12-15-2019

Understanding Unauthorized Access using Fine-Grained Human-Computer Interaction Data

Michael D. Byrd
University of Arizona, byrd@email.arizona.edu

Jeffrey L. Jenkins
Brigham Young University, jeffrey_jenkins@byu.edu

David Kim
University of Arizona, davidkim7@catworks.arizona.edu

Manasvi Kumar
University of Arizona, manasvik@email.arizona.edu

Andrew W. Schwartz
Brigham Young University, aws32@byu.edu

See next page for additional authors

Understanding Unauthorized Access using Fine-Grained Human-Computer Interaction Data

Research in Progress

Michael D. Byrd
University of Arizona
byrd@email.arizona.edu

Jeffrey L. Jenkins
Brigham Young University
jeffrey_jenkins@byu.edu

David Kim
University of Arizona
davidkim7@catworks.arizona.edu

Manasvi Kumar
University of Arizona
manasvik@email.arizona.edu

Andrew W. Schwartz
Brigham Young University
aws32@byu.edu

Joseph S. Valacich
University of Arizona
valacich@email.arizona.edu

Parker A. Williams
University of Arizona
parkerwilliams@email.arizona.edu

Ryan T. Wright
University of Virginia
rtw2n@comm.virginia.edu

ABSTRACT

Unauthorized Data Access (UDA) by an internal employee is a major threat to an organization. Regardless of whether the individuals engaged in UDA with malicious intent or not, real-time identification of UDA events and anomalous behaviors is extremely difficult. For example, various artificial intelligence methods for detecting insider threat UDA have become readily available; while useful, such methods rely on post hoc analysis of the past (e.g., unsupervised learning algorithms on access logs). This research-in-progress note reports on if the analysis of Human-Computer Interaction (HCI) behaviors, which have been empirically validated in various studies to reveal hidden cognitive state, can be utilized as a method to detect UDAs. To examine this, an experimental design was required that would grant the subjects an opportunity to engage in UDA events while tracking the HCI behaviors in an unobtrusive manner. Background, experimental design, study execution, preliminary results, and future research plans are presented.

Keywords

Unauthorized Data Access, Insider Threats, Human-Computer Interaction, Mouse Cursor Movements, HCI Dynamics, Deception Detection.

INTRODUCTION

All modern organizations are threatened by the menace of Unauthorized Data Access (UDA). UDA refers to the unsanctioned access of an organization's data and information resources (e.g., customer records, intellectual property, trade secrets, etc.) by employees, contractors, or outsiders. 69% of organizations reported one or more unauthorized theft or corruption of data by insiders in 2016 (McClimans, Fersht, Snowden, Phelps and LaSalle, 2016). Reports suggest that there are two basic types of individuals that engage in UDA (Upton and Creese, 2014): (1) non-malicious individuals engage in UDA out of curiosity, boredom, or even the desire for recognition, (2) malicious individuals engaged in criminal activities that are motivated by monetary gains or revenge against the organization. Non-malicious individuals tend to work on their own while the malicious actors have been known to hire hackers on the dark web that help them identify and sell valuable data (Scott and Spaniel, 2017).

Security experts believe that most UDA events are unknown to the organization (Protenus, 2017), and many of those that are known, go unreported out of fear of how such disclosure will damage the organization's reputation (Jackson, 2017). Nevertheless, highly visible reports of UDA are numerous and seemingly unending. This holds true for governmental agencies and contractors. Edward Snowden, for example, engaged in one of the highest-profile acts of UDA. In 2013, Snowden accessed and stole upwards of 1.7 million files while as a contractor for the NSA (Harding, 2014). More recently, the intelligence contractor Reality Winner printed and sent classified reports on Russia's interference in the 2016 election to the news outlet, The Intercept. Winner was caught, not because of access control logs or file tracking technology, but due to unnoticeable marks left on the document by the printer that were published online by the Intercept. These marks identified the serial number of the printer Winner used (Errata Security, 2017).

