1-1-2010

An Empirical Assessment of the CAN SPAM Act

Alex Conrad Kigerl
Portland State University

An Empirical Assessment of the CAN SPAM Act

by

Alex Conrad Kigerl

A thesis submitted in partial fulfillment of the
requirements for the degree of

Master of Science
in
Criminology and Criminal Justice

Thesis Committee:
Scott Cunningham, Chair
Warren Harrison
Danielle McGurrin

Portland State University

Abstract

In January 2004, the United States Congress passed and put into effect the Controlling the Assault of Non-Solicited Pornography and Marketing Act (CAN SPAM). The Act was set forth to regulate bulk commercial email (spam) and set the limits for what was acceptable. Various sources have since investigated and speculated on the efficacy of the CAN SPAM Act, few of which report a desirable outcome for users of electronic mail. Despite the apparent consensus of anti-spam firms and the community of email users that the Act was less than effective, there is little to no research on the efficacy of the Act that utilizes any significant statistical rigor or accepted scientific practices. The present study seeks to determine what, if any, impact the CAN SPAM act had on spam messages, to identify areas of improvement to help fight spam that is both fraudulent and dangerous. The data consisted of 2,071,965 spam emails sent between February 1, 1998 and December 31, 2008. The data were aggregated by month and an interrupted time series design was chosen to assess the impact the CAN SPAM Act had on spam. Analyses revealed that the CAN SPAM Act had no observable impact on the amount of spam sent and received; no impact on two of three CAN SPAM laws complied with among spam emails, the remaining law of which there was a significant decrease in compliance after the Act; and no impact on the number of spam emails sent from within the United States. Implications of these findings and suggestions for policy are discussed.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

The advancement of computing today has allowed human beings to automate many tasks that together make life easier. As technology continues to improve, more and more of our everyday activities are similarly improved, with better speed, ease, and functionality. Yet technology is a tool, the functionality of which can be given many different meanings. Technology together with information in this age is power, and power can be used for both good and ill.

With new ways to exploit technology, the law has always been there to reign in the means by which criminals take advantage of changing tools. The law, like technology, is not static. But in this age science and technology have known an exponential improvement of accelerating returns, where newer inventions emerge faster than they did before. The legal system does not seem to have matched this speed, however.

In the information age, knowledge is power. So too is the fraudulent representation of such knowledge similarly powerful. Computers have given fraud new meaning, and new avenues to locate victims in the millions. The abuse of information to mislead is likely as old as language itself. Long before digital computers, fraudulent messages were being carried out face to face.

The earliest commercial fraud in America centered on phony medical cures for the pilgrims that arrived at Plymouth Rock in the seventeenth century (Armstrong & Armstrong, 1991). Snake oil salesmen offered treatments ranging from false medicines, spiritual cures, and bloodletting to cure the sick. Flash forward to today with miracle cancer cures and fraudulent medicines sold online, and it seems that the crimes haven't changed (Armstrong et al., 1991). However, what has changed is the massive audience that can be reached with near anonymity by the perpetrator using the distance from one's target that the internet provides.

Coupled with the internet and newer means of facilitating transactions, fraudsters can do more than deceive a victim into a onetime payment for a faulty or nonexistent product. With the emergence of credit cards, social security numbers, and countless accounts scattered everywhere on the internet, a victim's identity can be stolen too. This allows the fraudster to make many transactions in the victim's name, rather than just one per victim.

Such fraud requires the misuse of information to mislead another for self gain. Since fraud involves the misuse of information, it can be seen why information technology might provide ways to better carry out fraud. Such techniques will always be dependent on the technology of the time. Identity theft used to be much more primitive than the multibillion dollar industry that is witnessed today, due to the more primitive technology available at the time.

Prior to 2004, before identity theft became more organized, fraudsters had to buy a card encoding machine, own or work at a business, and make an additional copy of their customers' credit card information (Anderson, Bohme, Clayton, & Moore, 2008). The fraudster might be a waitress, retrieving a credit card from her seated customers and taking it back to scan as a legitimate payment. Utilizing an additional encoding machine in secret, the magnetic strip of the card would be copied and kept by the fraudster. Later a clone of the card could be created from the copied data, and ATM withdraws could be made in the victim’s name. Another method might have been carried out by a corrupt banker who could collect password data for similar means of theft (Anderson et al., 2008).

These methods of identity theft were inefficient, less profitable, and less organized, and had fewer offenders who were able to participate in such schemes. Expensive card cloning hardware was required, or an existing position in a bank, or the ownership of a business. The circle of people who might make a living out of such schemes were limited, and most of the time the fraudster required a second job to maintain the primary illegal operations. This theft, however, was only limited because of the similarly limited technology of the time.

The emergence and prevalence of computers has allowed new and creative means for illegally stealing information, theft which has become available for more and more people, should they have the desire. The early 1990s was when such force began to be used, where students and anyone with the skill and time were known to electronically break into corporate and government systems (Stone, 2007). As with most newly developed crimes, such actions were more innocent than what exists today, where perpetrators merely left calling cards and tokens of their presence on the compromised systems. But by 2001, cybercriminals had learned that money can be made from such misappropriation of electronic resources. Keyloggers, password stealers, and crimeware entered the scene, harvesting personal data from users’ computers or commandeering online banking and other secure services (Stone, 2007).

However, force is not always the easiest, or best way to acquire stolen goods. Hours of effort and study in cracking encrypted channels, exploiting obscure and technical software vulnerabilities, and writing thousands of lines of malicious code takes more skill than is necessary. Often it is easier to just convince victims to give up their information willingly, albeit unaware of the consequences of their actions until it is too late. Here is where fraud and technology meet, where electronic communications are used to send misleading offers and solicitations to gain information or money.

The most likely candidate for such fraud is through email, or spam, as email is the most efficient way to send literally billions of messages at the press of a button, without so much as a dime being spent to produce so many transmissions. The most common purpose of malware today is to send spam from millions of infected computers controlled by the cybercriminal (Schiller, Binkley, Harley, Evron, Bradley & Willems, 2007).

Spam is the sending of unsolicited electronic mail messages to multiple recipients, usually for commercial purposes. Spam, like most technologies and the crime that followed such technologies, started small. The first recorded instance of spam email being sent out was in May, 1978, on Arpanet, the US government-run computer network that eventually became the internet (Kleiner, 2008). Gary Thuerk, a marketer at a now-defunct computer firm Digital Equipment Corporation, sent bulk email to 393 recipients on Arpanet. Thuerk made no effort to conceal his identity during the sending, and recipients complained directly to him. Thuerk was subsequently reprimanded by the Arpanet administrators (Kleiner, 2008).

Later, once the internet became available to the general public, spam emerged first on usenet discussion groups, rather than through email. The first was in 1994, where two Arizona lawyers posted messages on 6,000 separate discussion groups, advertising their services as immigration lawyers. However, such perpetrators of usenet spam found their inboxes flooded with angry responses from their recipients (Kleiner, 2008).

Spam was considered a nuisance from the very beginning, only spammers made less money, had less anonymity, and were not as organized. Spammers soon went underground, and spam itself has since been elevated to quite a great deal more than just a nuisance. It has now become a multibillion dollar industry (Kleiner, 2008) responsible for slowed internet traffic, wasted time and costly effort separating junk email from legitimate email, malware infections and spyware proliferation, stolen credit card and online account credentials, stolen identities, fraudulently sold commercial products, duped victims who are tricked into committing crimes for the cybercriminal, and in some rare cases, has cost individuals their lives (Smith, Holmes & Kaufmann, 1999).

As of this writing, spam in 2009 makes up over 90% of all email sent worldwide (McMillan, 2009), and seems to continue to increase. In 2003, only 45% of email was spam (McCain, 2003), but was, and continues to, rise as professional spammers grow in skill and reach further and further to make more money. An average of 120 billion spam messages are sent every day (Kleiner, 2008).

The foremost purpose of spam messages are to make money for the spammer (Schiller et al., 2007), and the primary means to do this is through fraudulent offers and deals. Among the most costly fraud includes phishing and advance fee scams. Phishing is the sending of email messages that masquerade as a source the user trusts, such as their online bank or Ebay account (Brody, Mulig & Kimball, 2007). The purpose of this tactic is to trick the target into revealing sensitive personal credential information, such as account logins, passwords, credit card numbers, or any piece of information that can be converted into stolen cash. This specific type of identity theft costs the United States 52.6 billion dollars a year, 10% of which includes consumer's money, the other 90% of the costs businesses must endure. The average loss for an individual victimized by phishing is $1,180, while it is a costly $10,200 for businesses (Brody et al., 2007).

When spammers don't have the technical ability to spoof a victim's familiar website, such as their Bank account, they can always resort to relying exclusively on persuasion. Advance fee fraud does just this, where the spammer sends email solicitations pretending to be a potential business partner or foreign character with an opportune offering. A fanciful story is described in which the victim stands to make millions of dollars, by laundering money, collecting lottery earnings, or making a business deal. The catch is that the victim must pay an advance sum of money before they can collect their reward. Of course, the reward never arrives and the victim is left empty handed (Smith et al., 1999).

The loss of your investment is the only cost incurred if one is lucky. In the case of some scams perpetrated in Nigeria, the masquerade becomes so complex that finalizing a transaction must be completed by traveling to Nigeria itself. The victims are lured by the fraudsters out of their country, where they are often kidnapped and held for ransom, to make even more money. Since 1992, 17 people have been killed in such Nigerian scams (Smith et al., 1999).

Such frauds are dangerous only because they reach so many people's inboxes. In the case of one spam operation, it has been observed that only 1 in 12.5 million spam messages receives a buyer (Larkin, 2008). Only a rare few recipients are fooled by these messages. However, this is enough for spam to cost billions in damages. But such damages can only be dealt if the spammer is able to send bulk email in millions of messages per day. The best means the spammer has to do such a thing is through malware infections.

Computers infected with malware and networked together in what is called a botnet are responsible for the majority of spam (Schiller et al., 2007). This is malware installed on a victim computer through trojans, viruses, or worms that can scan a host computer for user contacts, web browser data, and anything where email addresses might be stored. The botnet subsequently automates the sending of spam messages to these addresses, from the victim's own computer, and not the spammer's, sometimes posing as the victim. Spammers can control these computers remotely, sometimes owning botnet clients installed on millions of unsuspecting victims' PCs. Sometimes botnets are spread by spam itself, with scripts, attachments, or links to malicious sites that install malware on the recipients computer (Schiller et al., 2007). It is estimated that between 16 and 25% of computers are infected by a botnet (Weber, 2007).

Growing technologies and anti-spam laws have been in development continuously for almost as long as spam has existed. Yet spam continues to grow as a market, and the bulk of email sent worldwide continues to be predominantly spam. The difficulty in eliminating this problem seems to be partially due to the profitability and large underground market network that has emerged and has been built around spam.

Cybercriminals seem just as professional and organized as a legitimate industry, with a strict division of labor, investments, traded goods and hired consultants. Spammers contract botnet herders, who allow the spammer to rent botspace to send bulk email, and the spammer is hired by a malware writer, who writes phishing websites and scripts to install malicious code on victim computers (Anderson, 2008). Identities stolen from such operations can be bought and sold online, with a credit card selling for as little $0.50 each. Sellers of stolen identities receive reviews and ratings for the quality of their stolen goods, and the stolen identities are sold to launderers to conceal the origins of the stolen money. Some even offer to clone stolen credit cards and mail the spoofed card to buyers (Brody et al., 2007).

This lucrative and booming business is not going away anytime soon. The majority of the fight against spam is technological, with new filters, authentication software, antivirus, and antispyware being developed and improved to limit the amount of spam users receive or that can fool recipients. Aside from building these technological defenses, there are also offensive measures to combat spam, that involve illegalizing certain spamming techniques and enforcing such laws.

Deleting and blocking spam messages makes it harder for the spammers, if only a little. But arresting the spammers themselves ends their operations entirely. According to Spamhaus (an independent network which tracks the internet’s spammers, spam gangs, and spam services), 80% of spam received by users in North America and Europe is sent by less than 200 spam gangs, comprising some 500-600 professional spammers (Moustakas, Ranganthan & Duquenoy, 2005). If legal authorities could take out these kingpin criminals, it would surely make a dent in the world's spam rates.

Unfortunately spam crosses international borders, and no country has jurisdiction over all the spam that it receives. Anti-spam laws exist in many countries, but they are unequipped to target spammers not in their jurisdiction. In America, Federal anti-spam legislation went into effect in January 1, 2004, called the CAN SPAM Act. The act supersedes any state laws in existence regulating the sending of spam, replacing them with some basic requirements if businesses so choose to send commercial email messages (CAN SPAM Act of 2003).

The major laws outlined in the CAN SPAM Act include requirements for honesty and accuracy of the content of email messages, genuine identifying information about the sender of the email messages such as address and contact information, and an opt-out method that allows recipients to choose to no longer receive messages from a given sender. Harsher sentences exist for those who send spam from an unauthorized location, such as from a botnet on an unwilling person's computer.

Since the CAN SPAM Act went into effect in early 2004, efforts have been made to determine its efficacy in limiting spam. The CAN SPAM Act is enforced by the Federal Trade Commission, which subsequently produced a report on the success on its Act to Congress in 2005 (Majoras, Leary, Harbour & Leibowitz, 2005). Contained in the report was the conclusion that spam has stabilized since the creation of the CAN SPAM

Act, whereas before spam displayed a steady increase over time. The data used were the number of spam emails received per day by the FTC.

Other authorities have similarly analyzed spam volume to determine whether spam rates or content differed after the passing of the Act. According to MessageLabs, an anti-spam and security company, after one year following the CAN SPAM Act, email that was considered spam went up from 60% of all global email the year before to 80% by the end of the year (Zeller, 2005). When the FTC measured spam in absolute terms (amount received per day), spam seemed to have slowed. But when measured in relation to the amount of legitimate email sent, spam seemed to have increased.

It should be mentioned however that neither the Federal Trade Commission nor MessageLabs report the statistical significance of their findings. Although overwhelming consensus seems to be that spam has not decreased significantly since the creation of the CAN SPAM Act (Lee, 2005; Arora, 2006), such conclusions should be finalized with a little more statistical rigor.

Despite this, the actual volume of spam sent is not the only measure of the success of the CAN SPAM Act. Questions as to whether compliance with the specific requirements detailed in the act may have increased since the passing of the law have been investigated. In one such case, 1,100 unsolicited commercial emails were randomly selected from 5 email accounts, once six months after the passing of the CAN SPAM Act and another sampling two years after the act went into effect (Grimes, 2007). Each email message was rated as either complying with the CAN SPAM Act, or not complying with the act.

