AFOSR Project Final Report

Project Title: On Insider Threats, Deception, and User Modeling
PI: Eugene Santos Jr., Dartmouth College
AFOSR Grant No. FA9550-07-1-0050
AFOSR PM: Dr. Robert Herklotz

Summary of Project

There exists a critical gap in current insider threat technology. To date, efforts on insider threat have not seriously taken into account the impact of deception by the insider. Needless to say, without a clear understanding of this impact and mechanisms for deception detection, technology for handling insider threat attacks (beyond simple attacks) can only be reactive in nature that will be often too slow and too late to prevent or even correct the damage done. In this project, we have identified a number of potential technology and research avenues that can provide an essential avenue for developing a dynamic and proactive response to insider threats. The two primary technologies of interest are *user modeling* and *deception detection*. First, the application of user modeling technology in a novel manner provides unique capabilities in recognizing various classes of insider threats. User modeling in the past has typically been employed to assist the user, to capitalize on knowledge about his/her previous behavior and current roles to infer goals, motives, and intentions in order to anticipate (predict) and facilitate subsequent actions. We observed that such prediction can be used not only to anticipate a future course for the purpose of facilitating pursuit of that course, but also to detect deviations from that course. In the context of insider threat, a deviation of observed behavior from that predicted by a user model is a signal, one that might indicate that the nominal user has been supplanted, or is functioning under directions from someone else. Predicting the goals and intentions of the user's actions serves as the essential baseline critical to positively identifying deviations and to achieving an accurate determination of the nature and goals of the given insider threat situation. The second technology is the detection of deception, where different levels and types of deception and their indicators are modeled. At the most direct level of indicators, both physiological/biometric and behavioral traits have been used in various ways to recognize masquerades and other forms of deceptive behavior. However, they have been unable to identify the type and goals of the deception ranging from simple data access to operational disruption to misinformation and intelligence diversion. Deception goals and courses of action can be applied in conjunction with our knowledge of the potential user activities and user model. This will permit us to anticipate the potential courses of actions of the insider threat and to deal with them in an effective and timely manner. The merger of these two technologies provides a key element to properly securing systems against insider threats. Both user modeling and deception detection as we have described can be applied at any level of abstraction to ultimately, in the long run, determine the overall intent and goals of the deception and deliver an understanding of the user's

behaviors and actions. With such an analytic capability, we believe that efficient and effective real-time responses to insider threat can ultimately be achieved.

In what follows, we briefly describe our major research contributions for this effort.

1. Insider Threat in Intelligence Analyses

Our goal is to detect malicious insiders among a group of analysts in the Intelligence Community. The biggest challenge is in determining indicators of abnormal behaviors in analyst activities. The insiders manipulate the information they present in their reports, which are subtle malicious actions to characterize. The characterization of these malicious actions clearly requires an analysis of the contents of their reports. However, we must also measure the behavioral consistencies between the information the analysts have collected against the reports they have written. We have proposed a framework for intent-driven insider threat detection. The heart of the framework is the IPC user modeling technique which captures analyst's interests, knowledge context, and preferences over time. This technique allows us to describe analysts' behavioral consistencies in a quantitative way, which is key to addressing our main challenge. We tested our method on the APEX '07 test bed which contained eight benign analysts and five simulated malicious insiders. The empirical evaluation demonstrated that our framework was effective in identifying insider threats. The results showed so far that we were capable of identifying all five malicious insiders without raising any false positives.

Papers: [Santos et al., 2012a][Santos et al., 2009a][Santos et al., 2008]

2. Deception Detection in Human Reasoning

Deception detection plays an important role in safely and reliably using multientity advisory models such as multiagent intelligence systems. Unfortunately, deception detection is extremely challenging. The average detection rate by humans alone is only above chance, and the skill for detection has been shown to be difficult to improve even with training. In psychological studies, deception detection is typically based on examining a person's nonverbal cues and expressions such as facial expressions, gestures, and movements. Our approach instead is focused on the agent's reasoning process.

