Volume 10, Issue 2
August 2017
ISSN: 1946-1836

JOURNAL OF INFORMATION SYSTEMS APPLIED RESEARCH

In this issue:

The **Journal of Information Systems Applied Research** (JISAR) is a double-blind peer-reviewed academic journal published by **ISCAP,** Information Systems and Computing Academic Professionals. Publishing frequency is currently semi-annually. The first date of publication was December 1, 2008.

JISAR is published online (http://jisar.org) in connection with CONISAR, the Conference on Information Systems Applied Research, which is also double-blind peer reviewed. Our sister publication, the Proceedings of CONISAR, features all papers, panels, workshops, and presentations from the conference. (http://conisar.org)

The journal acceptance review process involves a minimum of three double-blind peer reviews, where both the reviewer is not aware of the identities of the authors and the authors are not aware of the identities of the reviewers. The initial reviews happen before the conference. At that point papers are divided into award papers (top 15%), other journal papers (top 30%), unsettled papers, and non-journal papers. The unsettled papers are subjected to a second round of blind peer review to establish whether they will be accepted to the journal or not. Those papers that are deemed of sufficient quality are accepted for publication in the JISAR journal. Currently the target acceptance rate for the journal is about 40%.

Questions should be addressed to the editor at editor@jisar.org or the publisher at publisher@jisar.org. Special thanks to members of AITP-EDSIG who perform the editorial and review processes for JISAR.

JOURNAL OF INFORMATION SYSTEMS APPLIED RESEARCH

Editors

2017 JISAR Editorial Board

The Effects of Perceived Functionality and Usability on Privacy and Security Concerns about Cloud Application Adoptions

Makoto Nakayama
mnakayam@depaul.edu
College of Computing and Digital Media
DePaul University
Chicago, IL 60604, USA

Charlie Chen
chench@appstate.edu

Christopher Taylor
taylorcw@appstate.edu

Computer Information Systems and Supply Chain Management
Appalachian State University
Boone, NC 28608, USA

Abstract

Privacy and security risk are two primary concerns for end-users to consider when adopting cloud applications. This study investigates two potential antecedents for these two concerns: functionality expectation and usability. In addition, this study tries to understand whether their relationships exist and are correlated positively or negatively. An online survey was sent to 211 college users asking about their experiences using Google Docs. Statistical tests were conducted and showed that functionality expectation and usability improve as the length of use increases. Improved usability perception has negative effect on privacy and security concerns, indicating that privacy and security concerns could be reduced over time. On the other hand, increased functionality expectation raises more privacy concerns but does not affect security concern. Academic and practical implications are drawn from the findings to conclude this study.

Keywords: Cloud Computing, Privacy, Security, Risk, Google Docs

1. INTRODUCTION

Cloud applications were initially not considered reliable and practical, as users had doubt and skepticism. A recent survey shows that 93% of its respondents are adopting cloud applications (Weins, 2015). The rapid adoption of cloud applications could be caused by the applications' improved features or users' improved perception. What changes the users' perceptions of cloud applications depends on many factors. However, it is worth asking how end-user perceptions change over time on the functionality and usability of cloud-based applications.

The end-user perception changes about different non-standard cloud applications would be difficult to examine, given that the details of each cloud application vary. However, it is more feasible to

assess a standardized, common cloud application than a non-standard, customized one. In this study, we focus on Google Docs as one example of end-user oriented popular cloud applications. Google Docs is "a cloud productivity suite and it is designed to make computer-mediated collaboration easy and natural so that users can access any document they own or that has been shared with them anywhere, any time and on any device" (Sun, Lambert, Uchida, & Remy, 2014, p. 234). Google Docs is easy to use for a wide range of students in different educational settings. A study (Moonen, 2015) reports its successful incorporation even into an elementary school curriculum. At the university level, professors would consider integrating Google Applications into their instructional strategies, provided the appropriate professional development and training (Cahill, 2014). These professors agreed that collaborative technology was an effective teaching tool and assisted students when working on group and individual projects (*ibid*.). However, Google Docs is not limited to educational uses. In fact, it is suited to facilitate collaborations between workers using word processor, spreadsheet, and presentation applications. A recent survey (BetterCloud, 2016) notes that more than 40% of cost savings are seen at small to large firms due to adoption of Google applications, including Google Docs. Given the interest and possible business impact, our main research question is twofold: How do functionality expectation and usability of cloud computing affect privacy and risk concerns of users?

The plan of the paper is as follows: We hypothesize that functionality expectation and usability perception differently affect privacy and security concerns of these cloud applications. After describing method and results, we discuss the implications and future research agenda.

2. THEORETICAL BACKGROUND AND HYPOTHESES

Google Docs is "a free Web-based office suite that allows users to collaborate and facilitate conversations as they create and edit live documents" (Woodard & Babcock, 2014, p. 2). Users of Google Docs may have concerns about intentional or unintentional disclosure of personal information, as well as the inconveniences or costs due to the temporary or permanent unavailability of documents. This means that users have concerns over privacy and risk.

Merriam-Webster defines privacy as "the state of being alone" or "the state of away from public attention." However, the meaning of privacy is contextual and varies among different academic disciplines (Paul A Pavlou, 2011; Smith, Dinev, & Xu, 2011). Privacy is categorized as value-based or cognate-based (Smith et al., 2011), with the former viewing privacy as a right or commodity and the latter as the state of limited information access. Since the study focuses on the perception of individual cloud-application users, we frame privacy concerns as those about "opportunistic behavior related to the personal information submitted" (Dinev & Hart, 2006, p. 64) through Google Docs.

Cloud computing has the flexibility of changing functionality and can do so at a potentially lower cost than dedicated infrastructure (Ali, Soar, &Yong, 2016). Thus, users have a higher functionality expectation for cloud computing. As the degree of functionality expectation for a cloud application becomes greater, the users essentially expect more interactions with the application. A study shows that cloud services with a transparent and adaptable interface can encourage users to spend efforts and time in provisioning privacy requirements before uploading their sensitive data into the services (Henze et al., 2016). Using a cloud application, the user may perceive a 1 in 100 chance of having a privacy violation. If the user keeps using the application in the same way more frequently, the same user would feel a higher chance of experiencing a privacy violation. The more the application delivers its functionality to the user through increased interactions, the higher the perceived chances of privacy violations. We therefore hypothesize:

H1a: The degree of functionality expectation is positively associated with the extent of privacy concerns.

Oxford Dictionary defines risk as "a situation involving exposure to danger." In our study, risk is contextual and depends on subjective perceptions similar to privacy. However, the key difference between privacy and risk relates to the fact that privacy is a perceived state of isolation, whereas risk hinges on the probability of outcomes. Adapting from Gefen and Pavlou (2012), we define security risk as the belief in a potential of suffering a loss while interacting Google Docs fellow users.

Based on this definition of security risk, we can make a parallel argument on the relation between increasing functionality expectation and security

risk, as with the hypothesized relation between increasing functionality expectation and privacy concerns (H1a). The more the user uses a cloud application, the higher the chance of some risk compromise everything else being equal. We therefore hypothesize:

H1b: The degree of functionality expectation is positively associated with the extent of risk concerns.

Advances in information technology bring tremendous benefits to society and yet they could also threaten information privacy and create security risk concerns. This digital dilemma has forced customers to think analytically about how much personal information to disclose in face of growing usability features. According to privacy calculus theory, consumers feel comfortable releasing personal information only when they feel that the benefits of doing so can outweigh potential threats (Milne, Rohm, & Bahl, 2004).

As technology acceptance grows, users realize how much they could be susceptible to privacy and security threats. For instance, as users contribute and share more personal information to Web 2.0 sites (Facebook), they are more likely to have rich user experiences (e.g. expanded personal network, relevant commercials & latest information about friends). However, the success of these rich online socializing experiences depends on the sharing of personal information (e.g. what one did with whom, what opinion one has on a sensitive subject, how one's health exam resulted). Fortunately, a growing number of usable features are easing the process of using Web 2.0 sites. Testing the password strength is now a prevalent feature to assist users in creating a new account. The single sign-on (SSO) feature enables users to access other unfamiliar Web 2.0 sites via their Facebook or Google accounts and passwords. All the contact information on Facebook and Google could be automatically released to other applications (e.g. instant messaging services). Phishing-detection applications with the built-in feature of blacklist-based and whitelist-based anti-phishing toolbars can increase perceived usability and reduce privacy and security concerns for users (Li et al., 2014). Scheduling a personal and business event can be synchronized across Google platform. All these features are integrated on a limited number of platforms with a more sophisticated SSO password. Such evidence shows that the increase of perceived usability is negating privacy and security risk concerns of users.

The perception of usability is based on how the user interacts with the application as opposed to what functions are used or how much the application is used (McNamara & Kirakowski, 2006). In online banking, better website usability leads to higher trust in the website (Casalo, Flavián, & Guinalíu, 2007). Higher trust can ease risk concerns (Kim et al., 2008). A study (Hart, Ridley, Taher, Sas, & Dix, 2008) on Facebook use notes the relation between better usability and more Facebook use, while privacy concerns can discourage more Facebook use. A study compares single-factor and two-factor authentication methods in automated telephone banking and finds that users have a higher degree of perceived security with the two-factor method (Gunson et al., 2011). However, the advanced security feature is harder to use and takes longer time for users to complete. Because of its lower perceived usability, users expressed in the study that they are less likely to use the system. This finding indicates that better usability has direct impact of the intention of system use. In addition, better usability has direct impact on satisfaction and trust (Flavián, Guinalíu, & Gurrea, 2006). Based on the popularity of e-commerce and Facebook, we can surmise that the impact of better usability has overall eased the privacy concerns. Thus, the last set of hypotheses are:

H2a: The degree of perceived usability is negatively associated with the extent of privacy concerns.
H2b: The degree of perceived usability is negatively associated with the extent of risk concerns.

Thus, our theoretical model is shown as Figure 1 below.

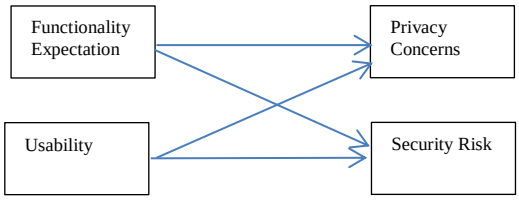

Figure 1. Theoretical model

3. METHOD AND RESULTS

Participants and Procedures

A total of 224 college students in the College of Business of a state university in the southeast region of the United States participated in the study. These students were taking an introductory management information systems course. Participation was voluntary. However, students could earn an extra credit (0.5% of their final grade) if they choose to participate. A final sample of 202 valid questionnaires was used in the present study.

Survey Instrument

We measured the functionality expectation of Google Docs users with a combination of two constructs, collaboration support (Park & Ryoo, 2013) and adoption intention (Gefen, Karahanna, & Straub, 2003). We assessed usability by testing usefulness (Burda & Teuteberg, 2015) and ease of use (Burda & Teuteberg, 2015) for cloud applications. The user's perceived privacy while using Google Docs was measured using three items adapted from Vannoy et al. (2013). To measure the perceived risk construct, we modified the original questions from Pavlou and Gefen's study (2004) into 3 items.

The partial least squares (PLS) (Fornell & Bookstein, 1982) analysis was conducted with the SmartPLS software, because it enables a small sample size. An additional benefit of conducting PLS is that it is nonparametric. Therefore, assumptions such as normality and independence are unnecessary (Chin & Newsted, 1999).

	Function	Privacy	Risk	Usability
Function	0.742			
Privacy	0.397	0.701		
Risk	-0.231	0.022	0.926	
Usability	0.594	0.144	-0.316	0.770

Table 2. Convergent and discriminant validity test results

After removing items with loadings less than 0.7, we conducted the Cronbach's alpha test. In addition, we conducted convergent and discriminant validity tests based on the average variance extracted (AVE) value for each construct reported (Yoo & Alavi, 2001). This test result indicates that all questions used to measure constructs in the model have high discriminant and convergent validities. Table 2 in the Appendix shows that the square root of these AVEs on the diagonal are larger than the correlations with other constructs. This test result indicates that all questions used to measure constructs in the model have high discriminant and convergent validities.

After confirming acceptance of the survey instrument's reliability and validity, we entered the data into the path analysis to test our hypothesized relationships. Table 3 shows the path analysis results, including path coefficients and their respective t-statistics. H1a was supported, given that functionality expectation increases privacy concern (β=-0.476; $p<0.1$). However, h1b was not supported since there was no effect of increased functionality expectation on security risk perception (β=-0.068; not significant). We consider possible reasons in the next section. H2a was weakly supported (β=-0.150; $p<0.10$), indicating that usability has a negative influence on privacy concern in cloud computing applications. H2b was supported, indicating that usability has a negative impact on security risk (β=-0.256; $p<0.05$).

Hypothesized Relationships	Path Coefficients (Beta)	T-Statistics
H1a: Functionality expectation → Privacy Concerns	0.476	6.208***
H1b: Functionality expectation → Risk Concerns	-0.068	1.166
H2a: Usability → Privacy Concerns	-0.144	1.778*
H2b: Usability → Risk Concerns	-0.256	2.929***

Table 3. Path analysis results

4. IMPLICATIONS

One major implication is that improved perceptions of functionality expectation and usability may change privacy and risk concerns. Security concerns will ease as the usability perceptions of standardized cloud applications improve through more frequent use of these applications. Contrary to H1b, the perceptual changes on functionality expectation do not have significant impact on security perceptions. This may be explained partly by the diminishing effect of consumer risk perception, and partly by the habituation effect (Amer & Maris, 2007) between Google Docs and its users. First, in consumer purchase decisions, risk perception generally continues to move from the beginning of product

purchase intention to post-purchase product evaluation (Mitchell & Boustani, 1994). This is because consumers use risk reduction strategy in their purchase process to minimize two types of uncertainties: knowledge uncertainty and choice uncertainty (*ibid*.). Cloud application users go through a similar process of initial application evaluation to post-adoption evaluation, just as consumers go through pre-purchase research to post-purchase evaluation. A survey in a past study shows that user experience affects trust (Beldad, de Jong, & Steehouder, 2010). Trust in turn lowers the degree of risk perception (Kim, Ferrin, & Rao, 2008). That is, as Google Docs users continue to use the application, they develop more trust on Google Docs and, in turn, have lower risk perception. These are driven by the user learning through continuous interaction with the cloud application over time.

Second, more use may increase security risks, but the habituation effect may ease security concerns at the same time. However, the model of this study posts that the usability improvement is likely to ease both privacy and risk concerns. A growing number of regulators and system developers are collaborating to develop systems by using the concept of "privacy by design" or "build in" privacy (Rubinstein and Good, 2013). This emerging concept further affirms the importance and impact of increased perceived usability on reducing security and privacy concerns.

For the developers of cloud applications, these results highlight the importance of continuous usability improvements that not only give the end-users better application experience but also accelerate the adoption of cloud applications by pacifying the concerns on privacy violations and risks. The developers should also be aware that the end-users are likely to better appreciate the functions of standardized cloud applications.

For researchers, the results of this study provide research opportunities to investigate our hypothesized relationships over time. Scholars of human computer interactions should further study how much influence habituations have on functionality expectation, usability of standardized, and non-standardized cloud applications.

One limitation is that the study is rooted in the use of Google Docs in the higher educational settings. However, the participants of the study were mostly adults. Future studies could use participants with broader profiles. Another limitation is rooted in the nature of Google Docs. It is a productivity suite as well as a collaboration tool (Sun et al., 2014). Future studies need to focus on other types of business and consumer applications.

5. CONCLUSION

This study examines the potential effect of functionality and usability on security and privacy concerns while using Cloud applications. Based on the survey of 211 users of Google Doc., this study finds that improved usability perception eases both privacy and security concerns. In contrast, increased functionality expectation raises more privacy concerns but does not affect security concern. These findings provide implications about promoting standardized cloud applications, such as Google Docs.

6. REFERENCES

Ali, O., Soar, J., & Yong, J. (2016). An investigation of the challenges and issues influencing the adoption of cloud computing in Australian regional municipal governments. *Journal Of Information Security And Applications*, 27-28(Special Issues on Security and Privacy in Cloud Computing), 19-34.

Amer, T., & Maris, J.-M. B. (2007). Signal words and signal icons in application control and information technology exception messages-hazard matching and habituation effects. *Journal of Information Systems, 21*(2), 1-25.

Beldad, A., de Jong, M., & Steehouder, M. (2010). How shall I trust the faceless and the intangible? A literature review on the antecedents of online trust. *Computers in Human Behavior, 26*(5), 857-869. doi:http://dx.doi.org/10.1016/j.chb.2010.03.013

BetterCloud. (2016). *The 2016 State of Cloud IT Report*. Retrieved from New York, NY: http://blog.bettercloud.com/google-apps-vs-office-365/

Brender, N., & Markov, I. (2013). Risk perception and risk management in cloud computing: Results from a case study of Swiss companies. *International Journal of Information Management, 33*(5), 726-733.

Burda, D., & Teuteberg, F. (2015). Understanding Service Quality and System Quality Success Factors in Cloud Archiving

From an End-User Perspective. *Information Systems Management, 32*(4), 266-284.

Cahill, J. L. (2014). University Professors' Perceptions About the Impact of Integrating Google Applications on Students' Communication and Collaboration Skills. *Journal of Research Initiatives, 1*(2), Article 7.

Casalo, L. V., Flavián, C., & Guinalíu, M. (2007). The role of security, privacy, usability and reputation in the development of online banking. *Online Information Review, 31*(5), 583-603.

Chin, W.W., & Newsted, P.R. (1991). Structural Equation Modeling Analysis with Small Samples Using Partial Least Squares. *Statistical strategies for small sample research, 1*(1), 307-341.

Dinev, T., & Hart, P. (2006). An extended privacy calculus model for e-commerce transactions. *Information Systems Research, 17*(1), 61-80.

Flavián, C., Guinalíu, M., & Gurrea, R. (2006). The role played by perceived usability, satisfaction and consumer trust on website loyalty. *Information & management, 43*(1), 1-14.

Fornell, C., & Bookstein, F.L. (1982). Two Structural Equation Models: LISREL and PLS Applied To Consumer Exit-Voice Theory. *Journal of Marketing Research*, *19*(74), 440-452.

Gefen, D., Karahanna, E., & Straub, D. W. (2003). Trust and TAM in online shopping: An integrated model. *MIS Quarterly, 27*(1), 51-90.

Gefen, D., & Pavlou, P. A. (2012). The Boundaries of Trust and Risk: The Quadratic Moderating Role of Institutional Structures. *Information Systems Research, 23*(3-part-2), 940-959. doi:doi:10.1287/isre.1110.0395

Gunson, N., Marshall, D., Morton, H., & Jack, M. (2011). User perceptions of security and usability of single-factor and two-factor authentication in automated telephone banking. *Computers and Security*, *30*(4), 208-220.

Hart, J., Ridley, C., Taher, F., Sas, C., & Dix, A. (2008). *Exploring the facebook experience: a new approach to usability.* Paper presented at the 5th Nordic conference on Human-computer interaction: building bridges.

Henze, M., Hermerschmidt, L., Kerpen, D., Häußling, R., Rumpe, B., & Wehrle, K. (2016). A comprehensive approach to privacy in the cloud-based Internet of Things. *Future Generation Computer Systems*, 56701-718.