Unfortunately, 14.3% of spam complied with legal requirements the first six months following the act, whereas a mere 5.7% of emails complied with it two years later (Grimes, 2007). However, a sample size of 1,100 might not be large enough for something done on as massive a scale as spam. Additionally, no baseline of compliance was established in spam messages before the CAN SPAM Act was instated.

Another possible measure that seems lacking in the literature includes the location from which spam is sent. If spam has not been affected by existing laws, it could be that spammers have moved their operations across borders, outside of the United States, where the CAN SPAM Act has no jurisdiction. CAN SPAM may not have decreased spam rates, but rather moved spam sending botnets to where the Act has no reach.

In light of these concerns, several research questions come to mind that could be answered by further investigation: (1) has the CAN SPAM Act affected spam rates over time, (2) has the CAN SPAM Act influenced compliance among spammers with CAN SPAM Act rules, and (3) have the primary locations of spam operations moved since the CAN SPAM Act went into effect?

Considering the enormity of the problem spam poses to the 1.6 billion people online (Internet World Stats, 2009), and some of the seeming impunity spammers enjoy since the majority of anti-spam practices have been defensive rather than offensive, existing spam law warrants extensive investigation. Anti-spam law has been given less emphasis than anti-spam technology, and it should be improved just as much as spam filters and intrusion detection systems. With the rigor of scientific analysis of existing laws, such as the CAN SPAM Act, we can hope to piece together what laws might work, and which ones could stand to see some improvement.

CHAPTER 2: SPAM

Spam is the sending of bulk unsolicited messages to multiple recipients, usually for commercial purposes. Based on this definition, there are many types of spam; text message spam, link spam, forum and chat spam, social networking spam, and of course, email spam. Probably the most concerning form of spam would likely be this last type. Email spam is the best means to reach as large an audience as possible. More people likely use email than forums, chat groups, and social networking sites, and sending massive amounts of text message spam might be harder to automate cheaply. Email is the best means of reaching potential buyers/victims. Because of this ease and efficacy, spam has caught on during its 30 year lifespan. It is an industry that is growing at an accelerated rate.

Spam rates have historically been rising over time with the further spread and reach of the internet. As more and more users acquire a connection to the internet and retain one or more email addresses, the market value of bulk commercial email grows. Spam has been rising steadily since its inception (Lee, 2005). In 2003 only 45% of all email sent and recorded by ISPs was considered spam (McCain, 2003), but in 2009 spam has become 90% of all email sent (McMillan, 2009). This volume of spam email is of such proportion that now 60% of all internet traffic, not just email traffic, is spam (Ananthaswamy, 2009). Even if spam filters were to catch all of this spam and they never reached their intended recipients, the energy expenditure alone from processing so many messages is costly.

WHY SEND SPAM

The reasons for sending spam are the same reasons for committing most forms of cybercrime, that being to make money (Paul, 2006). A theoretical foundation that might aid in the understanding for the reasoning behind the violation of spam law would be that of routine activities theory. Three conditions must be met for this theory to explain variation in crime: there must be a lack of a capable guardian against committing the crime, there must be a target suitable for the offender to misappropriate, and the offender him/herself must be motivated to commit the crime. The last condition, motivation to offend, is clearly explained by observation of the majority of cybercrime, spam crime included. Making money is a motive that stands out from all the rest, as cybercrime is an easy opportunity to earn money from the luxury of one's own home with skills that can be learned on the internet.

However, it takes a lot of spam to make money from it. In 2008, the average daily volume of spam sent worldwide was 120 billion messages per day (Kleiner, 2008). The considerable effort to transmit so many electronic messages begs the question: why send so much spam, especially when it seems users are savvier to such manipulative schemes. The spam business remains as profitable as ever however, with the majority of spam messages used as advertising (*Types of Spam*, 2009).

Spam must be profitable enough to maintain a growing industry and culture of professional spammers. However, due to the secretive nature of such a profession, it is difficult to determine just how much money can be made from full time spam operations. Due to the anonymous nature of the internet, and the remote distances that can exist between spammer and victim, not many spammers are caught and prosecuted. They keep few records, and move their operations all over the internet.

There are many examples and instances where spam profits can be analyzed. In the cases where spammers are caught or reveal themselves, more can be gleaned from the life of a spammer. In 2005, Levon Gillespie, 21, was served a court summons by lawyers from Microsoft, stating that he had violated both state and federal law by flooding Microsoft's internal and customer email networks with spam. Gillespie had been operating a website offering to provide shelter to bulk advertisers by hosting their operations on offshore servers to protect them from antispam laws.

Gillespie failed to show up during his court hearing, and was given the default judgment fine of 1.4 million dollars for his crimes. During a later follow up over phone by journalists, Gillespie said he was not aware of the judgment and that no one from the courts had showed up in response to it. Gillespie said he would definitely continue his spam operations because there was "way too much money involved." At its peak, Gillespie's spam business acquired for him a six figure income. Gillespie doubted that anyone would be deterred by existing antispam law since the business was so lucrative.

Spam as a business can be lucrative enough that the legal risks associated with it can seem minor in comparison to what can be gained from such operations. Gillespie chose to continue breaking the law, even after being caught, and even being willing to admit such continuation of his crimes to the press. Another such instance involved a former spammer who decided to quit his illegal trade in 2004 not because of the legal risks involved, but because of the social stigma attached to spamming adult material and male enhancement products to annoyed users (Sjouwerman & Posluns, 2004). The former spammer felt embarrassed revealing his identity to anyone asking him what he does for a living, so instead decided to write a book called *Inside the Spam Cartel: Trade Secrets from the Dark Side*. In his book he explains how readers themselves can take up the spamming profession.

The former spammer presented at the Spam Symposium in Europe in 2007, where he discussed his former business as a professional spammer. During his presentation he revealed that he sent an average of 40 million spam messages per week, with recipients clicking on his spammed links an average of 0.12% of the time. Of those, only one out of 200 made a purchase, meaning one in over 166,000 spam messages receives a buyer; that amounts to 240 transactions a week. After subtracting expenses (buying email lists, botnet space, server hosting, etc.), the former spammer once made $336,000/year from his trade (Sjouwerman et al., 2004).

This was prior to 2004, however, when spam rates were not quite as high. There is a higher volume of spam sent worldwide today; which is either an indication that there are more and better spammers, or that this higher rate is necessary in a world of savvier internet users. One buyer in over 100,000 advertisements is obviously a tough crowd to sell something too, but that rare buyer is clearly enough for spam to continue filling up the inboxes of the rest of email users. It is easy to educate most users to the nature of spam, but unless every last of the 1.6 billion internet users worldwide (Internet World Stats, 2009) refrain from buying in to such marketing, spamming will continue to be a profitable enterprise.

More recent investigations into the success rate of spammers can now be done without the cooperation of the spammers themselves revealing their identities. The majority of spam is carried out by botnets, multiple infected computers that can be remotely controlled by a spammer to send spam (Schiller, Binkley, Harley, Evron, Bradley, Willems & Cross, 2007). Sending spam from the spammers own computer would be costly and inefficient; it is better practice to carry out the bulk of these mass mailings from thousands of unsuspecting users' computers remotely. Not only are such activities harder to trace to the original source, but it allows spammers to send massive amounts of mail in parallel.

Since the spam is sent out from botnets, which infect as many computers as possible, these infected computers can be acquired or intercepted and studied by white hat spam researchers. By deliberately baiting or downloading malware that installs a bot client on a virtual machine, researchers can observe the signals the newly installed malware sends or receives from the bot master, usually from a command and control server that the bot master can use to send commands to his/her bots. One noteworthy example where this was done was on the Storm Botnet, a botnet primarily used for spam (Kanish, Kreibich, Levchenko, Enright, Voelker, Paxson & Savage, 2008).

Kanish et al. established personal servers used as a honeynet, a computer set up to bait bot infections for research purposes. Eight servers were used to install Storm Worm proxy bots, which are used as relays between worker bots that carry out the spam operations, and the master servers, which the botmaster uses to send commands to his/her workers. Since these proxy bots were owned by the researchers, transmissions between bot and master could be intercepted and analyzed.

There were 75,869 bots that connected to the proxy bot servers total (Kanish et al., 2008). The bots would send spam email templates and email address lists to be spammed. The Kanish et al. intercepted these spam templates and replaced the links contained within them to point to a fake pharmacy website owned by the researchers. This was done since the original website owned by the spammer would not be traceable, and it could not be determined how many mail recipients had visited or made a purchase there.

From their own fabricated pharmacy website, Kanish et al. could monitor how many email recipients clicked through to the site (response rate), and from there how many visitors attempted to make a purchase (conversion rate). For legal reasons, the website was written to result in a 404 page before the visitors could finally enter in their credit card information. The experiment lasted for 26 days, during which time 213,760,147 unique email addresses were spammed pharmaceutical advertisements. There were a total of 350 million email messages sent during that time, and only 28 sales resulted from those messages. The average purchase price for each of these sales was approximately $100. Taken together, these sales would have resulted in profit of about $2,731.88 a month. However, it was estimated that they had observed only 1.5 percent of the bot network. Given this, the actual revenues would be more along the range of $7,000 a month (Kanish et al., 2008).

Even with a conversion rate of one buyer in 12.5 million spam emails sent, spam continues to be profitable. However, Kanish (2008) mentions that of all the countries that received spam, the United States had the lowest response rate, even though the United States is often the biggest target of spammers. This may be due to better spam filtering technology in the United States, or perhaps because of greater familiarity with fraudulent spam practices. Despite this, spammers have a large profit motive to infect as many hundreds of thousands of victim PCs as possible to help carry out their illegal trade.

Without sufficient profits, there is little doubt that spam rates would decline into something significantly less than what we see today. But spam as a business is only one piece of the highly interconnected cybercrime marketplace as a whole. Spammers may often make their money by other means than just spam alone, and spammers often partner themselves with cybercriminals of other sorts. Cybercrime is organized crime, and spam is highly interdependent on the rest of the cybercrime markets.

SPAM’S PLACE IN THE CYBERCRIME HIERARCHY

One reason why spam may be rising in particular could be because it has become so much easier to send spam and make money from it. One basis of explanation for the observed rise in cybercrimes committed over time would be the anonymous nature of cyberspace, allowing for cybercriminals to avoid detection from law enforcement. Compound this with the vastness and immediacy of the internet, and any willful cybercriminal can find any suitable target he/she chooses given the limitless number of potential victims in cyberspace. These environmental conditions, that the internet has in large numbers, is explained by routine activities theory, which proposes that the risks of crime increase given easy targets, willful offenders, and lack of law enforcement.

Cyberspace makes committing crimes easier, and cybercriminals of all types have taken note. Taking advantages of the anonymity of cyberspace, spam crimes have become well developed and integrated with other forms of crime and even legitimate businesses. A potential spammer may lack all the skills necessary to send profitable amounts of spam by themselves, but may team up with, hire, or buy products and services from other cybercriminals of other varied skills to successfully facilitate their illegal business (Anderson, Bohme, Clayton & Moore, 2008).

Spam is a form of white collar crime, with the spammer fitted in the midst of a cybercrime chain of similar skilled offenders in a division of labor that allows all cybercrime in general to be a very organized business model. Spam and other forms of cybercrime can be classed under the typology of white collar crime as a combination of what Friedrichs (2009) calls contrepreneurial crime and technocrime. Contrepreneurial crime is the pairing of a legitamate business with illegal dealings and cons in addition to and through the legal business. Technocrime is the carrying out of criminal operations with the use of advanced forms of technology. Some spammers may be hired by legitimate businesses for the purposes of spamming advertisements for that given business (Saltzman, 2009). If paired with a legitimate business, spammers could fall under the category of contrepreneurial crime. All spammers would qualify as technocriminals, given the necessity of relying on technology to mass produce spam messages.

The network of cybercrime professionals of varying skills and classes can involve a combination of legal businesses, entirely illegal businesses, and independent contractors for hire, with varying degrees of collaboration and organization. Because of the interdependence and overlap of spam crime with other forms of cybercrime and business models, it becomes necessary to understand cybercrime as a whole to get to know spam better. Spammers tend to work or trade with other types of cybercriminals and dabble in various sorts of cybercrime (Anderson et al., 2008). Cybercrime itself has become something that is now similar to all markets. It has a division of labor, goods and services, and supply and demand. It has become a big business, with 85% of malware written with the intention of profit in mind (Paul, 2006). And unlike the risks associated with other forms of crime, only an estimated number of 5% of cybercriminals are caught and prosecuted (Paul, 2006). Low risks due to anonymity over the internet combined with high payoffs of reaching millions of potential buyers or victims a day with automated technology means cybercrime is a successful business enterprise.

Noteworthy of cybercrime is its division of labor. For example, phishermen who operate phishing websites specialize in tricking visitors into entering their bank account or other information on an online form masquerading as a site familiar to the victim. However, phishermen need to lure visitors to their phishing sites, so they contract spammers to send out bulk email solicitations to attract victims to the phishing site. The spammer may in turn rent a botnet from a bot master, which is thousands of victim computers remotely controlled via malware. The thousands of bots can be used to send out bulk spam. The spammer may also send email to solicit people to “work” for the spammer, called mules. Mules are duped into accepting bank payments from stolen information gathered by the phisher, and the mule wires the money onwards to another country via Western Union to launder its source. The mule takes on most of the risk, and the real cybercriminals take most of the profits (Anderson et al., 2008).

The criminals may amass so many stolen identities that they cannot cash them all in themselves. They may additionally sell the credential goods online to get rid of them, in one of many cybercrime forums or chat groups available on the internet. Stolen credit cards, called CCV2s by the cybercrime community, are sold for as little as fifty cents to ten dollars each, depending on what country they are from and how trusted the seller is (Giles, 2009). More expensive credential goods may have more information and are referred to as fullz, such as bank account and personal information (e.g. social security number). Stolen credit cards that also contain a copy of the magnetic strip or smart chip in the credit card can also be found, referred to as dumps (Giles, 2009). These can be cloned and ATM withdraws can be made. The sellers of these stolen goods accumulate customer feedback and ratings upon continued successful transactions. Sellers can eventually be designated as trusted or verified by the community, so that buyers can be confident they are not being scammed with poor quality or nonexistent goods (Giles, 2009).

The stolen identities are subsequently sold online to other cybercriminals who may use them for their own cybercrime operations. Paying for goods and services in the pursuit of criminal activities with one’s own credit card would not be advisable. It is far more economical and secure to purchase them with stolen credit cards, which can be cheaply and easily bought. Cybercriminals can use them to anonymously buy servers which they may rent out to spammers or be used as command centers for a botnet that sends spam (Brody, Mulig & Kimball, 2007). Thus the cycle continues.

