International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Designing a Framework for AI-Based User Authentication System through Behaviour Monitoring

Praveen Kumar Chandapeta [1], Dr. Ajay Raundale [2]

Submitted: 08/01/2024 **Revised**: 14/02/2024 **Accepted**: 22/02/2024

Abstract: In the ever-changing cybersecurity landscape, Behavior Identification has emerged as an essential aspect of protecting any digital assets. In recent years, authentication techniques have shifted to a multi-factor verification procedure. This involves proving a user's identity through many complimentary elements, which improves system security. The popularity of smart gadgets, including phones, tablets, and watches, enabled this growth. Smart devices can give a confirmation code to reinforce the primary authentication factor, often based on a user/password. Companies are looking for novel ways to improve their identification policies as data breaches and complex cyber threats become more prevalent. Designing a system with behavior identification is the need of today's hour. To detect imposters when an authenticated session is hijacked in a real-world, or to identify the imposter in a continuous authentication method was designed and tested, and a rejection test was finally carried out.

Keywords*: Cybersecurity, assets, behaviour, authenticated, tested*

1. Introduction:

To build a near-real-time, accurate, and scalable system that authenticates users using mouse dynamics in a transparent manner, several components must be properly organized. This study examines the effectiveness of using mouse behavioural attributes as a second-phase authentication method to continuously monitor user activity using AI algorithms. Further development, including parallel and distributed implementations and real-world integration, is outside the scope of this study and should be considered for future work. Authentication systems [2] are one of the most extensive security methods, and practically every computer system or service relies on them to verify users' identities and provide access rights based on their profile. This authentication process may be described as a two-step method in which an entity presents an identity to the system, and the system verifies the correspondence between the identification and the entity [5]. This procedure can prevent and detect identity theft by offering access and privileges to a legal user while refusing them to a dishonest user. However, using a common strategy for this purpose is impracticable since it would continuously interrupt the user sessions. An alternate or supplementary approach is to employ the user's behavior as a second-phase authentication method that focuses on continual monitoring of the user-system interaction. In this regard, typical input/output gadgets such as keyboard or pointing devices are the most obvious candidates for testing.

2. Literature Survey

F. Chong et al. [24] advocated that an identity be described as a set of qualities, attributes, and preferences that allow one to obtain tailored services online, on mobile devices, at work, or in a variety of other settings. It goes on to categorize the many types of identification traits into three main groups: biometrics, physical metrics, and pseudo metrics. Biometrics is the automated technique for measuring and analysing an individual's physical and behavioural traits. Fingerprint, facial, and iris scans are among examples. The physical metrics refer to what we have, and they encompass all physical-based credential tokens such as personal computers, mobile phones, and card-based credential tokens like smart cards. Good examples include passwords and personal information. In this article, the physical tokens' identifying properties shall be known to as device metrics. The device metrics include an IP address, an International Mobile Equipment Identifier (IMEI), Subscriber Identify Module (SIM), and a unique card identification number. The pseudo metrics encompass all find attributes that fall under the heading of "something you know." Password and personal identification number (PIN) are excellent examples. The three categories are utilized in this study to create a multiple-factor authentication system using information fusion, in which the user must input an identifying attribute from at least one of the three groupings. . Identity theft and identity fraud are phrases used to describe various forms of crimes in which someone illegally gets and utilizes another person's personal data for financial advantage. Cybercrime has emerged as one of the world's most rapidly expanding crimes. Identity fraud has become a big problem for both the public and corporate sectors, especially in relation to terrorism, money

[1] *Dr. A. P. J. Abdul Kalam University, Indore, M. P, India*
ORCID ID: 0009-0006-0590-7818
[2] *Dr. A. P. J. Abdul Kalam University, Indore, M. P, India*
ORCID ID: 0000-0000-0000-0000
** Corresponding Author Email: pkc06032013@gmail.com*

laundering, financial crime, drug trafficking, immigrant smuggling, and arms smuggling. To solve these security concerns, M. Hansen et al. [27] created virtual identities based on service sessions to safeguard the user's privacy from both service providers and access network providers. Another milestone is frameworks that allow users to track the history of how their identification information is treated once it is transmitted between domains of control. G. Hidehito et al. [29] propose a method for privacy-controlled exchange of identification characteristics in a federated setting. This research introduces statistical and artificial intelligence approaches in a multifactor authentication system to aid in the prevention of cybercrime.