Iyer, B., & Henderson, J. C. (2010). Preparing for the future: Understanding the seven capabilities cloud computing. *MIS Quarterly Executive, 9*(2), 117-131.

Kim, D. J., Ferrin, D. L., & Rao, H. R. (2008). A trust-based consumer decision-making model in electronic commerce: The role of trust, perceived risk, and their antecedents. *Decision Support Systems, 44*(2), 544-564.

Koay, K. L., Syrdal, D. S., Walters, M. L., & Dautenhahn, K. (2007). *Living with robots: Investigating the habituation effect in participants' preferences during a longitudinal human-robot interaction study.* Paper presented at the 16th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2007) Jeju Island, Korea.

Li, L., Berki, E., Helenius, M., & Ovaska, S. (2014). Towards a contingency approach with whitelist- and blacklist-based anti-phishing applications: What do usability tests indicates? *Behaviour & Information Technology*, *33*(11), 1136-1147.

Matsubara, N., Matsumoto, S., & Nakamura, M. (2011). *Characterizing user habituation in interactive voice interface: experience study on home network system.* Paper presented at the 13th International Conference on Information Integration and Web-based Applications and Services.

McNamara, N., & Kirakowski, J. (2006). Functionality, usability, and user experience: three areas of concern. *ACM Interactions, 13*(6), 26-28. doi:10.1145/1167948.1167972

Mitchell, V. W., & Boustani, P. (1994). A Preliminary Investigation into Pre- and Post-Purchase Risk Perception and Reduction. *European Journal of Marketing, 28*(1), 56-71. doi:10.1108/03090569410049181

Moonen, L. (2015). 'Come on guys, what are we really trying to say here?': Using Google Docs to develop Year 9 pupils' essay-writing skills. *Teaching History, 161*, 8-14.

Nansen, B., Vetere, F., Robertson, T., Downs, J., Brereton, M., & Durick, J. (2014). Reciprocal habituation: a study of older people and the Kinect. *ACM Transactions on Computer-Human Interaction (TOCHI), 21*(3), Article 18.

Park, S. C., & Ryoo, S. Y. (2013). An empirical investigation of end-users' switching toward cloud computing: A two factor theory perspective. *Computers in Human Behavior, 29*(1), 160-170.

Pavlou, P. A. (2011). State of the information privacy literature: where are we now and where should we go? *MIS Quarterly, 35*(4), 977-988.

Pavlou, P. A., & Gefen, D. (2004). Building Effective Online Marketplaces with Institution-Based Trust. *Information Systems Research, 15*(1), 37-59.

Petter, S., Straub, D., & Rai, A. (2007). Specifying formative constructs in information systems research. *MIS Quarterly, 31*(4), 623-656.

Rankin, C. H., Abrams, T., Barry, R. J., Bhatnagar, S., Clayton, D. F., Colombo, J., . . . Thompson, R. F. (2009). Habituation revisited: An updated and revised description of the behavioral characteristics of habituation. *Neurobiology of Learning and Memory, 92*(2), 135-138.

Rubinstein, I.S., & Good, N. (2013). Privacy by design: A counterfactual analysis of Google and Facebook privacy incidents. *Berkeley Technology Law Journal*, *28*(2), 1333-1413.

Sanchez, R., & Sudharshan, D. (1993). Real-time market research. *Marketing Intelligence & Planning, 11*(7), 29-38.

Smith, H. J., Dinev, T., & Xu, H. (2011). Information privacy research: an interdisciplinary review. *MIS Quarterly, 35*(4), 989-1016.

Sun, Y., Lambert, D., Uchida, M., & Remy, N. (2014). *Collaboration in the cloud at Google.* Paper presented at the 2014 ACM conference on Web science, Bloomington, Indiana.

Urbach, N., Smolnik, S., & Riempp, G. (2010). An empirical investigation of employee portal success. *Journal of Strategic Information Systems, 19*(3), 184-206.

Vannoy, S. A., Chen, C. C., & Medlin, B. D. (2013). Investigating the impact of differences in kind upon resource consumption in web-based social networks. *Social Network Analysis and Mining, 3*(3), 437-456.

Weins, K. (2015). Cloud Computing Trends: 2015 State of the Cloud Survey. Retrieved May 29, 2016, from http://www.rightscale.com/blog/cloud-industry-insights/cloud-computing-trends-2015-state-cloud-survey

Woodard, R., & Babcock, A. (2014). Designing Writing Tasks in Google Docs that Encourage Conversation: An Inquiry into Feedback and Revision. In R. S. Anderson & C. Mims (Eds.), *Handbook of Research on Digital Tools for Writing Instruction in K-12 Settings* (pp. 1-29). Hershey, PA: Information Science Reference.

Yoo, Y., & Alavi, M. (2001). Media and group cohesion: Relative influences on social presence, task participation, and group consensus. *MIS Quarterly, 25*(3), 371-390.

Editor's Note:

This paper was selected for inclusion in the journal as a CONISAR 2016 Distinguished Paper. The acceptance rate is typically 7% for this category of paper based on blind reviews from six or more peers including three or more former best papers authors who did not submit a paper in 2016.

APPENDICES

Variable	Construct	Reference
Length of Use	How long have you used Google Docs? [year]	
Functionality expectation α = 0.859	The extent of collaborative interaction among users is increased by using Google Docs. The extent of sharing information among team members is increased by using Google Docs. The openness to share data among team members is increased by using Google Docs. Overall, the extent of collaboration is increased by using Google Docs.	collaboration support (Park & Ryoo, 2013)
	I would use Google Docs to archive my class assignments. I am very likely to archive my class assignments using Google Docs. I intend to use Google Docs for archiving class assignments in the future.	adoption intention (D. Gefen et al., 2003)
Usability α = 0.863	Google Docs enables me to archive and retrieve my class assignments faster. Google Docs enhances my effectiveness in archiving and retrieving my class assignments. I find Google Docs useful for archiving my class assignments overall.	usefulness (Burda & Teuteberg, 2015)
	Google Docs is easy to use. It is easy to get Google Docs to do what I want it to do. Learning to operate Google Docs is easy.	ease of use (Burda & Teuteberg, 2015)
Privacy Concern α = 0.751	I need to think twice before providing personal information to Google Docs. It is my concern if Google Docs collects too much of my personal information. Google Docs should not disclose any personal information, unless they are explicitly given the right to do so. Google Docs should not use personal information for any reasons other than the only purpose of information sharing. Google Docs should never sell personal information from its database to any other organizations.	privacy (Vannoy et al., 2013)
Security Risk α = 0.917	There is a high potential for loss involved in using Google Docs for archiving class assignments. There is a considerable risk involved in using Google Docs for archiving class assignments. A decision to use Google Docs for archiving class assignments is risky.	risk (Paul A. Pavlou & Gefen, 2004)

Finding the "Radicalness" in Radical Innovation Adoption

Aditya Sharma
asharma@nccu.edu
School of Business
North Carolina Central University
Durham, NC 27707, USA

Dominic Thomas
dthomas@suffolk.edu
Sawyer Business School
Suffolk University
Boston, MA 02108, USA

Benn Konsynski
benn.konsynski@emory.edu
Goizueta Business School
Emory University
Atlanta, GA 30322, USA

Abstract

Prior conceptualizations of radicalness have been useful but are incomplete and have often assumed that term "radicalness of an innovation" is clearly understood and means the same for all researchers and managers. This however is far from truth. Different people characterize the same innovation as radical for very different underlying reasons and in some cases even as incremental. This lack of definitional clarity belies understanding the inherent attributes of radicalness for effectively understanding radical technologies and innovations. Researchers often face ambiguity in understanding and explaining the effects of radicalness on adoption and implementation decisions and outcome due to this lack of clarity, even though they may agree that something is special about "radicalness." This study addresses a conceptual gap and synthesizes existing research to define the perception of an innovation as radical by its adopters. By identifying the attributes that make an emerging technology innovation radical from the adopter's perspective, this study contributes a grounded construct for adoption research and attempts to clarify the current ambiguity concerning the application of the term "radicalness" regarding technology and innovation adoption. Using the context of Radio Frequency Identification (RFID) adoption by organizations, data from field interviews indicate technology radicalness in adoption is better understood and measured as a perceived and formative construct with five critical sub dimensions 1) embedded knowledge in the technology or product knowledge; 2) knowledge and prior experience in the application of technology or application knowledge; 3) changes in fundamental concepts of the activities to which it is applied or extent of concept change; 4) changes in the resources needed for the activities to which it is applied or extent of component change and 5) changes in the processes of the activities to which it is applied or extent of linkage change, each of which contribute to the degree of perceived radicalness of a technology.

Keywords: radical innovation, adoption, perceived radicalness, RFID, disruptive technology

1. INTRODUCTION

Multiple labels such as disruptive, breakthrough, revolutionary, discontinuous and radical have been used in prior literature, to represent innovations that may provide significantly new offerings and are perceived as providing significantly large benefits and rewards that alter the competitive position of the innovating firms (O' Connor & McDermott, 2004). Besides potential rewards these innovations are also associated with high degree of risk and uncertainty in their potential outcomes. These labels have been used interchangeably in many cases, but may mean very different things. Most labels such as breakthrough or disruptive are based on the perceived outcomes of the innovation and hence give rise to circular arguments which are true by definition (Sood & Tellis, 2005). For example, disruptive innovations have been characterized as those innovations which fundamentally alter the competitive landscape of a firm or disrupt the existing positions of the key market players. As Henderson and Clark (1990) rightly point out "the distinction between radical and incremental innovations has produced important insights but is fundamentally incomplete." There is ambiguity in their definitions and their operationalization are more categorical rather than on a continuous scale.

We address this literature gap and argue that the radicalness of a technology innovation is inherently related to technology adoption and will be understood more completely when we conceptualize it as a multi-dimensional formative construct including user perceptions and their application context along with the inherent technology attributes. The conceptualization of radicalness in technology adoption we present herein extends work by Sood and Tellis (2005); Chandy and Tellis (2000); Henderson and Clark (1990) on innovation attributes by incorporating technology-organization-context focused dimensions which, we argue, will enable radicalness to better explain when and why a technology will experience adoption resistance or success. We begin by discussing the role of technology radicalness in new technology adoption and making a case for its relevance in adoption studies. We follow it with a discussion on prior conceptualizations of technology radicalness in the innovation literature. We define perceived radicalness of a technology as a second order formative construct and present its five critical dimensions based on our data collected through semi-structured interviews. We conclude with a discussion of implications for research and practice.

2. TECHNOLOGY RADICALNESS AND ADOPTION LITERATURE

All technologies are not created equal and hence should not be treated the same. Differences in their adoption patterns exist based on their attributes and their perceived impact. This issue needs to be addressed by Information Systems (IS) researchers (Lyytinen & Rose, 2003).

Hage (1980) identified radicalness as one of the "most critical dimensions" along which an innovation may differ, however it remains to be thoroughly explored in innovation adoption literature and even more so in the interorganizational system adoption context. Radical technologies are very different from incremental technologies. Radical technologies are less frequently adopted than incremental innovations (Damanpour, 1996) and pose a greater challenge to the existing structure of political influence, causing more resistance during their implementation (Frost & Egri, 1991). Radical technologies are also more likely to fail than incremental technologies (Pennings, 1988). Radical technologies appear more complex to adopters and generate uncertainty about the resources required to use them effectively and hence have lower adoption likelihood (Gopalakrishnan & Damanpour, 1994). The perceptions of radicalness of a technology may hence influence its adoption by individuals and organizations, and therefore needs to be investigated (Ciganek & Zahedi, 2004).

Prior Conceptualizations of Radicalness

Radical innovations are likely to be competence destroying often making existing skills and knowledge redundant (Tushman & Anderson, 1986). Radical innovations often require different management practices (O' Connor, 1998). Dewar and Dutton (1986) recognize radical innovations with high degree of new knowledge embedded in them. According to them, the labels radical and incremental represent differences in degrees of novel technological process content embodied in the innovation. Also these innovations have been suggested as usually originating from scientists and are market push innovations where new features of the technologies and possibility of grasping new opportunities trigger the interest in their adoption (O' Connor, 1998) compared to incremental innovations which are more pull innovations triggered by market need either from customers' demand or a perceived need to stay competitive. Radical innovations are also likely to

open opportunities for follow-on incremental innovations (Ettlie & Rubenstein, 1987). Additionally, radical innovations whether they are new-to-the-world or new–to-the-firm, represent risky departures from existing business practices (Hage, 1980). Another characterization of radical innovations is based on the changes in behavior resulting from using the innovation (Schiffman & Kanuk, 1997) or having a customer orientation of providing greater value or benefits over existing products or technologies (Chandy & Tellis, 1998). These conceptualizations while useful do not adequately address the question: what makes a technology/innovation radical?

Similarly prior literature characterizes innovations dichotomously (i.e. product-process, administrative-technological and/or incremental-radical (Hage, 1980)), but little operationalization of these characterizations on continuous scales or testing them for mediating/moderating effects has been done at individual, organizational or inter-organizational levels.

3. PERCIVED RADICALNESS BY ADOPTERS

Radicalness of a technology has been studied mainly from its development and creation standpoint in the new product development and marketing literature. These characterizations are from the developer's perspective and suggest that radicalness of the technology is an objective characteristic, inherent to the technology. However, we argue that in the adoption of an innovation what matters is the radicalness of the innovation as perceived by its adopter. The adopter could be an individual, a group, a business unit or an organization. In each case, it is the perception of radicalness of the innovation by those that make decisions related to its adoption. In case of an organization it could be the CIO's/managers that make decisions related to whether a new technology is suitable for their organization. We propose that radicalness of an innovation would be better understood by viewing it as a combination of technology-organization-context focused dimensions which not only includes inherent attributes of the technology but the relative newness of the technology based on prior experience of the adopter and the application context within which the innovation is adopted.

Perception based on relative newness

Radicalness also has been suggested more as a perceived or subjective construct rather than an objective measure of an innovation. The perceptions of radicalness would vary based on the "newness-to-the-organization" or experiences and familiarity of the managers in adopting organizations with the innovation (Dewar & Dutton, 1986). The greater the prior experience with the innovation the more likely that knowledge embedded in the innovation would not be perceived as new and hence lower the perceived radicalness of the innovation. The degree of perceived radicalness would be related to prior experiences and existing skills and competencies in an organization that are relevant for the adoption of the innovation in question. For example, an innovation such as the Google search engine may have be considered as radical in the late 1990s for those who transitioned from library style sequential search using catalogs by one field to multiple field simultaneous search using an electronic search engine; however, the same innovation may have been considered less radical or more incremental for those who moved from a search engine such as AltaVista to an enhanced product such as Google. This difference in perceptions of radicalness of the same innovation comes from the fact that in the first case the new innovation may have required significant new conceptual knowledge in terms of how to use key words for search engines and the change that it mandates in established routines of library search. In the second case, the leap may be only slight in the perceived outcome of the result with minimal or limited new knowledge and changes in established routines. Hence, perceptions of radicalness of the same innovation may vary across organizations depending on its newness to the organization in question.

Perceptions based on Application Context

As discussed earlier, differences in perceptions of radicalness exist between development/creation of an innovation and its adoption and use. Certain innovations may be perceived as being highly radical in terms of creation but may not be perceived radical in their application and use. For example, replacement of vacuum tubes by transistors may have been perceived as a radical shift by radio manufacturers as it overturned existing concepts and components of the technology it was replacing but may or may not have been considered a radical change by its users as the only perceptible difference for them would be improved voice quality. Similarly, a certain innovation by itself may represent a new technological paradigm, but unless it is considered in its application context at the individual or business activity level and unless it requires drastic changes or alterations in the routines or replaces existing concepts underlying the individual or business activities it is likely to be perceived as being more incremental than radical. For example, a personal computer might

have been be a paradigm shifting invention for its creators because it overturned previously existing concepts of space and processing power. However, to a computer user it would have been a paradigm shift only if it overhauled the concepts of its application context, and redefined what could now be done with this machine as compared to what was done prior to its use. Hence, a user that considered a PC as a replacement of an electronic typewriter and used it for printing documents only may not have perceived it as being highly radical. On the other hand those users that made use of its high processing ability in tasks that were complex such as running computational models were using it in a context that required overhauling of what could and could not be done to accomplish the given task (i.e. the difference in terms of the changes that it may have mandated in their existing and established routines for modeling – computerized vs. hand executed) may have perceived it as being highly radical. This difference would reflect itself in the degree of new knowledge they needed to acquire and apply to accomplish the given activity and the changes that needed to be made at the concept, component and linkage level for the activities it was used for. Hence, there is in most cases an implicit comparison with the technology that is being replaced and with the context of its prior application. Same is true for the mobile smart phones replacing the traditional land line and even the voice based mobile phones earlier.

Similarly, a search engine such as Google based on new search principles, may have been a radical innovation for its developers because it overturned existing concepts about how the engine searches and requires different logic and but may not be radical for an adopter who already had been using other search engines because all they can perceive is the output which may not be very different from other search engines. Thus, we believe that an innovation idea in its development may be perceived as being radical but it may or may not be perceived as radical in terms of its adoption and use. This study focuses on adoption and use of innovations rather than their creation (inception and technological initial development)

4. PROPOSED CONCEPTUALIZATION OF PERCEIVED RADICALNESS

Ettlie et al. (1984) define an innovation as radical if it is new and introduces significant change. Consistent with Ellie et al and Lyytinen-Rose's (2003) work we go further and extend this definition to include embedded knowledge in the technology, prior experience of the adopting individual or organization and the application context changes (in terms of concepts, components and linkage changes of the individual or business activity to which it is applied.

It is to be noted here that the term "business activity" is used as a high level description of the application context and includes the business processes that are required to accomplish that activity. For example marketing a product can be considered as a high level business activity which subsumes various processes such as research, promotions and sales. Hence activities have been suggested as subsuming the processes that are needed to accomplish them.

We define radical innovations as requiring high degree of new knowledge about the product and its application and mandating substantial change in concepts, components and linkages in the context of its application.

Based on the conceptualizations in prior literatures in IS, marketing, strategic management, innovation management and other related disciplines and findings from data gathered from semi-structured interviews we define and conceptualize perceived radicalness of a technology as a five dimensional construct which includes 1) embedded knowledge in the technology or product knowledge; 2) knowledge and prior experience in the application of technology or application knowledge; 3) changes in fundamental concepts of the activities to which it is applied or concept change; 4) changes in the resources needed for the activities to which it is applied or component change and 5) changes in the processes of the activities to which linkage change.

Following is the discussion on how each of the sub-dimensions is defined and measured.

The new knowledge to adopt an innovation could entail two types of knowledge: 1) product knowledge and 2) application knowledge.

1) Product knowledge: This dimension captures new knowledge about the description of the product and features and how it could be potentially used by the adopter.

2) Application knowledge: This knowledge refers to the knowledge about the settings and contexts in which the product could be applied to potentially benefit the adopter. Hence, new knowledge for adoption of an innovation would be a combination of product knowledge of how the

product works and what it can do and knowledge about what individual and business activities it can potentially impact.

These dimensions capture the extent of new knowledge that needs to be acquired to adopt and apply the innovation in an individual or business activity setting (Hall & Andriani, 2002). This dimension is measured along a continuum from low to high and is an important dimension in the perception of radicalness of an innovation along a continuum.

