Association for Information Systems
AIS Electronic Library (AISeL)

ICIS 2010 Proceedings

International Conference on Information Systems (ICIS)

2010

Is It Spam or Ham? Testing Asynchronous CMC Deception Detection Theory

Fletcher H. Glancy Ph.D.
Lindenwood University, fglancy@lindenwood.edu

Surya B. Yadav Ph.D.
Texas Tech University, Surya.Yadav@ttu.edu

Is It Spam or Ham? Testing Asynchronous CMC Deception Detection Theory

Research-in-Progress

Fletcher H. Glancy, Ph.D.
School of Business & Entrepreneurship
Lindenwood University
209 S. Kingshighway
St. Charles, MO 63301
E-Mail: fglancy@lindenwood.edu

Surya B. Yadav, Ph.D.
Department of Information Systems and Quantitative Sciences
Rawls College of Business, Texas Tech University
Lubbock, Texas 79409
E-Mail: Surya.Yadav@ttu.edu

Abstract

This paper presents two empirical evaluations of Asynchronous CMC Deception Detection Theory (ACDDT) in the CMC genres of legitimate and spam e-mail. The ACDDT theorizes that deception detection in asynchronous text is indicated by the presence of the constructs of concealment, normality, and isolation. The ACDDT was tested on two common types of e-mail that attempt to deceive the reader/recipient, e-mail that attempted to steal the recipient's personal identification information and e-mail that attempted to steal the recipient's money. The e-mail was clustered using the text mining. The terms representing the clusters were analyzed using text analysis. The constructs developed from the text analysis provided measures of the hypotheses. Both tests confirmed that ACDDT provides a theoretical base for detection of deception in e-mail. The contributions of this work are support for the ACDDT and a development of a new and unique method of detecting fraudulent e-mail.

.

Keywords: Deception detection, Concealment, Isolation, Normality, Spam, E-mail, Asynchronous CMC, ACDDT

Introduction

The paper reports an ongoing work that empirically tests a new theory to detect deception in asynchronous Computer-Mediated Communication (CMC). The Asynchronous CMC Deception Detection Theory (ACDDT) (Glancy, 2010) proposes that concealment, normality, and isolation play a direct role in detecting deception in text documents. The ACDDT is unique because it was developed specifically for asynchronous text. This paper presents the empirical testing of ACDDT in the area of spam. Spam is unsolicited email (Androutsopoulos et al., 2000; Clark et al., 2003; Cormack & Kotcz, 2009). MessageLabs (2010) estimated that spam in January 2010 comprised 83.9% of all e-mail. Spam is estimated to cost e-mail users $130 billion annually (Jennings, 2009). This cost continues to rise (Internet Crime Complaint Center, 2009). The cost of spam is from the detection and maintenance of detection filters, the clogging of email storage, time users spend reading spam, and the direct cost to users that fall victim of the spam. This research addresses fraudulent e-mail, specifically identity theft and money theft. It has been estimated that these types of fraud cost users $2.6 billion in 2008 (Jennings, 2009; Synovate, 2007).

There is a large segment of research that concentrates on methods to distinguish spam from not spam, which is also called ham. Spam is often sent from programs that send out the emails to large lists of email addresses (Rigoutsos & Huynh, 2004). Our survey of existing spam filtering research indicated that spam-filtering methods are mechanical; the filters are constructed from existing and identified spam. New e-mail is categorized as legitimate or spam by comparison to the filter. Spam filters are not stable and require frequent rebuilding. An example (Rios & Zha, 2004) is that a filter may recognize the word Viagra, but not vi*gra or v!agra or <Viagr*. When the filter depends on a database for comparison or grading, the filter is always chasing the spammer.

In this work we empirically evaluated the Asynchronous CMC Deception Detection Theory (ACDDT) in two tests on spam. The research question for this work was: Does the ACDDT provide a theoretical base for detection of deception in e-mail? A method of filtering spam with a theoretical base for spam detection could potentially lead to the development of a unique method of spam detection. This work is organized as follows: Section 2 gives background on spam, spam detection, and the ACDDT; Section 3 discusses operationalization of the ACDDT on email; Section 4 discusses the testing and results; Section 5 discusses the research contribution; and Section 6 presents conclusions and future work.

Section 2 - Background

In this section, we briefly review spam, spam detection methods, and the ACDDT.