Businesses are also not immune to UDA insider threats. In fact, security experts are adamant that malicious insiders pose the greatest risk for data security threats (Giandomenico and Groot, 2017). The problem of UDA was so pervasive that Carnegie Mellon University had set up a database which tracks over 1000 publicly available insider incidents in the US (S.E. Institute, 2017). Clearly, UDA is a massive problem that can have a wide range of deleterious effects on virtually any organization.

Many organizations nowadays use cloud-based File Management System (FMS), also known as Enterprise File Synchronization and Sharing (EFSS) that allows the management of data files remotely over the Internet and permits access to the files to those who have been granted access, authorized users. An EFSS allows an organization to create, edit, delete, and share various files (e.g., text documents, spreadsheets, presentations, graphics, images, videos, code repositories, etc.) in an organized manner with individuals within the organization. With the onset of Agile and DevOps practices, teams change more rapidly than ever before. Not only do the administrators have to manage access to the ever-growing quantity of data, but they must also maintain proper access to the rapidly changing organization and teams. One study reports that 62% of business users have access to data they probably should not be able to view (Ponemon Institute, 2016). Maintaining proper access control to sensitive information is an unsolved problem.

With data proliferating on EFSS services that are outside the organizations typical network, the problem of access control has become practically impossible to manage. To mitigate the risks involved with UDA, organizations nowadays employ different techniques. First, most UDA technologies such as data encryption, remote backups, and media tagging focus on mitigating data loss so an insider cannot change or delete sensitive information (Bose, Avasarala, Tirthapura, Chung and Steiner, 2017). Thus, many existing approaches do not detect UDA but simply mitigate the downside risk of data loss. Recently, various artificial intelligence-based methods for detecting insider threat UDA have become available. These approaches are primarily employing unsupervised machine learning algorithms on access logs. For example, such analyses include the examination of unusual geolocation, excessive data transmission, usual device, or application access.

Although promising, the current UDA technology relies on a post hoc analysis of the past. None of these approaches can provide a likelihood estimate of an individual's intention in near real-time while interacting with the EFSS. Furthermore, according to Data Breach Incident Report published by Verizon from 2014 - 2018, more than 70 percent of the unsanctioned acts such as unauthorized access to a database takes at least a month and even years to be discovered (Verizon, 2019). In fact, the report also suggests that only 4 percent of the Insider and Privilege Misuse breaches were discovered via fraud detection, which is a typical prevention technique that many organizations employ.

To address the limitations with the currently existing preventative measures, we propose a novel experiment that utilizes Human-Computer Interaction (HCI) devices, which are mass deployable, cost-effective and empirically validated in various studies that investigate the cognitive process of individuals with malicious intent (Valacich, Jenkins, Nunamaker, Hariri and Howie, 2013; Hibbeln, Jenkins, Schneider, Valacich and Weinmann, 2014). Modern computing devices, including computer mouse, are equipped with an array of sensors that can be used to capture and measure the motor movements of users with fine detail and precision. For example, a computer mouse streams finely grained data at millisecond precision that can be translated into various statistical features reflecting changes in speed and movement trajectories that can be used to classify individuals with malicious intent. In particular, this preliminary report focuses on the context and experimental design to answer the following research question:

RQ How does the interaction behavior related to Unauthorized Data Access (UDA) differ from those that do not engage in such behavior?

To address this question, we have developed an EFSS simulator that allows the user to open and access different folders and files. By creating the simulated EFSS and having participants perform tasks where actual behavior and movements can be tracked, UDA behaviors can be examined at a granular level.

