December 2004

Fighting Identity Fraud with the Addition of Biometric Techniques

Benjamin Ngugi
New Jersey Institute of Technology

Marilyn Tremaine
New Jersey Institute of Technology

Michael Recce
New Jersey Institute of Technology

Fighting Identity Fraud with the Addition of Biometric Techniques

Benjamin Ngugi
New Jersey Institute of Technology
Department of Information Systems
Newark, NJ 07102
benjamin.ngugi@njit.edu

Marilyn Tremaine
New Jersey Institute of Technology
Department of Information Systems
Newark, NJ 07102
tremaine@njit.edu

Michael Recce
New Jersey Institute of Technology
Center for Computational Biology
Newark, NJ 07102
michael.l.recce@njit.edu

ABSTRACT

This paper reviews current levels of identify fraud and then addresses a key component in the efforts to combat this fraud, that is authentication. Two key problems exist with authentication. First, authentication systems treat identity as a localized entity when it is actually a distributed concept. Stolen identifiers are not detected because repudiating data for the real person is not locally available. Second, authentication systems use identifying information that is prone to theft. Password, token cards and PIN numbers can be stolen whereas retinal and handwriting patterns cannot readily be "lifted." We therefore propose that the solution to identify theft involves combining multiple identification methods, some with biometric measures. We describe our research in progress using keystrokes dynamics to identify the typist, perhaps of a personal identification number. Previous work uses time latencies. We are proposing incorporating both time and pressure measures to give us more discriminative and resilient identification.

Keywords

Identity fraud, identification, authentication, access control, keystrokes dynamics, computer crime.

INTRODUCTION

Identity theft involves "*stealing of another person's personal identifying information such as Social Security number, date of birth and mother's maiden name, and then using the information to fraudulently establish credit, run up debt or take over existing financial accounts*"(General Accounts Office USA, 2002). Identity theft has been the leading fraud in the USA for the third year in a row (Federal Trade Commission USA, 2004). More than twenty seven million American have been victims of identity theft in the last five years with an average loss of $4,800 from new accounts opened using other's personal information or from the misuse of existing bank and credit accounts (Synovate, 2003). This adds up to a loss of $50 billion for 2002. The loss has not leveled out but continues to increase. The time wasted by the victims to resolve the problems created by the misuse of their personal information and restore their credit worthiness was approximated to be 300 million hours. Even higher is the cost to the business community. The identity theft resource center approximates that the losses suffered by the business communities are about $279 billion dollars without considering recovery expenditures (Identity Theft Resource Center Inc., 2003). The identity theft problem is global and threatens many nations. In the United Kingdom, identity theft was estimated to cost more than 1.4 billion British pounds as early as the year 2000 and has been increasing at a rate of 54% per annum since then (United Kingdom Cabinet Office, 2002).

EXECUTION OF IDENTITY FRAUD

The most common examples of identity fraud include credit card, check, bank debit, telephone card and social benefits fraud (Federal Trade Commission USA, 2004). A closer look at the different types of identity fraud shows that the domains of application are different, but the criminal methods used are similar. In all cases, the offender either open new accounts using stolen documents (true application fraud) or takes over existing victim's accounts (account takeover fraud). Identity fraud is both an online and offline problem. Recent trends indicate that online fraud is becoming the more dominant fraud having moved from 42% in year 2001 to 55% of total fraud in year 2003 (General Accounts Office USA, 2002). These statistics have not separated out online identify theft. However, this same document notes that identify fraud forms 42% of all fraud both online and offline, so we can likely assume that online identify theft is a growing online problem especially considering the large number of financial transactions that are now done online. We posit that most identity thieves would like to avoid appearing physically at a primary financial institution where a teller can use an account holder's picture, signature or other recorded behavioral cues to detect fraud. A thief would rather shop online or make monetary transfers online. Thus the Internet offers a better alternative with less potential of being caught. In fact a thief need not reside in the country where the theft is being performed. Although there are many reasons why identify fraud is on the rise, a pertinent issue in managing this problem is one of catching the identity thief prior to the theft. Improving our authentication technology can do this.

PROBLEMS WITH AUTHENTICATION

Two key problems exist with current authentication systems. The first of these is the localization of the identification system. Identities are often easily copied because their verification data is local, e.g., it is easy for thieves to order credit cards in someone else's name, change their address, and use that person's good credit history because there is a lack of cross checking of house sales, tax records and vehicle registrations. We will not address this issue further in this paper. A second key problem with current authentication technology is that they are mostly token-based (e.g., a bankcard) or knowledge-based (passwords or PIN numbers). Magnetic cards can be easily scanned and counterfeited. Smart cards have higher data carrying capacities and are harder to counterfeit, but they can be broken using ionizing or microwave radiation (Smith, 1999). A diligent thief can readily find personal knowledge, often through web searches. Humans have a predilection to generate passwords that are easy to remember. Unfortunately, these passwords are also easy to crack. If they are difficult to remember, the user invariably writes them down where they can be stolen. Online data is vulnerable if users are not adept at maintaining firewalls and engaging in other security measures. A solution to this problem is to use "identities "that cannot be stolen or separated from the owner. This can be achieved by combining the current identifiers with biometric methods.