Spam

Spam is most general defined as unsolicited email (Clark, et al., 2003; Cormack & Kotcz, 2009; Cormack & Lynam, 2007). The purpose of most spam is to commit deception (MessageLabs, 2010); deception is defined as the intentional misleading of another such that they draw an inaccurate conclusion (Buller & Burgoon, 1996; DePaulo, et al., 2003). When the deception violates the law, it becomes fraud. The top ten types of Internet fraud reported by the Internet Crime Complaint Center (IC3) (Internet Crime Complaint Center, 2009) are: non-delivery, auction fraud, credit fraud, confidence fraud, computer fraud, check fraud, Nigerian letter fraud, identity theft, financial institution fraud, and threats/cyber stalking. E-mail was identified as the source of 74% of the Internet fraud in 2008. The direct loss by the individual from Internet fraud reported to the IC3 in 2008 was $264.6 million. The estimated cost of spam in 2009 was $130 billion worldwide and $42 billion in the United States (Jennings, 2009). The loss of productivity is estimated to be 85% of the total cost of spam. This includes the time spent deleting spam, looking for e-mail that has been mistakenly identified as spam, and reading spam. An estimated 62 trillion spam e-mails were sent in 2008 (ICF International, 2009). From this, we conclude that spam is a major problem for the Internet and users, and spam is a global problem.

Spam Detection

A spam filter is a program that attempts to identify and segregate spam from legitimate email. These programs can either stand-alone such as the open source SpamAssassin (Apache Software Foundation, 2010) or be incorporated as

part of an e-mail program as with Google's Gmail. Spam filters use a variety of classification methods. The earliest filters used heuristics and keywords (Hidalgo et al., 2000). Later classification methods used statistical methods to extract a set of features that are compared to a new email and used to decide if the email is spam or not (Hershkop & Stolfo, 2005; Zhang et al., 2004). These methods include naive Bayesian (Androutsopoulos, et al., 2000; Schneider, 2003), N-gram (Zhang, et al., 2004), support vector machines and random forests (Rios & Zha, 2004), and neural network (Clark, et al., 2003). The filter is created from a database of spam and legitimate e-mail. The filter may look at some or all of the features of the e-mail; these features include the sender's name, IP address, subject, e-mail body text (with or without punctuation and special characters), and images (Aradhye, Myers, & Herson, 2005). The design of the filter may rely on one method (Goodman & Yih, 2006) or combine methods into a single filter (Hershkop & Stolfo, 2005; Lynam & Cormack, 2006). Because these methods create a filter from a set of spam e-mails, as the spammers change methods the filter needs updating (Delany et al., 2005). Most of the filters are binomial. When the new email is compared to the filter, the answer is either spam or legitimate. The spam is either rejected or sent to the recipient's spam folder and the legitimate e-mail to their inbox.

It is generally believed that the cost of spam is low compared to the value of legitimate e-mail (Androutsopoulos, et al., 2000). A common measure of the utility of a spam filter is percentage of false positives; a false positive is an e-mail that is legitimate and has been mistakenly identified as spam (Yih, et al., 2006). Spam filters have false positive rates less than 0.1 percent. Maintaining low false positive rates allows spam to reach the user at rates as high as five percent. This means that of the 62 trillion spam e-mails, 3.1 trillion spam e-mails get past the filter. We tested the Asynchronous CMC Deception Detection Theory on this 5% of spam that got past a spam filter and were sent to the user as legitimate e-mail.

