SAND2015-2787C

Data Privacy and Security Considerations for Personal Assistants for Learning (PAL)

Elaine M. Raybourn[ab], Nathan Fabian[a], Warren L. Davis IV[a], Ray Parks[a], Jonathan T. McClain[a], Derek Trumbo,[a] Damon Regan,[b] Paula J. Durlach[b]

[a]Sandia National Laboratories
P.O. Box 5800 MS 1327
Albuquerque, NM 87185
{emraybo, ndfabia, wldavis, rcparks, jtmccl, dtrumbo}@sandia.gov

[b]Advanced Distributed Learning Initiative
13501 Ingenuity Drive, Suite 248
Orlando, Florida 32826
{elaine.raybourn, damon.regan.ctr, paula.durlach}@adlnet.gov

ABSTRACT

A hypothetical scenario is utilized to explore privacy and security considerations for intelligent systems, such as a Personal Assistant for Learning (PAL). Two categories of potential concerns are addressed: factors facilitated by user models, and factors facilitated by systems. Among the strategies presented for risk mitigation is a call for ongoing, iterative dialog among privacy, security, and personalization researchers during all stages of development, testing, and deployment.

Author Keywords

Personalized learning; privacy; security

ACM Classification Keywords

H.5.m. Information interfaces and presentation: Miscellaneous

INTRODUCTION

Learning opportunities today extend beyond traditional classroom education or on-the-job training. Educational technology such as distributed online content, serious games, simulations, collaborative virtual environments, Massively Open Online Courses (MOOCs), and intelligent tutoring systems, have accounted for much of this growth. A Personal Assistant for Learning (PAL) could further provide tailored learning content, provided it had sufficient knowledge of the learner's current operating context, and a dynamic model of the user's knowledge, experience, and learning goals. However, the need for ubiquitous data collection and user modeling introduces a number of privacy and security issues that must be considered.

PAL VISION

A future PAL, not unlike current iterations of virtual assistants, would make recommendations and monitor user activities in realtime. It has the ability to infer the user's needs and respond by making recommendations for learning, and is also ready to provide information in the form of assistance. One approach to achieving this vision is to foster interoperability among existing and future heterogeneous systems, all of which can contribute some knowledge of the user to a persistent user model [1, 2]. This user model would exist independent of any service or application, allowing those services and applications to both contribute and consume its knowledge about the user. As discussed by [1], there are many social and technical challenges associated with persistent user models. These include learners' ownership of their own data, and their ability to inspect them, as well as privacy control and protection, or security, of personal data. While privacy and security are not one and the same, both privacy *and* security need to be considered in the context of personalized learning as both impact the *trustworthiness* of a PAL. A PAL will need to address data integrity and authenticity in the context of how information is used and to what extent shared by the PAL and learners. For the PAL vision to succeed, users will need to feel that their data are safe with the PAL. They will also need to trust the PAL to use their information so that it meets their evolving requirements for privacy.

PAL SCENARIO

The following hypothetical scenario, based on a book chapter in which design recommendations were made for intelligent tutoring systems [3], introduces the desired functionality of a PAL used to support high consequence military missions. Consider the scenario of a Soldier in 2020. Captain Rivera is about to deploy to Haiti on a hurricane relief mission. Since her government-issued PAL is linked to the Army's personnel system, it already knows she is preparing to deploy and has recommended to her several relevant training exercises. Her PAL knew that she already had a badge for Level 2 French (limited working proficiency) according to the language competency criteria set by the Interagency Language Roundtable (ILR), and that she wants to get to Level 3 (professional working proficiency). Her PAL recommended some training and

IUI'15 Companion, Mar 29 - Apr 01, 2015, Atlanta, GA, USA
ACM 978-1-4503-3308-5/15/03.
http://dx.doi.org/10.1145/2732158.2732195

game-based scenarios so she could brush up on her old skills and learn some Haitian Creole. Her PAL knows that her main job while deployed will be managing the X22 search and rescue robot, so it has selected new vocabulary that will likely be useful in that situation. It also cued up the interactive technical manual for the X22. On the flight to Haiti from New Mexico, her PAL offers her various methods of informal learning. She selects a documentary about the impact of the 2017 election on the people of Port Au Prince, and she reviews a well-known report on the X22 sensors and their tendency to malfunction in humid conditions. The captain's PAL suddenly alerts her to a new blog post it has found discussing the same issue, with a solution to follow-up on later. She wishes she had thought to investigate this issue before she left and asks the PAL to load her personalized module for learning Haitian Creole.