COMBINING CURRENT IDENTIFIERS WITH BIOMETRIC METHODS

The performance of any biometric system is measured using both the false rejection rate (FRR) and false acceptance rate (FAR). FAR is the "*probability for an unauthorized user or a user who does not exist within the biometric systems to be falsely recognized as the legally registered user."* FRR *"is the probability of the legally registered user to be falsely rejected by the biometric system when presenting his or her biometric feature*" (Graeventiz, 2003). Unfortunately, reducing FAR increases the FRR, which frustrates the genuine user who keeps getting rejected. A compromise is required. Good biometric authentication systems should behave as follows (Polemi, 1990):

Operation: They should be easy and convenient to use with minimum time for enrollment, identification, and verification.
Technical: The device should be of reasonable size. It should be rugged and stable. It should be accurate with error rates within acceptable low levels.
Financial: The cost used to buy, install and maintain should be reasonable.
Manufacturing: The chosen biometric systems should be standardized and supported by the mainstream manufacturers. It should also be interoperable with current computer hardware and software.
Connectivity: We added the requirement that the authentication method should work online and offline.

We used the above criteria to compare existing physiological and behavioral biometric methods, namely the fingerprint verification, retina scan, iris scan, face recognition, hand geometry, speech analysis, handwritten signature verification and keystroke analysis methods. The finger print method is very accurate although associated with criminality by some people. The iris recognition method is very accurate, but has acceptance problems. The facial analysis method works well but is sensitive to facial angles and lighting. The hand geometry method is accurate and well accepted, but expensive. The hand written signature verification is also highly acceptable but not as accurate (Polemi, 1990). All of these methods are hard to implement in online systems. Speech analysis is mostly appropriate for voice-based systems. We conclude that the keystroke dynamic method has the potential of being most appropriate for our needs. This method is simple, cheap, convenient and transparent to the user. Keyboards are also commonly available and need not be online to be used. In the next section, we discuss work that has already been done on keystroke biometric measures and then present our plans for extending this work.

DESIGN AND IMPLEMENTATION OF THE KEYSTROKES DYNAMIC METHOD

It is believed that each user of a keyboard generates a completely unique typing pattern. Thus, the "*latencies between successive keystrokes, keystrokes' durations, finger placement and applied pressure on the keys can be used to construct a unique user signature/profile*" (Monrose and Rubin, 2000). Gaines et al. (1980) measured keystroke latency times for a succession of keystrokes from seven secretaries typing three passages of text. They were able to identify each secretary reliably (FAR= 0 %, FRR=4 %). Leggett et al. (1989) extended Gaines work to include more subjects and longer text (FAR=5 %, FRR=5.5). Mahar et al. (1995) improved this work with better outlier detection and the elimination of pooled variance. Garcia (1986) patented a method that uses mean latencies to form an electronic signature but did not report FAR and FRR statistics. Young and Hammon (1989) patented another method, which describes using time and pressure, but the implementation details are very scanty. Joyce and Gupta (1990) integrated the existing works to come up with a simple but elegant classifier using key digraphs (FAR=0.25 % FRR= 16.7%). Monrose and Rubin (2000) improved the sample replacement methodology used by Joyce and Gupta and obtained identification rates of 83.2%, 85.6 % and 87.2 % with the Euclidean, non-weighted and weighted Bayesian classifiers respectively. Ord and Furnelli (2000) used a neural network approach to develop a keypad authentication system (FAR= 9.9% FRR= 30%). Obaidat and Sadoun (1997) computed digraph similarities using k-means, cosine, minimum distance, Bayesian and potential function algorithms with the inter-key times, key hold times and their combination. The best identification accuracy of 100% was achieved using combined hold and inter-key time using a fuzzy network solution. Bleha et al. (1990) used a Bayesian minimum distance classifier (FRR= 8.1%, FAR= 2.8%). Bergadano et al. (2002) suggested a new method that measures the degree of disorder between two typed samples with authentication occurring only if the distance between the two samples is below a certain threshold (FAR = 4%, FRR = 0.01%). All of the above studies use time latencies. Like Young and Hammon, we propose that adding pressure patterns to the time latencies will give more discriminative and resilient identification. The next section describes our plans to implement this proposal.

