CDFAI

CANADIAN DEFENCE & FOREIGN AFFAIRS INSTITUTE

Biometric-Based Authentication for Cyberworld Security: Challenges and Opportunities

by M.L. Gavrilova

June, 2014

Research Paper

Biometric-Based Authentication for Cyberworld Security: Challenges and Opportunities

by M.L. Gavrilova

Associate Professor, Computer Science, University of Calgary
June, 2014

Prepared for the Canadian Defence & Foreign Affairs Institute
1600, 530 – 8th Avenue S.W., Calgary, AB T2P 3S8
www.cdfai.org

ISBN: 978-0-9737870-9-2

Acknowledgements

This paper presents a summary of invited talk presented at ISPIA'2013 workshop at the University of Calgary. Some research projects mentioned were conducted by staff and students at BTLab, sponsored by CFI, NSERC, MITACS and URGC Seed grants.

Executive Summary

This paper outlines the novel research directions that have recently emerged in the biometric security domain. First, it classifies three major research directions in biometric security domain. Secondly, it reviews the state-of-the art methodologies being used for better authentication and risk management in both the real and virtual worlds, based on multi-modal information fusion, template protection and social biometrics. Finally, the paper concludes with a number of challenges intrinsic and specific to the domain of security in the Cyberworlds.

Over the past decade, the security research domain has witnessed tremendous growth in respect to all aspects of information access and sharing. Notable progress has been made in developing successful approaches to tackle the problems of user authentication, password protection, network security, data encryption, and information privacy. In the field of security research, biometric-based authentication firmly established itself as one of the most reliable, efficient, and versatile tools for providing discretionary access control to a secure resource or system [5]. While state-of-the art methods for biometric authentication are becoming increasingly more powerful and better understood, the same unfortunately cannot be said about the security of users populating on-line communities or Cyberworlds.

Ensuring safe and secure communication and interaction among users and, respectably, their on-line identities presents unique challenges to academics, as well as industry and the public. Security breaches, credit card fraud, identity theft, criminal on-line activities, and cyberbullying are just some of the Cyberworld security issues that plague society. Despite the fact that those challenges are regularly making headlines in the news, government reports, and in the IT security domain, there is an appalling lack of effort to address this urgent problem. The efforts that do exist are currently limited to network security, password protection, encryption, database security and privacy policy-making efforts [5]. However, one of the most crucial components for ensuring on-line security – the relationship of online communication among users, and their identities in the real world – has been largely overlooked. A systematic study, and targeted effort to develop effective security solutions to this crucial concern is the main focus of this paper.

BACKGROUND

Rapid growth of biometric technologies and wide accessibility to biometric capturing devices has resulted in biometric systems becoming increasingly common in different consumer and industrial applications. They can be routinely found in consumer electronics (cell phones, ipads, and notebooks), authentication documents (passports, credit cards, and IDs), access providing documents (park passes, employee badges) and even toys (EEG controlled 3D maze, fingerprint identification kits etc). Biometric data usually includes any physiological or behavioural characteristic that a human can possess, including, but not limited to, facial images, fingerprints and palm prints, iris and retina scans (as examples of physiological traits), and voice, signature, gait, keystroke patterns (as more volatile behavioural traits). The state-of-the art methodology for biometric-based authentication consists of:

a) appearance-based and feature-based methods for recognizing and comparing similar patterns in different user's biometric data [21];
b) dimensionality-reduction techniques for extracting and learning the most significant biometric characteristic, such as Principal Component Analysis (PCA), k-mean clustering and Chaotic Neural Networks [19]; and
c) novel decision-making techniques based on information fusion, Markov chains, fuzzy logic, and cognitive informatics [5].

The metrics used to evaluate the efficiency of proposed solutions usually include comparing the performance of various techniques on standardized biometric databases, publicly available or provided by vendors wishing to test state-of the art academic research in industrial settings. False Accept Rate (FAR), False Reject Rate (FRR) and their mutual relationship through ROC (Receiver Operating Curve) are such commonly used metrics. The expectations on biometric

method performance largely depends on the size of the dataset, quality of samples, sensor quality, type and number of biometrics being used, and the type of application domains [5].