Most of the cybercrime roles discussed so far require certain skill sets to apply in making money. However, such skills are not always necessary to pursue an illegal business online. Having recently emerged, cybercrime services have made it easier for all involved to participate in computer fraud. Cybercrime services are the sale of crimeware software or services that buyers may use in their own illegal activities. Now anyone with enough money can purchase all they need to become a spammer, a phisher, or a bot herder (Wiedrick-Kozlowski & Stinchombe, 2008).

This has become a boon to some cybercriminals, as the selling of software, such as viruses and worms, is not necessarily in itself illegal. Using the malware to compromise computers or data is illegal, and the distributers of such products are not themselves responsible for what their customers do with the purchased goods. This low risk may create an incentive for malware writers to help make it easier for those who are willing to use malware illegally.

Unless one purchases these services with stolen credential goods, utilizing them does not always come cheap. Those with few cybercrime skills but money to spend can rent bullet proof hosting to operate an illegal business online. Bullet proof hosting is hosted server space located in countries with few laws to crack down on whatever illegal content is hosted on the server. One organization that specializes in bullet proof hosting is the Russian Business Network, which provides many cybercrime services (Krebs, 2007). Computer crime of every sort can be safely conducted on their rented servers, such as scam sites, child porn, and malware distribution.

For $300 a month and a $100 setup fee, buyers can rent their own servers to host a phishing website. To acquire spam sending software on the server to lure victims to their site is approximately $1,200 a month. Another $1,900 monthly charge can be used to buy a database of email addresses to spam, spam sending proxies, and other add-ons (Brody et al., 2007).

Cybercrime has become profitable, low risk (only 5% caught (Paul, 2006)), and easy to carry out even without many skills. Spam is a large part of the cybercrime problem as a whole, for it is used when spreading malware, furthering phishing and other scams, recruiting mules, and various commercial solicitations for legal, illegal, and fraudulent products. Without spam, the cybercrime business would take a substantial hit. Without spam many of the harms of visiting cyberspace would be lifted. However, the harms of spam are manifold; some harms are obvious, while others are less direct but still destructive. Spam may not receive as much media and law enforcement attention as other forms of crime, but its threat is still substantial.

HARMFUL CONSEQUENCES OF SPAM

Spam is generally considered a nuisance. It fills up inboxes and costs extra time to sort and remove. Finding the ham (desirable non-spam email) amidst the spam can often be laborious, and mistakes may occur where ham is deleted by mistake. Most recipients do not attempt to purchase any of the spamvertised products or services

(Larkin, 2009), or find themselves deceived by fraudulent spam, and so the most common problem of spam is extra time and effort sorting through it all.

The time consumption that goes into deleting unwanted emails affects businesses, not to mention individuals. An annual survey conducted by the Center for Excellence in Service at the University of Maryland’s School of Business and Rockbridge, a technology research firm, found that internet users in the United States spend an average of three minutes deleting spam each day they use email. Based on the 169.4 million online adults in the United States, and based on an average wage, this comes down to $21.58 billion in terms of lost productivity (Clayburn, 2005).

Yet the problems of spamming run deeper than this. While wasted time and unsolicited email messages are bothersome, the risks to both individuals and businesses that spam poses go beyond these concerns. Successful spam targets one’s money, identity, and even the victim’s own computer and network to further cybercrime schemes. And even in some rare cases, victims have lost their lives.

Spam Scams

If spam just sold legitimate products, the problem would be lessened. But often are there deceptions and outright fraud in email spam. Most spam is deceptive in some way, with two thirds providing false from addresses, misleading subject titles, or misleading message text contained in the body of the email (e.g. posing as a fake acquaintance of the recipient), based on a sample of 1,000 emails (MacFarlane, Harrington, Salsburg & Goodman, 2003). Spam will generally attempt to make money in one of four ways of solicitation. The four methods include requesting the recipient buy a product or service (spamvertising), wire the sender money in a proposed business or similar offer (advance fee fraud), enter sensitive personal information such as an account login (phishing), or help launder the spammers money (by becoming a mule). In just about all of these scenarios, fraud or deception is occurring.

When attempting to purchase spamvertised products, one of four things can happen: (1) the item you order arrives and is in good condition, (2) the item you order is late arriving and/or does not work or is of a different quality than that spamvertised, (3) the product never arrives and the spammer takes your payment anyway, and (4) you receive nothing and the spammer steals your credit card information you used to pay and empties your bank account with it (Saltzman, 2009). The proportion of spamvertisements that fall into each of these four categories is unclear, however, as buying a significant amount from randomly selected spam messages would be costly, as the average price is around $100 for each product (Kanish, Kreibich, Levchenko, Enright, Voelker, Paxson & Savage, 2008), not to mention the risk of having one's credit card stolen.

Yet the most problematic spam would be the outright fraudulent scams, such as phishing and advance fee fraud. There is not as much of these types of scams in relation to spamvertisements, but they can be significantly more costly. In 2003, 8% of spam emails were considered to be in these categories of scams, while in 2004 the number dropped to 6% (Hulten, Penta, Seshadrinathan & Mishra, 2004). In 2006, the number rose back up to around 9% (Evett, 2006).

Among the two most common scams is phishing, where the spammer sends email transmissions that masquerade as a source that is hoped to be familiar to the recipient.

The fraudulent email may appear to be from the target's bank, online retail account (e.g. Ebay, Amazon), email service provider, or even the administrator of an online computer game the recipient uses (Brody, Mulig & Kimball, 2007). Whatever website a user has an account on that could be sold on the cybercrime market, or whatever website that stores sensitive credential goods (e.g. credit card, social security number), the phisher may attempt to fabricate a phishing scheme off of it. Spotting these scams is sometimes easy, but not always. In the case of spear phishing, the phisher targets a potential victim that he/she already knows something about, such as what bank they use. Spear phishing is especially deceptive, as it creates a more plausible story and therefore a more slippery trap for the victim to fall into (Aycock, 2007).

The brunt of the damage of phishing scams is felt by businesses, which are especially vulnerable to spear phishers targeting their employees. The phishing losses to the United States every year totals $52.6 billion (Brody et al., 2007). Ninety percent of the victims are businesses and financial institutions, while the remaining 10% is composed of consumers. The average losses on a case by case basis for falling prey to a phishing scheme are $10,200 for businesses and $1,180 for individuals (Brody et al., 2007).

Once the phisher has the target stolen credentials, he/she will use them to open new credit accounts, apply for loans or benefits, file fraudulent tax returns, sell the information to other cybercriminals, (Brody et al., 2007), make purchases online with the stolen card (called carding), take out mortgages, and destroy the credit ratings of the victims who may be unable to get new jobs, buy houses, or get passports until they sort out the damages (Stone & Levy, 2005). Stolen credit cards are among the most common illegal goods sold on the cybercrime black market (Giles, 2009). Yet fabricating false websites that look authentic and gaining personal information to use to fool a target is not necessary to conduct a profitable scam over the internet. Sometimes the deception can be as easy as sending out a few carefully worded stories to fool someone into making what they believe is a lucrative exchange.

Called advance fee fraud, or 419 scams (named after the criminal code in Nigeria prohibiting fraud (Tive, 2006)), scammers in such cases rely more heavily on social engineering rather than technical expertise. Advance fee fraud is an attempt to trick a recipient into wiring the perpetrator money by deceiving the victim into believing that by cooperating they will be rewarded or, in some cases, spared some sort of threatened retaliation. The fraudulent solicitor usually creates some elaborate story as to why the potential victim is being contacted, such as a wealthy foreigner needing to transfer a large sum of money into the United States, or that the recipient has won some sort of prize, or that some attractive woman needs the recipient’s help. Whatever the case, the catch is that in replying to these emails an advance fee must be paid first before the solicitor can reward the recipient. Usually several fees are required to be paid, one after the other, as the fraudsters will attempt to extract as much payment as possible from their victims by elaborating new details into the story.

The different stories concocted by the fraudster are many and varied, although the goals are all the same. Some contact the recipient and inform him or her that they’ve won the lottery (Dryud, 2005), or that a wealthy investor, widow, orphan, etc. wants to transfer money into the victim's home country (Nigerian Advance Fee Fraud, 1997). Often the victim must pay various transaction fees. The victim can be strung along for months or years paying various fees and taxes before realizing that the money does not exist. Sometimes the victim can be further taken advantage of by recovery scams. The same fraudster may contact the victim sometime after successfully defrauding them. The cybercriminal might contact the victim posing as the police, and offer to recover the victim's stolen money. However, there is a small fee (taxes, etc.) that must be paid before the authorities can pursue the criminal (Dyrud, 2005). Of course, the stolen amount is never recovered, and the criminal successfully takes advantage of the same victim again.

The average loss to a victim from such scams is $3,864 (Dyrud, 2005). However, in some rare cases, money might not be the only thing a victim stands to lose. In the case of Nigerian scams (advance fee fraud from fraudsters living in Nigeria) Some 419 scams require the victim to travel to Nigeria to undertake further steps to complete a transaction. Once there, the offenders may hold them hostage, and demand more be paid to them in exchange for the victim's release. The victim is sometimes given a forged visa; making their stay in the country illegal and leaving them open to further acts of extortion by threatening to reveal them to the authorities. Some victims travel to Nigeria of their own accord, desiring to recover the funds stolen from them (Smith, Holmes & Kaufmann, 1999). Some people are even killed during these encounters; since 1992, 17 people have been murdered in Nigeria attempting to recover their stolen money. The US State

Department has documented over 100 cases where US citizens have been rescued from Nigeria (Smith et al., 1999).

Not all victims of email fraud always stand to lose money. In some rare cases, the victim is tricked into participating in helping the fraudster carry out crimes against other victims. Called mules, victims are talked into helping the cybercriminal launder their stolen goods. The mule is contacted by email with a business offer by the fraudster. The mule accepts checks mailed to him, whereby they are deposited into the mules own bank account. The mule then wires most of that money via Western Union or Liberty Reserve to the scammer, keeping a small fee for himself. Little does the mule know that the checks were actually stolen from other cybercrime victims, and that the mule is taking on considerable risks by cooperating with the spammer (Goodin, 2007).

For example, the cybercriminal may create a fake posting on Ebay selling some commercial good (dirt bike, laptop), and the buyer who bids for the product is sent a private message by the scammer that they can only accept payment by check through the mail (rather than securely through Ebay). The victim is given the mule’s address, and the check is sent there to be deposited and forwarded to the fraudster (Goodin, 2007). However, such relationships may be short, as one of the many victims reports or confronts the mule, and the bank overdrafts the amount stolen from the mule’s account, leaving him at a loss or in debt. The spammer then cuts off all contact with the used mule, and seeks new partnerships elsewhere.

Deceptive as spam may be, it would not be nearly as successful without the proper technology to mass produce the fraudulent emails and send them to as many potential recipients as possible. Most email users can see through the deception present in such scams; the trick lies in getting emails out to millions of potential victims a day, adding to the probability that just one recipient will be fooled. Such requirements for mass sending depend on more than just the perpetrator's own legitimate computational resources. These schemes largely depend on the availability of infected PCs working towards the spammer's ends in parallel, giving the spammer a wider audience than could otherwise be acquired. These infected PCs, called botnets, have given spammers exponential power for mass mailing purposes, and spam would be greatly lessened without them.

Botnets

A botnet is a network of computers connected to the internet infected by malicious software that allows them to be remotely controlled by a single hacker, called a bot master or bot herder. Each individual infected PC is called a zombie, and has a bot client installed on it. The goal of the bot master, for starters, is to collect as many zombie computers as possible, which he/she can then utilize for just about any cybercrime scheme imaginable. A bot herder can control up to millions of infected PCs simultaneously, which can all be used in parallel to accomplish large tasks that are beyond the computing power of just a single PC (Schiller, Binkly, Harley, Evron, Bradley, Willems & Cross, 2007).

Botnets can be used for various ends, such as distributed denial of service attacks (flooding a server or website with signals to slow it to a crawl), spyware to collect passwords and financial information, or they can be sold to other cybercriminals (some infect PCs just so they can sell the botnet bots). However, one of the most common purposes of the botnet is to send spam (Schiller et al., 2007). In fact, 80% of all spam is sent from botnets (IronPort Systems, 2006). Sending millions of emails a day would not be possible from the cybercriminals own computer, or even from several rented SMTP servers hosted specifically to send spam. The most effective means to mass mail victims is to command a distributed network of thousands of bot clients to send spam day and night, completely without the perpetrators own computer. Not only is this faster, but it means the spam is being sent from computers the spammer does not own, and is difficult to trace back to the spammer. It is estimated that about 15-25% of internet ready computers are infected with a botnet (Weber, 2007).

A botnet bot can facilitate spam operations in various ways. It can automatically register for multiple email accounts online to send spam from. It can crawl random internet sites to search for email addresses in plain sight that it harvests. It can even open the host user's contacts files and extract all email addresses from it. Spam might even be sent from the victim's own email account, sending messages to everyone in the victim's contacts list posing as the victim (Schiller et al., 2007). This can make for more deceptive fraud when a spam message is disguised as a trusted friend or coworker.

When a bot client has mined multiple emails from either contacts lists or the internet, the bot herder can then use them for one or more of three things. The spammer can send test messages (emails that are blank) to the email addresses to see which ones do not bounce (indicating that they belong to a real person). If the spammer has a large enough list of genuine email addresses for real persons, the spammer can also choose to sell the email list on the cybercrime market to other spammers. And of course, he/she can upload various email lists to all his bots and have them start spamming everyone on those lists repeatedly. Often botnets are modularized, meaning they can receive updates that the spammer chooses based on his desired requirements for that particular bot. For instance, a bot installed on a computer that did not have any spam functionality can be sent a spam module by the bot master so that it can then begin to spam multiple recipients (Schiller et al., 2007). Many botnets begin with stripped down functionality (for faster installs over a network), only updating themselves as needed.

Botnets give spammers a large advantage in their commercial pursuits. Without botnets, spam might be reduced by as much as 80% (at least at first) (Email metrics program, 2007). However, with as many as 25% of all computers connected to the internet infected with a bot (Weber, 2007), substantially reducing such a threat has remained elusive. But antispam solutions are available, and like any technology, there is a race to improve such antispam methods. The arms race between spam and antispam technology has seen large strides made by both sides of this conflict.

DEFENSES AGAINST SPAM

The problem for most email users with spam is the time it takes to sort through it all. It is easy to mass produce email spam by the spammer, it is much more difficult to go through daily received emails and determine which is spam and which is ham (desirable email) on a case by case basis. Naturally the most common defenses against spam are automating processes that decide for the user whether to classify an email as spam or not. Software can process data at a much higher rate, and so spam filters and authentication software have been created to sort incoming messages for the user. Technological defenses against spam are by far the more prevalent.

