ChatSafe (Patent Pending)

The Carmichael Group, LLC/James Carmichael

ABSTRACT

An urgent security problem is the creation of a consistently reliable identification and authentication system within the context of the Internet, text messaging, etc. Threats of misuse are diverse including sexual predators, identity theft, cyberbullying and others. We propose an innovative solution which draws on combining existing/ proven/readily available hardware, software, and security techniques to create a unique result. The process can be described as a multi-layered, biometric approach which provides high levels of protection from this wide range of threats *now* while accommodating new technology as it progresses from theory to proven, widely available status.

Keywords

Security, Identity, Biometrics, Malicious Intent Detection.

Functional Goals

☑ 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 – please specify. Limit identity theft, flag users existing on exclusionary databases; sexual predators, terrorist watch-lists, and others as available.

PROBLEM INTRODUCTION

Internet and related communications security revolves around ensuring that access to information or system partitions fits within patterns that are socially acceptable and safe. Further, in the context of the ISTTF, the problem involves identifying unacceptable content. Age verification has been the first focus. However, the problem expands to add confirmation that no user exists in an exclusionary database (such as sexual predators, or other watch-lists). Finally the problem expands to *user intent*. Is there intent to engage in illegal acts, attempt to gather identity information, or engage in creating bullying or slanderous content?

The presently used processes generally involve steps that require the user to provide some means of identification. Then the security process attempts to authenticate that the user is who they claim to be. For example, the user may be required to respond to a request for credit card information belonging to the person they claim to be.

However, the Internet (and related communication tools) are founded on anonymity. This, combined with human behavior, renders all tools existing in the Internet world unreliable. For example, passwords may be guessed, more complex "strong passwords" are likely to be kept in written form vulnerable to copying. Physical devices may be lost, stolen, or copied. Challenge data may be incomplete, known by a sibling or be overtly shared. In essence, human behavior may override even the best designed technology.

A differentiating factor for the proposed solution is including tools not commonly used in the Internet world so that the human behavior factor is strongly addressed. It also removes the "veil" of anonymity during the login process. This opens the door to much more accurate and specific validation processes. The end result is a single solution that can address multiple Functional Goals listed above.

Some may feel hesitance about removing anonymity, concerned about privacy issues. This will be addressed in the Proposed Solution.

PROPOSED SOLUTION

The ChatSafe (Patent Pending) process is based on using in coordination existing technology and evaluation techniques. It is not dependant upon theoretical technology. The process begins by removing anonymity during validation. At login, the user is required to have real-time biometric feed(s) active. Face and voice would easily be supplied by widely used webcams (web based cameras with sound). This step opens the door to validation process paths which give broad protection. (It may be possible to use a somewhat modified version for text messaging type services by requiring a near-to-real-time self captured face/voice clip to begin a session. Thus, voice response evaluation discussed below would remain viable, as would face matching against databases).

When video is recorded and played back, there are indicators in the signal, "artifacts" of having been recorded. Software already exists to detect these [1] ensuring that real-time data is being received.

Some have expressed concern over privacy if anonymity is removed. However, users of Social Networking sites commonly post pictures of themselves to be viewed openly. The proposed solution would remove anonymity only to the highly secured validation process. Any stored data may be structured in a storage system that is built to use features such as encryption, white listing of software, messaging

oriented processing, and closed system design (elimination of unused ports available for extracting data).

One path now open would directly check the facial biometric against sexual predator databases. (It should be noted that MySpace is currently using the services of Sentinel Holding to check pictures that users choose to post against a consolidation of several databases. This service can be kept in place using the live biometrics rather than posted pictures. Face matching is established technology in use by U.S. Government agencies [2,11] and can be expanded to include databases such as watch-lists).

Another path would perform voice analysis. Scripts designed with assistance of voice analysis experts would prompt users for voice responses to questions. LVA [4] technology (already in use by police authorities [6,12]) can perform real-time evaluation for intent to deceive. Thus, the scripts can be focused on specific goals such as age verification, intent to engage in cyberbullying, etc.