While biometric authentication of users became increasingly understood, the interaction between users within government organizations, financial institutions, employers, and even friends were changing. The rise of modern means of communication, powered by on-line communities on the World Wide Web, or Cyberworlds, present unique security challenges, while also affecting the way user authentication was previously viewed and addressed. To give just a few examples, cybercrime is rampant in virtual worlds that are populated by millions of avatars and operating multibillion dollar economies [20]. International teams of hackers assisted by semiautomatic hacking software agents have perpetrated numerous attacks against the Pentagon and other government agencies computers and networks, including during the recent conflict in Ukraine. The Financial domain is not immune, and in some cases is directly dependent on activities in the Cyberworlds. In 2014, Bitcoin suffered a dramatic drop in value due to the mysterious disappearance of a large sum of its virtual currency. Since bitcoin had been accepted for real-world transactions there was direct impact to individuals and the economy in general to a lesser degree.

To counter these threats, research has been dedicated to user authentication and on-line security over the last few years. The body of literature can be broadly classified into three distinctive branches: risk analysis and program understanding; socio-economic analysis of virtual world players; and behaviour-based recognition of user identity.

The first branch encompasses works on *program understanding and risk analysis*, scrutinizing the source code of the program with the goal of understanding the original purpose and ensuring it does not contain malicious code [15,18,20]. It also contains research on robot behaviour recognition and prediction, focusing on recognition of specific instances of behaviour and estimating the degree to which a threat may be present, and includes works on robot detection and robot self-recognition [6,15,18,20]. All these works are relevant but were never translated to research of on-line user behaviour identification.

The second branch examines *the relationship between social, economic, and psychological status of game players*. A recent poll of on-line gaming community users discovered that almost 40% of the respondents asked for additional security in virtual worlds [21]. More than half of the respondents admitted to being harassed (including imprisoning, stalking, gossiping and using inappropriate language), while one third indicated that certain actions should not be permitted in the Second Life [21]. Avatars, or the second identities of virtual world users, for the most part resemble their “owners” rather than being completely virtual creations. As the physical and the virtual world come closer, the distinction between the two begins to fade and the urgent need for a security system capable of working in both contexts arises. [21]

The third branch examines the problem of *behaviour-based authentication of identity*, and is closely related to forensics and authorship recognition. It includes research on vocabulary analysis, as well as profiling plain text, emails and source code [8,9,10,13]. It is based on the idea that written text or spoken word can be analyzed in terms of vocabulary and style to create a linguistic profile and determine its authorship. Many linguistic features can be profiled including lexical patterns, syntax, semantics, information content or item distribution through a text. Once linguistic features have been established, machine learning methods can be applied to

determine the authorship, which could lead to identity confirmation [2,7,11]. This line of enquiry, however, was never extended to on-line user communities.

BIOMETRIC RESEARCH AND MACHINE INTELLIGENCE

Work by the researchers and collaborators of Biometric Technologies lab demonstrate the potential of using machine intelligence and context-based biometrics in the design of new generation security systems. The recent book "Multimodal Biometrics and Intelligent Image Processing for Security Systems" published by IGI outlines a number of methodologies [5]. It argues for the use of multi-modal biometric system, rather than the traditional single biometric approach [14]. It has been well established over the last decade that individual biometrics have a number of deficiencies, including issues of universality, uniqueness, changes over time, behaviour state dependence, poor sample quality, and human error. Due to the fact that multi-modal biometric system can incorporate two or more individual biometric traits, the overall system recognition rate can increase significantly. This remains true even in the presence of erroneous, incomplete or missing data.

The typical design of a standard biometric system is shown in Figure 1 below.