The preferred and most highly rated defense against spam according to businesses would be spam filters (Siponen & Stucke, 2006). Filters apply some algorithm to classify a message as spam or not, and removes these undesirable messages before they can reach the user's inbox. The first and most common filters are Bayesian filters that determine the probability that a message is spam. This is done in sequential steps. For example, first the probability that an email is spam regardless of content is calculated, then the probability that a word used in the message is spam is calculated, then the probability that the message appears in any email is calculated, and so forth (Sahami, Dumais, Heckerman & Horvitz, 1998). Messages with a high probability of being spam are removed from the recipient's inbox, or marked in some way.

Spammers have since adapted to this technology, crafting spam messages that can better bypass filters. Some may attach an image in the spam email with text on it advertising a website (Chitu, 2007). Most filters can only read plaintext, so text displayed on an image is unreadable to some filters. Use of images preceded an increase of traffic on email filters by 334% in 2006 (Mosher, 2007). Spammers may also write software that automatically changes the ordering and spelling of words and sentences before sending out messages, so that almost every single message is unique (Kestenbaum, 2006). This can also confuse spam filters.

Another method of determining whether an email message is unwanted is checking identifying features of the email. Called authentication, filters check header information regarding a message and check to see if any of it has been spoofed or does not come from a trusted source. For instance, a third of all email has sender ID authentication applied to it, where the domain name in the from field of a message (sender's email address) is compared with the domain name servers (DNS, associates domains with IP addresses) for that domain (Lemos, 2006). If that DNS does not allow the IP address associated with the email message, then the from field has been spoofed. Another method includes comparing an email's return address with a list of trusted domain names registered with the spam service. If the return address is not on the trusted list, the message is rejected (Wong & Schlitt, 2006).

Perhaps one of the most effective and latest technological defenses against spam would be user generated spam filters. Google's email account service (Gmail) implements this functionality by recording in a composite list all messages ever marked as spam by any Gmail user. The more Gmail users mark a specific message as spam, the greater the probability that Google Mail will block any further transmissions of that particular message from reaching Gmail user inboxes (Chitu, 2007). One user can't hope to mark all spam in the world to be blacklisted, but all email users collectively over the internet may very well be able to do this.

Anti spam technologies attempt to prevent spam messages from ever reaching their intended recipients. However, these technologies do nothing to stop spammers from sending messages in the first place. While anti spam technologies have been under development since spam first took root, alternate anti-spam methods, such as stopping the spammers themselves through law enforcement, has received much less attention. While anti spam technology seeks to stop spam, anti spam laws seek to stop the spammer. Legal measures of combating spam are still relatively new in the world. The efficacy of legal measures has yet to be fully tested and improved upon.

CHAPTER 3: THE CAN SPAM ACT

The anti-spam laws that concern this research are those outlined in the United States Controlling the Assault of Non-Solicited Pornography and Marketing Act (CAN SPAM Act) of 2003. The legislation was passed by Congress in 2003 and subsequently went into effect on January 1, 2004. The laws detailed in the CAN SPAM Act list the rules and regulations commercial electronic mail senders and advertisers must adhere to in order to continue transmitting unsolicited emails. Content of emails must be accurate and truthful. Commercial emailers must also fulfill requests by recipients to no longer send them messages. Violators of the CAN SPAM Act can be fined for the damages or per email transmitted, and may also receive a prison sentence for aggravating factors.

Leading up to the passing of the CAN SPAM Act, the United States Congress found that the expansion and growing pervasiveness of the internet in both people's personal lives and in global commerce were considerations of growing importance for legal considerations. Deliberations in congress in 2003 led to some conclusions about the importance of the internet and electronic commerce. The CAN SPAM Act of 2003 states the congressional finding that the internet comes with it the right to free speech and expression. Because of this freedom of information, commerce has witnessed significant progress, and was and is expected to continue such growth. Commercial email was considered by congress to be a valid means by which to market and advertise one's business.

However, Congress found that in order for electronic commerce to continue its natural trajectory of growth, both users and ISPs must be guaranteed that their internet activities are able to run efficiently and predictably. Unsolicited commercial mail can have costs to both the intended recipients and internet service providers. Such bulk email may congest internet activity and compromise the speed of such services, or accrue additional costs for ISPs. Sorting through and managing unsolicited emails can also accrue costs in time and effort to the recipients of the mail, both with regards to the storage and retrieval of large amounts of electronic messages.

In addition to costs of the messages themselves, the content or nature by which they are sent can be problematic. The content of some electronic messages may be obscene or unsavory, in which case recipients may wish to disallow such messages from reaching them or their children. Some senders of bulk email do not provide avenues to request the recipient no longer receive any further messages from the sender. Many senders disguise their message headers to obscure the transmissions originating location or source. Some messages are also sent from illegally compromised (hacked) computers without authorization from the actual computer owner.

In light of these conclusions and concerns, it was decided that there was a substantial government interest in regulating commercial electronic mail. It was decided that bulk email senders should not be permitted to mislead their recipients, and that those recipients should be able to decline to receive further messages from the sender. The CAN SPAM Act was created to address these concerns (CAN SPAM Act of 2003).

Prior to the passing of the CAN SPAM Act, spam was formerly regulated by each state’s laws individually. Spam sent or received in one state was determined to be illegal or not by that state. The CAN SPAM Act preempts the majority of state laws that deal with electronic mail. The aspects of bulk commercial email that are left to the states to regulate include laws that enforce honesty in email. State laws that regulate the sending of email with false headers, obscured addresses or deception in the content of the emails themselves are still left up to the states, so long as they have such laws. Additionally, CAN SPAM does not supersede state laws that address fraud or computer crime (CAN SPAM Act of 2003).

While the CAN SPAM Act replaces most state anti-spam law, it is second to all Federal laws and offences. Federal laws on obscenity and the sexual exploitation of children are to take precedence over any regulations over such crimes the CAN SPAM Act may provide.

REGULATIONS OF SPAM

There are three levels of rules detailed in the CAN SPAM Act, with different gradations of severity considered for each of the offences. First there are the general requirements for the sending of commercial messages, which deals with the content of the emails sent. Second there are aggravating factors that can compound additional penalties when deciding on a sentence. These aggravations regulate the way in which electronic mail was sent. Lastly there is an additional penalty for the sending of pornographic content in email transmissions that does not adhere to CAN SPAM regulations.

Regulations of Commercial Email

The basic rules governing the sending of bulk email starts in the CAN SPAM Act with the rule with the strongest penalty of the basic regulations. This rule states that false headers are forbidden in the sending of commercial mail. Headers include the from address, to address, bounce address, and other routing information contained in each email sent. Senders may spoof their own email address, domain name, IP address, or anything else they think will help them conceal their originating location or mimic some source trusted by the recipient (e.g. a bank) to facilitate deception. Any such deception or tampering with email header information is forbidden.

Restrictions also apply to the content of an email's subject heading as well. Senders are not allowed to write subject headings that can be considered to mislead a recipient about the contents of an email message. Subject headings must be either descriptive or related to the contents in the body of the email. Email subjects, often being the first part of an email that is read before opening it, should indicate that the message is an advertisement so that recipients may choose not to open the message, saving them time.

The sender must also provide a real return address in the email. The from field in an email must be a genuine email address, must be the actual address of the sender, and the sender must be able to receive and check messages sent to that return address. The recipient must be able to reply to the advertised email, and be capable of sending a message to the sender in response to receiving email from the sender. Spam emails sent from an address that is not owned by the sender (from a botnet or hacked email account), or emails sent from an address that the sender has no intention of checking (such as hundreds of online email accounts used to automatically send spam in bulk but never used to receive mail), are disallowed.

Spammers must provide an opt-out method for those receiving their messages. While the spammer does not require opt-in from his recipients, they must be able to opt-out. An opt-in counterpart to this rule is affirmative consent beforehand that the recipient wishes to receive commercial email. Opt-out, on the other hand, is the ability of the user, and the willingness of the sender, to discontinue spamming the recipient upon request. How the recipients are supposed to opt-out is not detailed in the 2003 version of the CAN SPAM Act, just so long as those avenues are available. Emails could provide a link to a website where the user can enter and submit the email address they wish to no longer receive spam from. Or the user may merely be advised to reply to the message with a request to be taken off the spammer's listserv.

The sender must also provide a valid physical postal address in the body of the email where they can be contacted. This is to reduce more of the anonymity that electronic mail provides. Lastly, the message or subject must contain self identification that the email is an advertisement. There does not have to be a notice that the message is an advertisement, however, if the recipient gave prior consent to receiving the email.

Aggravating Factors of Illegal Spam

Aggravating offences committed can triple the maximum fine under the CAN SPAM Act. The first among these factors are address harvesting or dictionary attacks. These two methods used by spammers are employed to acquire as many email addresses as possible, regardless of whether they are real or not. Quantity is preferred over quality, and amassing email accounts is desirable so that they may be added to spam lists or sold to other spammers.

Address harvesting is the process of running automated software that “crawls” the internet looking for email addresses. Crawling is an act carried out by a bot that scans internet web pages, reads all text on those pages (both display text and the HTML code that renders it), and follows all links on those pages. If the pages that are linked to have more links, those are often followed by the crawler as well. Once the crawler finds a recognizable email address, the software saves it to a list to be spammed.

Dictionary attacks, on the other hand, are easier to carry out, but result in fewer genuine email addresses acquired. This involves software that generates email addresses from a dictionary. The software will use common email domain names (e.g. @gmail.com, @yahoo.com), and prefix a word from the dictionary onto that string as the user name (word@gmail.com). For the half a million words in the English dictionary, many of these email addresses are bound to belong to real people.

A second aggravation is the automated creation of multiple email accounts. This involves more software automation, except that the script registers for multiple free online email accounts that are later used to send spam. Sending thousands of emails from one free internet email account may not be feasible, as the server may deny the spammer so much traffic. But if the process can be automated across many email accounts from varying email service providers, then the goal of mass mailing can more easily be reached.

The third and last aggravation is the relaying of spam without authorization through either a protected computer or a network. This could be either botnets or other malware remotely installed on a protected computer or network, or open relays that transmit email from the sender to the marked destination, removing originating header information. The term protected computer here is taken from Title 18, Section 1030 of the Computer Fraud and Abuse Act of 2006. A protected computer is defined as any computer owned by the government, a financial institution, an international or interstate commerce institution, or a computer used in communications.

Sexually Explicit Material

Most of the violations detailed so far mostly come with a cost of a monetary fine for breaking them. However, Congress deemed sexually explicit material in electronic spam to warrant additional penalties, including a prison sentence. Any bulk emailer who transmits electronic messages with sexually explicit material (images, etc.) must provide a warning label in the subject heading notifying users to its contents before they open it. This warning label is not required if the recipient has opted-in to receive the message beforehand. Violations of this rule can result in a prison sentence of up to a maximum of five years.

PENALTIES FOR CAN SPAM VIOLATIONS

There are three bodies authorized to pursue criminal spammers legally for violations of the rules detailed in the sections above. They include (1) attorney general for within state violations, (2) internet service providers, and (3) the Federal Trade Commission. There are also rewards offered to any person who reports a spammer to one of the given authorities that leads to a successful conviction.

Enforcement by States

The attorney general may bring civil action towards a spammer on behalf of the victims residing in that state. The state may fine for the amount of damages the offender contributed to all victim residents of that state, or the amount determined by the number of emails sent to recipients within that state multiplied by $250. This total amount may not exceed two million dollars, unless the offender sent the messages with false header information. Then the amount is unlimited. The fines determined above may be tripled if it is found that the offender both willfully and knowingly transmitted the emails (as opposed to someone who hired the spammer without knowledge of his illegal methods). The fines may also be tripled for committing any of the aggravating factors in the CAN SPAM Act.

Enforcement by Internet Service Providers

Internet service providers (ISPs) are to file suit in a district court with jurisdiction over the defendant. The ISPs can sue for the damages cost to them from the spammer abusing their services with bulk email and malware transmission. Instead of suing for the damages cost to them, ISPs can opt to sue for the amount of $100 per email sent over their networks it the defendant used false headers, or $25 per email for any other violation other than false headers. The amount determined may not exceed one million dollars, unless there were false headers. As with the case of the attorney general, the fines can be triples if any of the aggravations were committed by the offender, or if the offender was knowing and willing in committing the offence.

Enforcement by the Federal Trade Commission

The Federal Trade Commission (FTC) may enforce any violations in the Federal Trade Commission Act with penalties in the Federal Trade Commission Act as if they were written in the CAN SPAM Act. The FTC has jurisdiction in any offences affecting interstate commerce, which includes spam. The legal actions by the FTC are detailed in Title 18, Section 1037 of the United States Code (Fraud and Related Activity, 2006). Any person that obstructs interstate commerce to send spam can fall under the FTCs jurisdiction. Penalties can include up to a maximum of five years in prison plus a fine if the defendant is a repeat offender, or committed the crime in furtherance of a felony. The fine is only a maximum of a three years prison sentence if the offender used unauthorized access to a computer, registered multiple email accounts to send spam, or accrued $5,000 or more in damages. Without any of the aggravations above, the sentence can only be up to one year imprisonment.

There is one additional law in the CAN SPAM Act that only the FTC may enforce. This regards affiliate spam, which is the legality of hiring or contracting another to send spam on one’s behalf. This is only illegal if the defendant knew the affiliate sent spam, received a monetary benefit from the illegal spam, and took no action to prevent the spammer from carrying out these illicit activities.

Rewards for Reporting Violations

While only ISPs, the FTC, and states may prosecute spammers, individuals may still find incentive to report spamming offences to the Federal Trade Commission. The criminal informant who knows of a suspected illegal spammer’s actions must identify the suspect to the FTC. The information must also lead to the successful collection of a civil penalty by the FTC. Given these conditions, the criminal informant is entitled to 20% of the civil penalty collected from the offender.

EFFICACY OF THE CAN SPAM ACT

The CAN SPAM Act of 2003 contains plans by the Federal Trade Commission to carry out studies of the Act's efficacy following its release. In addition to investigating whether the CAN SPAM Act had an impact on overall spam, it is also written that the FTC should investigate the merits of a do not email registry, similar to the do not call registry. It was soon determined that such a registry was unadvisable, since a do-not-mail list would be publicly available for spammers to find (Muris, Thompson, Swindle, Leary & Horbour, 2004). Such a list might increase spam, in fact.

By December, 2005, the FTC had finished with its report to Congress on the outcome of the Act. There was both good and bad news in the FTC's findings. Among the reasons reported to Congress of the Act's success included the fact that there had been over fifty prosecutions of illegal spammers at the time of the writing. It was concluded that many legitimate online marketers were now adhering to the CAN SPAM Act laws (Majoras, Leary, Harbour & Leibowitz, 2005). While the full-time cybercriminals and spammers might not be as likely to be deterred by the Act, those with legitimate businesses would be successfully directed to market their products in a legal manner.