We first detect deception by observing the correlations between agents, which can be used to make a reasonable prediction of the agents' reasoning processes. Our experiments demonstrate the effectiveness of this method and show the impact of different factors on detection rate. We further conduct some preliminary experiments to explore its performance at detecting both disinformation and misinformation and that of identifying more than one deceiver in the system.

Next, a novel method was developed to detect deception by identifying inconsistencies, explaining the reasoning behind the inconsistencies, and measuring the likelihood of deception based on cues in reasoning. The initial experiment demonstrated the effectiveness of the approach in identifying and explaining communications containing inconsistencies. Reasoning cues that can best discriminate deception from truth are further proposed, and aspects of the verification and measurement of such cues as possible future directions of work have been explored.

Papers: [Li and Santos, 2012][Santos and Li, 2011][Li and Santos, 2011][Santos, Li, & Yuan, 2008][Yuan, 2007]

3. Impact of Cognitive Styles

A user's cognitive style has been found to affect how they search for information, how they analyze the information, and how they make decisions in an analytical process. We have shown that we can use Hidden Markov Models (HMM) to dynamically capture a user's cognitive style by automatically exploring the sequence of actions and relevant information with respect to the content of the actions. The evaluation results show that our HMM model achieves an average of 72% recall with the APEX 07 collection. We also studied the link between a user's cognitive style and the various attributes relating to document content during an analytical process. The results show that the "analytic" group tends to focus on documents with significantly more specific information than the "wholist" group. The specific/general attribute of documents can help us in classifying a user's cognitive styles automatically. We have applied this notion of cognitive style to help better explain the variations in intelligence analysis which is critical to detecting both insider threat and deception.

Papers: [Nguyen et al., 2011][Santos et al., 2010][Nguyen et al., 2008]

4. Fusing Multiple (Potentially Conflicting) Source of Knowledge Under Uncertainty

This work addresses the challenges of information/knowledge fusion from multiple (possibly conflicting) sources. For example, consider that there are multiple experts (sources) providing knowledge-based models of the same scenario/situation and we wish to aggregate this information in order to assist in decision-making. There are several problems we may run into by naively merging the information from each source – the experts may disagree on the probability (uncertainty) of a certain event or they may disagree on the direction of causality between two events (e.g., one thinks A causes B while another thinks B causes A); the experts may even disagree on the entire structure of dependencies among a set of variables in a (probabilistic) network. The challenge here is to develop a semantically sound and computationally effective methodology that explicitly accounts for the uncertainty and conflicts. In our solution to this problem, we represent the knowledge-based models as Bayesian Knowledge Bases (BKBs) and provide an algorithm called Bayesian knowledge fusion that allows the fusion of multiple BKBs into a single BKB that retains the information from all input sources. This allows for easy aggregation and de-aggregation of information from multiple expert sources and facilitates multi-expert/source decision making by providing a framework in which all opinions can be preserved and reasoned over. The problem of fusing multiple conflicting sources occurs in many other domains from sensor/information fusion to intelligence analyses. This work establishes a mathematical foundation for hypothesizing about insider threat and deception intentions that underlies human reasoning. We have also extended the theory to account for time and uncertainty.

Papers: [Santos et al., 2012b][Santos, Gu, & Santos, 2011a][Santos, Gu, & Santos, 2011b][Santos & Jurmain, 2011][Santos, Wilkinson, & Santos, 2011][Santos, Wilkinson, & Santos, 2009][Santos, Li, & Wilkinson, 2009][Santos et al., 2009b]

Publications [5 journal articles, 1 book chapter, 12 conference papers, 1 MS Thesis]
[The publications below were supported in full or in part by this project.]

Li, Deqing and Santos, Eugene, Jr., "Deception Detection in Human Reasoning," *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, 165-172, Anchorage, AK, 2011.