As a support to all the identified paths, when a validation result appears questionable, screening staff already in place would be trained in rigorous behavior evaluation techniques. This is a live two-way process enabled by the real time communications capabilities of the Internet. A preferred technique is trade marked as Behavioral Pattern Recognition™. BPR™ has been in use at Israel's Tel Aviv airport for decades. In that time, there has been no successful attack of Tel Aviv facilities or aircraft [5,9]. It is now being integrated into Boston Logan International Airport's security processes.

Yet another path would make use of electronic age verification systems and other similar tools. As additional data sources become available results will improve. New tools are likely to become available that provide complementary checks. A proposed automated tool is the Biometric Daemon [3]. Such tools may require time to be developed and to penetrate the market place. ChatSafe (Patent Pending) is structured to be malleable to accommodate new tools when viable.

Suspicious/unacceptable results from the proposed evaluation paths may generate logon rejection, further investigation, or direct notification to a designated agency such as The National Center for Missing and Exploited Children.

Additional Data

The proposed solution has the attribute of using both established technology (webcams, face matching systems, existing and proposed databases, voice analysis systems) and the enhancement of behavioral observation. End user technology requirements are straight forward; installation of a webcam and its software. Face matching software can be executed on Social Networking servers. Alternatively, a major face matching software vendor (L-1 Identity Solutions)[7] provides large scale blade servers as a service to execute their ABIS system. Voice Analysis software can be described as a process of executing algorithms, therefore very "numeric". It can be performed on almost any platform or added to Social Networking web servers. Both face matching and voice analysis have existing deployments which demonstrate capability to perform rapidly and in large scale settings. For example, LD-1 Identity claims a processing time of under 4 seconds in U.S. Government installations where 20,000 matches are attempted per day against 85 million records each time. The key issue tends to be architecting the system for rapid throughput to the user.

The number of use cases possible is too numerous to exhaustively present due to the interplay of the various paths and the fact that ChatSafe (Patent Pending) addresses a wide range of functional goals. However, sample Use cases include the following:

A female (actually 14 years old) enrolls claiming to be 20 years old. Face matching finds no matches in any database. The automated tools produce no clear result. The voice analysis indicates intent to deceive. Result; rejection of allowance to enroll.

A male (actually 35 years old) enrolls claiming to be 14 years old. Face matching finds no matches in any database. The automated tools produce no clear result. The voice analysis indicates intent to deceive. Result; notification of internal staff. BPR™ type research is initiated. Further evidence of deceptive intention results in notification of authorities.

A male (actually 35 years old) enrolls claiming to be 14 years old. Face matching against predator database finds apparent match. The automated tools produce no clear result. The voice analysis indicates intent to deceive. Result; notification of internal staff and immediate notification of authorities.

A male (actually 15 years old) enrolls claiming to be 15 years old. Face matching against predator database finds no apparent match. The automated tools support the age claim. The voice analysis indicates intent to deceive relative to subject of adhering to rules for acceptable content. Result; notification of internal staff. BPR™ type research is initiated. Further evidence of malicious intention relative to content results in rejection or monitoring of the specific user's content.

A male (actually 23 years old) enrolls claiming to be 15 years old. Face matching against predator database finds no apparent match. The automated tools do not support the age claim. The voice analysis indicates intent to deceive relative to subject of rules for seeking to obtain personal identity information from other users. Result; notification of internal staff. BPR™ type research is initiated. Further

evidence of malicious intention results in rejection or monitoring of the specific user's exchanges with other users.

The proposed solution has capability to have high reliability as to age verification (thus limiting harmful contact between adults and minors). Both the automated tool path and voice analysis focus on this issue. Both are backed up by BPR™ type evaluation.