3) Extent of change in concepts: Engineering or fundamental scientific principles which determine the components that would be needed for a technology product have been defined as concepts by Henderson and Clark (1990). However, that definition of concepts was in context of product innovation creation. A product innovation when it is brought into a new setting for its adoption and use may mandate changes in concepts related to the individual or business activities where it is to be applied to derive benefits from it. These changes are more important from the adopter's perspective than the scientific principles behind the innovation. Hence, we extend that definition to an activity setting where the product is applied and define concepts as underlying principles which drive the routines and tasks of an individual or a business activity. For example, an RFID tag and reader enable the unique item-level identification, non-line of sight, real time and parallel processing of identification data. All of these scientific concepts are embedded in the technology. However, the use of RFID in business activities such as asset management would lead to a change in the concepts of how that activity is conducted and would mandate either change in components for the activity or the linkages between the components or both.

This dimension captures the extent of the change in the activity concepts in terms of whether the concept change is reinforcing existing routines or overturns them and requires unlearning of old routines and replacing them with new ones.
This change in concepts is measured as the degree of substitution of conceptual knowledge and varies from low to high on a continuum where low signifies reinforcement of existing concepts and high signifies overturning of existing concepts. Please note that there could be many concepts or principles involved in a business activity at different levels of the activity, however our focus is on the changes in fundamental principles that govern the activity.

4) Extent of change in components: Components have been defined as physical manifestation of scientific concepts embedded in the technology by Henderson and Clark (1990). This definition when extended to an individual or business activity setting in which the technology would be used, means components are resources which are mandated or required for the application of the concepts. Any improvements, replacements, additions or removals of existing resources would mean a change in components for the activities. Hence, in the context of RFID use, the readers and tags, other hardware, software, systems and sub-systems and people would be components associated with the RFID innovation required to execute a business activity. The level of change in components will be high when RFID technology is to be used to accomplish business tasks that were earlier manually performed because the innovation adoption may involve all of the above mentioned changes. Please note that changes in components may or may not involve a change in the fundamental concept but would involve a change in linkages at some level.

The extent of change in components dimension would measure the overall degree of improvement or alteration in the resources of the individual or business activity that the new innovation requires on a low to high continuous scale where low signifies similar resources with no improvements and high signifies new and improved resources with high level of improvements contributing towards higher perceptions of radicalness.

5) Extent of change in linkages: Linkages have been defined as the links (or connections) between the components that have been embedded in a technology according to Henderson and Clark (1990). We extend the definition of linkages from technology creation context to the activity context where a technology is applied and used. We define linkages as the connections or relationships between components or resources associated with the innovation for the individual or business activities. Hence, in context of RFID, it would mean how the tags, readers, other hardware, software, middle-ware, other systems and people are inter-connected to accomplish the business activity. Any change in the way components or resources are connected and interact with each other for accomplishing an individual or business activity would mean a change in linkages. When RFID is introduced, as discussed earlier it is likely to be compared to the technology it replaces in the business activity context and because it would require improvements or changes in components it would

also change the linkages between them and hence is likely to be perceived as radical. Please note that any change in components would reflect as a change in linkage at some level but any change in linkages may or may not require a change in components. Any change in linkages however would be a change in concepts at some level.

The extent of change in linkages dimension would measure the degree of restructuring in the existing linkages of the business activity that the new innovation requires on a low to high continuous scale where low signifies no or minimal change in the basic architecture of business activity and high signifies major restructuring of the business activity by changing the existing links. Hence, high levels of restructuring of linkages would contribute towards higher perceptions of radicalness.

5. DATA AND METHODOLOGY

We wished to open the radicalness "Blackbox" and explore the meaning of radicalness of an innovation from the adopter's perspective. For this purpose we utilized the context of RFID adoption by organizations to understand, why organizations perceive some innovations as more radical than others, and how radicalness may impact their decision to adopt and integrate a technology-based innovation. Prior literature showed inconsistent definitions and incoherence across fields in understanding radicalness in innovation adoption. In such a case, interpretive research focusing on exploring the unknown phenomenon best serves to initiate a valid and accurate line of inquiry (Yin, 1989), (Lee, 1991) precisely our underlying research goal. To accomplish the above-mentioned goals and to develop a better understanding of the adoption process, we conducted in-depth, semi-structured interviews using a convenience sample. The interviewees were executives and RFID program managers and supply chain managers across 10 organizations (12 interviews) involved in RFID initiatives at some level. We sampled from three perspectives in order to triangulate and, thereby, strengthen our understanding of radicalness of RFID adoptions. These perspectives were the adopter perspective (7 firms and 8 interviews in three industries: manufacturing, retailing, and logistics), the implementer perspective (1 top IT consulting firms and 2 interviews), and the vendor perspective (2 firms and 2 interviews).
The interviews were conducted over a period of three months (May-July, 2005) and were either face to face or over the phone, lasting between one and two hours. The questions for the interviews were a mix of open-ended questions and closed questions to allow both the flexibility of exploring new contexts but also to help maintain focus on some of the previously identified relevant themes. At the time of the interviews, we were not exploring radicalness as perceived or context dependent. These themes emerged from the data and were later developed conceptually, because of what we found from practice.

The interviews were recorded and later transcribed. The authors coded the interview data in an effort to extract key ideas underlying the concept of innovation radicalness for managers evaluating emerging technologies such as RFID. This coding process involved the first author identifying patterns and underlying themes that emerged from quotations in the raw text, excerpting them and bringing them to the other two authors for joint discussion and refinement over a period of 7 months and more than 20 hours of discussion.

6. RESULTS AND DISCUSSION

During the analysis phase of our study we became aware that all three perspectives were unified in seeing adoption radicalness for RFID as a continuous, context-dependent phenomenon with multiple dimensions. Prior conceptualization of radicalness as dyadic or non-perceptual does not fit these data from practice. The context dependency fits well if we expect radicalness would be perceptual for innovation adoptions.

Some of the key quotes of managers that were interviewed are presented in Table 1 (Appendix) as a representative sample that supports our multi-dimensional conceptualization of radicalness as perceived and depending upon relative newness/prior experience and application context. Table 1 also shows the major patterns and underlying themes found as a result of the coding and analysis process.

As can be seen from the interview data Organizations A, G and C made repeated mentions of "need for learning" in terms of features of the technology and of how the technology can be applied in their current processes. This related to the theme of Product and Business Application Knowledge. Organizations A, J and C mention the "need for high level of changes in business processes and infrastructure that could prove disruptive" which support the dimensions of product knowledge, business application knowledge, change in business linkages and business components.

Another important theme that emerged from the interviews and was mentioned by organization J was about paradigm shift in the way a particular business activity or process is conducted. This idea is also reflected in our proposed dimension of change in activity concepts.

This study addresses an important question i.e., why an innovation might be perceived as radical by its adopters? In doing so it also discusses what radicalness means and how perceptions of radicalness may influence adoption decisions.

The conceptualization of radicalness as a multi-dimensional construct has implications for both theory and practice. For the practitioners our conceptualization addresses the issue of "lack of definitional clarity" and enables managers to understand the inherent attributes of innovation radicalness. This will allow managers to effectively develop or respond to radical innovations. From the theoretical and academic perspective, our conceptualization opens the "black box" of radicalness by proposing a multi-dimensional construct. This will enable researchers to reconcile seemingly disparate results and aggregate their understanding of role of radicalness in innovation adoption.

7. CONCLUSION

In this paper, we defined technology radicalness as a second-order perceived construct formed of five dimensions. We presented prior literature showing that radicalness by itself is popular and exciting but confounding concept, often discussed without clear conceptualization and difficult to measure directly. By identifying the attributes that make an emerging technology innovation radical from the adopter's perspective, this study contributes a grounded construct for adoption research and attempts to clarify the current ambiguity concerning the application of the term "radicalness" regarding technology and innovation adoption.

Technology radicalness has objective characteristics inherent to the technology being adopted and the specific business processes to be changed, but these are only instantiated as radicalness in the perceptions of the individuals who must change within an organization. Thus, radicalness depends on prior experiences and competences of individuals, groups, and the adopting organization. If a technology-enabled radical innovation will be implemented in two different business units involving the same business processes, we could expect differential effects from radicalness of the technology because of its perceptual nature and how it can be applied differently across units and across time.

We presented the five dimensions of perceived radicalness that will enable future examinations of radicalness to examine it on a continuum rather than as dichotomous as in prior research. The ability to understand radicalness on a continuum contributes to current literature, better capturing the theoretical nature of radicalness while also encompassing what we know about radicalness in its five dimensions as one construct:

1) Product Knowledge to be acquired
2) Business application knowledge to be acquired
3) Extent of changes required in the activity concepts (concept change)
4) Extent of changes required in the activity components (component change)
5) Extent of change required in the activity linkages (linkage change).

Technology adoption provides a seductive and powerful means for accelerating and enabling business process change, which can lead to tremendous growth and competitive advantage (Collins, 2001). However, the radicalness of a technology –enabled innovation leads to uncertainty as to how to adopt a new technology and get the benefits from it. We believe the conceptualization of perceived radicalness construct from the adopter's perspective in this study helps understand and explain its role in the area of radical technology adoption and will forward research in this area.

8. REFERENCES

Attewell (1992). Technology diffusion and organizational learning: The case of business computing. Organization Science. 1992; 3(1).

Ciganek-Zahedi (2004). "Radical! The Influence of Perceived Radicalness on Technology Acceptance", Proceedings of the Tenth Americas Conference on Information Systems, New York, NY, August 2004.

Chandy-Tellis (1998). "Organizing for Radical Product Innovation: The Overlooked Role of Willingness to Cannibalize," Journal of Marketing Research, 35 (November)

Chandy-Tellis (2000). "The Incumbents Curse? Incumbency, Size, and Radical Product Innovation," Journal of Marketing, 64(July)

Damanpour (1991). "Organizational Innovation: A Meta-Analysis of Effects of Determinants and Moderators." Academy of Management Journal 34(3)

Damanpour (1996). Organizational Complexity and Innovation: Developing and Testing Multiple Contingency Models, Management Science, 42, 5

Dewar-Dutton (1986). The adoption of radical and incremental innovations: an empirical analysis. Management Science, 32, 11 (November 1986)

Ettlie-Bridges-O'Keefe (1984). Organization Strategy and Structural Differences for Radical versus Incremental Innovation, Management Science, 30(6)

Ettlie-Rubenstein (1987). Journal of Product Innovation Management, Volume 4, Number 2, June 1987, Blackwell Publishing.

Fichman (2004). Real Options and IT Platform Adoption: Implications for Theory and Practice, Information Systems Research, May 2004.

Frost-Egri (1991). The Political Process of Innovation, In Research in Organizational Behavior Cummings, L.L. and Staw, B.M. (eds.), JAI Press

Gatignon-Xuereb (1997). Strategic Orientation of the Firm and New Product Performance, Journal of Marketing Research, 34

Gopalakrishnan,-Damanpour (1994). Patterns of Generation and Adoption of Innovations in Organizations: Contingency Models of Innovation Attributes, J. Engineering and Technology Management, 11

Hage (1980). Theories of Organizations, John Wiley, NY.

Hall-Andriani (2002). "Managing Knowledge for Innovation", Long Range Planning, 35

Hall,-Andriani (2003). "Managing knowledge associated with innovation", Journal of Business Research, 56

Henderson-Clark(1990). Architectural Innovations, The reconfiguration of existing product technologies and failure of the firms. Administrative Science quarterly, 35.

Kauffman-Riggins-Curtin (2004/05): Working paper, University of Minnesota, Minneapolis.

Lee, A. S. (1991) Integrating positivist and interpretive approaches to organizational research, Organization Science, 2, 342-365.

Lyytinen-Rose (2003). "Disruptive Nature of IT Innovation", MISQ, 27(4)

O'Connor (1998). Market learning and radical innovation: a cross case comparison of eight radical innovation projects. Journal of Product Innovation Management, 15

O'Connor-McDermott (2004). The Human Side of Radical Innovation. Journal of Engineering Technology Management. Vol. 21

Pennings (1988). Technological Innovations in Manufacturing Organizations. International Studies of Management and Organizations, 1988, xvii, 68-89.

Schiffman-Kanuk (1997). Prentice-Hall.

Sood-Tellis (2005). Technological Evolution and Radical Innovation, Journal of Marketing(69), July.

Teo, Wei –Benbasat (2003). Predicting intention to adopt IO linkages, MIS Quarterly, 27(1)

Tushman-Anderson (1986). Technological Discontinuities and Organizational Environments, Administrative Science Quarterly (31).

Yin, Robert K. (1989). The Information Systems Research Challenge: Qualitative Research Methods, Vol. 1 (Eds, Cash, J. I., Jr. and Lawrence, P. R.) Harvard Business School, Boston, MA, pp. 1-6.

APPENDIX

Table 1. Key Quotes from Managers

#	**Key Quotes**	**Organization**	**Underlying Themes**
1	We find benefits but RFID is not on our priority list and we don't think we are ready as we *don't have the infrastructure and expertise to process huge amount of data* that would be generated by it and make sense out of it. Lack of standards and cost of tags and readers is prohibitive. Also RFID will be *a major change for our company in over hauling our business processes.*	A	Business Application Knowledge, Product Knowledge, Business Component Change, Business Linkage change
2	For RFID we could easily identify which tag would work and what device would work for our products, that didn't take very long, less than six months but now we are facing a *major issue as far as its application. How much changes you have to do to all the existing ERP systems and front end business applications required in its application, we are not clear as there may be a lot.*	G	Business Application Knowledge
3	Smaller organizations see RFID as an opportunity to make two leaps at once and hence displace some of the existing organizations. Also I believe that it is *more perceptual and determined by the business context* in which it is applied. For us, in terms of retail checkout at this point it is not a major change, as it does not fundamentally change the business process. But going into the future, when there is item level tagging, and automated checkouts. It may be a paradigm shift because it *Eliminates the basis of our business. We may have to kiss our scanning and retail business goodbye.*	J	Business Concept Change, Product and Business Application knowledge, Business Component, Business Linkage Change
4	RFID would require *altering our existing optical scanners infrastructure and processes currently* in place. A lot of *learning, major changes in infrastructure may be required*. This would be *disruptive* for the organization.	C	Product Knowledge, Business Application Knowledge, Business Component Change, Business Linkage Change

A Multi-Criteria Network Assessment Model of IT Offshoring Risks from Service Provider's Perspective

Doncho Petkov
petkovd@easternct.edu
Dept. of Business Administration,
Eastern Connecticut State University, Willimantic, CT, 06226, USA

Olga Petkova
petkovao@ccsu.edu
Dept. of Management Information Systems,
Central Connecticut State University, New Britain, CT, 06050, USA

Abstract

The paper proposes a multi-criteria framework for assessment of Information Technology (IT) offshoring risks from provider's perspective using the Analytic Network Process (ANP). The authors present an overview of current literature on IT risks in software project development, IT outsourcing and offshoring. Then the network evaluation framework of offshoring risks is outlined and justified. The model is illustrated on a real case of evaluating IT offshoring risks from the point of view of a foreign service provider. The conclusion outlines possible future research directions.

Keywords: IT offshoring risks, outsourcing, ANP, AHP, Systems Thinking.

1. INTRODUCTION

The importance of outsourcing as a topic has generated much research, focused originally on domestic outsourcing (see Dibbern et al., 2004) and for the last decade also on offshore outsourcing (see Gonzalez et al., 2013). Oshri, Kotlarsky and Wilcocks (2015:3) define "sourcing is the act through which work is contracted or delegated to an external or internal entity that could be physically located anywhere. It encompasses various insourcing (keeping the work in-house) and outsourcing arrangements such as offshore outsourcing (when the work is outsourced to a third party), captive outsourcing (when the work is performed by a subsidiary of the same organization located on another continent), nearshoring (when the work is performed in a neighboring country like Mexico) and onshoring (work is outsourced within the same country). According to Oshri et al. (2015) a conservative estimate for the global outsourcing contract value of business and Information Technology (IT) services exceeded US$700 Billion by the end of 2014 while it was only about US$10 Billion in 1989.

Davis et al. (2006:741) define offshoring as "the provision of organizational products and services from locations in other countries, whether they are actually overseas or not." Since 2005 there is a greater focus on offshore outsourcing (see Lacity et al. (2009), Peslak (2012), Persson and Schlichter (2015)) as opposed to traditional domestic outsourcing (onshoring). The most comprehensive analysis of outsourcing research and practice is presented in Dibbern et al. (2004). They have explored in depth the outsourcing decision (whether to outsource or not), the reasons for outsourcing, what business activities in IT are being outsourced, how firms outsource and the outcomes of outsourcing and their

measurement. A detailed analysis of the topics in the IT outsourcing literature between 1992 and 2013 is presented in Liang et al (2016). Similar research issues are applicable also to offshoring though there are some specific aspects to it. According to Gonzalez et al. (2013:230), "the geographical as well as cultural distance which often exists between clients and providers of these services leads to the emergence of several risks which are specific to Offshore Outsourcing, such as those derived from having to battle with various time zones, different legislations or additional security and privacy problems. For this reason, an enterprise will only decide to venture into this new business area if it has additional incentives…". Lacity et al. (2009:140) conclude that researchers have found that offshore outsourcing poses considerably more challenges than domestic outsourcing. These are associated with various risks, some of which are related to the factors listed above. A very detailed systematic literature review of the reference theories and major topics in IT offshoring research in recent years is presented in Strasser & Westner (2015).

Papers on evaluation of risks in IT offshoring have only occasionally been published. That is contrasting with the fact that the topic of IT offshoring risks is ranked as the second most often researched topic in the empirical Information Systems offshoring literature according to Gonzalez et al (2013).

Risk areas represent organizational contexts that include many related risk factors, which together possess a threat to a software development project's success (Boehm, 1991). Research on IT offshoring risks is quite diverse. Outsourcing and offshoring risks can be explained with transaction cost theory (see Ngwenyama & Bryson, 1999). Chatfield and Wanniniaka (2008) have investigated IT offshoring risks and governance capabilities. The cost of risk in offshore systems development is explored in De Hondt & Nezlek (2009). The nature of offshoring and the dangers from it are analyzed in Hirschheim (2006), Herath & Kishore (2009) and elsewhere. A framework for managing IT offshoring including risk mitigation is provided in King (2008). A detailed analysis of risks in global software engineering is provided in Venter et al. (2012). An investigation of the effects of different relational norms on the link between behavioral risks and offshore software development success is presented in Matthew & Chen (2013). Most of the research on IT offshoring risks is from the client perspective (e.g. Abdullah & Venter (2012) and very few authors are treating this problem from a provider's perspective (e.g. Aundhe & Matthew (2009). Some papers integrate both perspectives on sourcing risk (e.g. Bunker et al., 2015). Most of the papers on offshoring risks are based on empirical analysis but there are also case studies on managing risk areas in IT offshoring (e.g. Persson & Schlichter, 2015). The above list of references dealing with aspects of offshoring risks and their management is by no means comprehensive and many more sources can be found in review papers like Verner et al. (2012).

The publications on IT offshoring risks are often dealing with several risks based on expert opinions (e.g. Davis et al. (2005), King (2008)). Sometimes research on this topic results in uncategorized large lists of risks like in Sakhtivel (2007) which makes their use in real decision making by practitioners difficult. Some papers deal with offshoring risks from the point of view of the client while others are dealing with IT offshoring risks from the point of view of the service provider (see Taylor,2005). Sourcing risks have been also explored from practitioner perspectives as in Bunker et al. (2015). Other previous research has focused just on IS development risks or on operational risks only.