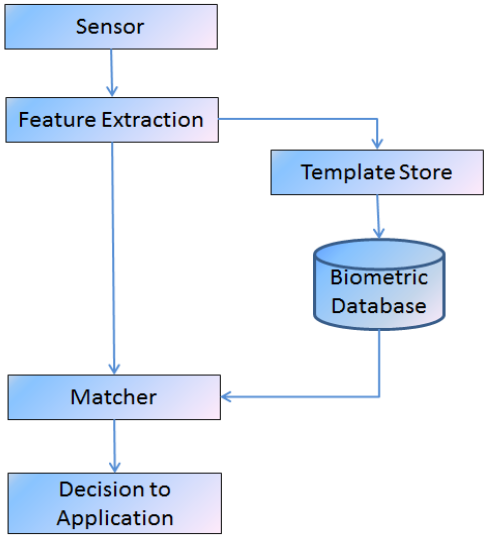

Figure 1: General block diagram of traditional biometric system [5].

The advantages of a multi-modal system over a traditional biometric system stem from the ability to analyze multiple sources of information. This means increased accuracy, fewer enrolment problems and enhanced security. All multi-modal biometric systems need a *fusion module* that takes individual data and combines it in order to obtain the authentication result: impostor or genuine user. The decision making process in a fusion module may be as simple as performing a logical operation on single bits or as complex as an intelligent system developed using principles of fuzzy logic and cognitive informatics [14,19].

Throughout the last decade, a large number of multi-modal systems have been developed in a quest to find the optimum combination of biometric characteristics and fusion approaches to minimize recognition errors. The drive to develop new methodologies based on a fusion of

features, algorithms and decision-making strategies can, on its own, be considered an intelligent approach to biometrics. Information fusion originates with decision-making strategies in a multi-sensor system that relies on a number of signals coming from varied sensors or devices; it has been quickly adopted to the biometric security domain of research. The intrinsic characteristics of biometric data and processing resulted in a variety of algorithms for biometric information fusion evolving. Combine this with standard methods relying on feature, sensor, match-level, rank-level and decision-level of information fusion and new hybrid approaches emerge. Thus, a combination of machine learning and information fusion in biometric systems has been explored recently. These systems rely on a unique combination of neural networks [3], cognitive systems [16,19], fuzzy logic [14] and adaptive learning [1]. The benefits of this approach are seen in increased performance, higher flexibility in matching biometric traits at different levels of decision-making, better circumvention (resistance to errors), fewer opportunities for intruders to undermine system security, and most importantly, the system has the ability to dynamically adapt to new information, which could include new photographs, video images, soft biometrics (height, weight, age, hair color etc), or a completely new trait such as infrared or vein biometrics. The methodology can be further extended to relevant domains, including multimedia, text and image categorization, multi-modal image retrieval and web-based social network analysis.

EMERGING METHODOLOGY

Three new approaches have been recently introduced to biometric technology with implications for Cyberworld security:

a) exploring the capabilities of multi-modal biometric fusion methods in the context of Cyberworld user identity recognition [4];
b) developing a set of metrics for identifying abnormal user behaviours through recognition of their physiological and behavioural traits [21]; and
c) introducing the notion of biometric cancellability in the context of Cyberworld authentication [16].

These new approaches will provide a powerful and unique methodology for enhancing user-security in on-line communities, and society as a whole.

A person's on-line appearance is often closely related to the user's real identity, and is relatively permanent over time. Knowing this, a novel content-based image retrieval approach based on the weighted features of colour and shape will be used as the backbone of the Cyberworlds multi-modal recognition system. Traditionally, Principal Component Analysis (PCA) or Linear Discriminate Analysis (LDA) methods have been used for visual recognition and facial feature extraction. The well-known, simple, and fast feature extraction algorithms: colour histogram and Haar Discrete Wavelet Transform (DWT) can be utilized to extract appearance features. Colour histogram method extracts the color features, and segmented matrix algorithm for Haar DWT can be applied to extract the shape attribute of an image. The weighted colour, texture, and shape features will then be combined in a single descriptor to reduce feature dimensionality and achieve reduction in required storage. To further adopt the proposed image retrieval system for on-line user authentication, texture feature can be included to boost the recognition accuracy. Next, the fusion of the three content-based low level features for appearance recognition in application to on-line user authentication takes place. University of Louisville, USA, has collected a large avatar database through both manual and automated generation. The database has provided the basic test cases of empirical studies for Artimetrics, or the study of

biometrics application to Artificial entities [20,21]. The reported results show that appearance-based methods, combined with multi-model fusion approach can be successfully used for validation of the developed methodology on avatar databases.

