Forsythand, E. and Martell, C., (2007) Lexical and Discourse analysis of Online Dialog, Paper read at proceedings of the International Conference on Semantic Computing, 17-19 September, pp:19 – 26.

Khan, F., Fisher, A., Shuler L., Tianhao, W and Porrenger W., (2002) Mining Chat room Conversations for Social and Semantic Intreractios, Technical Report, Lehigh University.

Kontostathis, A., Edwards, L. and Leatherman, A., (2009) *Text Mining and Cybercrime.* In Text *Mining: Application and Theory,* Michael W. Berry and Jacob Kogan, Eds., John Wiley & Sons, Ltd.

Kotsiantis, S. and Pintelas, P., (2005) Logitboost of Simple Bayesian Classifier, Computational Intelligence in Data mining, *Informatica Journal*, Vol. 29, Issue 1, pp. 53-59.

Krone, T., (2005) Queensland Police Stings in Online Chat rooms no. 301, Australian Institute of Criminology [online] www.aic.gov.au/documents/B/C/E/%7BBCEE2309-71E3-4EFA-A533-A39661BD1D29%7Dtandi301.pdf

McCallum, A. and Kamal N., (2009) A Comparison of Event Models for Naive Bayes Text Classification, [online], http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.65.9324&rep=rep1&type=pdf

Moore, A., (2009) Statistical Data Mining Tutorials, [online], www.autonlab.org/tutorials

Murphy, K., (2009) An Introduction to graphical models [online] http://people.cs.ubc.ca/~murphyk/Bayes/bayes_tutorial.pdf

Nash, K., (2009) "A peek inside facebook", [online], www.pcworld.com/businesscenter/article/150489/a_peek_inside_facebook.html

O'Connell, R., (2003) A typology of child cybersexploitation and online grooming practices, [online], Cyberspace Research Unit, University of Central Lancashire (UK), http://image.guardian.co.uk/sys-files/Society/documents/2003/07/24/Netpaedoreport.pdf

Olson, L., Daggs, L., Ellevold, B. and Rogers, T., (2009) Entrapping the Innocent: Toward a Theory of Child Sexual Predators' Luring Communication, *Communication Theory*, Vol. 17, Issue 3, pp: 231 – 251.

Pendar N., (2007) Toward Spotting the Pedophile Telling victim from predator in text chats, Read at Semantic Computing Conference, 17-19 September, pp: 235 – 241.

Rish, I., (2001) An empirical study of the naive Bayes classifier, [online], Workshop on Empirical Methods in Artificial Intelligence, www.cc.gatech.edu/~isbell/classes/reading/papers/Rish.pdf

Stanley, J., (2001) *Child abuse and the Internet*, [online], Australian Institute of Family Studies, Melbourne, http://aifs.org.au/nch/pubs/issues/issues15/issues15.pdf

Subrahmanyam, K., Smhel, K. and Greenfield, P., (2006) *Connecting Developmental Constructions to the Internet: Identity presentation and Sexual Exploration in Online Teen Chat Rooms*, National Science Foundation Grant.

Thorsten, J., (1998) *Text categorization with Support Vector Machines: Learning with many relevant features*, Paper read at Proceedings of ECML-98, 10th European Conference on Machine Learning.

Wolak, J, Finkelhor, D., Mitchell, K. and Ybarra, M., (2008) Online Predators and Their Victims: Myths, Realities, and Implications for Prevention and Treatment, *American Psychologist,* Vol. 63, Issue 2, pp: 111-128.

Xiong, F., Yong F. and Tao, T., (2005) Web chat monitor systems- Research and Implementation Proceedings of the Australian Undergraduate Students, [online], Paper read at *p*roceedings of the Australian Undergraduate Students′ Computing Conference, www.ics.mq.edu.au/~tao/Web-chat%20Monitor%20System-Research%20and%20Implementation.pdf