SafeChat: Using Open Source Software to Protect Minors from Internet Predation

Brett Thom
Department of Mathematics and Computer Science
Ursinus College
Collegeville, PA USA
(+1) 610-409-3789
brthom@ursinus.edu

April Kontostathis
Department of Mathematics and Computer Science
Ursinus College
Collegeville, PA USA
(+1) 610-409-3789
akontostathis@ursinus.edu

Lynne Edwards
Department of Media and Communications
Ursinus College
Collegeville, PA USA
(+1) 610-409-3789
ledwards@ursinus.edu

ABSTRACT

This paper describes the development of a plugin for Pidgin, an open source instant messaging client. The plugin, SafeChat, allows parents to protect their children from Internet predators. SafeChat gives parents a better alternative than other currently available tools by providing a communication-theory based tool.

Categories and Subject Descriptors

K.4.2 **[Computers and Society]** Social Issues

General Terms

Algorithms, Experimentation

Keywords

Cybercrime, Internet Predation

1. INTRODUCTION

Sexual predation is defined as an attempt by an adult to engage a minor in sexual activity. The Internet provides a means for perpetrators to meet potential victims, and the number of children targeted is continuing to grow. Thirteen percent of the youth said that they were solicited online and 31% of those solicitations were for offline contact [5].

The environment we targeted was Instant Messaging. Teenagers and other minors are relying on and using Instant Messaging services more and more. Most of the time, minors use these programs to communicate with real life friends; however some minors use public chat rooms and participate in one-on-one chats with complete strangers [1]. This gives predators the perfect environment to target underage victims and pursue them.

2. SIMILAR TOOLS

There are commercial and network-level tools that can be used to protect children from Internet predation. The most common alternative to SafeChat is a packet sniffer.

Parents, however, may have problems with these tools because:

1. They are too intrusive. Parents usually do not want to monitor all their children's chat data or be too intrusive into their children's social life.
2. They require parents to read through a lot of trivial chat data. If predation detection is not designed into a tool, the parent will need to read through the data by hand.
3. Tools that are currently available to detect predation are based on a simple keyword matching and not communication theory [3]. This brings the accuracy of these tools into question.

SafeChat is designed to overcome the limitations of other tools and provides a better alternative than the tools that are currently available.

3. SAFECHAT FEATURES

For SafeChat to achieve its goal, it needs a few core features.

3.1 Chat History System

For SafeChat to achieve its goal it needs to keep track of all of the user's interactions. This is achieved using an XML file system. The chat logs are stored in a file with the name of "<chatter>.xml". The software keeps track of every chat post between the local user and the other chatter.

3.2 Detecting Ages

The age of the chat participants is very important. If the chat participants do not include a minor and an adult, the situation cannot be a case of sexual predation. To tell the age of a chat participant, we developed rules. The rules were discovered by closely analyzing chat data from Perverted-Justice.com. Perverted Justice is an organization that catches Internet Predators by having volunteers pose as teens in chat rooms and respond to adults that approach them for a sexual relationship [4]. We examined how and when ages were exchange and derived the following rules:

- A two digit number, the age of the message sender, is in a post preceded by a post with the terms such as "asl", "a/s/l", "old", or "age".
- A two digit number, the age of the message sender, is in a post that also contains terms such as "age" or "old".

Using a set of 420 chat logs from Perverted Justice, we tested the age function to see how accurately the function identified the predator's actual age. When we ran the age function on these logs, it didn't detect any age in 68 of the logs and misidentified the age in 169 of the chat logs. It correctly identified 183 of the ages. Because predators are not deceitful that they are an adult, it is more important that the age function properly detects that the chat participant is 18 years or older. The age function properly classifies 326 out of 420 predators from our test data as adults.

WebSci '11, June 14-17, 2011, Koblenz, Germany.