Among the actual content of spam itself, the FTC also found some changes. The FTC concluded that there had been an observed decrease in sexually explicit content present in spam email. The FTC also claimed that spam rates appeared to have begun to level off, slowing in its ever steady increase over time. The FTC also acknowledged that the amount of spam received in inboxes had decreased due to better spam filtering technologies (Majoras et al., 2005).

Among some of the limitations of their findings, the FTC admitted that there was more malware attached in email spam, not just spamvertised products. Also, there had been no noticeable decrease in the amount of falsified information provided when registering domain names (Majoras et al., 2005). That is, registering for domain names with falsely identifying information to allow for anonymity of the registrant. It should be noted that none of the FTC's reports mention statistical significance.

The FTC was not the only authority to test the Act's efficacy. Other independent researchers mostly consisted of computer security firms and spam filtering technology companies. Two of the questions of most interest were whether spam rates had been affected or whether compliance with the CAN SPAM Act had gone up since the Act's arrival.

As to what spam rates looked like after the Act, it would seem that spam volume had in fact gone up. According to Scott Chasin, Chief Technology Officer of the spam and virus filtering firm MX Logic, spam had allegedly increased (Gross, 2004). According to MessageLabs, an anti-spam and security company, spam had increased from 50-60% to 80% one year after the CAN SPAM Act (Zeller, 2005).

As for compliance with the CAN SPAM Act's regulations, MXLogic reported that more than 99% of spam did not comply with one or more regulations in the CAN SPAM (Gross, 2004). This report was based on a random sample of 1,000 spam messages analyzed for compliance with the Act. Commtouch Software, another spam filtering company, customized software that analyzed millions of emails to measure compliance with CAN SPAM law, such as containing return addresses and meaningful subject lines. Commtouch also found that less than 1% of the emails complied. A third spam filtering vendor, Audiotrieve, found 10% compliance based on 1,000 analyzed messages (Gross, 2004). Yet, none of these studies mention what CAN SPAM compliance is down from, if it was down at all. There was no baseline of compliance rates before the Act. Additionally, these reports were only made during early 2004, shortly after the Act was instated.

Another account of compliance with a slightly longer follow up was done by Grimes (2007). During the first six months the CAN SPAM Act was in effect, five honeypot email accounts were created, which are email accounts used to bait spammers into emailing them by posting them on public forums, etc. for crawlers to harvest. After the six months following the Act, 1,100 unsolicited commercial emails were randomly selected from the five email accounts. A second follow up two years later was also carried out, this time with 800 randomly selected spam emails.

It was found that during the “baseline” of six months after the Act, only 14.3% of spam complied with the Act. The two year follow up found that that number had fallen to 5.7% compliance (Grimes, 2007). Unfortunately, there is no mention of statistical significance in this report. There were also no spam data gathered before the CAN SPAM Act went into effect, only six months after with a two year follow up.

All of the studies mentioned thus far seem to have a few limitations. None of them mention accepted statistical conventions used in science. Some of them have no baseline to contrast their findings with. And also lacking from the literature seems to be an address of the question of displacement of spam operations. It is possible that the CAN SPAM Act did not deter spam rates, but rather moved them outside of the United States beyond the CAN SPAM's influence. Spam rates and compliance with spam law are important, but these things could also be affected by the jurisdiction in which spam law, such as the CAN SPAM Act, applies.

CHAPTER 4: THEORIES OF SPAM

The theory chosen to help understand illegal spamming was routine activities theory, which generally has been used to predict cybercrime overall. Routine activities theory combines three factors that can predict whether a criminal act is committed. These factors are a combination of the environmental ease and desirability in which a crime can be successfully carried out, and the willingness of the offender to capitalize on such an opportunity. These factors are (1) suitable targets (e.g. to steal), (2) a motivated offender (e.g. willing to steal), and (3) absence of capable guardians (e.g. to guard something from being stolen) (Cohen & Felson, 1979).

The routine activities theory (RAT) has been easily applied to cybercrime since RAT relies both on the environment and the individual. Since the internet is an environment so large yet easily traversable that countless desirable or undesirable situations can be found with only a few searches and a few clicks. Since approximately 1.6 billion people use the internet now (Internet World Stats, 2009), many of them are guaranteed to be potential motivated offenders which can go just about anywhere on the environment of the internet. This is no less true for illegal spam, as many on the internet can both send and receive email. Even without email, there are other types of electronic spam, such as chat, social networking, search engine, and forum spam.

ROUTINE ACTIVITIES THEORY

The three predictors for routine activities theory have been used to describe and seek to understand crime. The three factors, motivated offenders, suitable targets, and absence of capable guardians, will be discussed in turn.

Motivated Offender

This describes someone willing to commit the crime in question in the first place. The other two factors, absence of a capable guardian and presence of a suitable target, matter little if the offender does not perceive there to be an absence of a capable guardian or the presence of a suitable target. This is crime according to the offender's point of view. This can be the potential offender's temperament, morality, upbringing, personality, attitudes, intelligence, or anything that describes him that might explain why the subject was willing to offend (Akers & Sellers, 2004).

Suitable Targets

A target is something of worth that someone might want to obtain (Cohen & Felson). This can be merchandise, money, or maybe even something less tangible, such as status or recognition. A target that is also *suitable* is one that can be easily acquired illegally, according to routine activities theory. Thus, suitable targets are ones than can be easily stolen instead of bought, or acquired in some antisocial way that is more enticing than other targets.

Felson and Claerke (1998) have broken the variable of suitable target into four constituent parts, known by the acronym VIVA. VIVA stands for value, inertia, visibility, and access. Value represents how much worth the motivated offender places on the target to determine whether it is suitable or not. If the target is worth a lot of money, for instance, then the offender might place more value on it. Inertia has to do with the ability to move the target. If the target is lightweight, then the offender may be able to quickly get it away before a guardian can spot the illegal act. Visibility is how easy the target is to spot. If a CD is placed in a window, then the offender is more likely to notice it. A target cannot be suitable if an offender does not know of it. Lastly there is the accessibility of the target, whether it is easy to steal or not, such as merchandise placed near a store exit. If the offender would have a hard time taking the object or acquiring the illicit goal, it is not likely to be considered a suitable target (Felson & Clarke, 1998).

Absence of Capable Guardians

A capable guardian is someone or something that is able to defend the suitable target from the motivated offender. This is often law enforcement, or absence thereof, involved in a crime being committed (Cohen & Felson, 1979). Guardians can be just about anything. Examples of capable guardians include police patrols, security guards, locks, fences, lighting, alarm systems, watchful neighbors, security cameras, or even potential victims able to fend off an assailant.

APPLICABILITY TO CYBERCRIME

With the advent of the internet, a new world opened up to countless participants. With billions of web pages and users available online, any number of them can be considered suitable targets. Any number of web pages or users can have poor security practices associated with their implementation, with any number of vulnerabilities to exploit. Likewise, any number of web pages or participants can be malicious, and willing to take advantages of the vulnerabilities available in cyberspace. The internet has few borders, so targets and offenders can easily meet up.

Motivated Offender

The lack of geographical limitations makes motivated offenders, no matter how rare in the world, a present danger to all users. Cybercrime used to require many skills to carry it out successfully. Now, phishing kits, crimeware, and cybercrime services are all available to anyone willing to pay for them, regardless of skill. Much of crimeware is even free and some is open source, making someone with just a little know how about programming and the internet capable of deploying illegal software to exploit a given target. Other crimes that require a certain skill set to carry out, such as securities or bank fraud, often require existing positions of authority to learn and find use for. Not so with cybercrime, anyone with the patience can learn how to commit low risk high yield cybercrime. Many carding forums (forums for buying and selling illegal goods) have a section of tutorials on everything from creating exploits to installing a botnet. The ease with which someone can become an offender is probably pretty motivating.

Also the anonymity of cyberspace may likely embolden those who may take criminal acts into consideration. It has been noted that interacting with other users over the internet contributes to what is termed the online disinhibition effect (Suler, 2004). This is the noticeable decrease in inhibitions when a user interacts with another online. People may feel less resistance to being rude, offensive, or even threatening. This disinhibition can also make cybercrime easier to someone who would not otherwise take advantage of someone when the interaction was face to face. Thus, online offenders might be more motivated than real life offenders.

Not only is there often low risk for committing crimes online, but the added motivation of the high potential for making money while doing it likely adds to the reasons to commit cybercrime or send spam. Considering the average spammer can make up to a six figure income from home on the internet (Kanish, Kreibich, Levchenko, Enright, Voelker, Paxson, & Savage, 2008), the motivations are obvious. Since spamming can be a high paying job that can be carried out from home and requires a range of skills where many people can find their niche, spam could be considered a very desirable career choice.

Suitable Targets

There are more people online every day, adding more content to the World Wide Web. This can mean that there are more people who use online banking, who shop online, and more people accustomed to entering sensitive information, such as credit card information, onto a web form. People are more likely to buy products online, which means more people willing to buy Viagra and other spamvertised products. This means there is more money, and more people willing to give and exchange money, on the internet. The type of suitable targets in cybercrime are many. They can include bank logins, credit card information, secret questions, virtual merchandise in virtual worlds such as Second Life and World of Warcraft, email accounts, software, server space and storage, bullet proof hosting, botnet installs, stolen physical goods to order, etc. As the internet continues to expand, there will be even more targets that offenders desire.

Absence of Capable Guardians

A capable guardian relevant to cybercrime can be at three different levels. That is, the street smarts of the potential victim to guard from being taken advantage of, the technology the potential victim may employ to guard against attack, and the laws in place that can be enforced should the victim fall prey to a cyber attack. Capable users may be knowledgeable enough about social engineering to abstain from responding to 419 scam emails. They may also be willing to use the proper technologies, such as antivirus and regular software updates. However, the technologies themselves need to be capable guardians as well. Patches need to be available as soon as possible, and different antivirus and anomaly detection applications need to be able to recognize something suspicious. Lastly, when these first defenses fail, law enforcement needs to be able to punish cybercriminals frequently enough that they decide cybercrime does not pay. If law enforcement is to properly deter would be cybercriminals, the probability of arrest and conviction for each crime committed must be high, as well as appropriately severe and timely (Mendes, 2004).

The online disinhibition effect may also apply here as well. People might not venture alone at night in dangerous neighborhoods, but venturing alone on the internet does not feel so intimidating. The disinhibition might result in victims not using antivirus, or not checking the domain name of the bank site requesting them to enter their credit card information. However, certainly most users know not to fall for phishing scams, and many more are conscious about malware. However, scammers on the internet just need one in thousands of potential victims to fall into their trap, and download their malware, buy their product, or enter their information.

Research on RAT and Cybercrime

The growing literature suggests that routine activities theory can be successfully applied to cybercrime. Some of the supported forms of cybercrime include phishing scams, malware infections, and even cyber bullying. In Holt and Bossler (2009), online harassment was taken as a form of cybercrime, and routine activities theory was the underlying theory to help describe it. Five hundred seventy eight college students were selected and surveyed on their internet activities and computer skills (suitable targets). Some of the items measured were unrelated to subsequent victimization by online harassment, including owning a computer or not, internet speed available, amount of internet activity (shopping, video games, etc.) computer skills, and the use of firewalls and antivirus. The activities that increased the risk of cyber bullying were hours spent in chat rooms, involvement in computer deviance (piracy, guessing other people's passwords, etc.), friends being involved in computer deviance, and being female. So perhaps time spent in chat rooms (suitable targets) where harassment may occur, or associating oneself with those involved in computer deviance (motivated offenders) contributes to the likelihood of a form of cybercrime being carried out.

Phishing as a form of cybercrime has also been addressed in the literature. Hutchings and Hayes (2009) sampled 104 subjects taken from a telephone directory aged 18 and over who use the internet. The subjects were surveyed via a telephone interview. The first research hypothesis proposed was that the lower the level of computer and internet experience and the higher the use of internet banking, the higher the risk for victimization by phishing attacks. Unfortunately, there were not enough subjects in the 104 sample that had actually been a victim of phishing to fully address this question.

There was conclusive evidence for the second hypothesis, however. The more computer use and online banking a subject engaged in, the increased risk for receiving a phishing email. Internet use was not related to receiving such an email. Lastly, the third hypothesis that mail filters reduce the risk of receipt of a phishing solicitation was not supported. While filtering software may not have been a capable guardian in this case, some internet and computer activities may make being the recipient of phishing scams more probable.

Similar research has investigated the susceptibility of users to become victims of a malware infection (Bossler & Holt, 2007). The sample was of 788 college students that were given a self report survey. The outcome measure was any loss of data or time due to malware infection. Routine activities were measured such as internet connection and speed, shopping habits, email and chat use, involvement in programming, or use of social networking sites. Most of the routine activities measures were unrelated to subsequent malware victimization. Guardianship such as using AV, protecting passwords, and computer skills was also measured. These forms of guardianship were also not related to subsequent victimization. It could be that antivirus was installed only after victimization during the time of the survey. Lastly, questions about deviance such as hacking, downloading porn, and piracy were also included on the survey. Deviant computer behavior tended to co-occur with malware victimization and supported the hypotheses. Hacking, unauthorized access to the internet, and pirating media were all weakly related to victimization. Having friends who similarly engaged in deviant behavior also increased the risk of infection. Similar to online harassment mentioned in Holt and Bossler (2009), deviant behavior, or association with deviants (motivated offenders), increases one's potential as a victim of cybercrime.

APPLICABILITY TO THE CAN SPAM ACT

The most noteworthy criterion for routine activities theory in its applicability to this research would be that of capable guardians. The CAN SPAM Act, hopefully serving as an adequate capable guardian, should be expected to deter the plethora of motivated offenders from pursuing targets of cybercrime. And if the CAN SPAM Act does not deter those offenders, it should be expected to subsequently punish them for their transgressions.

There are three research questions that this paper is aimed to answer. They are (1) whether CAN SPAM has affected spam rates, (2) whether CAN SPAM has affected spam compliance with spam law, and (3) whether CAN SPAM has affected the locality of originating spam emails. All three potential effects of the spam legislation involve the capability of the Act (capable guardian) to deter spammers in some way. If the first research question is found true, and that spam rates have decreased, it may be that the CAN SPAM Act deterred enough individual spammers so that some of them determined to discontinue their spam operations. If the second question holds, and that CAN SPAM compliance has gone up, then perhaps spammers decided to send spam legally and not take on the risks of violating the conditions in the Act. Lastly, if the third question holds, and spam is sent from within the United States less often following the Act, then perhaps spammers within the United States (where the CAN SPAM Act applies) had opted to discontinue their operations for fear of punishment. Or perhaps spammers attempted to move their botnets overseas, hopefully obscuring their source further.