The second direction will leverage the behavioural biometrics for abnormal behaviour detection and risk management in online community users. Biometrics has come close to avatar development and intelligent robots/software authentication many times before. In 1998 M.J. Lyons et.al. published "*Avatar Creation using Automatic Face Recognition*", where authors discuss specific steps that need to be taken in order for avatars to be created automatically from the human face [12]. A recently published paper demonstrated the possibility of using behavioural biometic strategies designed to recognize humans to identify artificially created intelligent software agents used to gain an unfair advantage by some members of multiplayer online communities [21]. The paper lays the theoretical groundwork for research in authentication of non-biological entities. Behavioural characteristics are even less likely to change than the avatar's facial appearance and clothes, as users typically invest a lot of time and money into creation of a consistent virtual image but would not so easily change their patterns of behaviour. The artificial intelligence learning methods based on chaotic neural networks can be successfully utilized to learn normal and abnormal user behavioural patterns [4].

The third direction is focused on *protecting user confidentiality in the Cyberworld*. Privacy, for an online user, is of paramount importance. While traditional identification and verification methods (ID, smart card, password) are commonly used for on-line user authentication, biometrics are frequently more convenient for users, and come with the added benefit of reducing fraud and being more secure. In a traditional system, if a password is compromised the user can usually easily change it. However, biometrics are unique to each user, and considered to be irrevocable until very recently. *Cancelable Biometrics* [16] has recently emerged as the solution to this problem. The cancelable template generation algorithm is used to generate the new biometric template on demand (i.e. if the previous template has been compromised). The discriminability of the original biometric is not degraded after the transformation. Similarly, the cancellable user template for authentication can be used to preserve user confidentiality in the Cyberworld. In the cancelable template generation, one of the main difficulties is keeping interclass variance of the features. It was recently discovered that interclass variations that are lost from the multi-fold random projection are recoverable through fusion of different feature subsets after projection. The resulting Cancelable On-Line Authentication System will enhance the interclass variability and thus improve the overall reliability of user authentication.

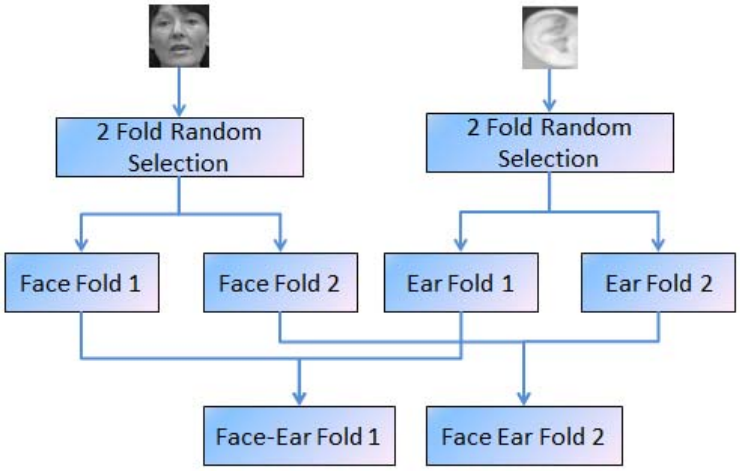

Figure 2: Block diagram of face and ear for multimodal cancellable biometric system [16].

CHALLENGES AND FUTURE RESEARCH DIRECTIONS

New research continues to emerge daily in the biometric security domain; however, a number of challenges remain that are both intrinsic and specific to Cyberworlds.