DATA PRIVACY & SECURITY CONSIDERATIONS

In the previous section we presented a vision and scenario describing the functionality for a personalized learning system that utilizes persistent user models and context-awareness to make recommendations. In the following sections we discuss some PAL privacy and security research areas under consideration with regard to user modeling and the PAL system. Our intention is not to provide an exhaustive description of all actions necessary to secure the PAL, nor to address all privacy challenges, but rather to draw attention to some research areas and provide high-level strategies for potential risk mitigation.

Factors Facilitated by User Models and Strategies for Risk Mitigation

User modeling is an important research area for PAL. Persistent models of learners populated with information that informs related prior knowledge, prior performance, current state of cognition and physiology, and whether one demonstrates evidence of being a self-directed learner will contribute to predictions of performance.

Open learner models are defined as user models that are accessible to the user being modeled or possibly to teachers, peers, or others who may be able to enhance the model [4]. In addition to improved accuracy, open learner models are thought to enhance metacognition, motivation, and collaboration and/or competition. Users may access data, add reflections, and edit, etc., which may ultimately enhance their trust in the system.

However, there are risks associated with open learner models, especially in terms of data manipulation and falsification. For instance, a user may falsify one's own performance data (e.g. test scores, certifications, grades, etc.) to manipulate course outcomes. Manipulations of this kind would present as statistical anomalies in the data and could therefore be detected through common analysis. A more sophisticated attack however, would require more sophisticated detection. A strategy for mitigating risks requiring more sophisticated detection such as the Counter Adversarial Data Analytics (CADA) research program is discussed below in the subsequent section.

Social media sites typically expose machine interfaces, which can be used to populate open learner models. Profile, career, education, awards, training history, and qualifications maybe captured from sites such as LinkedIn, etc. Daily experience, links, feedback from peers or instructors, and sources of reflection could be captured from sites such as Facebook and Twitter. Because PAL devices may have access to social media sites or multiple apps installed (not all of which are dedicated to PAL) they are another area of concern. For instance, social media sites have become a new vector for deploying malicious links. Often apps have loose restrictions on accessing data on the device, especially in the case of smartphones where many users simply ignore the security permissions requested by the app, allowing all requests. Attacks through desktop computers such as viruses or phishing links downloaded through email are a common concern and can lead to illicit access of PAL data through that system. Finally, accessing through a lightweight web interface, possibly both on a phone and regular computer, opens the PAL system to attack through cross-site scripting (XSS), XML external entity (XEE), SQL injection, and many others currently used to attack web-based applications.

There is a natural tension between the need to protect privacy and the desire to use technical innovations that create new ways to share historically personal information in more public ways. For example, the "privacy paradox" suggests that although users report being concerned about privacy, they may take no protective action [5]. We see this tension impacting PAL research particularly as it relates to data collection for user models surrounding social media, sensors, and location-based services [6]. Understanding how data can be harvested from sensors and social media for automated population of open learner models while addressing system security as well as user privacy is an active area of research.

Allowing users to select the level of control the system has over their data and which pieces are private even from the PAL system can help build the user's trust [7, 8]. If the user knows the system will ignore things the user does not want known, then s/he will trust the system more when and where they do want to use PAL [9]. Strategies for enabling users to set permissions concerning what applications can access their data, how, when, and with whom to share data from the PAL could be explored early to address the "control paradox," or the desire for full control of personal data, but not devoting full attention its management [5].

Factors Facilitated by Systems and Strategies for Risk Mitigation

An additional risk to a PAL is unauthorized alteration of training content. Online training could also be modified to include incorrect information, compromising the training

that a user would actually receive. This could be extremely detrimental if the training were mission critical or sensitive in nature.