3. Objectives:

A continuous authentication technique needs a strong feature set that enables precise identity verification of users almost instantly, in addition to a non-intrusive method of monitoring users. Therefore, these factors must be considered to construct a workable authentication method; otherwise, it will not be appropriate for use in real-world settings. The primary objectives that we purport to achieve in this study can be summed up as follows:

- Create an entirely transparent data collection process for the user to ensure that it does not impede their activity. As a result, data collection needs to be done efficiently so as not to interfere with user experience or add undue computing burden to the system. This implies that data collecting needs to be sufficiently light to be sent across the communications network effectively.
- Determine a collection of pertinent or essential characteristics that can be utilized to conclusively confirm users' identities based on how they behave when using a shared pointer device. The real owner of the active user account is the identity that needs to be validated, as our aim is to create a second-phase authentication system.
- Therefore, to identify any session hijacking, the activity that is being monitored during a given session can be compared to a pre-built profile of the user for the actual account owner. To avoid identity usurpation, the entire process—from data collection to identity verification—must be completed in close to real-time. As a result, even though the chosen techniques are computationally possible, they should not jeopardize the authentication process' correctness. Specifically, the process of extracting features and verifying identity using pre-made user profiles are crucial elements in terms of efficiency.

4. Proposed Flowchart

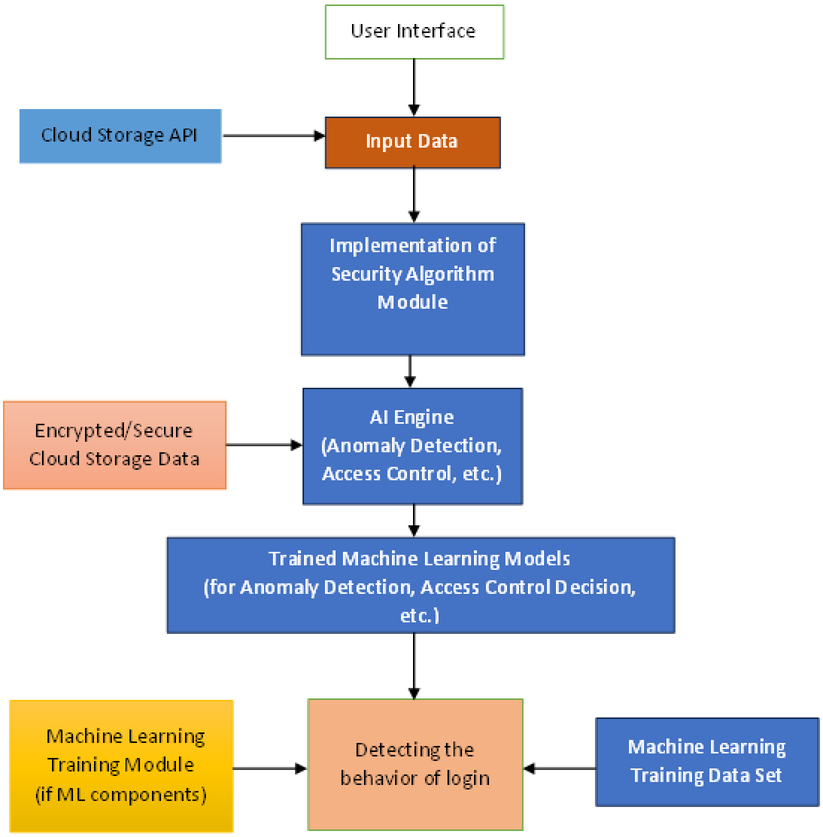
User Interface
Cloud Storage API
Input Data
Implementation of Security Algorithm Module
Encrypted/Secure Cloud Storage Data
AI Engine (Anomaly Detection, Access Control, etc.)
Trained Machine Learning Models (for Anomaly Detection, Access Control Decision, etc.)
Machine Learning Training Module (if ML components)
Detecting the behavior of login
Machine Learning Training Data Set

Fig. 1.1 Flow of the proposed system

5. Module Implementation

The raw data gathered during actual authenticate user and imposter must be analysed in order to extract the problem-relevant characteristics that may be handled effectively by the machine learning techniques. To do this, we examined many approaches in the literature that employed a variety of characteristics, mostly based on behaviour pattern such as number of times spelling mistakes was done, number of time alphanumeric character was not correct, capital and small letter placement and many more. As a result, we chose to focus on the set used for training, which allowed us to evaluate our methodologies using outside information that had not been collected by our system and in a non-guided environment. This dataset enables us to illustrate the capacity of the extracted characteristics and proposed techniques to carry out the authentication process in a real-world, general-purpose setting.