Asynchronous CMC Deception Detection Theory

The ACDDT has been developed as part of the coauthor's dissertation research (Glancy, 2010). The ACDDT asserts that deception can be detected in asynchronous text. Most extant research on deception has been in face-to-face communication with a concentration on the cues a deceiver gives that reveal the deception (DePaulo et al., 2003). The cues that were identified were verbal (both textual and oral), behavioral, and physical. The ACDDT (Glancy, 2010) is a theory of deception detection in asynchronous textual communication, which has limited types of deceptive cues (Lee et al., 2009). The first ACDDT concept is Deception Detection. Deception is the deliberate misleading of another so that they draw an incorrect conclusion (Buller and Burgoon, 1996, Carlson et al., 2004, DePaulo et al., 1996, DePaulo et al., 2003). Because deception is deliberate, the deceiver knows what s/he is doing, and does it with the intent to mislead. A deception can be through concealing facts, obfuscation of facts, creating false information, and through combinations of these (DePaulo et al., 2003). Detection is the identification of that what is not clearly seen. Deception detection occurs when a receiver of the deceptive message realizes that the message is deliberately misleading. The ACDDT is based on the assumption that deception in asynchronous communication is detectable because the writer knows or suspects that the communication he/she is writing is false, and it is written to deceive the reader. This knowledge causes both conscious and subconscious cognitive, affective and conative reactions in the writer (Bagozzi, 1992). The writer reveals these reactions in the text of the communication. The detectable concepts in the text of deceptive writing are concealment, normality and isolation. Concealment is the manifestation of the action oriented or conative concept. Normality is the manifestation of cognitive concept, and the Isolation concept is the affective reaction to the deception. Because the deceptive writing has these concepts present in the text, the deception can be detected. The first overall ACDDT proposition is that concealment, normality, and isolation are positively related to the detection of deception (Glancy, 2010). Table 1 lists the concepts, their constructs, and hypotheses.

Table 1. ACDDT Concepts, Constructs, and Hypotheses			
Concepts	**Propositions**	**Constructs (a + means the construct-presence increases with deception and a - means the construct-presence decreases with deception)**	**Hypotheses**
Concealment	Concealment is positively related to Deception	Collectives +; Ambivalence -; Denial -	H1. The presence of Collectives, Ambivalence, and Denial is positively related to deception detection.
Normality	Normality is positively related to Deception	Leveling +; Diversity -; Centrality -	H2. The presence of Leveling, Diversity, and Centrality is positively related to deception detection.
Isolation	Isolation is positively related to Deception	Aggression +; Exclusion +; Liberation +; Rapport -; Cooperation -; Motion -	H3. The presence of Aggression, Exclusion, Liberation, Rapport, Cooperation, and Motion is positively related to deception detection.

The second ACDDT proposition is that concealment is positively related to deception. The concealment could be from misstatement of facts or from omission of information. The constructs of increased Collectives, less Ambivalence, and less Denial contribute to the concept of concealment. Collectives are singular nouns that connote plurality and function to decrease specificity. These words include social groupings such as crowd and team, task groups such as congress and staff, and geographical entities such county and world. Ambivalence is indicated by words expressing hesitation or uncertainty; included in this group are words such as perhaps, might, allegedly, approximate, would, and almost. They imply the writer's reluctance to commit to what is being written. Denial is the use of terms that express negativity such as nor, not, nothing, none, and nobody. We hypothesize that the presence of collectives, ambivalence, and denial in a text document is positively related to the detection of deception.

The third ACDDT proposition is that normality is positively related to deception. The concept of normality is related to the writer's desire to give an impression that there is no deception occurring. This attempt whether conscious or unconscious causes the writer to use terms that express normality differently than would occur in non-deceptive writing. The constructs of increased Leveling, less Diversity, and less Centrality contribute to the expression of normality in asynchronous text. Leveling terms are words used to ignore differences and create a sense of completeness. These terms include words such as always, completely, everyone, each, unconditional, and absolute. Diversity is expressed through the use of words that identify individuals or groups that differ from the norm. Examples of these words include inconsistent, exceptional, illegitimate, dispersed, diffuse, deviant, and rare. Centrality refers to terms that denote regularities and/or agreement on core values. This includes terms such as native, basic, orthodox, ritualistic, standardized, conformity, unanimous, and reliable. We hypothesize that the presence of leveling, diversity, and centrality in a text document is positively related to deception detection.

The fourth ACDDT proposition is that isolation is positively related to deception. The concept of isolation is related to the writer's affective isolation from the norms of society by committing deception. Because the writer of deceptive text had some level of awareness of the deception, the writer felt that he was socially isolated. This isolation is expressed in the constructs of increased Aggression, increased Exclusion, increased Liberation, less Cooperation, less Rapport, and less Motion. Aggression is expressed in words that indicate competition and forceful action. Examples of these terms include crash, collide, conquest, violation, command, challenge, mastered, pushy, poke, shove, demolish, prevent, reduce, and defend. Exclusion terms express the sources and effects of social isolation and include words such as displaced, repudiated, secede, privacy, forsakes, loneliness, nihilism, and spurn. Liberation is expressed in words that maximize individual choice and reject social conventions. Examples of these terms include autonomous, options, radical, eccentric, liberty, freedom, disentangle, and loosen. Cooperation is expressed in terms that refer to group behavioral interactions. Examples of these words include unions, caucus, partner, comrade, friendship, teamwork, contribute, and network. Rapport is expressed in words that indicate