According to Nakatsu & Iacovou (2009:57), the risks in IT offshoring are often analyzed in papers just at the level of checklists. We may point that such an approach does not take into account the relative importance of risks and provides little opportunity for analysis of risks for the purposes of their management in the context of a specific project.

Gonzalez et al (2013) summarize findings from the literature on IT offshoring but do not investigate the nature of the risk factors and how they can be used in decision making. Their findings show that Decision Making is ranked only ninth in the list of 13 research topics on IT offshoring and that it is the subject of only 8 papers out of a total of 127 included in their analysis (see Gonzalez et al., 2013). That indicates the need for more research on that topic.

While a few published papers deal with prioritization of risks in outsourcing using statistical methods (see Gandhi et al, 2012), there are no papers dealing with a systemic evaluation of the importance of risks in offshoring through the Analytic Network Process (ANP) (see Saaty, 2005), a multi-criteria decision making (MCDM) approach. ANP was previously applied to outsourcing risks from a client perspective by Keramati et al. (2013). However their model does

not consider offshoring risks and it is developed with the unattainable goal to generalize the results which is not possible as they are heavily context dependent. Those authors incorrectly consider that a limitation of their work. We claim instead that the strength of ANP is based on its results being relevant for risk modeling in the context of a particular software project and hence it is suitable as a tool for systemic prioritization of offshoring risks. The previous paragraphs summarize the main motivations of this research.

The *goal of this paper* is to provide a systemic framework for assessment of IT offshoring risks from the provider perspective based on the Analytic Network Process. The *contribution of the paper* is in the formulation of an ANP model of the IT offshoring risks from a service provider's point which was not reported previously in the literature and in its practical demonstration for evaluation of risks in a specific project context.

Typically risk management involves three steps (Ghadge et al., 2013): risk identification, risk assessment and risk mitigation. Risk mitigation issues are outside the scope of this paper. The next section proceeds with an analysis of what can be learned from past research on software risks associated with IT outsourcing, systems development and offshoring. It is followed in the third section by an attempt to address the second step above through the formulation of a systemic framework for ANP assessment of IT offshore outsourcing risks from a provider's perspective. It is followed by a demonstration of the use of the ANP model and a conclusion.

2. ON IT OFFSHORING RISKS

Risks in Information Technology represent a multifaceted research area that is closely related to other fields like IT failure (including project development and operational failure), project success etc. IT offshoring project risks may be applicable to all types of projects and on the other hand may be specific only to specific offshore outsourcing projects depending on their context. IT offshoring risks overlap also with risks in some global or distributed software development projects. Sometimes the notion of risks is replaced by the notion of barriers for software project success but the meaning of that is very similar to risks. IT risks may play a role only in specific project contexts and hence there cannot be a universal list of risks applicable to every situation. Therefore IT offshoring risks are a very complex notion related to the more general notions of IT risks, IT outsourcing risks, IT project success, IT project failure, global or distributed software development and IT operations. IT offshoring risks are important because their understanding and evaluation can lead to better chances for their mitigation.

We will deal in this section with the identification of the types of IT offshoring risks. One possibility is to take as a leading point the broader area of IT development and operations. Another option is to treat that question starting from the point of IT Outsourcing or a third one is to follow a more narrow perspective associated with factors that relate only to offshoring. We will explore each of these separately below.

Risks derived from research in software systems development

Software engineering risk management emerged in the 1980s and its principles were summarized in Boehm (1991) and several earlier publications by the same author. A good review on general IT risks can be found in Pfleeger (2000). Further insights on the nature of IT risks are provided in Bahli & Rivard (2005) and elsewhere.

The first empirically validated list of risk factors in software development projects was generated through a Delphi survey by Schmidt et al. (2001). They were grouped in 14 categories. The risk factors were shown by rank order and that was another major difference of those results from prior findings of other authors. These authors claim to contribute to the unification of research on risk management and software project management. While the large group of experts included in their Delphi study is a positive aspect of their project, it has a possible limitation in the fact that they came only from three countries.

Wallace et al. (2004) analyzed the existing literature on software development risks and have conducted multivariate statistical modeling of the types of risks which allowed them to reduce the number of relevant factors grouping them into seven categories: organizational environment risk, user risk, requirements risk, project complexity risk, planning control risk and team risks. Their work is valuable for uncovering the relationships between various groups of risks.

The most exhaustive investigation on risk factors in global software project management is probably presented in the detailed report by Verner et al. (2012). They analyzed 24 systematic literature reviews of global software development and generated a list of risk factors in 10 groups. However, no justification is provided for the way how the groups were chosen and their results do not have the empirical validation of the findings

of Schmidt et al. (2001). Research on software development risks has influenced work on outsourcing risks.

IT risks derived from studies of IT outsourcing

An early important paper by Earl(1996) considers the following types of risks in IT outsourcing: possibility of weak management, inexperienced staff, business uncertainty, outdated technology skills, endemic uncertainty as IT project development and operations have been always uncertain, hidden costs, lack of organizational learning, loss of innovative capacity, dangers of an eternal triangle involving the client, the outsourcing provider and the business analysts serving as intermediaries in the project, technology indivisibility, and fuzzy focus of outsourcing only on the supply side of IT and not on other aspects like generating new application ideas or harvesting the benefits of IT.

A more elaborate list of 18 outsourcing risk factors grouped in 10 categories is presented in Dibbern et al. (2004) which extends the work of Earl (1996) with results from several other authors from the field of Management and other areas.

Bahli & Rivard (2005) divided IT Outsourcing risk factors into two groups: (a) factors associated with the transaction (Asset specificity; Small number of suppliers; Uncertainty; Relatedness between business units and functions; Measurement problems), and (b) factors related to the client and the supplier (Degree of expertise with the IT operation; Degree of expertise with outsourcing).

Taylor (2007) used the work of Schmidt et al. (2001) as a starting point to develop a list of factors affecting outsourcing projects from the provider's perspective and gathered opinions from a group of 22 experts from ten organizations to generate a broader set of categories of outsourcing risks. Her framework includes 42 risk factors, differentiated by source—vendor risks, client risks, and third party risks—and type—project management,

Lacity et al. (2009) provide a much larger list of 28 IT outsourcing risks based on analysis of published research in journals. While such a list is more informative about the types of outsourcing risks it is not very practical for decision making because of the lack of grouping of the factors. This issue is related to the difficulty of humans to differentiate between more than seven plus or minus two objects as was found by psychologist George Miller in 1956, a fact used by Saaty in the late 1970s to propose some of the concepts for structuring decision problems with the Analytic Hierarchy Process (AHP) and its extension, the Analytic Network Process (ANP) (see Saaty,1990).

The most comprehensive catalog of outsourcing risks to date is presented in de Sà-Soares, Soares & Arnaud (2014). It is again based on analysis of previously published research. They create a very detailed list of outsourcing risks, undesirable consequences and customer-related negative outcomes from outsourcing with the hope that those are initial steps in creating a theory explaining outsourcing risks. Those however are not reflecting well the specifics of offshore software development which will be discussed more in the next subsection.

Risks derived from studies of IT offshoring

Various aspects of risks in outsourcing and offshoring were investigated by Tafti (2005). They are summarized as 15 factors in four groups: Loss of Enterprise Knowledge, Privacy and Security, Hidden Costs and Outsourcing Contract. Some authors like Davis et al. (2005) and King (2008) provide small lists of IT offshoring risks based on expert opinion or on speculation or anecdote evidence, a feature of many publications as noticed by Nakatsu & Iacovou (2009:58). The first empirically validated list of IT offshoring risks through a Delphi study was developed by Nakatsu & Iacovou (2009). They investigated also outsourcing and software development as well.

Nakatsu & Iacovou (2009) investigated the project management literature and generated a summary of IT general risk factors derived from it. That list consists of 24 risk factors categorized in six groups: Team-related (Staff turnover, Lack of team communication, Lack of required technical and business knowledge, Lack of motivation, Team conflicts); Organizational environment (Lack of top management support, Organizational politics, Stability of organizational environment, Changes in organizational priorities); Requirements (Original set of requirements is miscommunicated, Continually changing system requirements, Unclear system requirements); Planning and control (Lack of project management know-how, Poor planning of schedules and budget, Poor change controls, Failure to consider all costs); User-related(Lack of adequate user involvement, Failure to gain user commitment, Failure to manage end-user expectations, Conflicts between user departments) Project complexity (Difficulties with

integration, Large number of links to other systems, Processes being automated are complex, Inadequate understanding of new technology).

Using as a starting point Earl (1996) and other published sources, the same authors summarize 36 IT outsourcing risks in the following 11 groups: Client capabilities, Vendor capabilities, Vendor–client communications, Contract management, Strategic risks, Legal/regulatory, Security, Financial, Geopolitical, Firm reputation/employee morale, Technology risks, Noncompliance with embraced development methodologies, Incompatible development tools.

The above findings were used by Nakatsu & Iacovou (2009) as a baseline for their Delphi study on risk factors in IT offshoring projects which identified 25 factors applicable to IT offshoring. As a result, they identified the following unique IT risk factors that are special to offshore outsourcing:

- Language barriers in project communications;
- Cross-national cultural difference;
- Constraints due to time-zone difference;
- Unfamiliarity with international and foreign contract law;.
- Political instability in offshore destinations;
- Negative impact on image of client organization;
- Currency fluctuation.

Since their Delphi study produced also the rankings of the various risk factors, Nakatsu & Iacovou (2008:64) concluded that with the exception of language barriers in project communications none of these risks were ranked very highly in importance by the panel of experts. Such findings are valuable for gaining general understanding of risks in software development but they do not apply strictly to the context of a specific software project. While the results of Nakatsu and Iacovou (2008) provide valuable insights into the different types of risks in outsourcing and offshoring, their lists of risks are not very suitable for decision modeling as they have not provided groupings in categories.

A comprehensive list of 18 IT offshoring risks and risk mitigation practices is discussed in Sakhtivel (2007). Another feature of that research is the comparison of the level of risk in two extreme cases of IT offshoring – having a single vendor as an outsourcing provider and own subsidiary located overseas as the offshore developer.

Chatfield & Wanninayaka (2008) used also previously published research to generate a list of risk factors in IT offshoring that are in three groups: 22 client related risks, 20 Vendor related risks and 6 inter-firm relationship risks. Abdullah & Verner (2012) analyzed offshoring risks based on the published literature and analyzed them through qualitative data analysis on a number of cases. Most of the research on offshoring risks is from client's perspective with the exception of the next paper.

Aundhe & Mathew (2009) have investigated the risks in IT offshoring from the provider's perspective on the basis of the published literature and have validated them using data gathered in five case projects. They produce the following list of risks and context factors:

Table 1. IT Offshore Risk Factors from provider's perspective (Aundhe & Mathew, 2009)

1 Macroeconomic risks
Government policy and regulations
Exchange rate
2 Relationship specific risks
Changes in client's corporate structure
Client's experience in offshoring
Client culture
Asset specificity
Client size
3 Project specific risks
Schedule and Budget Management
Staffing
Requirements capture
Knowledge transfer
Client expectations management
Testing
4 Context factors (not risks)
Relationship Maturity
Nature of contract
Nature of service
Nature of client

Through the analysis of the above risks in five case studies the authors have concluded that there is a strong interaction between relationship specific and project specific risks. The context factors however do not influence macroeconomic risks and are used just for understanding of the risks. Most of the factors identified by Aundhe & Matthew (2009) are general outsourcing and systems development risks while the following items from Table 1 were defined specifically as offshoring risks:

- *Knowledge transfer* resistance by the foreign client is an important risk factor especially when the project is about downsizing;
- *Client culture* of the client that considers the outsourcing relationship just as a transaction, i.e. pay the fees and get the service, results in risks for the provider as greater cooperation is better;
- *Client sizes* as bigger clients have higher bargaining power.
- *Exchange rate fluctuations.*
- *Government policy to offshoring.*

Aundhe & Mathew (2009) have concluded that the group of Relationship risks affects the category of Project related risks and vice versa. They have found also that there is no interaction between the relationship risk factors while Project schedule and budget management is affected by poor client expectations management, ambiguity in requirements capture, uncertainty in staffing and the risk of resistance to knowledge transfer by the client. Their results do not show a way to evaluate the strength of the above mentioned relationships and hence the need to provide a tool for modelling of offshoring risk factors in the context of particular software project which is proposed in the next section.

3. ON A SYSTEMIC MULTI CRITERIA FRAMEWORK FOR ASSESSMENT OF IT OFFSHORING RISKS FROM THE PROVIDER'S PERSPECTIVE

The proposed framework for assessment of IT offshoring risks from the perspective of a service provider is systemic because it fulfills the criterion for systemicity that all factors need to be considered with their relevant inter-relationships in the context of the particular software project (see Midgley, 2011). The systemicity of the framework will be supported by the choice of the Analytic Network Process. Since the latter enables the modeling of interdependencies like those discussed in the Aundhe & Matthew (2009) paper, it is a more powerful approach than the Analytic Hierarchy Process (AHP), a Multi Criteria Decision Analysis (MCDA) method (see Saaty, 1990 and Saaty, 2005). The features of MCDA as a systemic approach were analyzed in Petkov & Petkova (1998) and some aspects of its application to the selection of activities to outsource and outsourcing providers are discussed in Petkov & Petkova (2010). More details on the theory of AHP and ANP, their applications and suitability for various problems can be found in Saaty (1990) and Saaty (2005). We will mention here only a few characteristics that support the claim that AHP and ANP support systems modeling:

- Both AHP and ANP support decision models that aim at prioritizing the factors, in our case IT offshoring risks. Hence the models created with them support the purposeful system of assessing along multiple criteria the relative importance of IT offshoring risks in the context of a specific project.
- AHP models a problem in the form of a hierarchy, a useful construct to handle the complexity in systems, while ANP is used to model problems with interdependent elements as is the case of assessment of IT offshoring risk factors.
- Both AHP and ANP allow the measurement of pairwise importance of the IT offshoring risk factors involved in the models using a ratio scale that can convert both quantitative and qualitative variables to numbers representing human judgment about the risks involved.
- ANP is implemented in several software packages that hide the complex mathematics of the method from the user. We used Super Decisions Plus.

Both AHP and ANP use expert judgment about the pairwise comparisons of quantitative and qualitative factors in a model using a scale defined in Table 2. Those can be expressed as crisp judgements by a single individual or as a consensus judgment of a group of experts.

Other possible extensions of comparison modes in AHP/ANP include interval judgements or fuzzy judgments which however increase considerably the amount of effort in evaluating an ANP model (see Saaty, 2005) and hence that reduces their relevance for practical decision making about risk evaluation.

As a result of using this scale we get ratios representing the expert judgments about the quantitative or qualitative factors included in an AHP/ANP. These are organized in a matrix of comparisons whose elements are reciprocal with respect to the main diagonal. The local priorities of the factors from the matrix of comparisons are the elements of the principal right eigenvector of the matrix of comparisons corresponding to its largest eigenvalue (Saaty, 1990). Up to this point the procedure of ANP is overlapping with the AHP.

Table 2. The AHP/ANP pairwise comparison scale (Saaty,1990)

Intensity of importance	Definition when comparing two factors in AHP /ANP
1	Equal importance of the factors
3	Moderate importance
5	Strong importance
7	Very strong importance
9	Extereme importance
2,4,6,8	Intermediate values

The steps in ANP modeling involve:

- The network structure in ANP allows to model dependencies among elements in the model. When these dependencies are among clusters in the network model they are called outer dependencies. Some clusters have loops within themselves indicating inner dependence (Saaty, 2005:121). Paired comparisons are needed for all connections in the model. If there are inconsistencies in the comparisons the software allows to improve the judgments that are contributing to the inconsistency index (defined in Saaty, 2005:28). If its value is below than 0.1, it is considered that the provided judgments are reasonably consistent and do not violate the transitivity principle (Saaty, 1990).
- The priorities derived from pairwise comparison matrices are each entered as a part of some column of a supermatrix. The supermatrix represents the influence priority of an element on the left of the matrix on an element at the top of the matrix. The next step is to weight the supermatrix with the weights of the criteria in the control hierarchy that relate the criteria used in the model to the overall goal. The weights of the elements in the model are obtained as the limiting values of the columns of the weighted supermatrix raised to high powers as was shown in Saaty (2005).

The details of the mathematics of ANP can be found in Saaty (2005). These are not provided for space reasons and because manual calculations of the results without supporting software are too time consuming from a practical point of view. The steps in formulating an ANP model are outlined in Saaty (2005:90-92) and also in the online tutorials for the Super Decisions software package available at: http://www.superdecisions.com/category/support/support-2/. More details on AHP/ANP can be found in Subramanian & Ramanathan (2012) and in Sipahi & Timur (2010).

The proposed framework for assessment of IT offshoring risks from the service provider's perspective is presented in figure 1 below.

Exploration of IT project context and its stakeholders

Expert evaluation of the project characteristics, the relationship between the client and the service provider, the macroeconomic environment and the specific project context factors from the provider's perspective

Build the proposed network model of IT offshoring risk factors with the Super Decisions software and if relevant adapt it by adding or deleting some factors

Pairwise assessment of the risks in the network model and their prioritization

Fig.1 Proposed framework for assessment of IT offshoring risks in a particular project context

The understanding of the project context in the first step of the framework is developed through analysis of the stakeholders and their interests along the considerations provided in Petkov, Petkova & Andrew (2013) and Aundhe & Mathew (2009). The second step involves data gathering and traditional systems analysis activities about the nature of the offshoring work to be analyzed along the list of IT offshoring risk factors defined in Table 1. The third step is based on expert formulation in the Super Decisions software of the

ANP model of offshoring risk factors from the point of the service provider as it is defined in Figure 2. If necessary the model may include additional risk factors. The last step involves the ANP assessment of the set of relevant risks for the specific project.

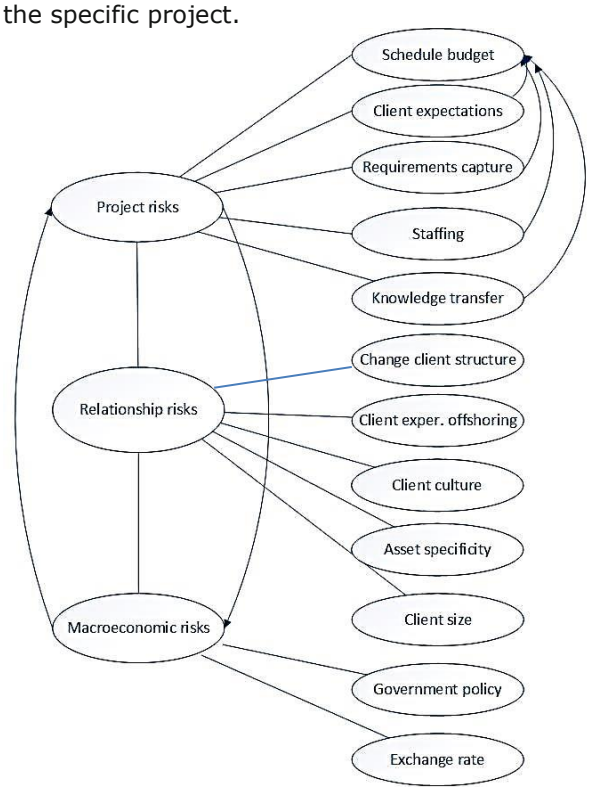

Fig. 2. Proposed ANP model of offshoring risk factors from the service provider's perspective (derived partly from the analysis of offshoring risks in Aundhe & Matthew (2009))

The ANP model includes the inner dependencies between project risks, relationship risks and macroeconomic risks within the Categories of risks cluster as well as the inner dependencies between schedule and budget management and the remaining Project related risk factors. It also shows no interference among the individual Relationship risk factors and the individual Macroeconomic factors following the previously mentioned findings of Aundhe & Matthew (2009).