Cognitive Process and Motor Movement

Recent neuroscience research has unequivocally demonstrated that linkages exist between cognitive processing (e.g., cognitive conflict, emotion, arousal, etc.) and hand movements. Motor movements were once thought to be the end-result of cognitive processing. However, a broad range of work has demonstrated that cognitive processing influences motor movements on an ongoing and continuous basis, even as mental processes are still unfolding. The "movement of the hand … offer continuous streams of output that can reveal ongoing dynamics of processing, potentially capturing the mind in motion with fine-grained temporal sensitivity… [revealing] hidden cognitive states that are otherwise not availed by traditional measures" (Freeman, Dale and Farmer, 2011).

Mouse cursor tracking as a scientific methodology was originally explored as a cost-effective alternative to eye-tracking to denote where people devote their attention in a human-computer interaction context (Byrne et al. 1999; Freeman et al. 2011; Williams et al. 2016). Dozens of studies have chosen mouse tracking for studying various cognitive and emotional processes. For example, viewing negative emotional images, increasing a person's stress level, viewing atypical information, and so on has been found to increase motor evoked potentials, hand and arm force production, and mouse movements (e.g., Dale, Kehoe

and Spivey, 2007; McKinstry, Dale and Spivey, 2008; Coelho, Lipp, Marinovic, Wallis and Rick, 2010; Freeman et al. 2011; Coombes, Naugle, Barnes, Cauraugh and Janelle, 2011). In the field of Information Systems (IS) alone, HCI devices such as facial and voice recognition system, mobile phone and computer mouse have been used to study deceptive behaviors and malicious intents (Freeman et al. 2011; Valacich et al. 2013; Hibbeln et al. 2017; Byrd, Jenkins, Valacich and Williams, 2018; Jenkins, Proudfoot, Valacich, Grimes and Nunamaker, 2019).

Many deception detection and fraud detection studies using HCI devices were conducted based on psychological findings that human's emotional response to a stimulus (e.g., feeling negative emotions, having a cognitive conflict) influences the level of cognitive load (Greene, Nystrom, Engell, Darley and Cohen, 2004). For example, when a person is induced with negative emotions while completing a goal-oriented task, some portion of the cognitive resources will be spent to react to the negative emotions. Thus, the total amount of cognitive resources that one can spend on a primary task will be reduced. On the other hand, the amount of cognitive effort that one needs to complete the task remains unchanged. As a result, the overall execution of the primary task will be impacted. In terms of HCI device usage, users will take additional time to complete the task while continuing to interact with the device (i.e., holding the mouse while having a cognitive conflict) and will deviate from the ideal trajectory path (i.e., dragging the mouse while hesitating to decide).

Hence, we propose that mouse movements can reveal hidden cognitive states during a UDA event. More specifically, when a person knowingly engages in a UDA event, the person is more likely to experience cognitive or moral conflict. Likewise, malicious people are more likely to experience hesitations as they reconsider their planned and current actions. Such cognitive and moral conflict can influence one's fine motor control, including hand movements (Freeman et al. 2011). For example, when a person is moving the mouse to engage in a UDA event, the person is much more likely to experience cognitive or emotional changes due to the thoughts related to the act itself.

In addition to an increase in predictable movement anomalies, malicious individuals will also increase the likelihood of engaging in various behavioral events that are indicative of illicit acts. For example, a person engaging in UDA may have a vastly different pattern of behaviors than a person carefully performing their work-related duties. The malicious user, for example, may quickly open and close a sequence of files as they promptly search for desired information. On the other hand, a non-malicious person will more carefully and deliberately navigate and select files to interact with. For this study in progress, we have focused on examining the difference in behaviors between the subjects that have engaged in UDA event and those who did not.

METHODOLOGY

To understand how unauthorized access impacts the interaction with HCI devices, an experiment was conducted. Participants were asked to complete various tasks listed in a google form by accessing an external file management system. They were explicitly asked not to open any files which were not relevant to the question. To entice the participants into engaging in UDA, various files with names suggesting potentially sensitive information (Credit card information, Social Security Number information, etc.) were placed in some of the folders (henceforth known as Misleading files). While the participants were performing the task, all interaction data, including clicks, coordinates, etc. were captured at millisecond precision. This data was later analyzed to see if there were any significant differences in behavior during authorized and unauthorized access.