attitudinal similarities among groups of people. These terms include congenial, companion, approve, tolerant, permission, and consensus. Motion is expressed by words that connote human movement such as job, leap, circulate, revolve, travel, ride, fly, zip, and lurch. We hypothesize that the presence of aggression, exclusion, liberation, cooperation, rapport, and motion in a text document is positively related to deception detection.

Section 3 - ACDDT Evaluation Process

A systematic process to evaluate ACDDT is shown in Figure 1. The ACDDT propositions are operationalized by formulation of specific and testable hypotheses for each test. A set of measures is developed. Domain specific text documents are collected, matched one-to-one, and a data set is created. The text documents are appropriately formatted before they are imported into the SAS Enterprise Miner software. The data sets are prepared for text mining. The text mining is performed, single value decomposition (SVD) (Albright, 2004) of the term-document matrix is generated, and the SVDs are clustered(Roiger and Geatz, 2003; de Ville, 2006). Terms representing the clusters are determined by binomial probability. The text documents were analyzed using a computational model for fraud detection (CFDM) (Glancy and Yadav, 2010). The CFDM supports the text mining steps of ACDDT evaluation process shown in figure 1. The top100 terms representing a cluster are analyzed using Diction 5.0 text analysis program. The text analysis step associates terms with constructs. Construct scores are computed and data analysis is performed to compute various statistics. The test results are compared to the hypothesis.

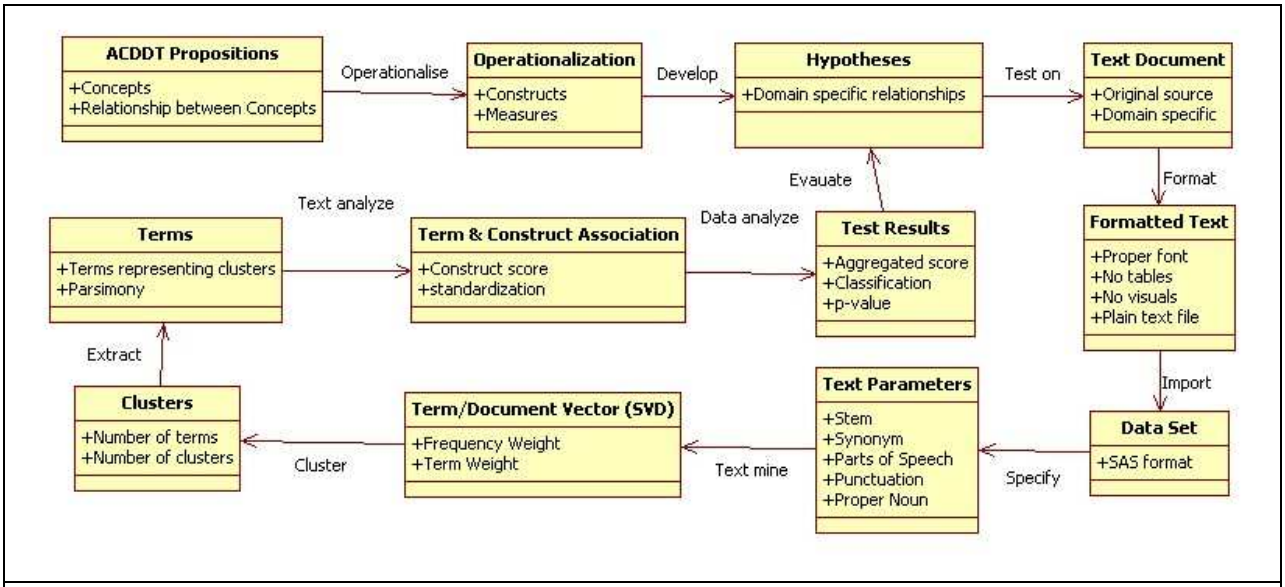

Figure 1. Theory Evaluation Process

Two separate and independent tests of the ACDDT were conducted on spam e-mail. Testing in these genres requires operationalization of the ACDDT's propositions and general hypotheses stated in table 1 and the development of measures.