In cases where an actor may want to erode trust or sabotage the system, with the right level of illicit access, s/he can post or share malicious content through the PAL with potential users of the system. This could result in negative training or poor recommendations for training. By inserting false information into a user's record, recommender models can be altered to provide incorrect information [10, 11]. This may result in wasted time and resources, with the user training in areas that would not truly be recommended. It also could result in obviously incorrect recommendations, which could degrade trust in the PAL in general [12].

Sandia National Laboratories investigates how to understand the magnitude of risks to recommender systems and mitigate their effects under the Counter Adversarial Data Analytics (CADA) research program. CADA takes a look at prediction and classification algorithms, and the types of algorithms that form the basis of many recommender systems. These systems typically learn models from training data, applying these models to aid in future recommendations. As such, the accuracy of the models is dependent on the validity of the training data. If this training data is corrupted the recommender models may suffer. Different algorithms are affected by corrupted data in different ways. Researchers on the CADA project are working to quantify and classify these different effects and suggest risk mitigation strategies [13].

Of particular concern with regard to intelligent systems, is the potential to glean some information from network traffic, without needing access to the training content. If organizations are hosting training information on their servers and allowing the PAL system to access these servers, it would be possible to observe, for instance, that an organization has suddenly increased communication with a particular device manufacturer giving indication that they are using or training to use that company's device. These kinds of analyses can be done even on encrypted data without knowing the content of the information, just by analyzing the characteristics of the data traffic going over a network [14]. For example, high traffic from particular users might indicate that these users are in preparation for a particular activity. That is, if X22 search & rescue teams using PALs are sending and receiving large amounts of training data, this may indicate an upcoming, coordinated group activity.

By watching the flow of information and timing this with information about real world events, information that was formerly anonymous can be gleaned. For instance, if activity levels in the PAL grow within a certain user community and then a few days later an operation is reported by media outlets, it may be possible to infer which community is responsible for which operation. If the pattern is repeated, it can become a predictor for which group is being tapped for an upcoming operation. Data models that correlate PAL activity with real world aspects such as deployment, shipment of supplies, or other missions can be built without knowing the particular nature of the training partners or training.

Researchers in the domain of cybersecurity have employed various techniques to tackle similar issues. An example is the deployment of a network architecture called Mutable Networks (MUTE). MUTE dynamically alters the network topology of a system, making it more difficult for hackers to target particular functions of a network with attacks like Denial of Service (DoS) [15]. Although focused towards network defense, it is reasonable that such dynamic network mutation techniques could be used to obfuscate the activities of a PAL network, making it harder to use network traffic information to gain insight into user activities.

CONCLUSION

Consideration for data privacy and security should not begin at deployment of a large-scale software system. Security and privacy considerations should be interwoven into the software development process from the very beginning. Allowing security to become an afterthought to be addressed once the software is complete is a recipe for failure. An example of an important security consideration when developing software systems is the construction and execution of SQL queries, and preventing unintended execution of SQL queries (known as SQL injection attacks).

This paper is not a prescription for securing a system against attack. It is a high level discussion of the kinds of risks that could be present in personalized, intelligent systems. In truth, it is not really possible to secure a system perfectly and indefinitely against motivated attacks. The questions instead are “how long is it protected?” and “how much is protected?” The answers will not be “forever” and “all.” Instead data privacy concerns and security strategies for PAL must be considered an ongoing effort.

The first step in considering the risk mitigation strategies discussed in this paper is to perform a risk assessment of the PAL as it currently exists, even as research prototype. Our discussion in the present paper considers which attack vectors on high profile information would be exploitable in the system. Furthermore, we ask the following questions: Are there design considerations at an early stage that can prevent later stage exploits from happening such as multiple redundant servers and databases, or authorization and encryption of modules? Who has access to the system now? Who will have access to the system after it is completed? Asking these questions early moves the design toward a more secure end point removing a need to do a complete design shift at the end.

Doing a risk assessment is not the end of the road; however, security and data privacy are ongoing concerns. Engaging

experts and researchers focused on securing the PAL and integration of the PAL subsystems into the final system will help address growing privacy and security concerns. Ongoing dialog allows iterative investigation into the details of particular design decisions (which have been glossed over for brevity here) and establishment of proper protections. This dialog should carry on from prototype construction all the way through final product to ensure that PAL systems are safe, secure, and trustworthy.