Experiment Tool

The participants of the experiment were provided a google form which gave them access to the "File Share" system (Figure 1). They were instructed to test the system by performing a series of tasks: (1) accessing files on an external "File Share" system (link provided) and (2) recording their answers on the Google form. The system and tasks were designed to ensure that participants operated in a simulated organizational context and could perform various tasks typical of such employees. The folders to navigate were present on the left side of the screen while their constituent files were displayed on the right side (similar to file management systems on most operating systems). The system also captured the timing and occurrence of events as well as real-time HCI dynamics of a user (e.g., fine-grained data that is used to create various types of metrics for analysis).

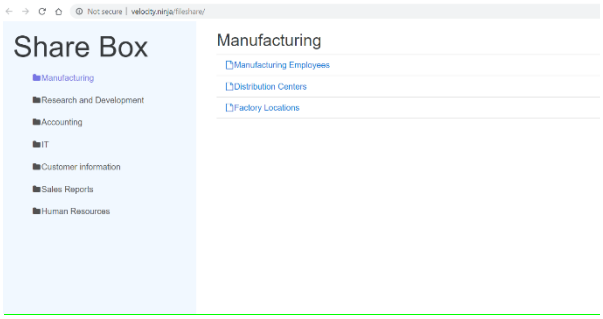

Figure 1. 'File Share' EFSS Simulator

A total of 15 tasks were allotted to each participant. Each task began with a question followed by detailed instructions on how to answer the question. The instructions suggested the participant to access a particular file within a specific folder to look for the answer. The answers consisted of numbers (sales figures, phone numbers, etc.) or text (URL, Address, etc.) and were

designed to be short to make it relatively easy to report them. There were a total of 7 folders in the File Share system. As the participant performed the assigned tasks, their HCI dynamics and behavioral patterns were captured for later analysis.

While all the tasks the participants performed were routine, they sometimes involved accessing folders which had the strategically placed Misleading files. For example, participants who accessed the folder named "Accounting" while performing task 8 had access to a file named "Credit Card Information," which was located in the same folder. A total of 5 such files were placed in various directories so that they could have been accessed while performing 7 of the 15 tasks. Thus, while participants were explicitly told not to look at other files, they were enticed to do so through the course of the experiment.

Participants

The participants of the experiment were undergraduate students at a private US University. The students at the University adhere to its strict honor code (stricter than most other US Universities). Participants were compensated through extra credit and were not allowed to retake the experiment. No personal information was collected as part of the experiment and the identity of the participants was kept anonymous.

Preliminary Findings

A total of 125 participants took part in the experiment. The data collected from the system was processed and analyzed using an in-house tool, and a total of 1894 accesses were determined. Of these, only 7 were unauthorized (0.37%). These 7 unauthorized accesses were performed by 6 different users (4.8%). Two of these 7 accesses were one of the "misleading files" – both of the same file named "Task Answers" (which contained answers to all tasks in the experiment). As there were very few UDAs, any analysis performed on the data wouldn't have much statistical significance.

Results from data collection showed that most participants of the experiment did not engage in UDA. One potential reason for observing this behavior could be attributable to the surrounding environment of the participants. More specifically, the students who took part in this experiment belong to a university whose honor code is stricter than most other US Universities. Another reason for observing a lower rate of malicious behavior could be due to a lack of incentives provided for a participant to engage in unauthorized behavior. In fact, the only misleading file participants performed UDA on was on a file named "Task Answers" which contained actual answers of the various tasks they were allotted (i.e., information that could benefit them to complete the task). However, as this file was placed in a folder to be accessed at the very last task of the experiment, the participants had less motivation to access the file.