Construct Measures

The ACDDT propositions relate the independent concepts of concealment, normality, and isolation to the concept of deception. These propositions must be operationalized so that they can be measured; they are indicated by the twelve constructs described above. The constructs are measurable in text through text analysis. While this can be achieved through coding of the texts, it is faster and more consistent to use a text analysis software program to code and relate the text to constructs. Diction 5.0 does this using a database that relates a word or term to a construct (Hart & Carroll, 2010). The database was created and validated on over 20,000 documents. The scores reported for each construct are standardized to allow comparison.

Of the twelve constructs described previously, five increase with deception and seven decrease with deception. The change is measured for the spam e-mail relative to legitimate e-mail. Constructs whose score increases with deception are represented as construct A. The score for construct A is a function (f) of the difference between the scores for the two clusters expressed as:

$$S(A) = f(S_{AD} - S_{AL}) \quad (1)$$

Where: $S(A)$ is the measure of construct A, which increases with deception

S_{AD} is the score of construct A for the deceptive spam

S_{AL} is the score of construct A for the legitimate e-mail

The constructs whose score decrease with deception are represented as construct B. The score for construct B is a function (f) of the difference between the scores for the two clusters expressed as:

$$S(B) = f(S_{BL} - S_{BD}) \quad (2)$$

Where: $S(B)$ is the measure of construct B, which decreases with deception

S_{BD} is the score of construct B for the deceptive spam

S_{BL} is the score of construct B for the legitimate e-mail

Section 4 - ACDDT Testing

Two types of spam were collected for the test. They were spam that attempts to steal the recipient's identity and spam that attempts to get the recipient to send them money fraudulently. These two types were selected because they are fraudulent and common (Internet Crime Complaint Center, 2009). The typical identity theft will spoof a financial institution and request the recipient to follow a link and put in his identity data to reactivate his account. The Nigerian letter is a typical example of a money theft spam. The recipient's assistance is requested to help the sender get money out of some country and is offered a percentage of the money. If the recipient responds, a request for money will be made in a subsequent e-mail. A newer spam of this type is sent to a recipient from someone he knows by a spammer who has gained access to that person's e-mail account. The request is for money to be wired to the spammer because of an emergency. If the recipient responds and sends money, there are often additional requests.

Data Collection

E-mail users determined the Deception Detection construct by selecting the two types of spam e-mails. The authors sent a request via Facebook for identity theft and money theft spam that had been sent to the recipient's inbox. Forty individuals responded and sent the authors over three thousand spam e-mails of each type. The authors obtained legitimate fund raising e-mails (money requests) from the Susan G. Komen Foundation, the March of Dimes, and the American Cancer Society for comparison with the money theft spam. Legitimate e-mail sent by financial and investment organizations was collected for comparison with identity theft e-mail. All e-mails were collected in November 2009.

Testing Method

The computational fraud detection model (CFDM) (Glancy and Yadav, 2010) was used for classification of the e-mails in each test. Previous work (Glancy, 2010, Glancy and Yadav, 2010) indicated that the CFDM effect size would be large. With a 0.20 effect size, and a two-tailed alpha of 0.01, the statistical power is 96% for a sample size of 110 if 69 of the 110 samples are correctly classified (Cohen, 1988) A sample size of 110 was selected for both tests. The e-mail included in the test was randomly selected from the each database of e-mail collected by using the random number generator in MS Excel. Fifty-five money theft e-mails were randomly selected and combined with fifty-five randomly selected legitimate money request emails to form a data set of 110 e-mails. This was repeated for the identity theft e-mail and matching legitimate e-mail. The document sets were text mined and each was

hierarchically clustered. The highest weighted one hundred terms representing each cluster were extracted. The terms were selected using binomial probability. These terms were analyzed with Diction 5.0 text analysis software. Diction 5.0 does this using a database that relates a word or term to a construct (Hart and Carroll, 2010). The score of each cluster construct was the standardized score reported by Diction 5.0. The construct measures were calculated using equations 1 and 2.

Findings

The e-mails clustered very well in both tests. The clustering correctly identified 107 out of 110 e-mails for the Identity theft test and 104 out of 110 for the Money theft test. The binomial *p*-values (Conover, 1999) in table 1 show that the results are quite significant. The statistical power is greater than 96% (Cohen, 1988).