Spam law could easily be considered a capable guardian. While this research does not account for the remaining two RAT criteria (targets and offenders), hopefully the CAN SPAM Act might have positive effects on these other two factors. If it does, targets might look less suitable and offenders might be less willing to transgress against anti-spam law.

CHAPTER 5: METHODS

Sample

The sample for the present study consisted of 2,071,965 email messages sent and received between February, 1998 and December, 2008. The sample was acquired from Untroubled Software, a software security and optimization website that also provides downloadable spam archives for researchers. The data were retrieved on January 3, 2009 from http://untroubled.org/spam, and consists of millions of text files. Each text file represents each individual email message sent. The files contain all text transmitted during the sending of each email, including body messages, headers, and even attachments (in the form of machine language in the text file). A single spammer could have sent multiple spam messages, and duplicate messages do exist in the data. Therefore, the data are not independent.

According to the author of Untroubled Software (http://untroubled.org/spam/), the spam archives were gathered via multiple bait email addresses, which are email accounts created with the sole purpose of baiting spammers to add the baited address to their spam lists. Baiting is most commonly achieved by posting the email address publicly online, usually on searchable forums or websites. Addresses in plain sight online then are at risk of being harvested by crawlers, software bots that scan the internet and record email addresses found; to later send spam to.

Since the date each message was received was important for the purposes of this research, some cases had to be eliminated due to false or missing received dates. There were 4,959 cases that were found to have no date contained in the email headers.

Additionally, much of the normal header information expected in emails was also missing in these particular emails, such as subject lines, hops/IP addresses, etc. Likely this was due to spoofing of the messages or general tampering by spammers.

There were also 2 more messages that had to be removed, since the dates were clearly wrong/tampered with. They both had values of “12/95/2005”, and thus could not be placed on a chronological continuum. After elimination, 2,067,004 spam messages remained in the sample.

The data were further aggregated by month for the purposes of time series analysis. Grouping the eleven years (1998-2008) of data by month resulted in a series of 131 months total. Of the remaining messages in the sample, there was an average of 15,779 messages received per month. Before the CAN SPAM Act was in effect, the average number of messages received per month was 1,373, whereas after the Act was passed an average of 32,825 messages were received per month.

There was an additional dataset acquired for this research from the Federal Trade Commission (FTC). The FTC had conducted their own analysis on the state of spam, and a subsequent Freedom of Information Act request allowed a portion of the FTC’s data to be used in this study. The data consisted of a summary of 479,701,868 emails collected by the FTC between the beginning of 2000 and the end of 2007. The data acquired for this research contains the number of emails received each day between the years 2000 and 2007.

Procedures

The sample of spam email messages was measured by a program written in PERL (Practical Extraction and Report Language). The script was written so that it would scan all files within the same working directory as itself and any files contained within folders and subfolders also in the current working directory as the PERL script. The spam messages themselves were divided up by year, with folders for each of the eleven years. Within the yearly folders, the emails were further divided into folders by month. The emails themselves were contained in text messages named with a timestamp each message was received followed by an ID number. The script written for this research subsequently scanned all email messages, starting by yearly folders and then by monthly folders, until all messages were scanned in chronological order one by one. While scanning each message, the script records eight variables for that given message in a comma separated values file as a single row. After completion of the scan, the CSV file can later be manipulated and imported into an application for statistical analysis.

The script also uses a database of world IP addresses to create a variable representing the country the spam message was sent from. The database contained IP address codes and a lookup of global information about each address type, including the country of origin for a given computer's IP address. The database was downloaded from WebNet77 (http://software77.net/cgi-bin/ip-country/geo-ip.pl) on March 20, 2009. The database was converted into a text file where IP address information could be read from a table contained in the file for each individual email message.

After running the script on the spam sample, the software output was tested for interrater reliability to determine if the script's judgment matched that of a human evaluator's. A random sample of fifty emails was taken from the spam collection. One problem that emerged with this sample was that two of the eight variables in the sample of fifty were constant values, making a test for reliability impossible to compute. The two variables were CAN SPAM compliance measures, which represent whether a given email complies with a certain CAN SPAM law. Few emails complied with CAN SPAM laws in general, and so two of four compliance variables were constant, with zero compliance found for all fifty cases.

To correct this and generate a sample with at least some variability, an additional sample of nine emails that specifically were rated by the software to have complied with one or more CAN SPAM laws were randomly selected. These nine emails replaced nine existing emails randomly eliminated from the existing sample. There were a total of fifty emails. An independent rater was selected to code four of the compliance variables for the sample of fifty emails. The interrater reliability of each of these four variables is discussed below.

Measures and Variables

Date Received. The script was set to record the first date it could identify when scanning each message, starting its scan from the top of the message headers. Email headers have a timestamp for each hop, or each time a message is routed through a network on its way to its destination. Information for each hop is appended to the top of the email headers. Thus, the most recent time a message was transmitted can be assumed to be the topmost date and time recorded in an email's headers. The topmost date was assumed to be the date the message was received. All other dates below the first were ignored by the script.

IP Address. This variable was only recorded just in case there was a need to further inspect the country of origin variable. The IP address was used to look up the country from which the message was sent from. The IP address the script was set to record was the lowermost IP address found in the message's headers. Each hop of a message records the IP address of the server which handled that transmission. Newer hops are appended to the top of the message. Therefore, the last IP address found in a message's headers can be assumed to be the originating mail server from which the email was sent from. These IP addresses were identified and recorded in the CSV file by the software.

Country of Origin. After the recording of an email message's originating IP address, the address is looked up in a database containing geographic information about world IP addresses. Details about the database can be found in the Procedures section above. If an address can be successfully identified in the database, the name of the country associated with that IP address can be identified. The full name of the country is then read from the database and saved to the CSV file.

Sent from within the United States. Country of origin was computed as a second, dichotomous variable. The value is "1" if the IP address of a given message was geolocated to the United States. Otherwise "0" for not in the United States was used. IP addresses that could not be identified because they were invalid or missing were excluded from this variable.

Opt-Out Compliance. Opt-out compliance is a measure of whether the email in question complies with the CAN SPAM Act regulation of providing a suitable opt out option for recipients. Opt-out allows recipients to notify the sender that they no longer which to continue receiving spam messages, followed by a ten business day requirement of the spammer to discontinue sending to that recipient. The message was assumed to have a valid opt-out option if any of the following keywords were found in the body of the email message: “opt-out”, “opt out”, or “unsubscribe”. The keywords to identify were not case sensitive. The opt-out variable was represented as a dichotomous measure of compliance, zero for no compliance, and a one if any of the three strings above were identified.

The opt-out compliance variable was tested for interrater reliability from a random sample of fifty emails. An independent coder rated each of the fifty emails in terms of whether each complied with the CAN SPAM opt-out requirement. The reliability of this measure was high (Cohen’s kappa = .9, $p < .001$), suggesting that the software’s judgment was consistent with that of the human coder’s.

Percent Opt-Out Compliance Per Month. Opt-out compliance was aggregated into a percentage of all emails each month that complied with the opt-out requirement in the CAN SPAM Act. The count of all emails in compliance with this law divided by the total emails received that month was computed. Percentage per month was calculated and added to a separate time series so that an impact assessment could be carried out for this particular variable.

Valid Mailing Address Compliance. A second CAN SPAM Act compliance measure attempted to determine the emails compliance with the requirement that the sender provide his own valid physical mailing address in the body of the email. The script used a regular expression that could identify any string with the following pattern: a number of any length, followed by one or more spaces, then any valid acronym for an addresses direction (“NE”, “SW”, etc.), followed by another series of spaces of any length, followed by any number of any characters so long as there was no line break, followed by one or more spaces, and ending with a street suffix of some sort (“ave”, “st”, “apt”, etc.) all without any line breaks found within the string. If a string was found in the body of the email message that matched this description, the address variable was recorded as “1”. Otherwise it would be “0”.

A sample of fifty emails was used to test the address compliance variable for interrater reliability. The interrater reliability of this measure was relatively high (Cohen’s kappa = .73, $p < .001$), indicating that the software and an independent human coder agreed on most of the items.

Percent of Valid Mailing Address Compliance Per Month. The variable representing valid address compliance of the CAN SPAM Act was further aggregated by month by dividing the number of emails each month that complied with the law with the total number of emails each month. Percent of address compliance per month was to be used as a separate time series for impact assessment.

Accurate Subject Heading Compliance. The CAN SPAM Act requires all commercial email to have a subject heading that is related to the actual content of the email message. The script identified this level of compliance by checking to see if any word in the subject line was also found in the body of the email message. First the subject line was recorded by the software. Then the subject line was stripped of any common words (e.g. “the”, “from”, “and”, “for”, etc.). The remaining words were then exploded into an array (a list). This list was then compared with each and every word contained within the body of the email. If one subject word matched any word in the body, compliance was assumed. Compliance was a true or false variable (“1” being compliance found, “0” being noncompliance assumed).

The accurate subject heading measure was tested for interrater reliability. The reliability for this variable was very low (Cohen’s kappa = .29, *p* = .003), suggesting the software was not consistently successful at identifying compliance with this particular law. However, given that the reliability was still significantly different from zero, this variable was kept in the dataset for use in subsequent analysis.

Percent of Accurate Subject Heading Compliance Per Month. The accurate subject compliance measure was further aggregated by month as a percentage of total emails each month that complied with the meaningful subject heading law. The purpose of aggregating subject heading compliance was to conduct an intervention analysis on it in a time series design.

Notice of Advertisement Compliance. CAN SPAM requires commercial emails to identify themselves as advertisements. An email was assumed to have complied with this regulation if any of the following strings were found in either the body of the email message or the subject line: “advertisement”, “ad” surrounded by at least one space, comma, or colon on each side, and “adv” surrounded by at least one space, comma, or colon on each side. The last two strings had to have been surrounded by spaces or similar characters to avoid false positives of words that contain those letters (“add”, “adverture”, etc.). If any of these three expressions were matched, “1” for compliance assumed was recorded. Otherwise a record of “0” for noncompliance was written to file.

The notice of advertisement measure was tested for interrater reliability. It was found that the reliability of this measure was not significant (Cohen’s kappa = .16, *p* = .241). This suggests that the ability of the software and the human rater to agree were on average no better than chance. Because of the lack of reliability, the advertisement compliance variable was excluded from further use in any subsequent analysis.

Scale of Compliance. The four dichotomous compliance variables described above were combined into a single scale variable of values between 0 and 4. All four compliance variables are binary zero or one values. On the scale of compliance, a “4” would be the highest level of compliance measured for a given email message. A potential lowest score possible of “0” would indicate that the email message complies with none of the four measures of compliance tested for.

However, to assess the reliability of the compliance scale, the index variable was tested to determine the degree to which each of the four compliance items were related to one another. The reliability of this measure was very low at Cronbach’s alpha = .06. This low reliability suggests that reasons for compliance on one of the four scale items were unrelated to reasons for complying or not complying with the remaining three compliance scale items. Given the extremely poor reliability of this index, the scale variable was removed from consideration in the analysis phase of the research.

Number of Messages per Month. To address each of the three research questions in this article, some of the variables had to be aggregated by month. Number of messages per month is the total number of spam emails received for each month. This measure will be used as a spam rate to test whether spam appears to change in frequency following the introduction of the CAN SPAM Act.

The Federal Trade Commission data were also needed to compare with this variable, and the FTC dataset had spam rate aggregated by month as well. The two spam rate variables were to be tested for their degree of correlation. The purpose of this analysis would be to determine if spam rates for both datasets are highly correlated. High correlation can be taken to mean that likely the two datasets measure the same thing, that being spam rates within the United States.

After inspection of the spam rate time series, however, there appeared to be an inordinately large spike in spam rates for three months between August and October 2006 (refer to Appendix A.1). The author who collected the spam data and provided it online for research purposes (http://untroubled.org/spam), makes note that the unusual spike was due to the use of a wildcard email address enabled in 2006. Wildcard addresses allow misspelling of an email address username (user@domain.com), to be successfully routed to the owner of the domain, regardless of the misspelling. This accounted for the inordinate increase in spam (mostly duplicates) in 2006. In 2007, the wildcard addressing was disabled, which was followed by a subsequent decrease in spam rates.

The abnormal spike is only observably present in the data for three months, between August and October of 2006. It is not discernable how much of the spike is attributable to the wild card address being enabled and whether other trends also played a role. There appears to be a moderate spike in the FTC's version of the spam data as well. Other factors may have helped exacerbate this abnormal shift in spam rates. Near the time of the spike, a new bot variant called the Mocbot worm was being spread via a technique that exploited a UPnP vulnerability of unpatched Windows servers near the time this spike was observed in the data (Stewart, 2006). A noticeable rise in spam at the time has been attributed to this botnet's spread at the time.

Regardless of the exact nature of this spike, the spam rate time series posed a problem of heteroscedasticity. This was a problem since the unusual spike was too violent of a deviation from the normal daily trends of spam in the sample. Before analysis, the spam rate variable was logarithmically transformed. This transformation allowed the data to be evened out while still maintaining the proportion of each month's spam rate relative to spam rates in contiguous months. The transformed data appeared to be sufficiently smoothed, which can be seen in Appendix A.2.

Percent from within the United States per Month. The measure of whether a message was sent from the United States was computed as a percentage of messages assumed to have been sent from within the United States for each of the 131 months. Percent US was calculated by dividing the number of IP addresses identified as coming from within the United States by the sum of this number plus the number of countries not of the United States. If a country could not be identified by the software (invalid IP address, no IP address found, etc.), then those cases were not computed in the percentage. There were a total of 107,971 email messages that could not be geolocated to a specific country. Of those that could be identified, 904,974 were found to be from within the United States, and 1,054,059 were determined to be from countries other than the United States.

Design

An interrupted time series design was used to test the three research questions. There were five time series to conduct, one for spam rates, three for each of the individual compliance related variables, and a final time series for the percentage of spammers within the United States.

The intervention point for each of the five models was January 1, 2004. On this day, the CAN SPAM Act first went into effect. The question of the first model regards whether the CAN SPAM Act affected spam rates over time. For the second through fourth time series design, each of the three measures of percentage of compliance per month were used as dependent variables. The three compliance measures were (1) percent of unsubscribe options provided in each email per month, (2) percent providing physical mailing addresses per month, and (3) percent of emails with descriptive subject headings per month. Lastly, for the fifth model, the dependent variable used was the percentage of messages sent from within the United States per month, to determine how this trend changes after the passing of the Act.