Saaty (1990) recommends the use of Benefit/Cost, Benefit/Risk or Benefit/ [Cost*Risk] ratios as a way of modeling risk in AHP/ANP. Millet & Wedley (2003) reject that idea and propose the direct use of risks as criteria in the prioritization process or the use of risk as an adjustment factor for costs or benefits. The proposed framework for ANP evaluation of IT offshoring risks and the corresponding ANP model in Fig. 2 prioritizes offshoring risk factors directly following Millet & Wedley (2003). The next section illustrates the practical application of the model.

4. PRACTICAL ILLUSTRATION OF THE ANP EVALUATION OF IT OFFSHORING RISKS FROM PROVIDER'S PERSPECTIVE

The model for assessment of offshoring risks from the service provider perspective was applied to a practical problem involving a small Bulgarian software company that operates since 2007. It provides web 2.0 services, e-commerce and related software to Bulgarian, American and other clients. The company is closely linked to an US software service provider from its inception. Its president while on a visit to the US in December 2016 provided the evaluation of offshoring risks related to a specific project about a custom based e-commerce application for an US client that involved also interactive web page design and customer relationship management components. He was assisted by the first author in implementing the model with the Super Decision software.

Following the first step of the framework in Fig.1 the analysis began with a discussion of the project context factors: Relationship Maturity, Nature of contract, Nature of service and Nature of client. The project is of medium complexity and it was the first instance when the client company was working with this software provider. The client company was quite big and that was giving it leverage in the negotiations about the contract. The exchange rate fluctuations were not considered problematic as the Bulgarian currency is linked to the Euro and its exchange rate to the US dollar does not fluctuate like some other currencies. On the other hand, possible future changes in US government policy towards offshoring were considered as a moderate risk. It was considered (in similarity to the findings of Aundhe & Mathew, 2009) that Macroeconomic risks were far less important than Project and Relationship related risks.

The inner dependencies among the nodes in the Risk Categories cluster were assessed separately against each of them considered as a criterion. The pairwise comparisons matrices and the local priorities for them are listed next:

Comparisons with resp. to Project risks

	R. r.	M. r.	Local pr.
Relationship risks	1	9	0.9
Macroecon. risks	1/9	1	0.1

Comparisons with resp. to Relationship risks

	P. r.	M. r.	Local pr.
Project risks	1	8	0.888
Macroecon. risks	1/8	1	0.112

Comparisons with resp. to Microeconomic risks

	R. r.	M. r.	Local pr.
Relationship risks	1	2	0.667
Project risks	1/2	1	0.333

The comparisons of the risk factors related to the categories of Project risk, Relationship risk and Macroeconomic risks and the local priorities derived from those matrices are shown in the Appendix. The resulting priorities of the risk factors are in the last column of Table 3.

Table 3. IT Offshoring risks for the specific project from the service provider's perspective

Risk category	Prior.from limit super supermatrix	Priorities normalized in clusters
Macroeconomic risks	0.044	0.096
Project risks	0.205	0.442
Relationship risks	0.214	0.462
Offshoring risk factors		
Asset specificity risk	0.010	0.019
Changes client corp.str.	0.036	0.067
Client culture	0.025	0.046
Client expectations mgt	0.045	0.085
Client exp. in offshoring	0.056	0.104
Client size	0.088	0.164
Exchange rate fluct.	0.007	0.014
Government policy	0.037	0.069
Knowledge transfer	0.061	0.114
Requirements capture	0.118	0.220
Schedule, budget mgt.	0.036	0.068
Staffing fluctuations	0.017	0.032

The Appendix contains also the comparisons between risk factors with respect to Schedule and budget management reflecting the inner dependencies in the Project related risk factors shown in Fig. 2. The Super Decisions software generated the unweighted and the weighted supermatrices and produced the limit supermatrix which are in the Appendix as well. Since any comparison between both clusters of Risk categories and Offshoring risks in our model (see Fig. 2) is not needed because the risk categories are just groupings of the risk factors and they should not be compared to each other, the sum of priorities for each cluster is equal to 0.5 as is evident from the second column of Table 3. These are used to generate the normalized priorities (their sum is equal to 1) within each cluster that are shown in the third column.

The risks with highest priorities are the danger of ambiguity in Requirements capture (22%), Client size (16.4%), Knowledge transfer on the problem by the client to the offshore provider (11.4%), client experience with offshoring (10.4%) and client expectations management (8.5%). Hence it was necessary to keep close contact with the client in the continuous verification of the project requirements and about the progress on the project as well as applying other possible mitigation strategies for those risks.

The least important risks were as follows: Exchange rate fluctuations (1.4%), followed by Asset specificity risk (1.9%), Staffing fluctuations for the developer (3.2%). That was due to the relative staffing stability of the provider, the fact that it had previous experience with similar projects for other clients and because historically the exchange rate of the US dollar to the Euro is stable (since the currency of the country of the provider is linked to the Euro).

The expert that provided the pairwise comparisons for the assessment of the risks for the particular project considered here found the results of the model adequate as they delivered a more precise quantitative expression of the importance of the risks associated with the project in comparison to the traditional approach for evaluation of risks based on perceptions.

5. CONCLUSION

We analyzed in this paper what is known from past research on IT offshore outsourcing risks which is a highly important topic in IT offshoring according to Gonzalez et al. (2013). The understanding of those risks was developed through investigation of findings of previous publications on software development project risks, IT outsourcing risks and from studies of IT offshoring risks with a focus of the service provider perspective as it is researched to a smaller degree compared to risks from the client perspective and there are no papers on prioritizing their interactions in the context of a specific project. A justification is provided for the use of the Analytic Network Process (see Saaty, 2005) for modeling such risks in a framework that is proposed in this paper. The practical application of the model is illustrated on the problem for modeling risks for a specific IT offshoring project

from the point of view of an Eastern European outsourcing provider serving US clients.

The theoretical validity of the model is supported by the fact that it was developed following the findings on offshoring risks from the provider's perspective by Aundhe & Matthew (2009). It is using the Analytic Network Model (see Saaty, 2005) which has been applied successfully in various problems according to Sipahi & Timor (2010).

The proposed framework for assessment of IT offshoring risk factors from the service provider perspective in the context of a particular project can be used for better understanding and management of risks in practice. To the best knowledge of the authors there is no published account of a systemic ANP framework for prioritizing of risks in IT offshoring risks from the provider perspective and hence the theoretical contribution of this paper.

Possible directions for further work include the practical application of the ANP framework for modeling and prioritizing of IT offshoring risks in additional situations developed both from the client and provider perspective. Another possibility is comparing the results from ANP models of offshoring risks with those obtained through unstructured text analysis as in Abdullah & Verner (2012), or through using Bayesian Networks or another technique for modeling of relationships between risks. The proposed framework and the corresponding Analytic Network Model are a step in improving the understanding of IT offshore outsourcing risk factors from a service provider's perspective.

6. ACKNOWLEDGEMENTS

The authors are grateful for the useful comments provided by the anonymous reviewers of an earlier version of the paper presented at CONISAR 2016.

7. REFERENCES

Abdullah, l. M. & Verner J.M. (2012). Analysis and application of an outsourcing risk framework. *The Journal of Systems and Software* ,85, 1930– 1952

Aundhe, M.D., & Mathew, S.K., (2009). Risks in offshore IT outsourcing: A service provider perspective, *European Management Journal*, 27 (6), 418-428.

Bahli, B., & Rivard, S. (2005). Validating measures of information technology outsourcing risk factors. *Omega, 33* (2), 175-187.

Boehm, B. W. (1991). Software risk management: Principles and practices. *IEEE Software*, *8* (1), 32-41.

Bunker, D., Hardy, C., Babar, A., Stevens, K.J. (2015). Exploring Practitioner Perspectives of Sourcing Risks: Towards the Development of an Integrated Risk and Control Framework, *Proceedings 2015 Australasian Conference on Information Systems,* Adelaide, South Australia.

Chatfield, A. T., & Wanninayaka, P. (2008). IT Offshoring Risks and Governance Capabilities. *Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS '08),* 436-445.

Davis, G., Ein-Dor, P., King, W. R.; & Torkzadeh, R. (2006). IT Offshoring: History, Prospects and Challenges, *Journal of the Association for Information Systems*, 7 (11), Article 32.

DeHondt II G. & Nezlek G. (2009). The Cost of Risk in Offshore Systems Development, *Proceeding AMCIS 2009.* AIS.

De Sà-Soares, F., Soares, D., & Arnaud, J. 2014. A Catalog of Information Systems Outsourcing, *International Journal of Information Systems and Project Management,* 2 (3), 23-43.

Dibbern, J., Goles, T., Hirschheim, R. & Jayatilaka, B. (2004), Information systems outsourcing: a survey and analysis of the literature, *Database for Advances in Information Systems*, 35 (4), 6-102.

Gandhi, S. J., Gorod, A. & Sauser, B. (2012) Prioritization of outsourcing risks from a systemic perspective, *Strategic Outsourcing: An International Journal*, 5(1), 39-71.

Ghadge, A., Dani, S., Chester, M., & Kalawsky, R. (2013), A systems approach for modelling supply chain risks, *Supply Chain Management: An International Journal*, 18 (5) 523–538.

Gonzalez R, Llopis J. & Gasco J. (2013) Information systems offshore outsourcing: managerial conclusions from academic

research, *International Entrepreneurship Management Journal,* 9:229–259.

Herath, T., & Kishore, R. (2009). Offshore Outsourcing: Risks, Challenges, and Potential Solutions. *Information Systems Management,* 26(4):312- 326.

Hirschheim, R. (2006). Offshore outsourcing: Challenge to the information systems discipline. In R. Hirschheim, A. Heinzl, & J. Dibbern (Eds.), *Information systems outsourcing. Enduring themes, new perspectives and global challenges* (2nd ed.). Berlin: Springer.

Keramati A., Samadi H., Nazari-Shirkouhi, S. (2013) Managing risk in information technology outsourcing: an approach for analyzing and prioritising using fuzzy analytical network process, *International Journal of Business Information Systems,* 12 (2), 210-241.

King, W (2008), An IS Offshore Outsourcing Framework: Emerging Knowledge Requirements for IS Professionals**,** *Journal of Information Technology Cases and Application Research* (JITCAR), 10 (4):7-31.

Lacity, M.C., Khan, S.A. & Willcocks, L.P. (2009) A review of the IT Outsourcing Literature: Insights for Practice, *Journal of Strategic Information Systems*,18(3), 130-146.

Liang H., Wang J-J., Xue Y., Cui X. (2016). IT outsourcing research from 1992 to 2013: A literature review based on main path analysis, *Information & Management* 53, 227–251.

Liu, L.B., Berger, P., Zeng A., & Gerstenfeld, A. (2008). Applying the analytic hierarchy process to the offshore outsourcing location decision, *Supply Chain Management: An International Journal,* 13/6, 435–449.

Mathew, S.K.& Chen, Y. (2013), Achieving offshore software development success: An empirical analysis of risk mitigation through relational norms*. Journal of Strategic Information Systems,* 22(4), 298–314.

Midgley, G. 2011. Theoretical pluralism in systemic action research. *Systemic Practice and Action Research*, 24 (1): 1-15.

Millet, I. & Wedley, W.C. (2002), Modelling Risk and Uncertainty with the Analytic Hierarchy Process, *Journal of Multi-Criteria Decision Analysis,* 11:97-102.

Nakatsu, R. T., & Iacovou, C. L. (2009). A comparative study of important risk factors involved in offshore and domestic outsourcing of software development projects: A two-panel Delphi study, *Information & Management*, 46(1): 57-68.

Ngwenyama, O.K. & Bryson, N. (1999), Making the information systems outsourcing decision: a transaction cost approach to analyzing outsourcing decision problems, *European Journal of Operational Research*, 115:351-367.

Oshri, I., Kotlarsky, J., Willcocks, L.P. (2015). *The Handbook of Global Outsourcing and Offshoring,* Third Ed., Palgrave MacMillan, London.

Persson, J.S. & Schlichter B.R. (2015), Managing Risk Areas in Software Development Offshoring: A CMMI Level 5 Case, *Journal of Information Technology Theory and Application*, 16 (1) paper 2, 5-24.

Pfleeger, S. L.(2000) Risky business: what we have yet to learn about risk management, *The Journal of Systems and Software*, (53), 265-273.

Peslak, A.R. (2012). Outsourcing and offshore outsourcing of information technology in major corporations, *Management Research Review,* 35 (1), 14–31.

Petkov D & Mihova-Petkova, O. (1998). The Analytic Hierarchy Process and Systems Thinking, in Stewart T.J. and van der Honert R.C. (Eds), *Trends in Multi-Criteria Decision Making*, *Lecture Notes in Mathematics and Mathematical Systems*, Springer, Berlin, Vol 465, 243-252.

Petkov D. & Petkova O. (2010). On Design Science, MCDM and IT Outsourcing Decisions, *Journal of Information Systems Applied Research (JISAR)*, Vol.3, No 20.

Petkov, D., Petkova, O. & Andrew, T. (2013). On Some Lessons from Modeling Contexts in Complex Problem Solving in Information Technology, *Journal of Information Technology Research (JITR)* 6(4), 55-74

Rottman, J. W. & Lacity, Mary C. (2006) Proven Practices for Effectively Offshoring IT Work *MIT Sloan Management Review. 47*(3), 56-63.

Saaty, T. L. (1990). *Theory and applications of the Analytic Hierarchy Process*, RWS publications, Pittsburgh.

Saaty, T. L. (2005). *Theory and applications of the Analytic Network Process*, RWS publications, Pittsburgh.

Sakthivel, S. (2007), Managing risk in offshore systems development. *Communications of ACM* , 50 (4), 69-75.

Sipahi, S. & Timor, M. (2010). The analytic hierarchy process and analytic network process: an overview of applications, *Management Decision*, 48 (5). 775 – 808.

Schmidt, R, Lyytinen, K, Keil, M, Cule, P. (2001), Identifying software project risks: an international Delphi study, *Journal of Management Information Systems* 17 (Spring (4)), 5–36.

Strasser, A. & Westner, M. (2015), Information Systems Offshoring: Results of a Systematic Literature Review, *Journal of Technology Management* 26 (2), 70-142.

Subramanian, N. & Ramanathan, R. (2012), A review of applications of Analytic Hierarchy Process in operations management, *International Journal on Production Economics*, 138: 215-241.

Super Decisions software package on ANP. http://www.superdecisions.com/

Tafti, M. (2005). Risk Factors Associated with Offshore IT Outsourcing, *Journal of Industrial Management and Data Systems*, 105(5), 549–560.

Taylor, H. (2007). Outsourced IT Projects from the Vendor Perspective: Different goals, different risks, *Journal of Global Information Management*, 15(2): 1–28.

Verner, J., Brereton, O., Kitchenham, B., Turner, M., & Niazi, M. (2012). *Evidence based Global Software Engineering Risks Extracted from Systematic Literature Review*, technical report, University of Keele.

Wallace, L, Keil, M. & Rai, A. (2004), How Software Project Risk Affects Project Performance: An Investigation in the Dimensions of Risks and an Exploratory Model, *Decision Sciences*; Spring 2004; 35 (2) 289-321.

8. Appendix: Further results from the ANP model of offshoring risks produced with the Super Decisions software

Table A1 Pairwise Comparisons of the risk factors related to Project risks

	C. e.	K. t.	R. c.	S. b.	S	Local priorities
Client expectations	1	1/3	1/4	1	1/3	0.11621
Knowledge transfer		1	1/3	2	4	0.23906
Requirements capture			1	2	6	0.42161
Schedule & budget mgt				1	7	0.17787
Staffing					1	0.04525

Table A2 Pairwise Comparisons of the risk factors related to Relationship risks

	A.s.	C.c.s.	C. c.	C.e.o.	C.s	Local priorities
Asset specificity	1	1/4	1/3	1/5	1/6	0.04718
Changes in client struct.		1	2	1/3	1/2	0.16728
Client culture			1	1/2	1/3	0.11561
Client exper. offshoring				1	1/3	0.25991
Client size					1	0.41002

Table A3 Pairwise Comparisons of the risk factors related to Macroeconomic risks

	E. r.	G. r.	Local priorities
Exchange rate fluctuations	1	1/5	0.16667
Government regulation to offshoring		1	0.83333

Table A4 Pairwise Comparisons of the risk factors related to Schedule & budget management as a result of the inner dependencies within the Offshoring risks cluster

	C. e.	K. t.	R. c.	S	Local priorities
Client expectations	1	3	1/2	2	0.29545
Knowledge transfer		1	1/2	2	0.16774
Requirements capture			1	4	0.42969
Schedule & budget mgt				1	0.10742

Table A5. Offshoring risks evaluation unweighted supermatrix (part1)

	Risk categories			**Various risks**		
	Macroec.	Project ris	Relationsh	Asset spe	Ch. Cl.str.	Cl. Culture
Macroeconomic risk	0	0.1	0.11111	0	0	0
Project risk	0.33333	0	0.88889	0	0	0
Relationship risk	0.66667	0.9	0	0	0	0
Asset specifity risk	0	0	0.04718	0	0	0
Changes in client corporate str	0	0	0.16728	0	0	0
Client culture	0	0	0.11561	0	0	0
Client expectatations	0	0.11621	0	0	0	0
Client experience in offshoring	0	0	0.25991	0	0	0
Client size	0	0	0.41002	0	0	0
Exchange rate fluctuations	0.16667	0	0	0	0	0
Government policy	0.83333	0	0	0	0	0
Knowledge transfer	0	0.23906	0	0	0	0
Requirements capture	0	0.42161	0	0	0	0
Schedule and budget managem	0	0.17787	0	0	0	0
Staffing	0	0.04525	0	0	0	0

Table A5. Offshoring risks evaluation unweighted supermatrix (Part 2)
Various risks (continued)

Cl.expecta	Cl.ex.offsh	Client size	Exc. rate f	Gov.policy	Knowl. Tr	Req.cap	Sch.budge	Staffing
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0.29545	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0.16744	0
0	0	0	0	0	0	0	0.42969	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0.10742	0

Notes:

- The above table is in two parts for page space reasons.
- The priorities in the weighted supermatrix are all equal to half of the values in Table A5 as the two clusters (Risk categories and Offshore risks) have the same weight.