A. Cloud Storage API

Authentication testing was done on cloud storage rather than implementing on local system. This API is designed and implemented to stored the data in authenticate cloud when the behavioural patterns match with authenticate user and if then pattern of authentication model doesn't matches with the user, then it will not allow to store any data into the cloud. An API library was used to integrate the data with cloud.

B. Security Algorithm:

In the context of AI, symmetric key cryptography is used for a recognition behaviour of the authentication and stored that record for further processing

Data Encryption:

Different components of an AI system, such as a client and a server or between nodes in a distributed computing environment and for that a prototype of sending and receiving was initiated after user authentication with proper behaviour sequence

Model Protection:

Model created through AI, particularly those built using proper behaviour identification was trained on sensitive original behaviour pattern and update the model.

Secure Access Control:

Symmetric key cryptography is used to detetct behavour changes and allows the validate user or component identities and control access to certain resources.

Secure API Communication:

When AI systems communicate via APIs, symmetric key cryptography may be used to encrypt API requests and answers, limiting unwanted access and data interception.

C. Ensemble Learning Algorithms:

Random Forest: Combines multiple decision trees to improve predictive performance and reduce overfitting. Proposing, assessing, and analysing a novel implicit authentication approach using the best feature selection strategy described in the preceding model. The comparison

and analysis revealed that this algorithm may overcome the aforementioned computing constraints of behaviour.

6. Results

A. User profile Authentication

Our main objective is to authenticate the user but authentication through its username and password is not only the perspective of the system, the main intension is to track the record of the user behaviour and make sure while proceeding the process the system should keep track on its behaviour. Flow for the user profile authentication is shown in figure.

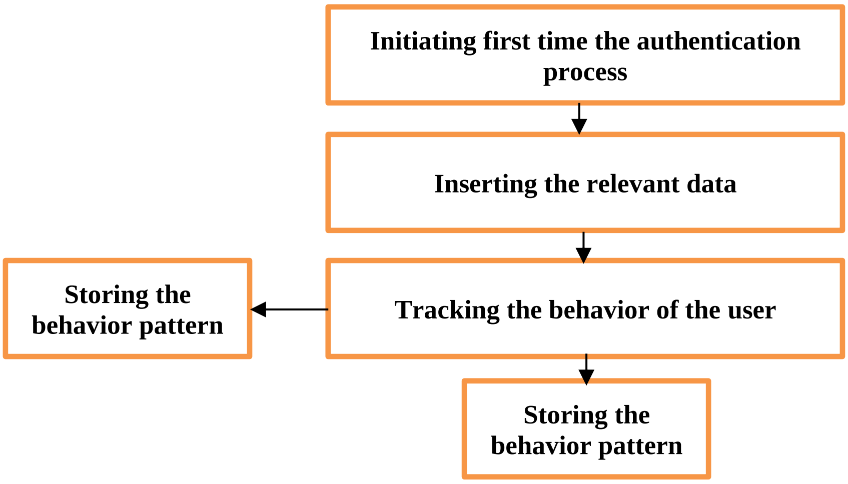
Initiating first time the authentication process
Inserting the relevant data
Tracking the behavior of the user
Storing the behavior pattern
Storing the behavior pattern

Fig. 1.2 User Profile Authentication

B. Identifying the unauthorised access

During the process of unauthorised authentication, the user tries to enter the system by validating the credentials but moreover the pattern and behaviour was changed and this change was detected by the already stored analysis and block the user and inform about its unauthorised access with retrying option.

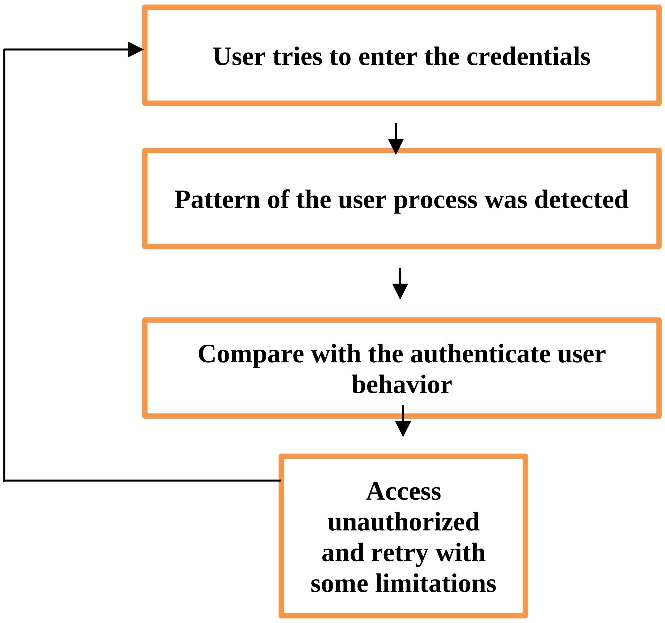
User tries to enter the credentials
Pattern of the user process was detected
Compare with the authenticate user behavior
Access unauthorized and retry with some limitations