Li, Deqing, and Santos, Eugene, Jr. "Argument Formation in the Reasoning Process: Toward a Generic Model of Deception Detection," *Proceedings of the EACL 2012 Workshop on Computational Approaches to Deception Detection*, 63-71, Avignon, France, 2012.

Nguyen, Hien, Santos, Eugene, Jr., Jacob, Russell, and Smith, Nathan, "Evaluation of the Impact of User-Cognitive Styles on Assessment of Text Summarization," *IEEE Transactions on Systems, Man, and Cybernetics: Part A* 41(6), 1038-1051, 2011.

Nguyen, Hien, Santos, Eugene, Jr., Jacob, Russell, and Smith, Nathan, "Evaluation of the Effects of User-Sensitivity on Text Summarization," *Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, 927-931, Sydney, Australia, 2008.

Santos, Eugene, Jr., Gu, Qi, and Santos, Eunice E., "Tuning a Bayesian Knowledge Base," *Proceedings of the 24th International FLAIRS Conference*, 638-643, Palm Beach, FL, 2011a.

Santos, Eugene, Jr., Gu, Qi, and Santos, Eunice E., "Incomplete Information and Bayesian Knowledge-Bases," *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, 2989-2995, Anchorage, AK, 2011b.

Santos, Eugene, Jr. and Jurmain, Jacob, "Bayesian Knowledge-driven Ontologies," *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, 856-863, Anchorage, AK, 2011.

Santos, Eugene, Jr. and Li, Deqing, "Deception Detection in Multi-Agent Systems," *IEEE Transactions on Systems, Man, and Cybernetics: Part A* 40(2), 224-235, 2010.

Santos, Eugene, Jr., Li, Deqing, Santos, Eunice E., and Korah, John, "Temporal Bayesian Knowledge Bases – Reasoning about uncertainty with temporal constraints," *Expert Systems with Applications* (2012b), http://dx.doi.org/10.1016/j.eswa.2012.05.002

Santos, Eugene, Jr., Li, Deqing, and Wilkinson, John T., "A Framework for Reasoning Under Uncertainty with Temporal Constraints," *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, 448-454, San Antonio, TX, 2009.

Santos, Eugene, Jr., Li, Deqing, and Yuan, Xiuqing, "On Deception Detection in Multi-Agent Systems and Deception Intent," *Proceedings of the SPIE: Defense & Security Symposium*, Vol. 6965, Orlando, FL, 2008.

Santos, Eugene, Jr., Nguyen, Hien, Wilkinson, John T., Yu, Fei, Li, Deqing, Kim, Keumjoo, Russell, Jacob, and Olson, Adam, "Capturing User Intent for Analytic Process," *Lecture Notes in Computer Science 5535: Proceedings of the User Modeling, Adaptation, and Personalization, 17th International Conference (UMAP 2009)*, 349-354, Trento, Italy, 2009a.

Santos, Eugene, Jr., Nguyen, Hien, Yu, Fei, Kim, Keum Joo, Li, Deqing, Wilkinson, John T., Olson, Adam, and Jacob, Russell, "On Intent-Driven Insider Threat Detection for Intelligence Analyses," *IEEE Transactions on Systems, Man, and Cybernetics: Part A* 42(2), 331-347, 2012a.

Santos, Eugene, Jr., Nguyen, Hien, Yu, Fei, Li, Deqing, and Wilkinson, John T., "Impacts of Analysts Cognitive Styles on the Analytic Process," *Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, 601-610, Toronto, Ontario, Canada, 2010.

Santos, Eugene, Jr., Nguyen, Hien, Yu, Fei, Kim, Keum Joo, Li, Deqing, Wilkinson, John T., Olson, Adam, and Jacob, Russell, "Intent-Driven Insider Threat Detection in Intelligence Analyses," *Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, 345-349, Sydney, Australia, 2008.

Santos, Eugene, Jr., Wilkinson, John T., and Santos, Eunice E., "Fusing Multiple Bayesian Knowledge Sources," *International Journal of Approximate Reasoning* 52(7), 935-947, 2011.

