Internet Safety Technical Task Force
Technology Submission

Infoglide Software Corporation / Douglas Wood
http://www.infoglidesoftware.com

ABSTRACT
In today's world of internet anonymity, known sex offenders have unlimited access to children through social networking sites, chat rooms, and other online methods. An internet site's ability to know its customers is paramount for public safety. In the absence of a system and method for determining possible matches between users and known sex-offenders, the safety of children is compromised. Infoglide Software (Infoglide) proposes using its patented similarity search technology [1] to match users attempting to access social network sites and immediately send a real-time alert (if applicable); thereby preventing sex offenders and other "watch list" individuals from posing a threat to children's safety via the internet.

Keywords
verification, validation, similarity, searching, attributes, identity, authentication, identification, forensics, blocking

Functional Goals
Please indicate the functional goals of the submitted technology by checking the relevant box(es):

☒ Limit harmful contact between adults and minors
☐ Limit harmful contact between minors
☐ Limit/prevent minors from accessing inappropriate content on the Internet
☐ Limit/prevent minors from creating inappropriate content on the Internet
☐ Limit the availability of illegal content on the Internet
☐ Prevent minors from accessing particular sites without parental consent
☒ Prevent harassment, unwanted solicitation, and bullying of minors on the Internet
☒ Other – Prevent known sex offenders from accessing social network sites, even through attempts to mask their identities

PROBLEM INTRODUCTION
Child predators are using social networking websites as a means to prey on children. The internet serves as a good medium for these types of predators because they are able to alter their identity and manipulate user access by:

- Creating artificial information or modifying existing information in order to open a new user account/log in
- Using multiple user names
- Deploying internet 'bots' that automatically create multiple user accounts
- Creating a user on another site after being suspended from one site

Infoglide proposes a solution that prevents predators from accessing a social networking site regardless if someone attempts to use the methods of deception listed above. The solution utilizes Infoglide's Identity Resolution Engine™ (IRE) to perform a similarity search for new user identity attributes across multiple databases, which can include searching against a database that includes the attributes of suspended users from multiple sites. A positive match can result in the user being denied access to use a site or the positive match can be flagged for further review.

PROPOSED SOLUTION
The power of this solution relies on both the inherent advantage in using Infoglide's patented Similarity Search algorithms and the solution being centrally hosted, such that there is a collaborative network amongst different social network sites.

Infoglide's solution would be deployed on a centralized server, which could be hosted at Infoglide's headquarters or setup at any other remote location desired. This server runs on a JBoss application server, which allows for easy scalability depending on the load of requests it is expected to handle. By using JBoss, Infoglide's server adheres to all of the J2EE and other similar industry standards. This also means that the solution is platform independent. Both JBoss and Infoglide's technology uses Unicode, which means that it will work both domestically and with international data.

Any social network site could access the solution by a variety of APIs including Web Services (SOAP and REST), JMS, TCP/IP, and HTTP(S). This means that almost any website provider in need of this type of solution can easily integrate their solution with ours. The solution would apply both to new users attempting to register for an account, and a periodic batch check on all existing users for newly identified sex offenders.

The solution would receive all of the registration information available by the social network site and would use all of this information in determining the potential risk of the user. This would be accomplished by:

- Searching the name and date of birth of the user against all known sex offenders

- Searching the IP address against any lists of known restricted IP address ranges (i.e., prisons, mental health institutes, IP addresses known to belong to sex offenders, etc.)
- Searching all of the registrant information against a list of known bad information (i.e., "offender list) generated from any other social network site
- Searching against a white list (or positive list) of users who have been flagged as potential sex offenders but have proven that they are not
- Determining relationships between the "friends list" of a known bad user to find potential rings of offenders

All of the searches described above would be done not simply using exact matching but using patented similarity searching techniques, specific to the data fields. Using similarity searching can detect several name variation types, including the user's nicknames (including cultural nicknames), misspellings, switching multiple name fields (i.e., John Bob Smith is the same as Bob John Smith), reversing first name and last name, etc. Additionally, Similarity Search algorithms from Infoglide are available to match against other data fields such as date-of-birth, email, IP addresses, street addresses, etc.