Table 1. Identity Theft and Money Theft Test Results

Concept	**Construct**	**Identity Theft Construct Score**	**Identity Theft Test Hypothesis Score**	**Money Theft Construct Score**	**Money Theft Test Hypothesis Score**
Concealment – H1			2.08		2.62
	Collectives	1.35		2.18	
	Ambivalence	0.52		2.94	
	Denial	0.21		-2.50	
Normality – H2			6.39		2.65
	Leveling	5.12		-2.35	
	Diversity	0.00		5.00	
	Centrality	1.27		0.00	
Isolation – H3			4.71		1.55
	Increased Aggression	4.33		1.63	
	Increased Exclusion	-0.36		2.24	
	Increased Liberation	1.91		-3.54	
	Less Cooperation	-1.17		-3.65	
	Less Rapport	0.00		4.87	
	Less Motion	0.00		0.00	
Clustering Results	Correct Classification	107	104		
	False Positive	1	2		
	False Negative	2	4		
	p-value	3.325×10^{-28}	3.300×10^{-24}		

The first test was on identity theft spam. The results confirmed H1. The sum of measures of collectives, ambivalence, and denial in the test e-mail clusters is positively related to detection of identity theft spam e-mail. The scores for the three individual constructs were positive. The results confirmed H2. The sum of the measures of

leveling, diversity, and centrality is positively related to detection of identity theft spam e-mail. The diversity score was zero, and both leveling and centrality were positive. The leveling score was very positive. This indicates that the spammers made a strong effort to impress the recipient that the message was legitimate by using words that indicate assurance. This was not necessary for the legitimate e-mail because it was from a legitimate source. The results confirmed H3. The sum of the measures of exclusion, liberation, aggression, cooperation, rapport, and motion is positively related to detection of identity theft spam e-mail. The rapport and motion scores were zero indicating there was no difference between spam and legitimate e-mails for these constructs. The cooperation score was negative. The aggression score was very high. This can be explained by the spammer's trying to get the recipient to click on a link and input identity information. The spammer uses aggressive words to convince the recipient of the danger of not following the spam instructions.

The second test was on money theft spam. The results confirmed H1. The sum of measures of collectives, ambivalence, and denial is positively related to detection of money theft spam e-mail. The score for denial was negative. This can be interpreted that for spam e-mails that try to get the recipient to send money, the sender used more negative words. One explanation may be that most of this type of e-mail contained a strongly worded disclaimer as an attempt to convince the recipient of the legitimacy of the request. The results confirmed H2. The sum of the measures of leveling, diversity, and centrality is positively related to detection of money theft spam e-mail. The leveling score was negative and centrality was zero; the score for diversity was very high. The diversity score indicates that very few words that would indicate differences between individuals or groups were used in the spam compared to the legitimate e-mail. The results confirmed H3. The sum of the measures of exclusion, liberation, aggression, cooperation, rapport, and motion is positively related to detection of money theft spam e-mail. The aggression, rapport, and exclusion scores were positive. The rapport score was surprisingly high indicating a lack words that would build rapport. Because the spammer is trying to get the recipient to send money, we had expectations that the spammer would try to build a rapport with the recipient. The cooperation score was negative indicating the spammer did try eliciting the recipient's cooperation.

All of the hypotheses in both tests were confirmed. Both tests empirically support the ACDDT.

Section 5 - Research Contribution

The contribution of this research work is two empirical tests of a new and unique theory that presents an explanation of why deception is detectable in asynchronous CMC. The ACDDT proposes the concepts of Concealment, Normality, Isolation, and Deception Detection. The concepts relate to the individual's conative, cognitive, and affective functions; and deception is detected, because the individual unconsciously reveals the deception in the attempt to prevent detection by the receiver. The ACDDT propositions explain the relationships between the four concepts. The ACDDT is the first theory that goes beyond how deception is achieved and attempts to explain why the deception is detectable.

The support provided by these tests is a major demonstration for a deception detection theory in asynchronous CMC. The e-mail that was used in the test had already been identified as legitimate by conventional spam filters, which increases the confidence in the support for the ACDDT. Finding support for the ACDDT in two types of spam e-mail provides an indication that the ACDDT may be generalizable to other types of e-mail and possibly other genres.