EXPERIMENTAL APPROACH

We are currently running a set of experiments using a representative sample of keyboard users. Each user is given the task of typing a four-digit number into a computer. They are trained on typing in this number until their time performance scores indicate that their learning curve has asymptoted. We collect data from each subject using a pressure-sensitive keyboard, we have implemented. We have developed a driver to sample data from this keyboard and software to extract a file of pressure readings. Our next step is to model these pressure readings using wavelets analysis in order to obtain the best discriminating wavelet coefficients between a set of keystroke patterns generated by a group of volunteers. We will then apply neural net learning algorithms to the data until we obtain the best discrimination. This next step is underway and we have been able to achieve discrimination between four pilot users. Once we feel that we have developed a set of viable coefficients, we will use Bayesian modeling to combine the pressure patterns with the time patterns generated by our subjects to determine if the combination of the two patterns are even more robust in lowering our FAR and FRR rates.

Once we obtain a good set of discriminating algorithms, we will run more subjects on the keyboard using a variety of four-digit numbers, a variety of keying environments and a wider range of subject types. The focus of this research will be on determining how robust our authentication methods are under a wide range of typical conditions. Once this effort is completed, we will then combine our time-pressure biometric method with the more traditional methods (e.g., pin number) to ascertain whether this combined approach can reliably identify a person as the “true” person.

CONTRIBUTION

This paper reviews and synthesizes the current efforts going on in the fight against identity fraud. We identify some of the pertinent issues that need to be addressed. We propose a biometric measurement solution to the issue of inadequate identification and authentication technology and present an experiment and data analysis plan for implementing such a solution.

REFERENCES

1. Bergadano, F., Gunetti, D. and Picardi, C. (2002) User Authentication Through Keystrokes Dynamics, *ACM Transaction On Information And Systems Security,* **5,** 367-397.
2. Bleha, S., Slivinky, C. and Hussien, B. (1990) Computer-Access Security Systems Using Keystrokes Dynamics, *IEEE Transaction On Patterns Analysis And Machine Intelligence,* **12,** 1217-1222.
3. Federal Trade Commission USA (2004) National and State Trends in Fraud & Identity Thefts for Jan-Dec 2003, Federal Trade Commission, USA.
4. Gaines, R., Lisowski, W., Press, W. and Shapiro, N. (1980) Authentication By Keystroke Timing: Some Preliminary results. Rand Report R-256-NFS, Rand Corporation, Santa Monica, CA, 1980.

5. Garcia, J. (1986) Personal Identification Apparatus-Patent Number 4621334, US Patent and Trademark Office; Washington D.C, USA.
6. General Accounts Office USA (2002) Identity Theft: Greater Awareness and Use of Existing Data are Needed, General Accounts Office USA, Washington D. C.
7. Graeventiz, G. (2003) Biometrics in Access Control, *A& S International Taipei,* **50,** 102-104.
8. Identity Theft Resource Center Inc. (2003) Identity Theft: The Aftermath 2003, Identity Theft Resource Center Inc., San Diego , California ,USA.
9. Joyce, R. and Gupta, G. (1990) Identity Authentication Based On Keystroke Latencies, *Communications Of The ACM,* **33,** 168-176.
10. Leggett, J., Glen, W. and Umphress, D. (1989) Verification Of User Identity Via Keystrokes Characteristics. *Human Factors In Management Information Systems*.
11. Mahar, D., Napier, R., Laverty, W., Henderson, R., Hiron, M. and Wagner, M. (1995) Optimizing digraph-latency based biometric typist verification systems: Inter and Intra typist differences in digraph latency distributions., *International Journal of Human-Computer Studies,* **43,** 579-592.
12. Monrose, F. and Rubin, A. (2000) Keystroke Dynamics As A Biometric For Authentication., *Future Generation Computer Systems,* **16**.
13. Obaidat, M. and Sadoun, B. (1997) Verification Of Computer Users, Using Keystrokes Dynamic, *IEE Transaction On Systems, Man And Cybernetics Part B,* **27**.
14. Ord, T. and Furnelli, S. M. (2000) User Authentication for Keypad-Based Devices Using Keystroke Analysis, *Proceedings Of The Second International Network Conference*, Plymouth, UK.
15. Polemi, D. (1990) Biometric Techniques: Review And Evaluation Of Biometric Techniques For Identification And Authentication, Including An Appraisal Of The Areas Where They Are Most Applicable; Report prepared for the European Commission DG XIII-C.4 on the Information Society Technologies http://www.cordis.lu/Infosec/Src/Stud5fr.htm.
16. Synovate (2003) Federal Trade Commission USA-Identity Theft Survey Report: September 2003, Federal Trade Commission, USA. http://www.ftc.gov/OS/2003/09/Synovatereport.pdf.
17. United Kingdom Cabinet Office (2002) Identity Fraud: A Study, UK Cabinet Office. http://www.Homeoffice.gov.uk/Docs/Id_Fraud-Report.pdf.
18. Young, J. R. and Hammon, R. W. (1989) Method And Apparatus For Verifying An Individuals Identity. Patent Number 4,805,222 US Patent and Trademark Office Washington DC; USA.