The effectiveness of Infoglide's name matching technology has been tested several times by an independent third party organization and has been determined to be one of the top name and date-of-birth matching technologies in the industry. The testing is performed by creating a dummy watch list, a list of records to be searched against the dummy watch list, and a set of expected results. The list of records is then passed through the server, and all records are searched against the dummy watch list. The actual results of the engine are compared against the expected results and evaluated based on expected matches found, expected matches not found, and extra matches not expected but which were found.

Some of the actual use cases would include:

- A new user signs up:
  - A CAPTCHA box is utilized to ensure a real human user and not a "bot."
  - The user's information is sent to the centralized server, and all of the searches are performed.
  - If there is no suspected threat, an email is sent to the user's email address to verify that the email address is valid and to complete the user registration process.
  - If there is a suspected threat, then the relevant information is sent to someone for manual adjudication of the results.
  - If the user is determined to be a genuine risk, the user request is rejected, and all of the user's information is stored in a bad user database.
- A periodic check of a social network's user list:
  - Same as above with the exceptions that there is no CAPTCHA box and no email verification.
  - Additionally, if after manual review a user is determined to now be a genuine risk, the user's friend's list is added to a known associates database and further scrutinized to determine any potential sex offender ring.
- A user who has been revoked submits further information to prove innocence:
  - The newly supplied information is validated against more rigorous searches including searching of public data to verify:
    - the physical address supplied is a real address
    - the name really does match the public information of the person living at the address
    - the name and email match the ISP's information for the person
  - After all of these searches determine a user is not the sex offender, then the user is provided with his account, and added to a white list.

The main weakness of the solution comes from the limited number of data fields available for searching. The more registration data that can be provided, the more accurate of a picture the solution could portray. However, Infoglide has successfully developed a solution at the TSA's Secure Flight program, which relies on just the name and date-of-birth, and can accurately determine the likelihood of a passenger matching a terrorist watch list while at the same time not significantly impeding other travelers.

To accomplish this requires more of a manual review process to validate the automated matches and make the final determination on whether these should in fact be matches. A similar process would have to be invoked in this situation as well, in order to insure that only relevant users are suspended or revoked.

The solution would include sufficient audit logging throughout the system to be able to provide information both for a Freedom of Information Act (FOIA) request, as well as in the event that someone does slip through the cracks. Infoglide has met all of the National Archives and Records Administration (NARA) schedules for storing and purging of personally identifiable information (PII) in the Secure Flight program, so it is familiar with all of the associate laws and strictly abides by them.

Because of the flexibility of the solution, there have been no failures of the product. The solution can scale to meet any data specific requirements. Our technology supports over 3.5 million requests a day at Secure Flight. Our technology has searched hundred million name lists with sub-second response time. The algorithms themselves can be tuned to meet any customer specific requirements in terms of reducing false positives and/or false negatives.

EXPERTISE

Infoglide Software has extensive background in government, law enforcement, and insurance. Infoglide Software's applications continue as the platform for passenger screening within the Secure Flight program of the Transportation Threat and Credentialing (TTAC) department of the Transportation Security Administration (TSA). In addition, Infoglide Software has deployed solutions for U.S. Customs and Border Protection, U.S. Citizenship and Immigration Service, the U.K. Ministry of Defence, and a Regional Intelligence Unit (RIU) of the U.K. Police.

In recent years, Infoglide Software's business has grown to include commercial as well as government solutions. The company now offers its identity focused solutions for retail, banking, insurance, government, and law enforcement. Commercial customers include The TJX Companies, eBay, Iowa Insurance Fraud Bureau, Maryland Automobile Insurance Fund, State Farm Insurance, and MetLife.