The current issues facing the biometric security domain are:

a) More efficient ways for intruder detection and prevention;
b) "Big data" challenges that demand real-time performance with high recognition rates on very large data repositories;
c) Need for better privacy policies and their enforcement to protect user confidentiality;
d) Changes in databases over time might require more sophisticated training and learning methods;
e) Ability to use contextual information obtained in parallel to main biometric features; and
f) Development of more advanced information fusions methods that have the ability to adapt to biometric system changes (such as the addition of a new module or data source) while maintaining a required level of precision.

Some questions remain open in the Cyberworld community. Areas requiring investigation include:

a) Detection of normal vs abnormal on-line user behaviour through advanced pattern analysis of appearance-based and behavioural biometrics (typing patterns, voice, text, blogs);
b) Investigating on-line social network activities as new type of biometric traits, i.e. "social biometrics" (Twitter, Wikepedia, Facebook and LinkedIn social networks etc);
c) Emerging research into spatio-temporal biometrics for on-line communities;
d) Further development of artificial biometrics domain (Artimetrics) with new data obtained on avatars, bots and other entities in virtual worlds;
e) Understanding mechanism that contribute to creating successful on-line communities; and

f) Studying leadership and popular personality traits in Cyberworlds

While work on some of the problems listed above has commenced, the majority remain untouched and present exciting future research directions in the field of biometric security.

REFERENCES

1. P. Bhattacharya and M. Gavrilova Density-Based Clustering Based on Topological Properties of the Data Set, in Generalized Voronoi Diagram: A Geometry-Based Approach to Computational Intelligence, 197-214,
2. P. Y. Chen and E. C. Liao, A new algorithm for Haar discrete wavelet transform, IEEE International Symp on Intelligent Signal Processing and Communication Systems, pp. 453-457, 2002
3. M. Gavrilova and K. Ahmadian, On-Demand Chaotic Neural Network for Broadcast Scheduling Problem, J of Supercomputing, vol 59, 811-829, 2012
4. M Gavrilova and R Yampolskiy Applying biometric principles to avatar recognition Transactions on computational science XII, 140-158, 2011
5. M. Gavrilova and M. Monwar Multimodal Biometrics and Intelligent Image Processing for Security Systems, IGI, book, 2013
6. K. M. Godby and J. A. Lane, Robot Self-Recognition Using Conditional Probability-Based Contingency Twenty-First National Conference on Artificial Intelligence, Massachusetts, 2006.
7. D. Joachim, K. Jorg, L. Edda and G. Paass, Authorship Attribution with Support Vector Machines, Applied Intelligence, pp. 109-123, 2003
8. P. Juola and J. Sofko Proving and Improving Authorship Attribution CaSTA-04 Face of Text 2004.
9. U. Kaufmann, G. Mayer, G. Kraetzschmar and G. Palm, Visual Robot Detection in RoboCup Using Neural Networks, Lecture Notes in Computer Science, pp. 262-273.
10. B. Kjell, Authorship attribution of text samples using neural networks and Bayesian classifiers, IEEE International Conference on Systems, Man, and Cybernetics, San Antonio, TX, pp. 1660-1664, 1994
11. Kitayama, M., Matsubara, R. and Izui, Y. Application of Data Mining to Customer Profile. 2002, Power Engineering Society Winter Meeting, Vol. 1, pp. 632-634, 2002
12. M. Lyons, A. Plante, S. Jehan, S. Inoue and S. Akamatsu, Avatar Creation using Automatic Face Recognition, ACM Multimedia 98, Bristol, England, pp. 427-434, 1998
13. M. R. Lyu, I. King, T. T. Wong, E. Yau and P. W. Chan, ARCADE: Augmented Reality Computing Arena for Digital Entertainment, IEEE Aerospace Conference, Big Sky, MT 5-12, pp. 1-9, 2005
14. M. Monwar and M. Gavrilova, Multimodal Biometric System Using Rank-Level Fusion Approach, IEEE Trans. on System, Man and Cybernetics, pp.867–878, vol. 39, no. 4, 2009
15. D. Ourston, Program Recognition, IEEE Expert, pp. 36-49, 1989
16. P. Paul and M. Gavrilova Multimodal Cancellable Biometric, 10th Int C on Cognitive Informatics & Cognitive Computing ICCI*CC 2012, IEEE, 43-50, 2012
17. N. Ratha, S. Chikkerur, J. Connell, and R. Bolle, Generating cancelable fingerprint templates, IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 4, pp. 561–752, 2007
18. S. J. Stolfo, S. Hershkop, K. Wang, O. Nimeskern and C.-W. Hu, A Behavior-based Approach to Securing Email Systems, Mathematical Methods, Models and Architectures for Computer Networks Security, 2776, pp. 57-81, 2003
19. Y. Tian, Y. Wang, M. Gavrilova and G. Ruhe, A Formal Knowledge Representation System for the Intelligent Knowledge Base of a Cognitive Learning Engine, Int. J. of Software Science and Computational Intelligence IJSCCI, IGI, 1-17, 2012
20. R. V. Yampolskiy and V. Govindaraju, Behavioral Biometrics for Verification and