Table A6. Offshoring risks evaluation – the limit supermatrix (part1)

	Macroec.	Project ris	Relationsh	Asset spec	Ch. Cl.str.	Cl. Culture
Macroeconomic risk	0.0443	0.0443	0.0443	0	0	0
Project risk	0.2051	0.2051	0.2051	0	0	0
Relationship risk	0.21412	0.21412	0.21412	0	0	0
Asset specifity risk	0.0101	0.0101	0.0101	0	0	0
Changes in client corporate str	0.03582	0.03582	0.03582	0	0	0
Client culture	0.02476	0.02476	0.02476	0	0	0
Client expectatations	0.04539	0.04539	0.04539	0	0	0
Client experience in offshoring	0.05565	0.05565	0.05565	0	0	0
Client size	0.0878	0.0878	0.0878	0	0	0
Exchange rate fluctuations	0.00738	0.00738	0.00738	0	0	0
Government policy	0.03692	0.03692	0.03692	0	0	0
Knowledge transfer	0.06125	0.06125	0.06125	0	0	0
Requirements capture	0.11782	0.11782	0.11782	0	0	0
Schedule and budget managem	0.03648	0.03648	0.03648	0	0	0
Staffing	0.01712	0.01712	0.01712	0	0	0

Note:
Part 2 of Table A6 is not provided here for space reasons as all its columns contain zeroes just like the last three columns in part 1 of the table. Note that non-zero columns in the limit supermatrix in ANP have the same elements as are the first three columns here. They provide the priorities of all the elements in the clusters of the ANP model.

Downloading Mobile Applications – Are Students Protecting Themselves?

Adnan A. Chawdhry
chawdhry_a@calu.edu
California University of Pennsylvania
California, PA

Karen Paullet
paullet@rmu.edu

David M. Douglas
douglas@rmu.edu

Robert Morris University
Moon Township, PA

Joseph Compimizzi
jcompomizzi@fau.edu
Florida Atlantic University
Boca Raton, FL

Abstract

Mobile applications (apps) are taking the world by storm. Currently, end users have downloaded over 225 billion apps on their mobile devices. Security concerns surrounding the downloading of apps are often overlooked. The apps on our smart phones can be accessed by the tip of our fingers or the sound of our voice. One must think about the interactive risks to our privacy and the security concerns that can affect our digital lives. This study explores awareness and security risks associated with downloading mobile apps. A total of 124 students were surveyed at two mid-Atlantic Universities. The study found that many students are downloading mobile apps without fully understanding the security risks associated with such action.

Keywords: Mobile security, mobile applications, apps, mobile device

1. INTRODUCTION

Mobile applications could be considered a scourge or savior to human interaction with our smart phones depending on who is asked. Each day many new or improved mobile applications are being created. These App creators can be found in all age groups, cultures and from all social economic backgrounds. Some are designed to make our life easier (location and directional) and less stressful (reminders and flashlight). It appears there is an App for all needs both real and perceived. According to Statistica (2016), there has been an upward trend in mobile app usage. In 2011, there were 22 billion free app downloads and 2.9 billion paid app downloads. As of June 2016, people have downloaded over 211 billion free apps and 13.49 billion paid apps showing the significant rise in mobile app usage.

These App creators, both young and old create for fun, profit, or perhaps most importantly to fill a "void" in the ever expanding catalogue of must have "apps." These apps, also known as mobile applications, are designed, or so they say, to improve our lives. Perhaps they do in some respect, but one of the unintended consequences is a more complacent and indolent mobile community especially in regards to cyber security and the oversharing of information both private and public.

However as with all things in life, there are unintended consequences. We live in a brave new world of the Internet of Things (IoT) and smart phones. The applications (Apps) on our smart phones are at the tip of our fingers or the sound of our voice. Knowing and unknowingly we often overshare many aspects of our personal information in cyberspace. Once shared, we can never retrieve or change this cyber data. The information is now beyond our grasp and control. One wrong click or one wrong tap of our finger on the wrong button or link can change a life instantly. This lapse of judgment or mistaken "click or send" can allow a miscreant hacker or rouge agency to gain access to financial and personal aspects of our digital lives. One must think about the interactive risks to our privacy and the security concerns that can affect our digital lives.

2. LITERATURE

Recent years have witnessed an explosion in the acquisition and use of mobile computing devices known as smart technology. According to the February 2013 Federal Trade Commission Report, 217 million smartphones were purchased in the fourth quarter of 2012 alone (Mobile Privacy Disclosures). Consumers of smart devices are using the technologies offered by these smart devices for a multitude of functions from waking up with alarm applications to lunch ordering and purchase to monitoring traffic for the commute home, not to mention the more mundane daily tasks of the texts, calls and emails completed through personal mobile devices.

As the functions of a mobile device become more complex, so too do their operating systems and development of their applications. And with this increased complexity of functionality, comes complexity with understanding: namely security and privacy understanding. Theoharidou, Mylonas, and Gritzalis (2012) explain that mobile apps are both an asset and threat for users. While the social, financial and business benefits of an app are numerous, they can act as a security attack access point for users. These security threats range from spoofing, to cloning, to unauthorized access, to disablement, to phishing to malware injection all related to permission access rights and authentication violations (Theoharidou, Mylonas, and Gritzalis, 2012).

One of the characteristics of how we conduct our mobile communication activities in 2016 is recognizing some of the more perilous aspects and unforgiving consequences of our more than casual acceptance of the "Terms of Service Agreement" before downloading any given application (APP). For many people, including some of the authors of this document, we are guilty of blindly checking "I accept" the terms of service for any given App without a hint of even reading the first sentence (Boyles, et.al, 2012).

This blind acceptance often permits the creator and/or carrier of the mobile application full access to many features of our mobile devices, including photos, contacts, and location to name just a few. Indeed, it is a frightening and somewhat unsophisticated Orwellian circumspection of our time and place in history. In short are we willing oversharing personal information about ourselves and those connected to and imbedded on our mobile devices.

We are at last finally comprehending just how much total and complete access to every aspect of our personal information we are blinding giving to a plethora unknown third parties to do with as they wish with our full and unequivocal consents. However, all is not lost as mobile device users are awakening to the fact that they do not want these third party terms of agreement unknowns to have control and access to their personal information. As our adoption of mobile technology cultivates and our acceptance of sharing our personal aspects of our life increases it would seem reasonable that we accept and welcome the apps that seemingly make our lives easier (Boyles, et.al, 2012).

Koved, Trewin, Swart, Singh, Cheng, Chari (2013) discussed the risks associated with the adoption of mobile devices regarding its authentication and authorization on network services. Their research especially focused when these devices were relied on to input or share sensitive information. Mobile devices such as smartphones, tablets, and other "mobile platforms" are now commonly used for banking and shopping. Accordingly, they have identified several risks. They include the possibility of that the user's action will be observed and allow an

unauthorized authentication or "impersonation" on a different device. Understandably, when devices are stolen or lost the risks of exposing sensitive information is increased. "In particular, mobile device applications, including the web browsers, are caching authentication credentials, enabling an attacker to exploit them. Modern smartphones can enable multi-factor authentication by using sensors such as cameras and microphones to capture biometric data" (Koved, 2013).

Concerning third party applications commonly referred to as mobile apps, distribution marketplaces such as Apple's Appstore offer two types: paid apps and free apps. Understanding the difference between the two provides a foundation to pivot a discussion on security issues with mobile technology devices. Free apps, with no surprise, are more popular than paid apps. According to Petsas, Papadogiannakis, Polychronakis, Markatos and Karagiannis (2013), "paid apps usually have more advanced functionality and do not include advertisements" (p. 285). According to the study conducted by Compomizzi (2013), of the college student participants with iPads, 54.2% indicated that they paid for a few apps while 20.5% indicated that they didn't pay for any. Further, participants in this study indicated that the apps they purchased were related to academic uses specifically to complete study tasks like note-taking app's, for academic tools like calculator and dictionary apps, and for course requirements like e-book apps and video apps.

Given that free apps rely on advertisements, learning about the usage patterns by mobile device operators yields additional information that leads to a more thorough examination of the issue of security. In the study by Petsas, et al. (2013), 55,000 free apps from the Google Play Store were categorized, tracked and examined. The analysis of data collected in the study disclosed that the top 10 categories accounted for 60% of the apps. These app categories included tools, entertainment, brain apps like puzzles, lifestyle, business, books, travel, education and casual. Of the 55,000 apps examined, 46,000 as for the android permission to access the network. Further, of these 46,000 apps, 19,000 were connected to at least one advertisement library (Petsas, 2013).

As a result, skepticism and mistrust about the use of personal information by platform hosts, app developers and advertisers are increasing among smart device owners. A 2012 study by Boyles, Smith, and Madden revealed that "more than half of app users have uninstalled or decided to not install an app due to concerns about personal information". In fact, of the 2,254 participants in their study, Boyles, Smith and Madden reported that 49% of users between 18 and 29 indicated that they decided not to install an app based on personal information concerns; of those in the same age bracket, 29% report uninstalling an app due to concerns about personal information sharing. Interestingly, their study also revealed that "app users with at least some college experience are somewhat more likely than those with a high school education to choose not to install an app over privacy concerns (Boyles, Smith and Madden, 2012).

With this understanding of mobile technology, system operations, user behaviors, and app interfaces, Theoharidou, Mylonas, and Gritzalis (2012) explain the mobile apps are both an asset and threat for users. While the social, financial and business benefits of an app are numerous, the app itself may need protection and can act as a security attack access point for users. These security threats range from spoofing, to cloning, to unauthorized access, to disablement, to phishing to malware injection all related to permission access rights and authentication violations (Theoharidou, Mylonas, and Gritzalis, p. 450). As Koved, Trewin, Swart, Singh, Cheng, and Chari (2013) write, "In particular, mobile device applications, including the web browsers, are caching authentication credentials, enabling an attacker to exploit them" (p. 1).

The good news is that advances in mobile technology and user protection continue in development. Secure passwords are only the beginning. Mobile and smart technology are incorporating camera and voice detection sensors. Biometrics with fingerprinting and retinal recognition are also advancing to counteract privacy and security concerns. The bad news is that these additional security features are often in direct contrast to mobile operators' expectations of easy to use, fast, and on-the-go technology. Users often view these additional security steps as burdensome. In a study conducted with IT professionals who also teach at the college level by Compomizzi, D'Aurora, and Rota (2013), of 90 question responses received regarding security practices, 76 indicated regular practice of low tech methods of protection such as password authentication and using multiple browsers for different computing functions while only 14 employed high tech methods of security protection like biometrics.

The literature concerning how mobile technology is perceived and used by operators is ever-growing. Interesting definitions of a mobile device continue to emerge. Likewise, the uses of mobile technology continue to grow, placing demand upon more flexible, available and integrated computing capabilities and mobile applications. With this expansion in mobile technology, security risks are also increasing. While software and hardware developers forge ahead with progressed security solutions, users may perceive them as burdensome; thereby opening the door to information invasion and attack.

3. METHODOLOGY

The study surveyed students from two small mid-Atlantic Universities from March to April 2016. For this study, the population chosen comprised of undergraduate and graduate students enrolled in on-campus or online programs. This population was chosen to ensure students surveyed would be 18 years or older. A total of 124 students completed the survey. The researchers utilized Survey Monkey, an online survey tool, to collect data, which were then imported into SPSS for organization and analysis. As part of the analysis, the researchers used a Chi-square analysis with a statistical significance at the .05 margin of error with a 95% confidence Level. The study addressed the following two research questions.

1. What actions are students taking to reduce privacy / security concerns when downloading applications on their mobile devices?

2. Is there a statistical significance among age, gender, and level of education with the actions student take to mitigate the risks of privacy / security with downloading applications?

The survey administered to students consisted of 22 closed-ended questions and one open-ended question for further understanding of the participant's responses. The first three questions focused on student demographics to include age, gender, and level of education. The remaining questions focused on whether students were aware of security and privacy concerns that exist with downloading mobile applications. The questions primarily focused on responses of "Yes" and "No", while a few questions provided additional options for students to select the type of mobile device they use, applications they use on their phone, and how many apps they have downloaded.

4. RESULTS

The survey presented seven scenarios where it prompted the participant to respond with a "Yes" or "No" answer, one open ended question for further analysis, and a multiple choice question with predefined responses including an "Other" option to include additional responses. These questions were designed to understand what actions students take to reduce security and privacy concerns when downloading mobile applications. These questions included what the use of anti-malware software, backing up phone content, clearing browsing history, disabling location services, uninstalling an application and why, and choosing to uninstall / not install an application once they were aware of the security and privacy impacts. The summary of the Yes / No results are provided in Table 1. Additionally, the researchers thought it would be important to understand how many applications downloaded on average. The highest response rate was between 11-20 applications with 37.90% followed by 1-10 applications at 22.40%. The breakdown of these results can be seen in Table 2.

Table 1: Survey Questions

Scenario	Yes	No
Downloaded Mobile Apps	96.64%	3.36%
Disabled Location Services	84.48%	15.52%
Clear Browsing / Search History	74.14%	25.86%
Backup using 3rd Party Software	34.21%	65.79%
Installed Anti-Malware	29.31%	70.69%
Uninstalled / Not Installed App	94.71%	5.29%
Not Installed after discovering how much personal information is shared	77.00%	23.00%
Uninstalled after discovering how much personal information is shared	64.60%	35.40%

Table 2: Number of Downloaded Applications

Number of Mobile Apps Downloaded	Percent
0	1.70%
1-10	22.40%
11-20	37.90%
21-30	19.00%
31-40	5.20%
More than 40	13.80%

Additionally, the researchers were interested to further analyze the student responses on reading the terms of use for an application compared to their awareness that applications have access to their phone's content. Approximately 83.19% of the students were aware that mobile applications have access to their content while only 14% were unaware of this. Additionally, only 33.61% of students responded that they have read the terms of use before downloading an app. The highest percentage of 51.26% was found where students did not read the terms of use but were aware that applications have access to their phones content. The breakout of these results can be found in Table 3.

Table 3: Reading Terms of Use Vs Awareness

Read Terms of Use	Aware that Apps have access to Phone context		
	Yes	No	Total
Yes	32.76%	1.72%	34.48%
No	52.59%	12.93%	65.52%
Total	85.34%	14.66%	100.00%

While understanding the actions students took in regards to protecting their mobile devices from security and privacy concerns is important, knowing the reasons behind their decisions to uninstall an app, choose to not install an app, or disable location services may provide additional insights. The survey asked why students chose to uninstall a mobile application. The most common reason was because the application was collecting personal information with a response rate of 37.5%. The least common was security concerns. Table 4 below shows the breakdown of responses including an option to choose "Other".

Table 4: Reasons to Uninstall or not Install

Reasons to uninstall App	Percentage
Privacy Concerns	18.80%
Security Concerns	12.50%
Collecting personal Information	37.50%
Other	31.30%
Total	100.00%

The survey provided a supplemental question if students selected "other." Below are responses from those participants.

- Didn't use the app
- Either too large or didn't use it often
- Privacy and security concerns as well as collecting personal information
- The app is not useful for me anymore
- The app was not what I had thought it was.

While it was important to understand why they chose to uninstall an app, we thought it was also important to note the reasons they may have chosen to disable locations services for apps they decided not to uninstall. The responses included the following:

- Not necessary for the app to function
- Battery Life
- Told to disable it
- Tracking me
- Privacy / Security Concerns
- Don't trust it
- Used too much data
- Feeling insecure

Additionally, the participants were asked which applications they chose to disable location services for. Below is a summary of those responses.

- All applications
- Social Media Sites
- Banking
- Retail / Shopping
- Unpopular Apps
- Weather
- Maps
- Games
- News
- Calendar
- Photos

Lastly, you will find a chi-square analysis performed on these participant responses against age, gender, and level of education to understand any statistical correlation that may have existed. Only values of .05 or less were considered

statistically significant. These results can be found in Tables 5-7. Age had a statistical significance with clearing the browsing / search history, backing up the phone's content, and using anti-malware software. Gender was statistically significant with clearing the browsing / search history and using anti-malware software. Level of Education did not illustrate a statistical significance with any of the response.

Table 5: Chi-Square Analysis with Age

Action to Protect Security and Privacy	Age (df = 6)
Disabled Location Services	0.704
Clear browsing / search history	0.016
Backup phone contents with third party app	0.05
Use anti-malware	0.028
Read Terms of use / service	0.197
Uninstalled / Not Installed App	0.856
Not Installed after discovering how much personal information is shared	0.375
Not Installed after discovering how much personal information is shared	0.933

Table 6: Chi-Square Analysis with Gender

Action to Protect Security and Privacy	Gender (df = 1)
Disabled Location Services	0.362
Clear browsing / search history	0.035
Backup phone contents with third party app	0.925
Use anti-malware	0.002
Read Terms of use / service	0.201
Uninstalled / Not Installed App	0.771
Not Installed after discovering how much personal information is shared	0.26
Not Installed after discovering how much personal information is shared	0.191

Table 7: Chi-Square Analysis with Level of Education

Action to Protect Security and Privacy	Level of Education (df = 5)
Disabled Location Services	0.98
Clear browsing / search history	0.234
Backup phone contents with third party app	0.506
Use anti-malware	0.234
Read Terms of use / service	0.249
Uninstalled / Not Installed App	0.265
Not Installed after discovering how much personal information is shared	0.622
Not Installed after discovering how much personal information is shared	0.454

5. DISCUSSION

Mobile applications can access a good amount of information on your phone which can lead to security and privacy concerns. Most of this is outlined in the terms of use, but the question is how often do we really read it? Even more important was do we take action if we read it or do we choose to take action just knowing there are general concerns in terms of security or privacy. The survey revealed that 96.64% of the respondents have downloaded apps on their mobile devices. Additionally, 84.48% have disabled location services on their device. These two numbers were interesting because it illustrated that while a high percentage do download apps, they took the first step of disabling location services to protect their privacy. Another important metric was that 94.71% of the participants have chosen to uninstall or not install an Application on their phone. A majority of the responses indicated a concern around privacy, security, or the application collecting too much data. However approximately 31% responded "other" with additional feedback that they no longer used the app or that the app did have the functionality they were looking for.

Another important piece to understand was if the participants took an action to uninstall or not install an application once they realized how much personal information may be shared. Of the

participants, 77% stated they chose to not install an application after discovering how much personal information was being shared. From the same sample, 64.6% stated they chose to uninstall the application once they realized the amount of personal information was being shared. These large responses indicated that the participants were worried about security and privacy and they took an action after understanding the risks an application posed. However, the study also asked if the participants read the Terms of Use and only 35% responded that they have. This low response compared to the prior question indicate that either participants were informed of the risks through a different channel, possibly through general knowledge, another person informing them, or just a pop-up that asked permission for the application to access some content on their mobile device.

As mentioned earlier, 84.8% of the respondents chose to take an action of disabling location services on their phone to mitigate certain security and privacy concerns. The researchers assumed the majority of responses were related to security and privacy concerns but they asked two follow up questions to understand other reasons they may have done this and what applications they may have done this to. Some of the responses included extending battery life, they felt location services were not needed for the application, lack of trust and sharing too much data, and feeling of insecurity. Additionally, respondents stated they have turned location services off for applications in the categories of social media, banking, retail, weather, games, news, calendar, and photos. Given these results, it not only seems that users are taking general actions for protecting their privacy, but also that they have done so on specific applications that they felt impede on their security or privacy.