Limiting (1) harmful contact between minors (2) limiting bullying/harassment (3) limiting harmful content or (4) solicitation of private information are strongly controlled by combining voice analysis, (which may trigger monitoring of text content for inappropriate words or phrases), and the backing of BPR™ type evaluation address these issues. The strength of LVA (Layered Voice Analysis) has been demonstrated [4,6]. It should be expected that the scripts used to elicit voice responses will be refined over time for this application. Further, the algorithms themselves are subject to enhancement. We fully expect that new biometric tools will appear that can act as complements to Voice Analysis and thus improve the accuracy of results in this path.

The proposed solution does not limit minors from accessing inappropriate content or websites. The intent to create illegal content may be detected but illegal content created on other sites is not prevented. When there is intent by a minor (when logging in) to visit web sites not approved by parents, it may be detected.

Electronic Age Verification tools and those of similar nature can be tested by having an already known set of persons execute them, provide the data asked for, and evaluating the returned result against the already known correct response. However, human behavior used to thwart such systems may never be objectively testable. Testing may also include response time tests, especially with concern to high volume use. Data on user volume/response time may already be available from vendors in this market niche. QA type tools are also available to simulate various levels of simultaneous use.

Testing the second path (validation against databases such as sexual predator databases) can be evaluated by looking at the work being done by Sentinel Holding on behalf of MySpace. Active use by major U.S. agencies such as the F.B.I., numerous security agencies, and state Departments of Motor Vehicles provide real life measures of both performance and their viewpoint toward accuracy.

The third path starts with Voice Analysis. A provider considered "best of breed" by The Carmichael Group, LLC is Nemesysco [8]. This type of system is difficult to test in a planned setting. By its very nature, it is designed to detect voice patterns occurring in the "real world". A reference is provided of already complied accuracy results [10]. Specific scripts and refinements may be evaluated in the real world by having well trained screeners connected during logins and comparing their conclusions versus the Voice Analysis results. On this path the ChatSafe (Patent Pending) process proposes the availability of screeners trained in behavior observation techniques. A preeminent technique is the method known as Behavioral Pattern Recognition or BPR™. The strength and proven history of this technique have already been addressed above.

ChatSafe (Patent Pending) anticipates that in high volume systems, some of the validation may be done in a "batch" mode (performed as a group, during periods of low on-line use).

ChatSafe's strength is in using multiple complimentary validation processes including elements to deal with human behavior. The likelihood of being thwarted is greatly reduced over a single tool approach. The possible weakness is in failing to detect Intent to Deceive/Malicious Intent via Voice Analysis. Artificially creating "real world" situations is difficult. We expect that scripts for voice responses will require refinement by comparing logged results of any failures detected by other means, such as reports of harmful content from other users. This can lead to script refinement. Use of ChatSafe (Patent Pending) in the world of text messaging is more difficult. Users likely would be comfortable with only relatively brief validation. Near real-time capture of face and voice announcement would not always be convenient. However, parents may be concerned enough about safety to "opt in" to a service that requires the same of all allowed to exchange text messages.

Implementation for uses requires only the addition of a webcam, if not already in place. The implementation of the preferred Layered Voice Analysis can be at Social Networking/Chatroom providers web servers. Alternatively, some providers have arrays of multi-cpu blade server "farms" to perform high volume processing of their biometric systems.

To deal with end user attitudes an "opt in" structure may be the best way to introduce this enhanced security. Many parents will likely welcome a system which segregates their children from those not yet willing to opt-in.

ISO/IEC Joint Technical Committee 1 (JTC 1) Subcommittee 37 (SC 37) – Biometrics is a major standards body involved in biometric issues presented in this proposal. Standards for sexual predator databases are not yet in place.

As the various paths of validation are performed negative results must be dealt with. We propose close involvement with agencies such as The NCMEC to craft policy as to appropriate action in specific exception situations. By establishing a standard policy, adherence to law is ensured.

No difference in viability of the proposed process is seen whether deployment is U.S. or international. The main anticipated need is having multi-lingual staff as new language groups are added. LVA is not language dependant.