Recognition of Malicious Software Agents, Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense VII. SPIE, 16-20, 2008.

21. S. Yanushkevich, M. Gavrilova, P. Wang and S. Srihari, Image Pattern Recognition: Synthesis and Analysis in Biometrics, World Scientific Publishers, book, 2007

About the Author

Prof. **M.L. Gavrilova** holds Associate Professor with Tenure appointment at the Department of Computer Science, University of Calgary, Canada. Prof. Gavrilova research interests lie in the areas of machine intelligence, biometric recognition, image processing and GIS. Prof. Gavrilova publication list includes over 150 journal and conference papers, edited special issues, books and book chapters, including World Scientific Bestseller of the Month (2007) – "Image Pattern Recognition: Synthesis and Analysis in Biometric," Springer book (2009) "Computational Intelligence: A Geometry-Based Approach" and IGI book (2013) "Multimodal Biometrics and Intelligent Image Processing for Security Systems". She has received support from CFI, NSERC, GEOIDE, MITACS, PIMS, Alberta Ingenuity, NATO and other funding agencies. She is an Editor-in-Chief of Transactions on Computational Sciences Springer Verlag Journal series and on Editorial board of seven journals.

Prof. Gavrilova received numerous awards and her research was profiled in newspaper and TV interviews, most recently being chosen together with other five outstanding Canadian scientists to be featured in National Museum of Civilization, National Film Canada production, and on Discovery Channel Canada.

Canadian Defence & Foreign Affairs Institute

CDFAI is the only think tank focused on Canada's international engagement in all its forms - diplomacy, the military, aid and trade security. Established in 2001, CDFAI's vision is for Canada to have a respected, influential voice in the international arena based on a comprehensive foreign policy, which expresses our national interests, political and social values, military capabilities, economic strength and willingness to be engaged with action that is timely and credible.

CDFAI was created to address the ongoing discrepancy between what Canadians need to know about Canadian international activities and what they do know. Historically, Canadians tend to think of foreign policy – if they think of it at all – as a matter of trade and markets. They are unaware of the importance of Canada engaging diplomatically, militarily, and with international aid in the ongoing struggle to maintain a world that is friendly to the free flow of goods, services, people and ideas across borders and the spread of human rights. They are largely unaware of the connection between a prosperous and free Canada and a world of globalization and liberal internationalism.

In all its activities CDFAI is a charitable, nonpartisan organization, supported financially by the contributions of foundations, corporations and individuals. Conclusions or opinions expressed in CDFAI publications and programs are those of the authors and speakers and do not necessarily reflect the views of the Institute staff, fellows, directors, advisors, or any individuals or organizations that provide financial support to CDFAI.