Lastly, the researchers wanted to understand if there existed a statistical significance among the three demographics (age, gender, and level of education) versus the actions taken to mitigate the security and privacy risks. Of the 8 scenarios, level of education did not have any statistical significance (a chi-square value of less than .05), while age had three and gender had two. For both Age and gender, the researchers found a statistical significance with clearing their browser / search history having chi-square values of .016 and .035, respectively. Using Anti-Malware software had a .028 chi-square value with age and a .002 chi-square value with gender. Additionally, age found another statistical significance with backing up the phone contents using a third party application while having a chi-square value of .05.

6. CONCLUSIONS

Mobile application can collect information from our mobile devices for a variety of reasons. While awareness is a key factor of ensuring that end users make informed decisions in order to stay safe while using their mobile devices, it is equally important to understand what actions these users take to protect their security and privacy. Using tools like anti-malware had a low response rate, participants illustrated that they were concerned about their security and privacy by their actions. Some had chosen to uninstall or not install an application once they learned of how much personal information would be shared. Others chose to keep the application but limit features like locations services to minimize the security and privacy risks. Given the low response rate for people who stated they read the terms of use, but the high response of some action being taken, it was clear that the participants were informed through another channel of the risks they pose. It was important to understand if users really cared about their security and privacy concerns and their actions certainly illustrate that they do. Since awareness is a key factor in protection, it would also be important to understand where they are getting their awareness from or if they are just generalizations about overall security.

7. REFERENCES

Boyles, J.L., Smith, A. and Madden, M. (2012). Privacy and data management on mobile devices. Retrieved 9/10/2016 from http://pewinternet.org/Reports/2012/Mobile-Privacy.aspx

Canalys. (2011) Smart phones overtake client PCs in 2011. Retrieved 6/11/2016 from http://www.canalys.com/newsroom/smart-phones-overtake-client-pcs-2011

Compomizzi, J. (2013). The influence of iPad technology on the academic and social experiences of veteran and military students: Academic preparation, collaboration socialization, and information access. ROBERT MORRIS UNIVERSITY.

Compomizzi, J., D'Aurora, S., & Rota, D. P. (2013) Identity theft and preventive measures: the cost is all yours. Issues in Information Systems. Vol. 14, Iss. 1, pp. 162-168.

Federal Trade Commission. (February, 2013). Mobile privacy disclosures: Building trust through transparency. Retrieved September 10, 2016 from www.ftc.gov/os/2013/02/130201mobileprivacyreport.pdf

Petsas, T., Papadogiannakis, A., Polychronakis, M., Markatos, E. P., & Karagiannis, T. (2013, October).

Rise of the planet of the apps: A systematic study of the mobile app ecosystem. In Proceedings of the 2013 conference on Internet measurement conference (pp. 277-290). ACM.

Salesforce. (2014). Mobile Behavior Report. Retrieved 6/9/2016 from http://www.marketingcloud.com/resource-center/digital-marketing/2014-mobile-behavior-report

Statistica, (2016). The Statistics Portal. Number of free and mobile app store downloads worldwide from 2011 to 2017 (in billions). Retrieved from www. Statistica.com/statistics/271644/worldwide-free-and-paid-mobile-app-store-downloads/

Theoharidou, M., Mylonas, A., & Gritzalis, D. (2012). A risk assessment method for smartphones. In Information security and privacy research (pp. 443-456). Springer Berlin Heidelberg.

Tongaonkar, A., Dai, S., Nucci, A., & Song, D. (2013, March). Understanding mobile app usage patterns using in-app advertisements. In Passive and Active Measurement (pp. 63-72). Springer Berlin Heidelberg.

Proposal for Kelly Criterion-Inspired Lossy Network Compression for Network Intrusion Applications

Sidney C. Smith
Sidney.c.smith24.civ@mail.mil
Computational Information Sciences Directorate
U.S. Army Research Laboratory
Aberdeen Proving Ground, MD 21005, U.S.A

Robert J. Hammell II
rhammell@towson.edu
Department of Computer and Information Sciences
Towson University

Towson, MD 21252, U.S.A

Abstract

This paper describes a proposal for a Kelly criterion inspired compression algorithm to be used in distributed network intrusion detection applications. Kelly's algorithm instructs a gambler how much to bet based upon the chance of winning and the potential payoff. There has been a significant amount of research into anomaly detection algorithms that will provide some indications of the maliciousness of a network session. We propose to combine expert knowledge, data mining, and best of breed anomaly detection algorithms to determine the likelihood that a session is malicious. Further, we propose using a Kelly criterion inspired algorithm to select which sessions and how much of each session to transmit. We expect that this will minimize the total amount of traffic we transmit while maximizing the amount of malicious traffic we transmit.

Keywords: lossy compression, network intrusion detection, Kelly criterion, anomaly detection

1. INTRODUCTION

Distributed Network Intrusion Detection Systems (NIDS) allow a relatively small number of highly trained analysts to monitor a much larger number of sites; however, they require information to be transmitted from the remote sensor to the central analysis system (CAS). Unless an expensive dedicated NIDS network is employed, this transmission must use the same channels that the site uses to conduct their daily business. This makes it important to reduce the amount of information transmitted back to the CAS to minimize the impact that the NIDS has on daily operations as much a practical.

One popular strategy for implementing a distributed NIDS is to do all of the intrusion detection on the sensor and send only alerts to the CAS. (Roesch, 1999) (Paxson, 1999) A second strategy might be to use lossless compression to reduce the size of the data returned to the CAS. A third strategy is to implement some form of lossy compression algorithm to send back relevant portions of traffic.

There are three problems with sending only alerts to the CAS. The first is that it has the potential to over burden the sensor's CPU and introduce packet loss. The impact of this packet loss has been discussed by Smith et al. (Smith, Hammell, Parker, & Marvel, 2016) (Smith & Hammell, An Experimental Exploration of the Impact of Sensor-Level Packet Loss on Network Intrusion Detection, 2015) (Smith, Wong, Hammell, & Mateo, 2015) The second problem is that the alerts by themselves often do not contain enough

information to determine whether the attack was successful. The third problem is that these systems are most often implemented with signature based intrusion detection engines. Signature based systems may be tuned to produce few false positives; however, they are ineffective at detecting zero-day and advanced persistent threats. (Kemmerer & Vigna, 2002)

Another alternative is to use lossless compression; however, one of the most widely used is deflation which is a variation of the LZ77 algorithm described by Ziv and Lempel. (Ziv & Lempel, 1997) Compressing the 2009 Cyber Defense Exercise dataset (Sangster, et al., 2009) with GNU Zip provides a ratio of 56.4%. Years of providing computer network defense services has taught us that to minimize the impact of NIDS on day-to-day operations, compression ratios of less than 10% are required. Lossless compression alone will not provide a reasonable solution.

The alternative that we will pursue is to use a lossy compression strategy to provide a solution. We may consider network traffic to be composed of sessions that span spectrums from known to unknown and malicious to benign as illustrated in Fig. 1. Quadrant III, the known malicious quadrant, is the domain of intrusion prevention systems as described by Ierace, Urrautia, and Bassett. (Ierace, Urrutia, & Bassett, 2005) We are most interested in quadrant II, the unknown malicious quadrant, because that is the quadrant where we will find evidence of zero-day and advanced persistent threat attacks. We assume that malicious traffic makes up a small amount of the actual traffic on the network. In 2004, Kerry Long described the Interrogator Intrusion Detection System Architecture. (Long K. S., 2004) In this architecture, remotely deployed sensors, known as Gators, collect network traffic and transmit a subset of the traffic to the analysis level. Interrogator employs "a dynamic network traffic selection algorithm called Snapper." (Long K. S., 2004) Long and Morgan describe how they used data mining to discover known benign traffic that they excluded from the data transmitted back to the analysis servers. (Long & Morgan, 2005)

In this research, we propose to combine expert knowledge, data mining, and best of breed anomaly based NIDS solutions to compute a maliciousness factor. We then propose to feed this maliciousness factor into a Kelly criterion (Kelly, 1956) inspired algorithm to compute the amount of traffic in each session that will be transmitted to the CAS. This should produce a lossy compression of the network traffic designed to reduce the amount of benign traffic and maximize the amount of malicious traffic being sent to the CAS.

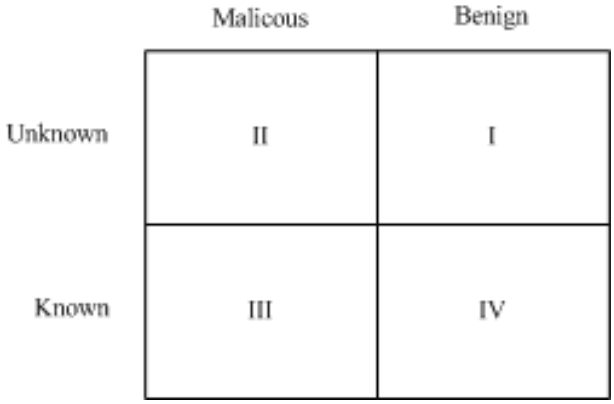

Figure 1 Network Traffic Composition

The remainder of this paper is organized into the following sections. Section 2 provides background. Section 3 will outline the approach chosen to address this problem. Section 4 will provided expected and preliminary results. Finally, Section 5 will conclude by restating the goals and approach of this research.

2. BACKGROUND

This research is broken down into to 2 basic questions: 1) How to rate the maliciousness of traffic and 2) How to use this rating to decide how much of each session to send back to the CAS. We will answer the first question by exploring expert knowledge, data mining and anomaly detection solutions. We will answer the second question by exploring the application of the Kelly criterion. We submit that the review of the literature presented demonstrates a wealth of knowledge in each of these areas that we hope to leverage for our maliciousness factor.

Session Rating

Data Mining

Lee and Stolfo used RIPPER (Cohen, 1995) on Tcpdump (Jacobson, Leres, & McCanne, 1989) data in their paper, "Data Mining Approaches for Intrusion Detection." (Lee & Stolfo, 1998) The dataset they used from the Information Exploration Shootout (Grinstein, Laskowski, Wills, & Rogowitz, 1997) contained only the header information for the network traffic and no user data. Lee and Stolofo cooked the network traffic down into records that look very much like Cisco NetFlow (Claise, 2004) records. Then they were able to feed this information in to RIPPER to generate rules. Their initial efforts were

unsuccessful; however, once they added a time window into their analysis they were able to achieve promising results. Since their data only contained Internet Protocol header information, and the positions of the exploits were not available to them, they were not able to assess the accuracy of their results.

While developing the Intelligent Intrusion Detection System at Mississippi State University, Bridges et al. integrated fuzzy logic, association rules, and frequency episodes data mining techniques to increase the flexibility of the system. (Luo, 1999) Genetic algorithms were employed to tune the membership functions of the fuzzy logic. (Bridges & Vaughn, 2000)

Dokas et al. addressed the problem of skewed class distribution in mining data for network intrusion detection. This problem exists because malicious activity compromises less than 2% of the network traffic. Their solution was to apply several boosting strategies to classification algorithms for rare classes as part of the Data mining in Minnesota Intrusion Detection System (MINDS). (Dokas, et al., 2002)

In the US Army Research Laboratory technical report, ARL-TR-4211 "Using Basic Data Mining Techniques to Improve the Efficiency of Intrusion Detection Analysis (Long & Morgan, 2005)**"**, Long and Morgan describe mining the Interrogator database to discover known benign traffic to be excluded from the traffic transmitted to the CAS. Their strategy was to exclude the most common day to day traffic flowing to and from the most popular trusted sites. (Long & Morgan, 2005)

Anomaly Based Network Intrusion Detection

In their history and overview of intrusion detection, Kemmerer and Vigna confirm a long standing belief that although anomaly detection techniques are capable of detecting unknown attacks, they pay for that capability with a high false positive rate. (Kemmerer & Vigna, 2002) In traditional NIDS, high false positive rates drain valuable time for the analysts.

In the computation of a maliciousness factor, false positives simply increase the amount of traffic transmitted. This is a cost to be considered; however, it is a much smaller price to pay than that paid by generating an alert for someone to analyze. This means that a significantly higher false positive rate can be tolerated in this application, making algorithms that would be unusable for detection attractive for rating the likelihood that traffic is malicious.

There has been a significant amount of work using anomaly detection in NIDS applications. Garcia-Teodoro et al. reviewed various types of anomaly-based detection techniques categorizing them as either statistics-based, knowledge-based, or machine-learning based. (Garcia-Teodoro, Diaz-Verdejo, Macia-Fernandez, & Vazquez, 2009)

In 1994 Mukherjee et al. provide a survey of intrusion detection technology titled, "Network Intrusion Detection." (Mukherjee, Heberlein, & Levitt, 1994) By today's standards the title is somewhat deceiving because almost all of the systems they surveyed are what would now be called host-based intrusion detection systems. These systems tend to examine the individual system's audit logs looking for intrusive activity. The notable exception is Network Security Monitor (NSM). NSM employs a System Description Language which is roughly modeled after a programming language and is used to describe the complex relationship which may be inferred from observable objects. These complex objects are analyzed using behavior-detection functions. NSM implements isolated object analysis and integrated object analysis. (Heberlein, et al., 1990) (Heberlein, Levitt, & Mukherjee, 1991) (Heberlein, Mukherjee, Levitt, Dias, & Mansur, 1991)

Sekar et al. describe their experiences with specification-based intrusion detection. They created behavioral monitoring specification language that they compiled into detection engines (Sekar & Uppuluri, Synthesizing Fast Intrusion Prevention/Detection Systems from High-Level Specifications, 1999) (Uppuluri & Sekar, 2001) (Sekar, et al., 2002)**,** validating their approach using the DARPA dataset. (Lippmann, et al., 2000)

Eskin et al. describe an unsupervised anomaly detection framework where network connections are mapped to a feature space and either cluster-based, k-nearest, or support vector machine-based algorithms are used to find anomalies in the sparse spaces. One of the key advantages to their approach is that it does not required labeled or known normal data to train the engine. (Eskin, Arnold, Prerau, Portnoy, & Stolfo, 2002)

Kruegel et al. developed a service specific anomaly detection engine. This engine contained a packet processing unit and a statistical processing unit. The packet processing unit pulled packets from the network and reassembled them into service requests. The statistical processing unit measured the type of request, length of request, and content of the request. It then computed values that ranged from 1 to 15 for each of these aspects, such that greater deviation translated into higher numbers. These

values were then combined to provide an anomaly score. This score was compared against a standard that the author suggested should be set, so that the system produces no more than 15 false positives a day. Because the deviation in type, length, and content varies significantly between services and even the types of requests, the statistical data must be partitioned by service and the length and content by type; however, the algorithms may be used without change by any service. Although the packet processing unit may need to be adjusted per service. (Krugel, Toth, & Kirda, 2002)

Ertoz et al. describe the MINDS. (Ertoz, et al., Detection and summarization of novel network attacks using data mining, 2003) (Chandola, Eilertson, Ertoz, Simon, & Kumar, 2007) (Ertoz, et al., Minds-minnesota intrusion detection system, 2004) MINDS uses Cisco NetFlow (Claise, 2004) data to collect statics for sixteen different features; half observed and half computed for each session. For each session the local outlier factor is computed. Sessions with features that contain very large local outlier factors are considered anomalous. These sessions then undergo associated pattern analysis which provides a summary of highly anomalous traffic for the security analyst. (Ertoz, et al., Detection and summarization of novel network attacks using data mining, 2003)

Munz et al. describe anomaly detection using K-means clustering. (Munz, Li, & Carle, 2007) Similar to Mukherjee et al. they separate the analysis for each service or port. Similar to Ertoz et al. they work with Cisco Netflow data. (Claise, 2004) Unlike the solutions mentioned above, this one requires both normal and attack training data to establish initial clusters. New traffic is then compared to the established clusters. (Munz, Li, & Carle, 2007)

Yassin et al. describe an approach which combines K-means clustering and naive Bayes classification called KMC+NBC. They were able to validate their algorithm against the ISCX 2012 Intrusion Detection Evaluation Dataset (Shiravi, Shiravi, Tavallaee, & Ghorbani, 2012) with strong positive results. (Yassin, Udzir, Muda, & Sulaiman, 2013)

In these references we can see a considerable amount of research has been using both data mining and anomaly detection to discover malicious network traffic. It is our intention of evaluate these techniques and use one or more to compute a maliciousness score for each session in the network traffic.

Session Selecting

In 1956 while working for Bell Telephone Laboratories, Kelly was developing a way to assign a value measure to a communication channel. He described a hypothetical illustration of a gambler who received advance notice about the outcome of an event through a communication channel with a non-negligible error rate. By doing this, Kelly was able to assign a cost value to the communication achieving his original goal. At the same time, he developed a formula based upon the probability of winning and the rate of pay off that would provide an amount to bet l that, if bet consistently over time would achieve and maintain greater wealth than any other value of l. We can see this in Eq. 1. where l is the fraction of wealth to bet, p is the probability of winning, and b is the net odds of the wager. (Kelly, 1956)

$$l = \frac{p(b+1)}{b} \quad (1)$$

Breiman uses the Kelly's work while discussing optimal gambling systems. (Breiman, 2012) He considers the problem of how much to bet on a series of biased coin tosses. To maximize returns on each toss one would bet their entire fortune; however, this will ultimately ensure ruin. In order to maximize winning and avoid ruin, some fixed fraction of wealth will be bet at each iteration. He uses Kelly's work to discover that fixed fraction. (Breiman, 2012)

Thorp first wrote about applying mathematical theory to the game of Black Jack in the 1960 paper, "Fortune's Formula: The Game of Blackjack." (Thorp E. O., Fortune's formula: The game of blackjack, 1960) Later Thorp published the book, *Beat the Dealer*, where he referred to what he called, "The Kelly Gambling System." (Thorp E. O., Beat the dealer, 1966) Although he mentions using the Kelly criterion as the optimal way to bet in his research for *Beat the Dealer* in his later work, (Thorp E. O., Understanding the Kelly Criterion, 2012) he mentions it only once in passing in this book. (Thorp E. O., Beat the dealer, 1966) The bulk of this book discusses the rules of Blackjack and methods to determine when one has an advantage over the dealer and how great that advantage might be. The Kelly criterion would be used to calculate how large of a bet to place based upon the size of the advantage. Instead of directly using the Kelly criterion, he talks about placing big bets and little bets. (Thorp E. O., Beat the dealer, 1966)

In his paper "Understanding the Kelly Criterion", Thorp mentions the application of the Kelly

criterion to the stock market and his previous book *Beat the Market* (Thorp E. O., Understanding the Kelly Criterion, 2012); however, the Kelly criterion is not mentioned at all in *Beat the Market*. Instead Thorp concentrates on how the market works, what short selling and warrants are all about, and how to determine the relative value of a stock or a warrant. (Thorp & Kassouf, Beat the Market: A Scientific Stock Market System, 1967) Thorp goes into greater detail about how the Kelly criterion would be used in Blackjack and the stock market in "Optimal Gambling Systems for Favorable Games." (Thorp E. O., Optimal gambling systems for favorable games, 1969) Thorp goes into even greater detail in his later work, "The Kelly Criterion in Blackjack, Sports Betting, and the Stock Market" where he graphically illustrates how the log for wealth is maximized to maximize the growth of wealth over time. (Thorp E. O., 1998) He specifically applies the criterion to the stock market in "The Kelly Criterion and the Stock Market." (Rotando, 1992) Studying Thorp's works, it appears that although having a formula to calculate the optimum bet is useful, clearly understanding the game is far more important.