ChatSafe (Patent Pending) is based on already proven technology. Its effectiveness has been addressed as the technologies have been presented above. The main need anticipated is refinement of voice response scripts. Electronic Age Verification is known to have failures, sometimes due to incomplete data and sometimes due to human behavior. Face matching has a high rate of reliability, but is not 100%. Voice Analysis has an approximate 90% rate of reliability[10]. It is expected that, performing *more than one,* reliability of the end result will rise.

EXPERTISE

The Carmichael Group, LLC Technology Division is a small entity developer. Main expertise is in major systems analysis and project management. While at first this may seem to be a disadvantage, in actuality we have been positioned to approach this critical issue without any vested interest in existing software or services.

COMPANY OVERVIEW

The Carmichael Group, LLC has key founder James H. Carmichael. Mr. Carmichael has 20 years experience in information technology, including of global scale projects. As a small entity, capital has been generated from other LLC endeavors. Mr. Carmichael has cultivated informal relationships with some of the preferred providers cited and positioned to formalize these to move this proposed solution forward.

BUSINESS MODEL OVERVIEW

The Carmichael Group, LLC intends to pursue this solution via licensing or patent buyout. The costs to end users are about $110 U.S. for a high quality streaming video, echo cancellation capable webcam. For providers of services such as Social Networking, Chatrooms and the like the major costs are licensing on a usage basis of face matching, voice analysis, and electronic age verification software, as well as use of a service such as Sentinel Holding for sexual predator database processing. It may prove to most cost effective (and viable for small/start up service providers) to create a centralized service to avoid redundancies and obtain most favorable licensing structures.

MORE INFORMATION

See Contact Information below.

CONTACT INFORMATION

James H. Carmichael, The Carmichael Group, LLC 1784 Panay Circle, Costa Mesa, CA 92626. Phone 714-751-2181. E-mail jamescarmichael@firstteam.com.

CERTIFICATION

See certification at the end of the submission.

USE OF THIS DOCUMENT

Public use accepted.

REFERENCES

1. Ablavsky, V. Image Processing. 2002. Proceedings. 002 International Conference on Volume 2, Issue, 2002 Page(s): II-317 - II-320 vol. 2.
2. Automated Imaging Association. Posted April 6, 2004. Available:http://www.machinevisiononline.org/public/articles/archivedetails.cfm?id=1988.
3. Briggs, P. CHI2008. Biometric Daemons: Authentication via electronic pets. Available: http://www.chi2008.org/altchisystem/login.php?action=showsubmission&id=146.
4. Department of Psychiatry University of Tsukuba, Japan. Available:http://www.nemesysco.com/Tsukuba_LVA100907.pdf.
5. Kaplan, E., Targets for Terrorists: Post-9/11 Aviation Security. Par. 11. Available: http://www.cfr.org/publication/11397/targets_for_terrorists.html.
6. Kleinhausen, J. October 15, 2007. YouTube Demonstration of Layered Voice Analysis. Available: http://www.youtube.com/watch?v=4c6Md3TgGCE.
7. L1-Identity Solutions services. Available: http://www.l1id.com/.
8. Nemesysco Services. Available at http://www.nemesysco.com/.
9. Ron, R. President, New Age Security. Biography Available:http://www.halldale.com/Assets/Files/WATS%202007/Biographies/Rafi%20Ron%20biography.doc.
10. Van Dame, Dr. Guy, Prof. of Forensic Criminology, University of Durban Westville, South Africa. MLVA Research January 29, 2008. Available at http://www.nemesysco.com/partners/FILES/MLVA%20Research.pdf.
11. Walsh, Trudy. Government Computer News. Available: http://www.gcn.com/online/vol1_no1/36185-1.html
12. Williams, C. KNBC, Sid Heal, Los Angeles County Sheriff's Department. Available: http://www.youtube.com/watch?v=g7CMOnjsJrg

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