COMPANY OVERVIEW

Infoglide Software Corporation is a Private Corporation founded in 1991. Since 1999 the company's finances have been audited by independent public accounting firms (PriceWaterhouseCoopers and PMB Helin Donovan) in accordance with Generally Accepted Accounting Principles. Financing for the company has been through institutional venture capital investors since 1996. The lead investors are Intersouth Partners of Durham, North Carolina; Stonepoint Ventures of Greenwich, Ct.; and CCP Equity Partners of Hartford Ct.

Infoglide Software develops and markets identity resolution software. Infoglide has offered the Bladeworks family of products since 2002 with the latest IRE offering released in 2007. A multi-year recipient of Inc. Magazine's prestigious Inc 500 Award, Infoglide Software has sustained impressive growth rates by providing powerful software that can resolve fraudulent identities and non-obvious relationships across disparate data sources.

Michael Shultz, President and CEO, provides strategic and operational leadership for Infoglide Software's corporate growth and development and is responsible for all aspects of the business, including future acquisition and financing opportunities. A seasoned technology executive with over 30 years of multinational experience, Mr. Shultz's background includes sales management, marketing, software development, finance, research and development, general management, and mergers and acquisitions. He is a veteran of change management and has extensive experience with financing through traditional and non-traditional methods.

Douglas Wood, Senior Vice President of Global Sales, is responsible for all sales activities in the government and commercial markets. He has over 20 years international experience in high-profile sales and marketing roles.

John Ripley, Chief Software Architect, is one of the identity resolution industry's key innovators and the holder of multiple U.S. patents, including Patent No. 7007174 [1]. Mr. Ripley has over 15 years of software development and architecture experience. Mr. Ripley has been a full-time employee of Infoglide Software for over 10 years. He is credited with designing and developing many of the features and functionalities that form the core of Infoglide's Bladeworks product suite. Mr. Ripley also played a key role in the design and creation of IRE. This includes defining Infoglide's Web Services strategy and accomplishing a significant software platform change from a proprietary server process to an open-source J2EE application server (JBoss).

Neil Stickels, Senior Solutions Architect, has over 10 years of Software Engineering and Architecture experience. Mr. Stickels' areas of expertise include designing, implementing, and deploying solutions meeting customer's demands based on Infoglide Software components. He served as a key member in architecting and implementing the Secure Flight solution and continues to serve as a member of the program's architecture team. His experience includes developing a Rules Engine and successfully integrating it into Infoglide Software's products. Mr. Stickels has been with Infoglide Software for six years and provides a vast amount of experience developing architectural systems for complex applications with custom modules and configurations.

BUSINESS MODEL OVERVIEW

Infoglide can deploy IRE with a great deal of flexibility and cost-effectiveness. IRE can be sold as either a perpetual or annual software license. Additionally, for smaller non-profit organizations, IRE can be deployed in a Software as a Service (SaaS) model.

For budgeting purposes only:

1. Perpetual IRE software licenses begin in the lower six figures and are licensed/limited by scope of use. Depending upon the number of data sources being analyzed, software license costs may increase.

2. Annual software licenses begin at approximately 40% of perpetual licenses and are also licensed/limited by scope of use.
3. SaaS models are deployed within Infoglide's secure facilities with an upfront start-up fee and a negotiated usage rate, usually a set fee per transaction.

For all three models, Infoglide charges for implementation services and ongoing annual maintenance. IRE is designed to be implemented in a matter of weeks (versus months), and Infoglide charges standard professional services rates. The annual software maintenance rate is 20%.

MORE INFORMATION

White papers and brochures are available at the following URL: http://www.infoglide.com/documents.htm

CONTACT INFORMATION

Name: Douglas Wood
Email: dwood@infoglide.com
Phone: (512) 532-3550
Address: 6500 River Place Blvd.
Building 2, Suite 101
Austin, TX 78730

CERTIFICATION

I certify that I have read and agree to the terms of the Internet Safety Technical Task Force Intellectual Property Policy.

REFERENCES

1. Infoglide Corporation. US Patent 7007174 - System and method for determining user identity fraud using similarity searching. (2006) Available at: http://www.patentstorm.us/patents/7007174.html