LIMITATIONS AND FUTURE WORK

To address the low cheating rates, a few changes in experimental design are being considered. Changing the participant pool to students from a US university with regular honor codes could help in improving the generalizability of the data and results. As mentioned earlier, there are two types of individuals who engage in UDA, those driven by malicious intent for monetary gain and those driven by curiosity, boredom etc. The low cheating rates obtained suggest that not many participants were curious enough to view the misleading files. To improve cheating rates, it is imperative to provide greater incentive for some participants to cheat (e.g., financial rewards for UDA). Previous work (Byrd el al. 2018) also indicates that offering additional aligned incentives to participants for cheating can help increase cheating rates. Various approaches to provide monetary incentives to cheat are being considered.

CONCLUSION

Assessing internal threats within an organization through UDA is very challenging. This research in progress note aims to address this issue by examining changes observed in interaction behavior when subjects commit malfeasant acts. An experiment was designed to allow for participants to operate in a simulated organizational context and perform typical tasks while providing opportunities to perform UDAs at various stages unobtrusively. Initial data collection from a pool of 125 participants revealed very few instances of UDA (0.37% of all instances). We posit that participants didn't engage in malicious behavior due to a couple of reasons – lack of incentive to cheat and a strict surrounding environment (reinforced by a strict honor code in their parent institution). To encourage participants to engage in UDA, a change in experimental design is proposed. Once redesigned, the experiment will be reconducted and results from the data collection will be reported in the future.

ACKNOWLEDGEMENTS

The authors would like to thank the participants of the experiment for their time and effort.

REFERENCES

1. Böse, B., Avasarala, B., Tirthapura, S., Chung, Y. Y., and Steiner, D. 2017. "Detecting Insider Threats Using RADISH: A System for Real-Time Anomaly Detection in Heterogeneous Data Streams," *IEEE Systems Journal*, vol. PP, no. 99, pp. 1-12.

2. Byrd, M., Jenkins, J. L. Valacich, J. S., and Williams, P. A., "Creating a Realistic Experimental Scenario for HCI-Based Deception Detection Research with Ground Truth and Unsanctioned Malicious Acts" (2018). *SIGHCI 2018 Proceedings*. 18.

3. Byrne, M. D., Anderson, J. R., Douglass S., and Matessa M. 1999. "Eye Tracking the Visual Search of Click-Down Menus," the *Proceedings of the*

SIGCHI conference on Human Factors in Computing Systems.

4. Coelho, C. M., Lipp, O. V., Marinovic, W., Wallis, G., and Riek, S. 2010. "Increased Corticospinal Excitability Induced by Unpleasant Visual Stimuli," *Neuroscience Letters*, vol. 481, no. 3, pp. 135-138.

5. Coombes, S. A., Naugle, K. M., Barnes, R. T., Cauraugh, J. H., and Janelle, C. M. 2011. "Emotional Reactivity and Force Control: The Influence of Behavioral Inhibition," *Human Movement Science*, vol. 30, no. 6, pp. 1052-1061.

6. Dale, R., Kehoe, C., and Spivey, M. J. 2007. "Graded Motor Responses in the Time Course of Categorizing Atypical Exemplars," *Memory & Cognition*, vol. 35, no. 1, pp. 15-28.

7. Errata Security 2017. "How the Intercept Outed Reality Winner," *Errata Security.* (https://blog.erratasec.com/2017/06/how-intercept-outed-reality-winner.html#.XX6PdWlKiUk).

8. Freeman, J. B., Dale, R., and Farmer, T. A. 2011. "Hand in Motion Reveals Mind in Motion," *Frontiers in Psychology*, vol. 59, no. 2, pp. 1-6.

9. Giandomenico, N. and Groot, J. de. 2017. "Insider vs. Outsider Data Security Threats: What's the Greater Risk?" *Digital Guardian.* (https://digitalguardian.com/blog/insider-outsider-data-security-threats).

10. Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., and Cohen, J. D. 2004. "The Neural Bases of Cognitive Conflict and Control in Moral Judgment," *Neuron* (44:2), pp. 389–400.