The ACDDT provides a theoretical base for developing advanced deception detection tools. The CFDM (Glancy & Yadav, 2010) is an example of one possible deception detection tool. These tools may allow early detection of financial reporting fraud. Early detection can limit financial loss from fraud, and it may in the future provide a deterrent to financial fraud.

Section 6 - Conclusions and Future Work

A systematic evaluation process was followed to test the ACDDT in two genres—identity theft and spam. Both tests supported the hypotheses and provided empirical evidence for the ACDDT validity. In addition, the coauthor's dissertation (Glancy, 2010) has also tested the ACDDT on the Notes to the annual financial statements of companies using the evaluation process supported by CFDM and text analysis software Diction 5.0. The Notes contain the accountant's explanations of the financial statements. The Security Exchange Commission (SEC) requires disclosure

of anything that has the potential to significantly impact the future results of a business. The notes are prepared by and/or reviewed by auditing accountants. The test results on Notes also provided a strong support for the ACDDT.

The testing could be criticized for using a small sample and selecting e-mails in two categories that are recognized as fraudulent. The successful categorization of spam and legitimate e-mail does indicate that the effect size is large for the ACDDT and that the sample size was appropriate. There is no claim made at this juncture that all spam is deceptive. Further and extensive work would be needed to verify that deception is prevalent in most spam.

The CFDM tool, based upon ACDDT, has the potential to be the basis of a predictive model for fraud detection in financial reporting. The ACDDT and CFDM open additional research areas: deception in business-to-consumer websites, deception in consumer-to-consumer websites, and increasing the understanding of the mechanisms present in asynchronous text deception. Further work analyzing the singular value decomposition and the respective document and text vectors may lead to improved understanding of the mechanisms present in deceptive communications.

References:

Albright, R. (2004) Taming Text with the SVD. In: *SAS Institute White Paper*, pp. SAS Institute, Cary, NC.

Androutsopoulos, I., Noutsias, J., Chandrinos, K. V., & Spyropoulos, C. D. (2000). *An Experimental Comparison of Naive Bayesian and Keyword-based Anti-Spam Filtering with Personal E-mail Messages.* Paper presented at the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Athens, Greece.

Apache Software Foundation (2010). SpamAssassin Retrieved February 15,2010, 2010, from http://spamassassin.apache.org/.

Aradhye, H. B., Myers, G. K., & Herson, J. A. (2005). *Image Analysis for Efficient Categorization of Image-based Spam E-mail*. Paper presented at the Eigth International Conference on Cocument Analysis and Recognition.

Bagozzi, R. P. (1992). The Self-Regulation of Attitudes, Intentions, and Behavior. *Social Psychology Quarterly, 55*(2), 178-204.

Buller, D. B., & Burgoon, J. K. (1996). Interpersonal Detection Theory. *Communication Theory* (6:3), pp. 203-242.

Carlson, J. R., George, J. F., Burgoon, J. K., Adkins, M., & White, C. H. (2004). Deception in Computer-Mediated Communications. *Group Decision and Negotiation* (13:1), pp. 5-28.

Clark, J., Koprinska, I., & Poon, J. (2003). *A Neural Network Based Approach to Automated E-mail Classification*. Paper presented at the International Conference on Web Intelligence.

Cohen, J. (1988). *Statisical Power Analysis for the Behavioral Sciences* (Second Edition ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Conover, W. J. (1999). *Practical Nonparametric Statistics* (Third Edition ed.). New York: John Wiley & Sons, Inc.

Cormack, G. V., & Kotcz, A. (2009, July 19-23). *Spam Filter Evaluation with Imprecise Ground Truth.* Paper presented at the SIGIR 09 Boston, MA, USA.

Cormack, G. V., & Lynam, T. R. (2007). Online Supervised Spam Filter Evaluation. *ACM Transactions on Information Systems* (25:3).

Delany, S. J., Cunningham, P., Tsymbal, A., & Coyle, L. (2005). A case-based technique for tracking concept drift in spam filtering. *Knowledge-Based Systems,* (18:2005), pp. 187-195.

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., Wyer, M. M., & Epstein, J. A. (1996). Lying in Everyday Life. *Journal of Personality and Social Psychology 70*(5), 979-995.

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to Deception. *Psychological Bulletin,* (129:1), p74-118.

de Ville, B. (2006) *Decision Trees for Business Inteligence and Data Mining: Using SAS Enterprise Miner,* SAS Institute Inc., Cary, NC.