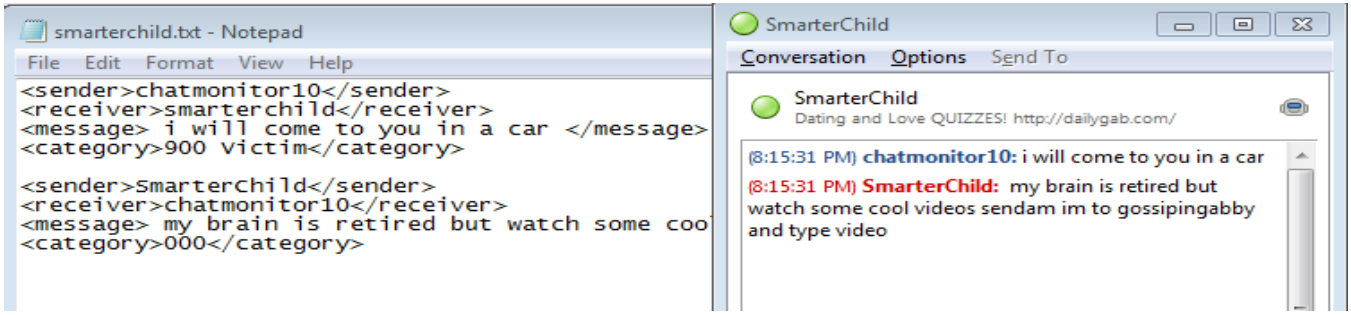

Figure 1. SafeChat Categorizing a Post

3.3 Rule-Based Categorization

SafeChat uses a rule-based approach for detecting predators. This is done by classifying chat posts into four different categories. The posts are categorized when they match a pattern. The patterns are based on term types.

The first category is when the exchange of personal information such as age, gender, location, likes, dislikes, former romances, and family life. The predators use this to build common ground with the victim and figure out what kind of support system the victim has. Grooming is the second category. This category involves the use of sexual terms. Predators use these terms in direct ways such as discussing virginity or using misspelling words on purpose to acquaint the victim with their use (i.e. "cum here" instead of "come here"). Reframing, which is the redefinition of non-sexual content into sexual terms (i.e. "I can help you become a woman") would also be identified as being in this category. The third category is when the predator attempts his approach. Examples of this are when a predator attempts to acquire a victim's address or arrange a meeting. The fourth category encompasses everything else and usually consists of innocent posts that move the conversation along [2].

The accuracy of this rule-based system was tested to see how consistent it was with three people categorizing the same 10 chat logs. Accuracy is defined as:

$$m/N$$

Where m is the number of times the computer system categorizes the predator posts the same as the human and N is the total number of posts in a chat log. This system achieved an accuracy ranging from 51.95% to 83.90% and with an average of 68.11% on 10 chat logs [2].

3.4 Predation Detection

It is preferable for SafeChat to wrongly flag something as predatory then to miss flagging a predator. If SafeChat detects that a chat participant is an adult and has a post flagged as 600 or 900 category, there is a good chance that he is a predator. SafeChat should, therefore, take necessary steps to protect the minor.

4. CONCLUSION AND FUTURE WORK

SafeChat is superior to current chat-monitoring tools because it uses an age function and rule-based predation detection based on communication theory. SafeChat allows parents to avoid reading through a lot of unimportant data, and allows them to still protect their children, while at the same time allowing their children to have privacy.

SafeChat detects an adult chatter 77% of the time and then runs a rule-based approach that categorizes the posts with 68% accuracy. SafeChat provides a strong foundation for future research on the detection of Internet Predators.

Several features need to be added before SafeChat can be released. Most importantly, it needs a mechanism for emailing parents or blocking predators. We also plan to add a feature that will collect chat data and send it to our project team for research purposes.

5. ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 0916152. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

6. REFERENCES

[1] Grinter and Palen. Instant Messaging in Teen Life. Proceedings of the 2002 ACM conference on Computer supported cooperative work. 2002.

[2] I. McGhee, J. Bayzick, A. Kontostathis, L. Edwards, A. McBride and E. Jakubowski. Learning to Identify Internet Sexual Predation. In International Journal of Electronic Commerce. To appear.

[3] Kontostathis, April, Lynne Edwards, and Amanda Leatherman. (2009). Text Mining and Cybercrime In Text Mining: Application and Theory. Michael W. Berry and Jacob Kogan, Eds., John Wiley & Sons, Ltd. 2009.

[4] Perverted Justice. www.perverted-justice.com.

[5] Wolak, Mitchell, and Finkelhor. Online Victimization of Youth: 5 Years Later. National Center for Missing and Exploited Children. 2006.