11. Grimes M. G., Jenkins J. L., and Valacich J.S. 2013. "Exploring the effect of arousal and valence on mouse interaction," in *International Conference on Information Systems*.

12. Harding, L. 2014. "How Edward Snowden went from loyal NSA contractor to whistleblower," *The Guardian.* (https://www.theguardian.com/world/2014/feb/01/edward-snowden-intelligence-leak-nsa-contractor-extract).

13. Hibbeln M., Jenkins J. L., Schneider C., Valacich, J.S., and Weinmann M. 2014. "Investigating the Effect of Insurance Fraud on Mouse Usage in Human-Computer Interactions" in *35th International Conference on Information Systems ICIS 2014,* Association for Information Systems.

14. Hibbeln, M., Jenkins, J. L., Schneider, C., Valacich, J. S., and Weinmann, M. 2017. "How Is Your User Feeling? Inferring Emotion Through Human-Computer Interaction Devices" *MIS Quarterly*, vol. 41, no. 1.

15. Jackson, D. 2017. "The Case for Disclosing Insider Breaches," *Dark Reading.* (http://www.darkreading.com/vulnerabilities---threats/the-case-for-disclosing-insider-breaches/a/d-id/1328996).

16. Jenkins, J. L., Proudfoot, J. G., Valacich, J. S., Grimes, M. G., and Nunamaker Jr., J. F. 2019. "Sleight of Hand: Identifying Concealed Information by Monitoring Mouse-Cursor Movements*." Journal of the Association for Information Systems* vol.20 pp. 1–32.

17. McClimans, F., Fersht, P., Snowden, J., Phelps, B., and LaSalle, R. 2016. "The State of Cybersecurity and Digital Trust 2016," *Accenture.*

18. McKinstry, C., Dale, R., and Spivey, M. J. 2008 "Action Dynamics Reveal Parallel Competition in Decision Making," (in English), *Psychological Science*, vol. 19, no. 1, pp. 22-24.

19. Ponemon Institute. 2016. "Closing Security Gaps to Protect Corporate Data: A Study of US and European Organizations," *Ponemon Institute.* (https://info.varonis.com/hubfs/docs/research_reports/Varonis_Ponemon_2016_Report.pdf).

20. Protenus. 2017. "Hacking Incidents Down, While Some Insider Health Data Breaches Took 5+ years to Discover," *Protenus.* (https://www.protenus.com/blog/hacking-incidents-down-while-some-insider-health-data-breaches-took-5-years-to-discover).

21. S. E. Institute. 2017. "Insider Threat Database," *Carnegie Melon University.* (http://www.cert.org/insider-threat/research/database.cfm).

22. Scott, J. and Spaniel, D. 2017. "In 2017, The Insider Threat Epidemic Begins," *Institute of Critical Infrastructure Technology.* (http://icitech.org/wp-content/uploads/2017/02/ICIT-Brief-In-2017-The-Insider-Threat-Epidemic-Begins.pdf).

23. Upton, D. M. and Creese, S., "The Danger from Within," *Harvard Business Review*, vol. 92, no. 9, pp. 94-101, 2014.

24. Valacich J. S, Jenkins J. L, Nunamaker, J. F., Hariri S., and Howie J. 2013. "Identifying Insider Threats through Monitoring Mouse Movements in Concealed Information Tests," *Hawaiian International Conference on Computer and Systems Sciences 2013.*

25. Verizon. 2019. "Data Breach Investigation Report," *Verizon*. (https://enterprise.verizon.com/resources/reports/2019-data-breach-investigations-report.pdf)

26. Williams P. A., Jenkins, J. L., and Valacich, J. S. 2016 "Real-Time Hand Tremor Detection via

Mouse Cursor Movements for Improved Human-Computer Interactions: An Exploratory Study," *Workshop on the SIG HCI*