Fig. 1.3 Identifying the unauthorised access

C. Account locked after Identifying unusual login behaviour

During the process of continuous try it was observed and detected by the system that the user making number of entries but cannot able to match the behaviour of the authenticate user, the system blocks the user for accessing the account and need to consult with administrator for the change and if the administrator confirms its identity, then only the user will able to access the system.

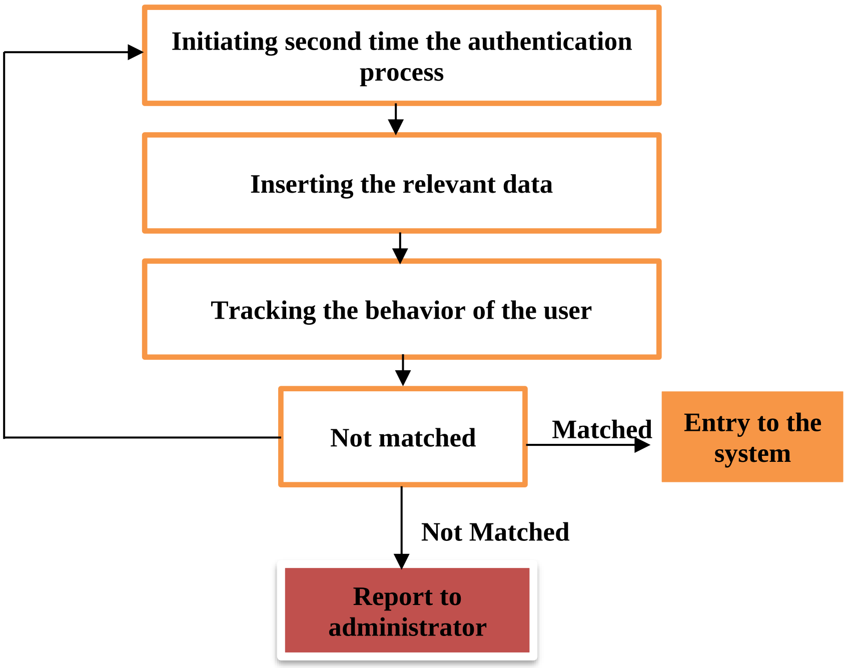
Initiating second time the authentication process
Inserting the relevant data
Tracking the behavior of the user
Not matched
Matched
Entry to the system
Not Matched
Report to administrator

Fig. 1.4 Account locked after Identifying unusual login behaviour

Table 1.1 Behaviour inputs detected

Behaviour	Spelling Mistake (%)	Spaces Mistake (%)	Alphnumeric Mistake (%)	Capital and small (%)	Repetition done(%)	Allowed	Not Allowed
User 1	**10**	**15**	**10**	**5**	**0**	100	-
User 2	**25**	**20**	**8**	**12**	**5**	100	
User 3	**35**	**23**	**2**	**18**	**6**	-	20
User 4	**42**	**26**	**9**	**23**	**3**	-	20
User 5	**48**	**28**	**4**	**26**	**4**	-	20
User 6	**75**	**65**	**7**	**29**	**2**	-	20
User 7	**95**	**23**	**6**	**27**	**1**	-	20
User 8	**25**	**28**	**5**	**56**	**8**	100	-
User 9	**28**	**29**	**1**	**28**	**9**	100	-
User 10	**29**	**27**	**1**	**82**	**7**	-	20