ACKNOWLEDGMENTS

The authors would like to thank Bart Knijnenburg and the reviewers for their feedback. Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. The views expressed are those of the authors and do not necessarily represent the views or policies of the Advanced Distributed Learning (ADL) Initiative.

REFERENCES

1. Kay, J. and Kummerfeld, B. Lifelong learner modeling. In P. Durlach & A. Lesgold (Eds.), *Adaptive Technologies for Training and Education*. Cambridge University Press (2012), 140-164.
2. Kuflik, T., Kay, J. and Kummerfeld, B. Challenges and solutions of ubiquitous user modeling. In A. Kruger & T. Ksvi (Eds.), *Ubiquitous Display Environments*. Springer (2006), 7-30. Retrieved January 14, 2015 from: http://sydney.edu.au/engineering/it/~judy/Homec/Pubs/2012_Ubiquitous_User_Modeling.pdf.
3. Regan, D., Raybourn, E.M. and Durlach, P.J. Learner modeling considerations for a personalized assistant for learning (PAL). In R. Sottilare, A. Graesser, X. Hu, & H. Holden (Eds.), *Design Recommendations for Intelligent Tutoring Systems*. U.S. Army Research Laboratory (2013), 219-227.
4. Bull, S. and Kay, J. Student models that invite the learner in: The SMILI:() open learner modelling framework. *International Journal of Artificial Intelligence in Education* (2007), 17(2), 89-120.
5. Compañó, R. and Lusoli, W. The policy maker's anguish: Regulating personal data behavior between paradoxes and dilemmas. Eighth Workshop on the Economics of Information Security, London, UK (2009), 1-16.
6. Raybourn, E.M. and Regan, D. Exploring e-portfolios and independent open learner models: Toward Army Learning Concept 2015. In I/ITSEC Proceedings, Interservice/ Industry Training, Simulation and Education Conference (2011), Orlando, Florida, USA.
7. Knijnenburg, B. P. and Kobsa, A. Increasing sharing tendency without reducing satisfaction: Finding the best privacy-settings user interface for social networks. Thirty Fifth International Conference on Information Systems (2014), Auckland, NZ, 1-21.
8. Malhotra, N. K., Kim, S. S. and Agarwal, J. Internet users' information privacy concerns (IUIPC): the construct, the scale, and a causal model. *Information Systems Research* (2004), 15(4), 336-355.
9. Brandimarte, L., Acquisti, A. and Loewenstein, G. Misplaced confidences privacy and the control paradox. *Social Psychological and Personality Science*, (2012), *4*(3), 340-347.
10. Douceur, J.R. The sybil attack. In Revised Papers from the First International Workshop on Peer-to-Peer Systems, IPTPS '01, Springer-Verlag (2002), 251-260.
11. O'Mahony, M.P., Hurley, N.J. and Silvestre, G.C.M. Promoting recommendations: An attack on collaborative filtering. In A. Hameurlain, R. Cicchetti, and R. Traunmuller (Eds.), *Database and Expert Systems Applications*. Lecture Notes in Computer Science, 2453, Springer Berlin Heidelberg (2002), 494-503.
12. Wang, W. and Benbasat, I. Trust in and adoption of online recommendation agents. *Journal of the Association for Information Systems* (2005), Vol. 6(3) 72-101.
13. Colbaugh, R., Glass, K.L. and Zage, D.J. Increasing cyber resilience via predictability based defense. International Symposium on Resilient Cyber Systems (2013), San Francisco, CA http://www.osti.gov/scitech/biblio/1115033
14. Wright, C.V., Monrose, F. and Masson, G.M. On inferring application protocol behaviors in encrypted network traffic. *J. Mach. Learn. Res*. (2002), (7), 2745-2769.
15. Al-Shaer, Ehab. Toward network configuration randomization for moving target defense. *Moving Target Defense: Creating Asymmetric Uncertainty for Cyber Threats* (2011), 153-159.