Hypotheses

The three research questions are whether the CAN SPAM Act has affected (1) spam rates, (2) spam compliance, or (3) spam IP addresses. The two hypotheses of the first two research questions are as follows:

1. Spam rates will not have decreased after the passing of the CAN SPAM Act.
2. Level of spam compliance will not have increased after the passing of the CAN SPAM Act.

Existing research suggests that spam has done just about everything except decrease since the passing of the CAN SPAM Act. Some spam filtering and security companies report that spam has since increased after the passing of the Act (Gross, 2004; Zeller, 2005). The Federal Trade Commission conducted its own analysis and reported to Congress that the Act was a success due to the flattening of spam rates following the passing of CAN SPAM, a flattening which appeared to have ended a preceding and historical increase in spam (Majoras, Leary, Harbour, & Leibowitz, 2005). Although these studies were carried out only a year after the passing of the Act, other reports suggest spam is at an all time high of over 90% of all email in 2009 (McMillan, 2009). Therefore, the first research question is estimated to be answered with anything but a decrease in spam rates.

The second hypothesis is informed by existing literature that concludes compliance has not increased after the CAN SPAM Act (Gross, 2004, Grimes, 2007). Grimes (2007) reports a drop in compliance between six months after and two years after follow up after the passing of CAN SPAM. Gross (2004) reports a compliance percentage of between 1-10%. None of these studies relied on any data collected before the CAN SPAM Act passed. However, it is assumed that, whatever compliance existed before the Act, it will not have increased after the Act.

The last question has less existing research to inform it, and therefore there is no hypothesized outcome that is assumed to be likely to occur before actually conducting an impact assessment. The lack of findings on this subject in the literature could be due to the poor ability to measure the actual originating IP address of wherever the spam was initially sent from. The originating IP address could be of a botnet that sent the spam, a proxy or open relay from which the spam was sent through, or the initial routing information in an email header could be entirely spoofed (Conner, 2008). There is little literature on how much of spam has falsified routing information vs how much actually represents a genuine IP address somewhere along the email's route to the recipient's inbox. Because of this uncertainty, no assumptions can be made about the percentage within the United States per month variable used in this analysis. The percent of US spam may be a measure botnets in the United States, or it may be a measure of how often spammers want the recipient to think spam was sent from within the United States, and the proportion of spam that meets either of these descriptions is not known as of this writing. Therefore, there are no assumptions made about the findings when analyzing the IP address time series.

CHAPTER 6: RESULTS

Monthly Spam Rates

A time series model was developed to examine the effect the introduction of the CAN SPAM Act of 2004 had on the amount of spam received per month. To test the veracity of the spam rate time series itself, a Spearman's correlation was conducted between the monthly spam data and the Federal Trade Commission's own monthly spam data. The two samples included the spam archives of spam collected between 1998 and 2008, and the Federal Trade Commission data collected between 2000 and 2007. The two datasets were found to be strongly related ($r(96) = .81$, $p < .001$). High correlation can be taken to mean that the two datasets measure the same spam activity within the United States.

However, after inspection of the spam rate time series, there appeared to be a large spike in spam rates between the months of August and October 2006 (see Appendix A.1). This indicated a non-constant variance, and so a logarithmic transformation of the time series was necessary. The logarithmically transformed series can be seen in Appendix A.2. An Augmented Dickey-Fuller test revealed that the spam rate series was not trend stationary ($t = -2.63$, $p = .089$), meaning that there was high serial dependency in the time series data. After regular differencing, the data were identified to be sufficiently stationary ($t = -10.75$, $p < .001$).

With the parameters for the model estimated, an ARIMA(0, 1, 0) model was identified. Diagnostic checks of the residuals were conducted to test for the presence of autocorrelation. A Ljung-Box test was not found to be significant (*p* = .922), indicating no autocorrelation between the residuals. Refer to Appendix B.

A dummy variable for the intervention component of the model was created, with the intervention point starting in January 1, 2004, when the CAN SPAM Act first went into effect. The impact parameter was tested on the logarithmically transformed spam rate time series. The intervention coefficient (-.132) was found to be nonsignificant (*t* = -.395, *p* = .694, R^2 = .001), suggesting that there was no change in the underlying trend for spam rates in January 1, 2004. Thus the CAN SPAM Act had no influence over the volume of spam sent after the passing of the Act.

Percent Compliance with Unsubscribe Option in Spam Per Month

A second time series model was developed to determine the effect the CAN SPAM Act of 2004 had on the percent of spam that provided an opt-out choice for recipients. An Augmented Dickey-Fuller test revealed that the percent of compliance with an opt-out method series was trend stationary (*t* = -3, *p* = .038), and therefore met the underlying assumptions of the model.

With the parameters for the model estimated, an ARIMA(1, 0, 0) model was identified. Diagnostic checks of the residuals were conducted to test for the presence of autocorrelation. A Ljung-Box test was not found to be significant (*p* = .845), indicating no autocorrelation between the residuals. Refer to Appendix B.2.

A dummy variable for the intervention component of the model was created, with the intervention point starting in January 1, 2004, when the CAN SPAM Act first went into effect. The impact parameter was tested with CAN SPAM compliance with an opt-out requirement as the dependant variable. The autoregressive parameter (.81) was significant (t = 15.21, p < .001).

The intervention parameter (-.05) was not found to be a significant predictor (t = -1.8, p = .075, R^2 = .77), indicating that patterns of spammers providing an unsubscribe option in emails were not affected by the CAN SPAM Act.

Percent Compliance with Providing Physical Mailing Address in Spam Per Month

A third time series model was created to test the impact the CAN SPAM Act of 2004 had on the percentage of emails that provided a physical mailing address in the body of the message. An Augmented Dickey-Fuller test revealed that the percent of address compliance series was sufficiently trend stationary (t = -5.52, p < .001).

With the parameters for the model estimated, an ARIMA(3, 0, 0) model was identified, with one autoregressive parameter at lag 4 and another at lag 36. Diagnostic checks of the residuals were conducted to test for the presence of autocorrelation. A Ljung-Box test was not found to be significant (p = .81), indicating no autocorrelation between the residuals. Refer to Appendix B.3. The estimated constant parameter (< .001) was found to be nonsignificant (t = .08, p = .936), and therefore had to be eliminated from the model.

An intervention point for the model was created starting in January 1, 2004, when the CAN SPAM Act first went into effect. The impact parameter was tested with percent of compliance with an address requirement set as the dependent variable. The autoregressive parameter at lag 1 (.33) was significant (t = 3.98, p < .001), the parameter at lag 4 (.32) was significant (t = 4.34, p < .001), and the final autoregressive parameter at lag 36 (.13) was also significant (t = 2.83, p = .006).

The intervention parameter (.002) was not significant (t = .94, p = .35, R^2 = .56), suggesting that the intervention of the CAN SPAM Act had no noticeable impact on compliance with the Act's address requirement.

Percent Compliance with a Descriptive Subject Heading in Spam Per Month

A forth time series model was created to test the effect the CAN SPAM Act of 2004 had on whether spam used descriptive wording in their subject lines. An Augmented Dickey-Fuller test revealed that the average compliance per month series was not trend stationary (t = -1.35, p = .605). After regular differencing, the data were identified to be sufficiently stationary (t = -14.17, p < .001).

With the parameters for the model estimated, an ARIMA(0, 1, 1) model was identified. Diagnostic checks of the residuals were conducted to test for the presence of autocorrelation. A Ljung-Box test was not found to be significant (p = .806), indicating no autocorrelation between the residuals. Refer to Appendix B.4. The estimated constant parameter (-.003) was found to be nonsignificant (t = -1.41, p = .162), and therefore had to be eliminated from the model.

The intervention component of the model was created, with the intervention point starting in January 1, 2004, when the CAN SPAM Act was first instated. The impact parameter was tested with compliance with a descriptive subject per month as the dependent variable. The local moving average parameter (-.31) was significant (t = -3.77, p < .001).

The intervention parameter (-.09) was found to be a significant predictor (t = -2.32, p = .022, R^2 = .09), indicating that the intervention parameter had a negative impact on average compliance with CAN SPAM requirements for a meaningful subject heading in emails. The average percentage of compliance before the CAN SPAM Act was about 81.34% of emails that had accurate subjects. After the intervention, compliance appeared to have dropped by 9%.

Percent of Spam from Within the United States

A fifth and final time series model was developed to examine the effect the CAN SPAM Act had on the amount of spam with IP addresses that appear to be from within the United States. An Augmented Dickey-Fuller test found the percentage of United States spam to be stationary (t = -3.27, p = .018), suggesting no significant serial dependency of the time series data.

The parameters for the model were estimated and an ARIMA(1, 0, 0) model was chosen. Diagnostic checks of the residuals were conducted to test for the presence of autocorrelation. A Ljung-Box test was not found to be significant, indicating no autocorrelation between the residuals. Refer to Appendix B.5.

The interval chosen was the point starting in January 1, 2004, when the CAN SPAM Act first went into effect. The autoregressive parameter (.895) was found to be significant (t = 26.69, p < .001). The intervention parameter (.026), however, was not found to be significant (t = .43, p = .667, R^2 = .85). The results indicate that the CAN SPAM Act had no impact on which country spam appears to be originating from.

CHAPTER 7: DISCUSSION

The current study attempted to determine whether the CAN SPAM Act had any kind of significant impact on the behavior of spammers and the outcome and content of spam messages. Five interrupted time series models were constructed to assess the potential effect CAN SPAM had on either spam rates, spam's legality, or spam's apparent originating IP address. This research found that no impact could be discerned on the spam rate or the IP address time series. Nor were two of the three spam law compliance series significantly predicted by the CAN SPAM Act. It was found that CAN SPAM may have been followed by a drop in spam complying with a specific CAN SPAM requirement, that being that emails must have descriptive subject headings in emails.

This research attempted to fill in for the dearth of significant scientific findings on the efficacy of the CAN SPAM Act. Various entities have set out to identify the Act's limitations and strengths. The Federal Trade Commission, tasked with enforcing the Act, reported to Congress that spam rates appeared to have leveled off since the Act's introduction (Majoras, Leary, Harbour, & Leibowitz, 2005). Independent anti-spam firms found the opposite to be true, that spam had since increased (Gross, 2004; Zeller, 2005). These reports were made only at most a year after the passing of CAN SPAM. This present research found no impact on spam rates; no increase or any decrease. This is consistent with this article's first hypothesis: that spam rates will not have decreased after the passing of the CAN SPAM Act.

Spam is a relatively new form of crime, and is also a relatively new form of communication and technology. Inspecting the visual change in spam rates over time on a line graph, both on the spam rates collected for this research and that collected by the Federal Trade Commission (see Appendix A), spam clearly started out small in the early preintervention period of both spam rate time series datasets. Perhaps a baseline for something as early as before 2004 might have been during a time when there was too little spam to begin with to properly contrast with what spam is like today. Looking at both figures of spam rates over time, spam clearly starts out small and unnoticeable and then expands considerably over the decade. This increase in spam was likely inevitable, as the beginning of both time series are near a time when spam was new. Like any new form of crime or technology, it takes some time before it can grow into more stable levels. Perhaps if the CAN SPAM Act was released later, we might have witnessed some sort of noticeable change in spam trends following the Act. Fortunately, the CAN SPAM Act has since been amended in 2008, called the CAN SPAM Act of 2008. If a preintervention prior to 2004 truly is not a time we'd consider to be a proper baseline measurement of spam rates, then maybe future research could assess the impact of the CAN SPAM Act of 2008.

Considering that spam rates remained unaffected by the CAN SPAM Act of 2004, this could be evidence that the CAN SPAM Act was not a sufficient deterrent to sending bulk commercial emails illegally. It was expected that a result like this would be found, based on existing research and the logical reasons why a spammer would refuse to change their spamming techniques to comply with the United States Code. Profitability for a spammer requires quantity over quality; the more spam messages the better. A spammer must send millions of email per month, and to have to make those emails comply with spam law would make profitability nearly impossible. Decreasing the volume of spam would not be in the spammer's best interest.

But the CAN SPAM Act did not make spam illegal, it only regulated what kinds of spam is allowable. Certainly a spammer could send just as much spam as he/she did before, with only some substantial adjustments to make his electronic messages comply with the United States Code. If spammers sent just as much spam as before the Act, only making their messages comply with its regulations, then perhaps Congress would be wise to make all spam, regardless of how compliant it is, illegal.

As the present study revealed, noncompliance with spam law was similarly not deterred by the passing of the CAN SPAM Act. The first step towards mitigating the spam problem would be to actually affect spammers in some way, the result of which would hopefully be combined with both compliance and a decrease in mass produced spam. Illegalizing all spam, regardless of compliance, would likely have produced a result little different than that observed already. Most spam already is illegal anyway, not because of how much is sent, but because of the numerous violations present in the messages and headers themselves.

Other research has also attempted to measure spam's level of compliance with CAN SPAM regulations. All of it reports that a majority of spam does not comply with one or more measures outlined in the CAN SPAM Act (Gross, 2004; Grimes, 2007). However, none of this research utilized a baseline before the CAN SPAM Act was passed to contrast postintervention compliance levels with. Since this current research had a significant baseline, we can now be certain that compliance appears to have been either unaffected or to have actually gone down after the CAN SPAM Act.

This finding is consistent with the second research hypothesis of this article: that spam compliance with spam law will not have increased after the passing of the CAN SPAM Act. Not only was it found that compliance failed to increase, but there appeared to be a statistically significant decrease in compliance with one of the three spam laws tested. It seems unlikely that the CAN SPAM Act had some sort of causal impact that resulted in a drop in compliance with the truthful subject law. Rather, compliance was likely going to show a decrease regardless of whether any anti-spam legislation was passed.

Inspecting the line graph of CAN SPAM compliance with the meaningful subject law depicted in Figure A.5, compliance looks to begin dropping at around early 2002, well before the CAN SPAM Act was passed. Compliance plummets and evens off at the beginning of 2005. It could be that the CAN SPAM Act came at a bad time in history when commercial emails were becoming more fraudulent and less considerate of their recipients.

In the Federal Trade Commission’s report to Congress, it was described that legitimate businesses were showing an increase in compliance with CAN SPAM regulations (Majoras et al., 2005). Regardless, it is clear that the majority of spammers are not running legitimate businesses. The reasons for this trending drop of spammers using descriptive subjects overall are not particularly surprising when one considers how such deception in emails might be useful for advertising.

While the meaningful subject requirement of spam was the most common item spammers complied with, it is still likely a useful trick in the spammer’s book. If someone cannot tell an email is spam just by reading the title of the subject, then maybe they will open the message and read it before deciding to delete it from their inbox. Certainly more spammers might adopt this trick as time progresses and spammers become savvier to turning a profit through spam. As spammers share more ideas online and learn from each other, certain spam techniques will be adopted, and the more successful ones will likely persist over time. Perhaps around 2002, when compliance drops off for the meaningful subject law, spammers were just starting to learn the utility of this technique. If it is found to be a successful means of influencing recipients to open an email, or maybe even an attachment, then surely that technique will continue to increase in prevalence among spam message over time.