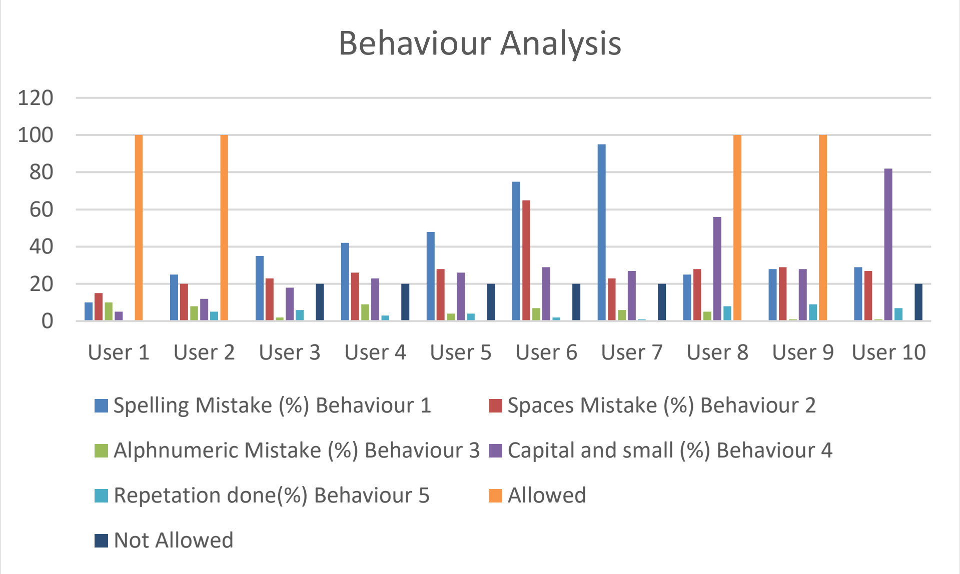

Fig. 1.5 Graphical representation of the system performance

7. Conclusion

Following the proper configuration, the models' performance was assessed using a variety of commonly used metrics, including recall, precision, and the F1 score. To begin with, a cross-validation process was run to get the baseline findings that show the suggested approaches can authenticate users that are legitimate. Subsequently, a denial test was conducted on the authentic user models constructed in the preceding phase to assess their potential to identify and avert outsider user session takeover. In this work, a comprehensive experiment was carried out to evaluate the capacity of behavioural of login procedure attributes to verify the user's identity using AI approaches. To conduct the experiment in a directed scenario, we first constructed and distributed a non-intrusive login model program participant. In this scenario, each user was given a series of tests including various mouse operations. Then, using three distinct AI techniques—MLP, SVM, and DL—a vast range of mouse movement data were retrieved, examined, and analysed to create unique user profiles based on the data that was acquired.

8. Future of Ai and Authentication

There will be further advancements and sophistication in risk-based authentication. AI-powered authentication will probably eventually transition from supervised learning—where the dataset contains the outcomes—to unsupervised learning, where the AI looks for new patterns to utilize in order to provide predictions that it may not have found in the human world. The accuracy and range of AI-based authentication services can be enhanced by cross-referencing several machines learning algorithms, applying pattern recognition, and utilizing time-series-based prediction algorithms. Simultaneously, developers will want to provide IT departments additional oversight over the AI system. This control might include the ability to ascertain precisely which data was used in a particular decision, modify the number of criteria considered, and customize the system to the specific environment of their firm. One area that businesses, such as OneLogin, are currently looking at is the capacity to use data from other parties. A stolen credential check that leverages third-party data on compromised or exposed credentials is part of OneLogin's Smart Factor Authentication. Furthermore, several cross-industry initiatives are in motion to facilitate improved data-sharing, allowing one organization's knowledge of a possible threat to be instantly shared with other businesses.

References

[1] S. Mike, Unify and Simplify: Re-thinking Identity Management, Network Security. 2006(7) (2006), 11-14. doi: dx.doi.org/10.1016/S1353-4858(06)70411-1

[2] R. Dhamija, and L. Dusseault, The Seven Flaws of Identity Management: Usability and Security Challenges, Security & Privacy, IEEE. 6(2) (2008) 24-29. doi: dx.doi.org/10.1109/MSP.2008.49.

[3] S. Clare, Digital identity - The Legal Person? Computer Law & Security Review, Elsevier. 25(3) (2009) 227-236. doi: dx.doi.org/10.1016/j.clsr.2009.03.009.

[4] P. Geraint, The benefits and drawbacks of using electronic identities, Information Security Technical Report, Elsevier. 13(2) (2008) 95-103. doi: dx.doi.org/10.1016/j.istr.2008.07.002.

[5] G. Goth, Identity management, access specs are rolling along, Internet Computing, IEEE. l9(1) (2005), 9- 11. doi; dx.doi.org/10.1109/MIC.2005.16.

[6] B. Geoff, The use of hardware tokens for identity management, Information Security Technical Report 9(1) (2004) 22-25. doi: dx.doi.org/10.1016/S1363-4127(04)00012-3.