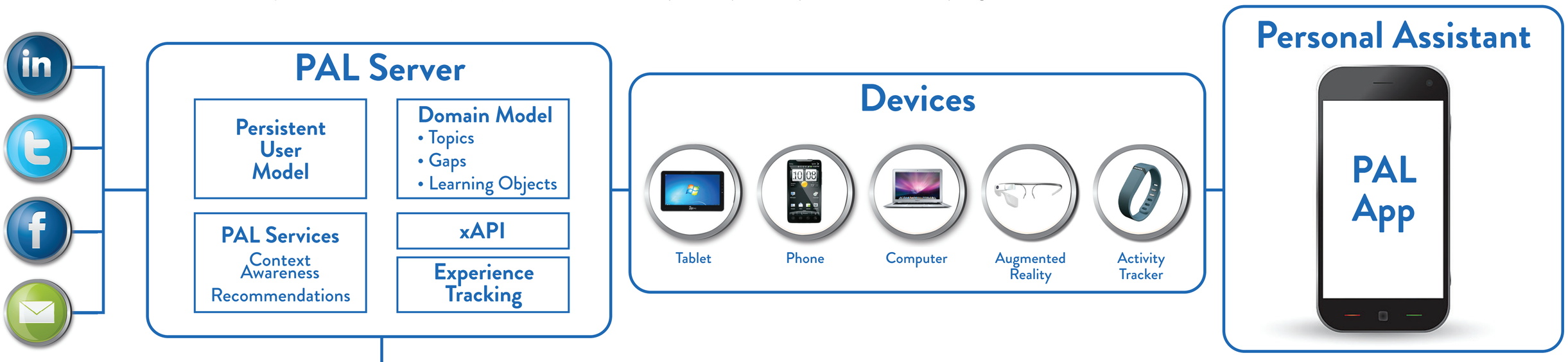
Data Privacy and Security Considerations for Personal Assistants for Learning (PAL)

Elaine M. Raybourn[ab], Nathan Fabian[a], Warren L. Davis IV[a], Ray Parks[a], Jonathan T. McClain[a], Derek Trumbo[a], Damon Regan[b], Paula J. Durlach[b]

INTRODUCTION

A future PAL, not unlike current iterations of virtual assistants, would make recommendations and monitor user activities in realtime. It has the ability to infer the user's needs and respond by making recommendations for learning. One approach to achieving this vision is to foster interoperability among existing and future heterogeneous systems, all of which can contribute some knowledge of the user to a persistent user model [1, 2]. There are many social and technical challenges associated with privacy control and protection, or security, of personal data. For the PAL vision to succeed, users will need to feel that their data are safe with a PAL. They will also need to trust the PAL to use their information so that it meets their evolving requirements for privacy.

This poster identifies some typical risks associated with user models and intelligent systems, and proposes mitigation strategies. In truth, it is not really possible to secure a system perfectly and indefinitely against motivated attacks. The questions instead are "how long is it protected?" and "how much is protected?" The answers will not be "forever" and "all." Instead data privacy concerns and security strategies for PAL must be considered an ongoing effort. The risks and proposed mitigations below are areas of research under consideration for PAL.



User Model Components

Components	Description
Profile	Basic information of the user (e.g., user identification, favorite subjects, hobbies, aspirations, goals)
Education & Training	Education and training history, grades, and feedback
Career	User's activities that demonstrate capabilities
Qualification	Official evidence data (e.g., academic transcripts, professional/vocational qualifications, certificates, licenses, and letters of recommendation)
Experience	Extra-curricular activities (e.g., clubs, internships, volunteer activities)
Outcome	Digital and non-digital artifacts that resulted from learning experience (e.g., documents, photos, animations, videos, audio files, images)
Feedback	Feedback from instructors, peers, and others from the learning process
Reflection	Personal descriptions (e.g., comments, explanations, etc.) about learning or teaching activities including perceived strengths and weaknesses

USER MODELS

Open learner models are defined as user models that are accessible to the user being modeled or possibly to teachers, peers, or others who may be able to enhance the model [3]. In addition to improved accuracy, open learner models are thought to enhance metacognition, motivation, and collaboration and/or competition. Profile, career, education, awards, and training history, and qualifications may be captured from sites such as LinkedIn, etc. Daily experience, links, feedback from peers or instructors, and sources of reflection could be captured from sites such as Facebook and Twitter.