Santos, Eugene, Jr., Wilkinson, John T., and Santos, Eunice E., "Bayesian Knowledge Fusion," *Proceedings of the 22nd International FLAIRS Conference*, 559-564, Sanibel Island, FL, 2009. **(Runner-Up Conference Best Paper Award)**

Santos, Eunice E., Santos, Eugene, Jr., Wilkinson, John T., and Xia, Huadong, "On a Framework for the Prediction and Explanation of Changing Opinions," *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, 1146-1452, San Antonio, TX, 2009b.

Yuan, Xiuqing, "Deception Detection in Multi-Agent System and War-Gaming," MS Thesis, Thayer School of Engineering, Dartmouth College, 2007.

Personnel Supported

Dr. Eugene Santos Jr.
Dr. Hien Nguyen
Dr. Keumjoo Kim

Students –
Xiuqing Yuan (completed MS 07)
Deqing Li
Fei Yu
John Wilkinson
Qi Gu
Adam Olson
Jacob Russell
Richard Detsch

Interactions & Transitions

"Insider Threat and Deception Detection in Intelligence Analysis," Invited Speaker, Information Assurance Analysis (IAA) Skill Community Speaker Series, National Security Agency, Baltimore, MD, 2010.

"Intent-Driven Cybersecurity," Invited Speaker, Cyber Security Workshop, The Scientific and Technical Intelligence Committee (STIC) and MITRE Corporation, McLean, VA, 2009.

DARPA AIR Advisory Group, Boston, MA, Invited Participant, Nov 2008.

Intelligence Science Board Forum on Understanding Hearts and Minds, Washington, DC, Invited Participant, Oct 2008.

"On Threats from Intelligent Adversaries," Keynote Speaker, 1st International Symposium on Resilient Control Systems (ISRCS), Idaho National Laboratory, ID, 2008.

"Situation/Threat Assessment Research," Invited Speaker, AFRL Workshop on Situation/Threat/Impact Awareness, Washington, DC, 2008.

"On Insider Threats, Deception, and User Modeling," Invited Speaker and Participant, ARO/FSTC Workshop on Insider Attack and Cyber Security, Arlington, VA, 2007.

Awards

2012 IEEE Fellow
2009 Runner-Up for Best Paper Award at The 22nd International FLAIRS Conference, Sanibel Island, FL

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*

2. REPORT TYPE
FINAL REPORT

3. DATES COVERED *(From - To)*
01 DEC 06 – 30 NOV 11

4. TITLE AND SUBTITLE
ON INSIDER THREATS, DECEPTION, AND USER MODELING

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-07-1-0050

5c. PROGRAM ELEMENT NUMBER
61102F

6. AUTHOR(S)
Dr Eugene Santos Jr.

5d. PROJECT NUMBER
2311

5e. TASK NUMBER
FX

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Dartmouth College
Thayer School of Engineering
130 Cummings Hall
Hanover, NH 03755

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Office of Scientific Research
875 North Randolph Street
Suite 325, Room 3112
Arlington, VA 22203-1768

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-OSR-VA-TR-2012-0503

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approve for Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT- There exists a critical gap in current insider threat technology. To date, efforts on insider threat have not seriously taken into account the impact of deception by the insider. Needless to say, without a clear understanding of this impact and mechanisms for deception detection, technology for handling insider threat attacks (beyond simple attacks) can only be reactive in nature that will be often too slow and too late to prevent or even correct the damage done. In this project, we have identified a number of potential technology and research avenues that can provide an essential avenue for developing a dynamic and proactive response to insider threats. The two primary technologies of interest are user modeling and deception detection. First the application of user modeling technology in a novel manner provides unique capabilities in recognizing various classes of insider threats. User modeling in the past has typically been employed to assist the user, to capitalize on knowledge about his/her previous behavior and current roles to infer goals, motives, and intentions in order to anticipate not only to detect deviations from that course.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(Include area code)*