Glancy, F. H., & Yadav, S. B. (2010). A Computational Model for Financial Reporting Fraud Detection. *Decision Support Systems, Pre-publication*.

Glancy, F. H. (2010). *Deception Detection in Asynchronous Computer-mediated Communication.* Unpublished Dissertation, Texas Tech University, Lubbock.

Goodman, J., & Yih, W.-T. (2006). *Online Discriminative Spam Filter Training*. Paper presented at the CEAS 2006 - Third Conference on Email and Anti-Spam.

Hart, R. P., & Carroll, C. (2010). Diction 5.0 Retrieved February 15, 2010, from http://www.dictionsoftware.com/.

Hershkop, S., & Stolfo, S. (2005). *Combining Email Models for False Positive Reduction*. Paper presented at the KDD '05.

Hidalgo, J. M. G., Lopez, M. M., & Sanz, E. P. (2000). *Combining Text and Heuristics for Cost-Sensitive Spam Filtering*. Paper presented at the CoNLL-2000 and LLL-2000, Lisbon, Portugal.

ICF International (2009). The Carbon Footprint of Email Spam Report. 28. Retrieved from http://resources.mcafee.com/content/NACarbonFootprintSpam.

Internet Crime Complaint Center (2009). *2008 Internet Crime Report*. Retrieved February 10, 2010. from http://www.ic3.gov/media/annualreport/2008_IC3Report.pdf.

Jennings, R. (2009, February 10). Cost of Spam is Flattening - Our 2009 Predictions. http://www.ferris.com/?p=322011.

Lee, C.-C., Welker, R. B., & Odom, M. D. (2009). Features of Computer-Mediated, Text-Based Messages that Support Automatable, Linguistics-Based Indicators for Deception Detection. *Journal of Information Systems, 23*(1), 5-24.

Lynam, T. R., & Cormack, G. V. (2006, August 6-11, 2006). *On-line Spam Filter Fusion.* Paper presented at the SIGIR, Seattle, WA.

MessageLabs (2010). MLI Report January 2010. *January 2010*, 12. Retrieved from http://www.messagellabs.com.sg/resources/mlireports.

Rigoutsos, I., & Huynh, T. (2004). *Chung-Kwei: a Pattern-discovery-based System for the Automatic Identification of Unolicited E-mail Messages (SPAM).* Paper presented at the First Conference on Email and Anti-Spam. Retrieved August 22, 2009, from http://research.microsoft.com/users/joshuago/conference/papers-2004/153.pdf.

Rios, G., & Zha, H. (2004, July 30-31). *Exploring support Vector Machines and Random Forests for Spam Detection.* Paper presented at the First Conference on E-mail and Anti-spam, Mountain View, CA.

Roiger, R. J. & Geatz, M. W. (2003) *Data mining: a tutorial-based primer,* Addison-Wesley, New York.

Schneider, K.-M. (2003, April 10-17). *A Comparison of Event Models for Naive Bayes Anti-Spam E-Mail Filtering.* Paper presented at the The tenth Conference on European chapter of the Association for Computational Linguistics, Budapest Hungary.

Synovate (2007). *Federal Trade Commission – 2006 Identity Theft Survey Report*. Retrieved February 10, 2010. from http://www.ftc.gov/os/2007/11/SynovateFinalReportIDTheft2006.pdf.

Wittel, G. L., & Wu, S. F. (2004, July 30-31). *On Attacking Statistical Spam Filters.* Paper presented at the First Conference on Email and Anti-Spam (CEAS), Mountain View, CA, USA.

Yih, W.-T., Goodman, J., & Hulten, G. (2006). *Learning at Low False Positive Rates*. Paper presented at the CEAS 2006 - Third Conference on Email and Anti-Spam.

Zhang, L., Zhu, J., & Yao, T. (2004). An Evaluation of Statistical Spam Filtering Techniques. *ACM Transactions on Asian Language Information Processing*, (3:4), pp. 243-269.

Zhou, L., Burgoon, J. K., Nunamaker, J. F. J., & Twitchell, D. (2004). Automating Linguistics-Based Cues for Detecting Deception in Text-based Asynchronous Computer-Mediated Communication. *Group Decision and Negotiation*, (13:1), pp. 81-106.