Risk: A user may falsify one's own performance data (e.g. test scores, certifications, grades, etc.) to manipulate course outcomes.

Mitigation: Manipulations of this kind would present as statistical anomalies in the data and could therefore be detected through common analysis.

Risk: PAL devices may have access to sensors, social media sites, or multiple apps installed (not all of which are dedicated to PAL).

Mitigation: Understand how data can be harvested from sensors, social media, and apps for automated population of user models while addressing system security as well as user privacy. Strategies for enabling the user to set permissions concerning what applications can access their data, how, when, and with whom to share data from a PAL could be explored early to address the privacy and control paradoxes [4].

SYSTEMS

The primary user experience we assume for a PAL is as a mobile app. The architecture would be client-server, where the mobile app acts as a lightweight front-end client to one or more servers providing context awareness and recommendations. The mobile app would be lightweight in the sense that its primary responsibility is sending sensor and user input information to the server and reporting results back from the server to the user through the mobile UI. In some cases, it may cache information locally, but otherwise contains no real actionable data on the device.

Risk: Online training could be modified to include incorrect information, compromising the training that a user would actually receive. By inserting false information into a user's record, recommender models can be altered to provide incorrect information [5, 6].

Mitigation: Different algorithms are affected by corrupted data in different ways. Researchers in the Counter Adversarial Data Analytics (CADA) program are working to quantify and classify these different effects [7].

Risk: It is possible to gain some information from network traffic, without needing access to the training content.

Mitigation: A network architecture called Mutable Networks (MUTE) dynamically alters the network topology of a system, making it more difficult for hackers to target particular functions of a network [8]. Dynamic network mutation techniques could be used to obfuscate the activities of a PAL network, making it harder to use network traffic information to gain insight into user activities.

REFERENCES

1. Kay, J. and Kummerfeld, B. Lifelong learner modeling. In P. Durlach & A. Lesgold (Eds.), Adaptive Technologies for Training and Education. Cambridge University Press (2012), 140-164.
2. Kuflik, T., Kay, J. and Kummerfeld, B. Challenges and solutions of ubiquitous user modeling. In A. Kruger & T. Ksvi (Eds.), Ubiquitous Display Environments. Springer (2006), 7-30. Retrieved January 14, 2015 from: http://sydney.edu.au/engineering/it/~judy/Homec/Pubs/2012_Ubiquitous_User_Modeling.pdf.
3. Bull, S. and Kay, J. Student models that invite the learner in: The SMLI:() open learner modelling framework. International Journal of Artificial Intelligence in Education (2007), 17(2), 89-120.
4. Compañó, R. and Lusoli, W. The policy maker's anguish: Regulating personal data behavior between paradoxes and dilemmas. Eighth Workshop on the Economics of Information Security, London, UK (2009), 1-16.
5. Douceur, J.R. The sybil attack. In Revised Papers from the First International Workshop on Peer-to-Peer Systems, IPTPS '01, Springer-Verlag (2002), 251-260.
6. O'Mahony, M.P., Hurley, N.J. and Silvestre, G.C.M. Promoting recommendations: An attack on collaborative filtering. In A. Hameurlain, R. Cicchetti, and R. Traunmuller (Eds.), Database and Expert Systems Applications. Lecture Notes in Computer Science, 2453, Springer Berlin Heidelberg (2002), 494-503.
7. Colbaugh, R., Glass, K.L. and Zage, D.J. Increasing cyber resilience via predictability based defense. International Symposium on Resilient Cyber Systems, San Francisco, CA (2013), http://www.osti.gov/scitech/biblio/1115033.
8. Al-Shaer, Ehab. Toward network configuration randomization for moving target defense. Moving Target Defense: Creating Asymmetric Uncertainty for Cyber Threats (2011), 153-159.

[a]Sandia National Laboratories
P.O. Box 5800 MS 1325
Albuquerque, NM 87185
{emraybo, ndfabia, wldavis, rcparks, jtmccl, dtrumbo}@sandia.gov

[b]Advanced Distributed Learning Initiative
13501 Ingenuity Drive, Suite 248
Orlando, Florida 32826
{elaine.raybourn, damon.regan.ctr, paula.durlach}@adlnet.gov

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.