[7] H. Marit, P. Andreas and S. Sandra, Identity management throughout one's whole life, Information Security Technical Report, Elsevier. 13(2) (2008) 83-94. doi: dx.doi.org/10.1016/j.istr.2008.06.003.

[8] EconomyWatch, List of Commercial Banks; Available (January 2011)

[9] Wikipedia, The free encyclopedia, Civil service; Available (January 2011): http://en.wikipedia.org/wiki/Civil_service

[10] Top University, QS World University Rankings; Available (January 2011): http://www.topuniversities.com/universityrankings/world-university-rankings, 2011

[11] Thomson Reuters, 100 Top Hospitals; Available (January 2011): http://www.100tophospitals.com/

[12] Google Double Click Ad Planner; Available (April 2010): http://www.google.com/adplanner/static/top1000/#

[13] Wikipedia, The free encyclopedia, List of social networking websites; Available (January 2011): http://en.wikipedia.org/wiki/List_of_social_networking

[14] L. Anthony. AntConc: Design and Development of a Freeware Corpus Analysis Toolkit for the Technical Writing Classroom, in Professional Communication Conference Proceedings, (2005), pp. 729. doi:10.1109/IPCC.2005.1494244

[15] C. Greaves, ConcApp Version 4 Concordancer, Edict Virtual Language Centre, Available (November 2010): http://www.edict.com.hk/PUB/concapp/.

[16] TextSTAT Corpus, Available (November 2010) on: http://www.edict.com.hk/PUB/concapp/ Published by Atlantis Press Copyright: the authors 429

[17] Neural-fuzzy multifactor authentication system 17. R. Togneri, and C. J. S. DeSilva, Fundamentals of Information Theory and Coding Design, (Chapman & Hall/ CRC Press, FL, 2005).

[18] M. Negnevitsky, Artificial Intelligence: A Guide to Intelligent Systems, 2nd edn. (China Machine Press, 2005)

[19] M. Fazle Azeem, M. Hanmandlu, N. Ahmad, Structure identification of generalized adaptive neuro-fuzzy inference systems, Fuzzy Systems, IEEE Transactions. 11(5) (3003) 666–681. doi:10.1109/TFUZZ.2003.817857.

[20] J. I. Agbinya, R. Islam and C. Kwok, Development of Digital Environment Identity (DEITY) System for Online Access, in Broadband Communications, Information Technology & Biomedical Applications, Third Int. Conf., (Australia 2008), pp. 23-26, doi: 10.1109/BROADCOM.2008.52.

[21] M. He, et al, Performance Evaluation of Score Level Fusion in Multimodal Biometric Systems, Pattern

[22] Recognition. 43(5) (2010) 1789-1800. doi: 10.1016/j.patcog.2009.11.018.

[23] L. Nanni, A. Lumini, S. Brahnam, Likelihood Ratio Based Features for a Trained Biometric Score Fusion, Expert Systems with Applications. 38(1), (2011) 58-63. doi: 10.1016/j.eswa.2010.06.006.

[24] F. Chong, Identity and Access Management, Microsoft Architect Journey. (2004).

[25] European Technology Assessment Group, RFID and identity management in everyday life, Available online (October 2010) at: http://www.europarl.europa.eu/stoa/publications/studies/s toa182_en.pdf

[26] The National Electronic Commerce Coordinating Council (NECCC), Identity Management, Presented at the NECCC Annual Conference, (New York, 2002).

[27] M. Hansen, A. Schwartz and A. Cooper, Privacy and Identity Management, Security & Privacy, IEEE. 6 (2008). doi: 10.1109/MSP.2008.41.

[28] M. Barisch, Modelling the Impact of Virtual Identities on Communication Infrastructures, Conference on Computer and Communications Security, in Proc. of the 5th ACM workshop, Digital identity management, (Chicago, Illinois, 2009) pp. 45–52. doi: http://doi.acm.org/10.1145/1655028.1655040.

[29] G. Hidehito, User-Centric Identity Governance Across Domain Boundaries, Conference on Computer and Communications Security, in Proc. of the 5th ACM workshop, Digital identity management, (Chicago, Illinois, 2009) pp. 35–44. doi: http://doi.acm.org/10.1145/1655028.1655038.

[30] V. Avram, Defining metrics to automate the quantitative analysis of textual information within a web page, in Int. Conf. of Application of Information and Communication Technologies, (AICT 2009), pp.1-5. d