Nekrasov created a formula for implementing the Kelly criterion in multivariate portfolios as seen in Eq. 2. Consider a market with n correlated stocks S_k with stochastic return r_k and a riskless bond with return r. An investor puts a fraction u_k of his capital in S_k and the rest is invested in bonds. The following formula may be used to compute the optimum investments where $\hat{\bar{r}}$ and $\hat{\Sigma}$ are the vector of the means and the matrix of 2nd mixed noncentral moments of the excess returns. (Nekrasov, 2014)

$$\overrightarrow{u^*} = (1+r)\left(\hat{\Sigma}\right)^{-1}\left(\hat{\bar{r}} - r\right) \qquad (2)$$

The interest of Thorp and others in the Kelly criterion indicate its usefulness is selected out much of the available resources to invest. Nekrasov's work extends this across multiple options in a collection that might resemble sessions in network traffic. Although the differences between our specific requirements are different enough from the requirement of those cited and we will need to start from first principles to create our Kelly criterion inspired formula, their work is close enough to demonstrate the feasibility of our approach.

3. APPROACH

This research effort breaks down into 2 research questions and 2 phases. The first question, which will be addressed in phase 1, is how to know what traffic is most likely to contain malicious activity. The second question, which will be addressed in phase 2, is how to select the traffic most likely to contain malicious activity for transmission to the analysis servers.

Phase 1

In phase 1, we plan to combine expert knowledge, data mining, and best of breed intrusion detection in order to compute a maliciousness rating. The first step of this phase will be to discover the relevant facts that may be gleaned from expert knowledge (e.g. when the Heart Bleed vulnerability was discovered, an expert could have caused the system to rate secure socket layer traffic higher; and when a known malicious internet protocol address or domain is discovered, an expert could cause the system to rate traffic including that IP or domain higher.) The second step of this phase will be to discover the relevant facts that may be mined from the Interrogator data store (e.g. Long and Morgan mined Interrogator to develop a white list of web servers to be excluded and instances of new servers to be included. (Long & Morgan, 2005) This could be expanded to rate traffic more malicious which contains addresses and ports associated with alerts or incidents.) The third step of this phase will combine best of breed anomaly detection algorithms to form a maliciousness rating (e.g. MINDS collected, computed, and assigned a local outlier factor to 16 different features (Chandola, Eilertson, Ertoz, Simon, & Kumar, 2007) (Ertoz, et al., Detection and summarization of novel network attacks using data mining, 2003) (Ertoz, et al., Minds-minnesota intrusion detection system, 2004) KMC+NBC uses K-Means clustering and Naïve Bayes Classification to detect anomalies in network traffic. (Yassin, Udzir, Muda, & Sulaiman, 2013) Again a measure of abnormality could factor into the session rating. The fourth step of this phase will be to develop a formula to combine all of these into a single score. Phase 1 corresponds to the top half of Fig. 2 where unrated sessions are captured by the sensor and flow into the session rater which uses expert knowledge, mined data, and anomaly algorithms to rate each session. The green sessions are known benign, the red sessions are known malicious, and the other colors are meant to represent the continuum in between.

Phase 2

In phase 2, we plan to develop a Kelly criterion based formula that takes the scores generated from phase 1 as input and produces as output a fraction of the available network traffic that

should be invested in each session. Kelly proved that there exists an amount to bet *l* being some portion of the total wealth *G*, that if the gambler bets it consistently, *G* will obtain and maintain a level greater than any other possible value for *l*. (Kelly, 1956) This may be seen in Eq. 1 where *l* is the fraction of wealth to bet, *p* is the probability of winning, and *b* is the net odds of the wager. Thorp applied the Kelly criterion to the game of blackjack. (Thorp E. O., Beat the dealer, 1966) Smoczynski and Tomkins applied the Kelly criterion to horse racing. (Smoczynski & Tomkins, 2010) Separately Thorp and Nekrasov applied the Kelly Criterion to the stock market. (Thorp & Kassouf, Beat the Market: A Scientific Stock Market System, 1967) (Nekrasov, 2014) Using this generalization, one would consider network flows to be stocks and rate of return to be the maliciousness score of the session. Phase 2 corresponds to the bottom half of Fig. 2 where the rated sessions flow into the algorithm and the session selector feeds those ratings into the Kelly criterion (Kelly, 1956) inspired formula to determine how much traffic to invest in each session. The fatter sessions represent more traffic being invested in the session and the skinnier sessions represent less traffic being invested in the session.

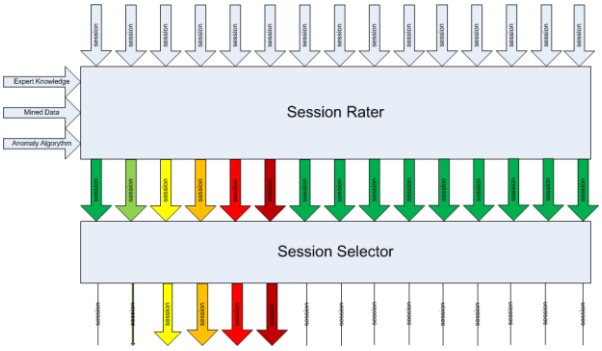

Figure 2 Kelly Compressor Diagram

We will use Nekrasov's formula in Eq. 2 to illustrate how this might work. To apply this to our problem we will substitute the returns for the maliciousness score and the investment for the amount of available traffic to assign to each session. Since a riskless bond makes no sense in our problem, we will set the value to zero simplifying the equation shown in Eq. 3. This leaves us with only one variable because the 2nd noncentral moment is a function of the maliciousness rating over time. Remember it is unlikely that Nekrasov's formula will work as given. This is because of the fixed nature of our investments. Correctly selecting malicious sessions does not increase the bandwidth available, and incorrectly selecting benign session does not decrease the bandwidth available. Further, there is no chance of ruin. We need to start from the same starting point that Kelly did to retrace his steps to construct a formula for this specific application.

$$\overrightarrow{u^*} = (\hat{\Sigma})^{-1}(\hat{r}) \quad (3)$$

Once the session rater and session selector algorithms are developed, they will be incorporated into a prototype which will be tested against open sources datasets to include those used by Smith et al. in their theoretical exploration. (Smith, Hammell, Parker, & Marvel, 2016)

4. RESULTS

Many of the data mining and anomaly detection techniques have settings that will increase the sensitivity creating more false positives and fewer false negatives or decrease the sensitivity creating fewer false positives and more false negatives. As we complete our research, we expect to tune these settings until we get the appropriate amount of compression and an acceptable level of false negatives. We will illustrate this by applying entropy to remove compressed and encrypted data.

As we interviewed experts in network intrusion detection, we discovered that there is very little value in transmitting encrypted or compressed data back to the CAS. Encrypted data is not very valuable because decrypting it is prohibitively expensive and beyond the capabilities of most network defense analysts. Compressed data is of little value because it is very difficult to decompress the file unless every packet of the session containing the compressed file is available. Network file carving is more efficiently done on the sensor and a cryptographic hash is sufficient for most network intrusion detection applications. The entropy of data may be used to detect if data are encrypted or compressed because this data has a much higher entropy than clear text data (Shannon, 2001).

We can illustrate the kinds of results that we expect to obtain by conducting an experiment where we drop packets with entropy values greater than a given threshold and pass the abridged data to Snort (Roesch, 1999) for analysis. We repeated this process lowering the entropy values from 7.9 to 4.0 in increments of 0.1. Fig. 3 plots the size of the datasets for each iteration and the alert loss rate for each iteration. Notice that at an entropy value of 7.0 the data

has been compressed to 27% of its original size, but has only lost 0.6% of the alerts.

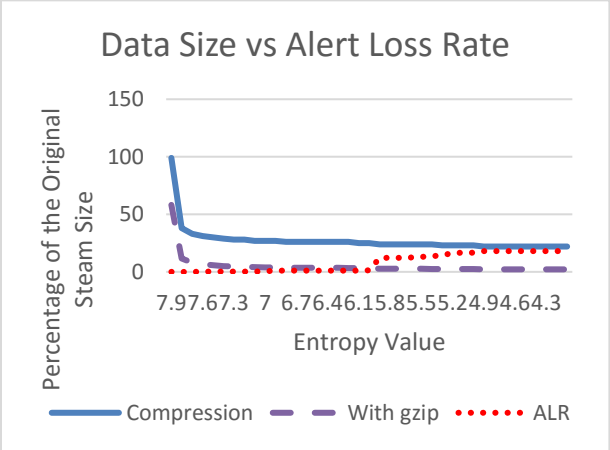

Figure 3 Lossy Compression Using Entropy

An interesting property of low entropy data is that it compresses very well. Applying GNU Zip lossless compression to the dataset that has been compressed using the entropy based lossy compression we get a file that is 4% of the original size of the dataset which is well within our bandwidth budget of 10%. These results are anecdotal and certainly may not be typical, but they do illustrate the feasibility of the approach.

5. CONCLUSIONS

In a distributed NIDS environment, it is necessary to transmit the right data back to the central analysis servers to provide analysts with the information necessary to detect and report malicious activity. Bringing back all of the data would double the bandwidth requirements of the site and require that the analysis servers have massive bandwidth available to receive it all. Standard lossless compression is not sufficient to reduce this traffic to an acceptable level. The goal of this research is to develop a lossy compression algorithm that will ensure that the traffic lost is the least likely to contain malicious activity. The approach is to use an algorithm based upon the Kelly criterion to allocate the limited bandwidth available, coupled with best of breed anomaly detection, to assess the maliciousness of the traffic. These two technologies will be combined into a packet capture tool which will produce data compliant with the standards used by existing NIDS tools. Preliminary results show a compression ratio of 96%. Although these results were obtain from a dataset that is unlikely to reflect real word traffic, they demonstrate the feasibility of the approach.

6. REFERENCES

Breiman, L. (2012). Optimal gambling systems for favorable games. In *The Kelly Captial Growth Investment Criterion: Theory and Practice* (pp. 47-60). New Jersey NJ: World Scientific.

Bridges, S. M., & Vaughn, R. B. (2000). Fuzzy data mining and genetic algorithms applied to intrusion detection. *Proceedings of the 12th Annual Canadian Information Technology Security Symposium.* Ottawa, Canada.

Chandola, V., Eilertson, E., Ertoz, L., Simon, G., & Kumar, V. (2007). MINDS: Architecture & Design. In *Data Warehousing and Data Mining Techniques for Cyber Security* (pp. 83-108). New York, NY: Springer.

Claise, B. (2004). *Cisco Systems NetFlow Services Export Version 9.* Fremont, California, United States: Internet Engineering Task Force (IETF).

Cohen, W. W. (1995). Fast Effective Rule Induction. *Proceedings of the Twelfth International Conference on Machine Learning* (pp. 115-123). Lake Taho, CA: Morgan Kaufman.

Dokas, P., Ertoz, L., Kumar, V., Lazarevic, A., Srivastava, J., & Tan, P.-N. (2002). Data mining for network intrusion detection. *Proc. NSF Workshop on Next Generation Data Mining.* Baltimore, Maryland, United States.

Ertoz, L., Eilertson, E., Lazarevic, A., Tan, P., Dokas, P., Srivastava, J., & Kumar, V. (2003). *Detection and summarization of novel network attacks using data mining.* Minneapolis, Minnesota: Army High Performance Computing Research Center.

Ertoz, L., Eilertson, E., Lazarevic, A., Tan, P.-N., Kumar, V., Srivastava, J., & Dokas, P. (2004). Minds-minnesota intrusion detection system. In *Next Generation Data Mining* (pp. 199-218). Cambridge, MA: MIT Press.

Eskin, E., Arnold, A., Prerau, M., Portnoy, L., & Stolfo, S. (2002). A geometric framework for unsupervised anomaly detection. In *Applications of data mining in computer security* (pp. 77-101). New York, NY: Springer.

Garcia-Teodoro, P., Diaz-Verdejo, J., Macia-Fernandez, G., & Vazquez, E. (2009). Anomaly-based network intrusion detection: Techniques, systems and challenges. *Computer & Security, 28*(1), 18-28.

Grinstein, G., Laskowski, S., Wills, G., & Rogowitz, B. (1997). Information Exploration Shootout Project and Benchmark Data Sets (Panel): Evaluating How Visualization Does in Analyzing Real-world Data Analysis Problems. *Proceedings of the 8th Conference on Visualization '97.* Los Alamitos, CA, USA.

Heberlein, L. T., Dias, G. V., Levitt, K. N., Mukherjee, B., Wood, J., & Wolber, D. (1990). A network security monitor. *Research in Security and Privacy, 1990. Proceedings., 1990 IEEE Computer Society Symposium on.* Oakland, CA.

Heberlein, L. T., Mukherjee, B., Levitt, K., Dias, G., & Mansur, D. (1991). *Towards detecting intrusions in a networked environment.* Davis, CA: U. of Calif., Davis.

Heberlein, L., Levitt, K., & Mukherjee, B. (1991). A method to detect intrusive activity in a networked environment. *Proceedings of the 14th National Computer Security Conference.* Washington, DC, USA.

Ierace, N., Urrutia, C., & Bassett, R. (2005). Intrusion Prevention Systems. *Ubiquity*, 1530-2180.

Jacobson, V., Leres, C., & McCanne, S. (1989). The tcpdump manual page. *Berkley (CA): Lawrence Berkley Laboratory*.

Kelly, J. L. (1956). A New Interpretation of Information Rate. *IRE Transactions on Information Theory, 2*(3), 185-189.

Kemmerer, R. A., & Vigna, G. (2002). Intrusion detection: A brief history and overview (supplement to computer magazine). *Computer*, 27-30.

Krugel, C., Toth, T., & Kirda, E. (2002). Service specific anomaly detection for network intrusion detection. *Proceedings of the 2002 ACM symposium on Applied computing.* Madrid, Spain.

Lee, W., & Stolfo, S. J. (1998). Data Mining Approaches for Intrusion Detection. *Proceedings of the 7th USENIX Security Symposium.* San Antonio, TX.

Lippmann, R., Fried, D., Graf, I., Haines, J., Kendall, K., McClung, D., . . . Zissman, M. (2000). Evaluating intrusion detection systems: the 1998 DARPA off-line intrusion detection evaluation. *DARPA Information Survivability Conference and Exposition* (pp. 12-26). DISCEX '00.

Long, K. S. (2004). *Cathing the Cyber Spy: ARL's Interrogator.* Army Research Laboratory, Computational Information Systems Directory. Aberdeen Proving Ground: DTIC.

Long, K. S., & Morgan, J. B. (2005). *Using Data Mining to Improve the Efficiency of Intrusion Detection Analsysis.* Aberdeen Proving Ground (MD): Army Research Laboratory (US).

Luo, J. (1999). *Integrating fuzzy logic with data mining methods for intrusion detection.* Starkville, Mississippi, United States: Mississippi State University.

Mukherjee, B., Heberlein, L. T., & Levitt, K. N. (1994). Network Intrusion Detection. *IEEE Network, 8*(3), 26-41.

Munz, G., Li, S., & Carle, G. (2007). Traffic anomaly detection using k-means clustering. *GI/ITG Workshop MMBnet.* Hamburg, Germany.

Nekrasov, V. (2014). *Kelly Criterion for Multivariate Portfolios: A Model-Free Approach.* Rochester, NY: Social Science Research Network.

Paxson, V. (1999). Bro: a System for Detecting Network Intruders in Real-time. *Computer Networks, 31*(23), 2435-2463.

Roesch, M. (1999). Snort: Lightweight Intrustion Detection for Networks. *Proceedings of the 13th System Administration Conference LISA '99. 99*, pp. 7-12. Seattle WA, US: USENIX.

Rotando, L. M. (1992). The Kelly criterion and the stock market. *American Mathematical Monthly, 99*(10), 922-931.

Sangster, B., O'Connor, T., Cook, T., Fanelli, R., Dean, E., Adams, W. J., . . . Conti, G. (2009). Toward instrumenting network warfare competitions to generate labeled datasets.

Proc. of the 2nd Workshop on Cyber Security Experimentation and Test (CSET09). Montreal, Canada.

Sekar, R., & Uppuluri, P. (1999). Synthesizing Fast Intrusion Prevention/Detection Systems from High-Level Specifications. *Proceedings of the 8th USENIX Security Symposium.* Washington, DC.

Sekar, R., Gupta, A., Frullo, J., Shanbhag, T., Tiwari, A., Yang, H., & Zhou, S. (2002). Specification-based anomaly detection: a new approach for detecting network intrusions. *CCS '02: Proceedings of the 9th ACM Conference on Computer and Communications Security.* Washington, DC, USA.

Shannon, C. E. (2001). A Mathematical Theory of Communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 3-55.

Shiravi, A., Shiravi, H., Tavallaee, M., & Ghorbani, A. A. (2012). Toward developing a systematic approach to generate benchmark datasets for intrusion detection. *Computer & Security, 31*(3), 357-374.

Smith, S. C., & Hammell, R. J. (2015). *An Experimental Exploration of the Impact of Sensor-Level Packet Loss on Network Intrusion Detection.* Army Research Laboroatory (US). Aberdeen Proving Ground: Army Research Laboroatory (US).

Smith, S. C., Hammell, R. J., Parker, T. W., & Marvel, L. M. (2016). A Theoretical Exploration of the Impact of Packet Loss on Network Intrusion Detection. *International Journal of Networked and Distributed Computing*, 1-10.

Smith, S. C., Wong, K. W., Hammell, R. J., & Mateo, C. J. (2015). *An Experimental Exploration of the Impact of Network-Level Packet Loss on Network Intrusion Detection.* Aberdeen Proving Ground: Army Research Laboratory (US).

Smoczynski, P., & Tomkins, D. (2010). AN EXPLICIT SOLUTION TO THE PROBLEM OF OPTIMIZING THE ALLOCATIONS OF A BETTOR'S WEALTH WHEN WAGERING ON HORSE RACES. *Mathematical Scientist, 35*(1).

Thorp, E. O. (1960). Fortune's formula: The game of blackjack. *Notices of the American Mathematical Society, 7*(7), 935-936.

Thorp, E. O. (1966). *Beat the dealer.* New York, NY: Random House.

Thorp, E. O. (1969). Optimal gambling systems for favorable games. *Revue de l'Institut International de Statistique, 37*(3), 273-293.

Thorp, E. O. (1998). The Kelly Criterion in Blackjack, Sports Betting, and the Stock Market. *Finding the Edge: Mathematical Analysis of Casino Games, 1*(6).

Thorp, E. O. (2012). Understanding the Kelly Criterion. In *The Kelly Capital Growth Investment Criterion: Theory and Practice* (pp. 511-525). New Jersey, NJ: World Scientific.

Thorp, E. O., & Kassouf, S. T. (1967). *Beat the Market: A Scientific Stock Market System.* New York, NY: Random House.

Uppuluri, P., & Sekar, R. (2001). Experiences with specification-based intrusion detection. *Recent Advances in Intrusion Detection.* Davis, CA.

Yassin, W., Udzir, N. I., Muda, Z., & Sulaiman, M. N. (2013). Anomaly-Based Intrusion Detection through K-Means Clustering and Naives Bayes Classification. *Proceedings of the 4th International Conference on Computing and Informatics (ICOCI).* Sarawk, Malaysia.

Ziv, J., & Lempel, A. (1997). A Universal Algorithm for Sequential Data compression. *IEEE Transactions on Information Theory*, 337-343.