Spammers might think twice about installing a spam bot or sending spam from within the United States considering the penalties set forth in the CAN SPAM Act. No research prior has addressed this possibility, although it might be considered that there would be fewer spammers in the US, or even less spam sent from the US, after the Act went into effect. It is difficult to determine what proportion of spam originated in the United States by examining the headers alone. It is even more difficult to determine this proportion merely by geolocating the first IP address found in spam, as this present research did. The findings in this research that relate to this question, while predicting that the CAN SPAM Act had no noticeable impact on where IP addresses in emails appear to be coming from, is still inconclusive. Determining the originating IP address of an email message that is considered to be spam requires a careful inspection of the headers to identify and eliminate obviously false routing insertions or invalid IP addresses. The software used to gather the present data was not equipped to inspect the headers with such precision.

If there had been a noticeable trend in the percentage of US IP addresses per month, then perhaps the time series might be considered to be meaningful of some underlying force, perhaps representative of spammer or cybercriminal behavior. Inspecting the time series of US spam contributions in Figure A.4, there does not seem to be a consistent direction to which US spam percentage takes. It may be that the United States CAN SPAM Act was not a sufficient deterrent for local spammers, or also likely is that the originating IP address of spam has little to nothing to do with the actual IP address of the spammer. Given the lack of significant findings with this data, it is difficult to say anything substantive about the underlying reasons for the results of the data.

Whatever the case, this research has concluded that the CAN SPAM Act has not significantly deterred spam. It could be that such legislation was too early to have an impact in the nascent and growing spam volume over time. In that case, the CAN SPAM Act of 2008 should be considered for a follow up. It could also be the difficulty of arresting and prosecuting known spammers. Spam, like any form of cybercrime, has a high degree of anonymity. Not even an IP address can be used to track the location of the offender. Purchases of spam or scam products are often done with merchants outside US jurisdiction, or anonymously through wiring services like Western Union or Liberty Reserve.

The CAN SPAM Act is just written laws and regulations, and those regulations are nothing if not enforced. Any given law must have a body authorized to enforce that law, otherwise there is not a sufficient or capable guardian to make any subsequent violations of the law unprofitable. The Federal Trade Commission, tasked with enforcement in areas where spam law applies, may not be presently equipped to take on these issues. By many accounts the FTC is underfunded and understaffed. Twenty years ago, the FTC was staffed by some 2,000 full-time employees hired to protect consumers. But since then, significant cut backs have been made, and there are only 1,000 or so full-time workers available.

The risks associated with committing any form of cybercrime, spam or otherwise, are clearly not high enough, especially if only 5% of malware writers and other cybercriminals are ever caught (Paul, 2006). But actually catching such criminals can be a difficult endeavor given the high degree of anonymity provided by the internet. Compound this with cybercrime crossing international borders, and there are problems of enforcement with the mix of jurisdictions involved.

The United States ought not just create and enforce local cybercrime laws; there must be some protocol in place to allow the collaboration with the governments of other countries to help bring offenders across borders to justice. There are already some rudimentary measures put in place to accomplish just this, such as the Tripartite

Memorandum of Understanding on Spam Enforcement Cooperation. This is an agreement between the United Kingdom, the United States, and Australia to enforce laws against cybercrime violators (Mustakas, Ranganthan & Duquenoy, 2005).

It will take more than the cooperation of just these three countries before anti-spam laws can pose a significant deterrent to offenders everywhere in the world. Since there are almost no borders on the internet, enforcing laws such as CAN SPAM will have to be done under greater agreement about the illegality of spam of all nations. While cybercriminals have made their operations more effective by collaborating and becoming more organized with other like minded groups and individuals, so too must law enforcement be similarly organized. If two jurisdictions from more than one country do not agree on the illegality of spam, then spam will surely continue.

And successful deterrence of illegal spam may be no more difficult than the simple collaboration among multiple national jurisdictions. While cybercrime can affect all users who have email and use the internet across the world, there may be only a limited number of cybercriminals in total. According to Spamhaus, which tracks internet spam, 80% of spam received in North America and Europe is sent by less than 200 spam groups comprising some 500-600 individual spammers (Register of Known Spam, 2009). If the certainty and swiftness of punishments for the sending of illegal spam can be increased, it could decrease what might be a limited number of spammers in total worldwide. Perhaps such a crackdown would finally allow users to see a historically novel decrease in spam rates, rather than the inevitable and persistent upward trend in spam over the decade.

Despite the possible solutions, nothing is final as of yet, and this research is not without its limitations. Of concern might be the abnormal spike in spam rates in late 2006 (see Figure A.1). While the FTC data has a modest spike near that time as well (see Figure A.3), the increase clearly does not dominate the time series as it does for the data used in this research. The wildcard address used by the author of the spam archives in 2006 is the biggest suspected reason for this abnormal increase. While not affecting the other time series models, since each case was a percentage or otherwise constrained by total emails sent each month, the spike in spam rates in late 2006 was not entirely representative of actual spam sent during this time worldwide.

Also of concern might have been the results of testing the compliance scale used in this research. Having only four items that were not correlated with one another may not only suggest a scale that does not properly measure compliance, but that each individual compliance item was similarly not a reliable measure of compliance. And sure enough, two of the four compliance items were either no better than chance when tested for interrater reliability, or at least very low in interrater reliability.

It may be of concern that the only significant finding in the impact assessment phase was of a time series model with very low interrater reliability. The meaningful subject requirement of the CAN SPAM Act was apparently exacerbated after the passing of the CAN SPAM Act. While it seems unlikely that CAN SPAM actually influenced spammers to be more deceptive with writing email subject lines, the drop in compliance may be in question entirely, regardless of spam law, because the software was so unreliable in identifying deceptive subject lines.

One clear reason for the limited ability of the software to identify deceptive subject lines is the poor ability to program software to understand semantic meaning of the spoken or written English language. Software is excellent at identifying keywords and making thousands of perfect string comparisons, but it is often not good at understanding the meaning of each word in the context of the sentence it is used in. Future research might be advised to write a script that is better able to extract such meaning.

Another problem with the software, which might be more easily corrected by future research, was its inability to match the same words when they are spelled differently. One word may appear in the subject and in the body of an email message, but the software of the current study would be unable to positively match the two if they differed from each other in spelling, such as different tense or plurality. Not to mention the software would be unable to match other words deliberately spelled differently to fool spam filters, such as the hundreds of ways Viagra is misspelled. It would not be so difficult to write software in future research that can match such strings.

Another area in which the software could be improved would be its measure of compliance with a notice of advertisement. As was determined by this research, the compliance variable used was unreliable as a measure of whether the notice of advertisement law was complied with, and therefore was not used in this study. A reason for the script's inability to properly classify the data appeared to have been the three keyword matches chosen. Of them were “advertisement”, “adv”, and “ad”. Of the random sample of fifty emails, virtually all of the messages that complied with this law had the keyword “ad” identified in them by the software. Upon inspecting the emails, “ad” was clearly not used in the context the script identified it to be in. More often than not, the term “ad” was found in emails composed entirely of gibberish or a random sequence of ASCII characters (likely to fool spam filters).

Upon further consideration, it seems unlikely an email would notify the user that the message is an “ad”. Thus, “advertisement” and “adv” would be the more common word choice likely to be used. Future research would do well to leave out this third keyword from the script. Perhaps then the data could be determined to be sufficiently reliable to the extent that it could be used in a time series model.

Lastly, the measure of each spam message’s country of origin was in no way a reliable measure of the origins of the spammer. Most spam is sent from botnets, and the remaining spam is still likely sent from rented or carded SMTP servers that can’t be traced back to the original spammer. The idea that spammers may begin infecting computers with spambot malware outside of the United States just because of the CAN SPAM Act seems unlikely. If it were easy to geolocate the IP address of the actual spammer’s residence, then there likely wouldn’t be much of a problem with spam to begin with. Determining the actual country the spammer is a resident of is exceedingly difficult.

Even if positive results were found in the research, a definitive explanation of why spam was sent less or more often in the United States would still be lacking. It cannot be determined from the data whether the majority of IP addresses in the sample represent botnets, mail servers, relays, proxies, actual spammer locations, hijacked wifi hotspots, or entirely made up IP addresses that cannot be geolocated to any place at all. The most important solution to this problem in future research would be a more careful consideration of the email headers. Software might attempt to geolocate all IP addresses found in the message's headers, and contrast whether each hop could reasonably have been a valid member in the family of related routing hops in the email's travel to its destination. Perhaps the software could eliminate invalid or unusual routing lines, and select the IP address that is both a combination of the earliest in the routine chain and the most likely to have not been spoofed.

If somehow IP addresses could be separated according to spoofed vs real ones, the sample would be much more meaningful. Even if what remains cannot be distinguished as to whether the message was from a botnet or the spammer's own house, the fact that it is unlikely to be a made up IP address could reduce a large amount of uncertainty about the meaning of the direction such a time series might take.

Despite the abnormal spike in the data used here in late 2006, it is fortunate that the spam rate data were strongly related to the Federal Trade Commission data (r = .81). With the three potential outliers of the spam rate spike in late 2006 removed, the relationship becomes even stronger (r = .86). This lends credibility to the data used for this research, suggesting that the means by which it was gathered and the resulting flood of spam collected matches that which the FTC acquired, and likely matches the entire population of spam rates in the United States during that time.

It also may be considered that the low reliability of the CAN SPAM Compliance scale might not say anything undesirable about the measure itself at all. That is, the fact that there was low reliability among the scale items is meaningful in and of itself. It may say something bad about spammers in general, rather than anything bad about the measure of compliance. It can be taken to mean that reasons for complying with one CAN SPAM requirement had nothing to do with the other compliance requirements. Rather, compliance with one item was likely a coincidence, and in all likelihood the only measure complied with was the meaningful subject item. Thus, compliance had nothing to do with the CAN SPAM Act, the reasons for complying with any item only known to the individual spammer him/herself.

Considering the profitability of spam, the negative findings of this research, and the low risks involved in sending spam, spam is a sound business strategy, with a low risk to reward ratio. In order to sufficiently deter spammers, punishments may have to be more probable for each cybercriminal that sends spam. However, in order to do such a thing, law enforcement may need more than just new laws. They may simply need better law enforcement, staffed with security experts and white hat hackers that are better able to track and apprehend cybercriminals. Spammers evolve and adopt new technology to boost their business with alacrity and eager readiness. Government agencies and other bodies granted the authority to pursue spammers legally may not be so adept at utilizing the internet and technology to make sure those in cyberspace comply with their laws. It may be necessary, in order to catch spammers, that government agencies begin to think like spammers themselves.

REFERENCES

Akers, R. L., & Sellers, C. S. (2004). In 4th Ed. *Criminological theories: Introduction, evaluation, and application* (pp. 33). Los Angeles, California: Roxbury.

Ananthaswamy, A. (2009, August 15). Defeat worms, send them to quarantine: The best way to stop highly virulent worms wreaking havoc on the internet is to give it an immune system. *New Scientist*, 2721, 16-17.

Anderson, R., Bohme, R., Clayton, R., & Moore, T. (2008). Security economics and the internet market. Retrieved July 30, 2008, from *ENISA: European Network and Information Security Agency.* Web site: http://www.enisa.europa.eu/doc/pdf/report_sec_econ_&_int_mark_20080131.pdf.

Armstrong, D., & Armstrong, E. M. (1991). The great American medicine show: Being an illustrated history of hucksters, healers, health evangelists, and heroes from Plymouth Rock to the present. *Prentice Hall.*

Arora, V. (2005). The CAN-SPAM Act: An inadequate attempt to deal with a growing problem. *Columbia Journal of Law and Social Problems*, 300-330.

Aycock, J. (2007, September). A design for an anti-spear-phishing system. *Virus Bulletin Conference September 2007.*

Bossler, A. M. & Holt, T. J. , 2007-11-14 *Examining the utility of routine activities theory for cybercrime.* Paper presented at the annual meeting of the American Society of Criminology, Atlanta Marriott Marquis, Atlanta, Georgia.

Brody R. G., Mulig, E., Kimball, V. (2007). Phishing, pharming and identity theft. *Academy of Accounting and Financial Studies Journal*, 11, 43-56.

CAN SPAM Act: Controlling the Assault of Non-Solicited Pornography and Marketing Act of 2003, 15 U.S.C. Sec 7701 (2005).

Chitu, A. (2007). *How Gmail blocks spam.* Retrieved May 18, 2009, from Google Blogspot: http://googlesystem.blogspot.com/2007/10/how-gmail-blocks-spam.html

Clayburn, T. (2005, February 3). *Spam costs billions: The cost of spam in terms of lost productivity has reached $21.58 billion annually.* Retrieved September 11, 2009 from InformationWeek Web site: http://www.informationweek.com/news/security/vulnerabilities/showArticle.jhtml?articleID=59300834

Cohen, L. E., & Felson, M. (1979). Social change and crime rate trends: A routine activity approach. *American Sociological Review*, 44(4), 588-608.

Computer Fraud and Abuse Act, 18 U.S.C. Sec 1030 (2006).

Dredze, M, Gevaryahu, R. & Elias-Bachrach, (2007). *Learning fast classifiers for image spam.* Proceedings of the Conference on Email and Anti-Spam (CEAS).

Dyrud, M. (2005). "I brought you a good news": An analysis of Nigerian 419 letters. Proceedings of the 2005 Association for Business Communication Annual Convention.

Email metrics program: The network operators' perspective (2007). *Messaging Anti-Abuse Working Group (MAAWG),* 6, 1-3.

Evett, D. (2006) *Spam statistics 2006.* Retrieved September 11, 2009, from TopTenReviews Web site: http://spam-filter-review.toptenreviews.com/spam-statistics.html

Federal Trade Commission Act, 15 U.S.C. Sec 41 (2007).

Felson, M., Clarke, R. V. (1998). Opportunity makes the thief: Practical theory for crime prevention. *Policing and Reducing Crime Unit: Research, Development and Statistics Directorate,* 98, 1-36.

Fraud and Related Activity in Connections with Electronic Mail, 18 U.S.C. Sec. 1037 (2006).

Friedrichs, D. O. (2009). Enterprise crime, contrepreneurial crime, and technocrime. In fourth Ed., *Trusted criminals: White collar crime in contemporary society* (pp. 192-218). Belmont, CA: Cengage Learning.

Giles, J. (2009, May 25). *How much is your identity worth?* NewScientist, 202(2709), 36-39.

Goodin, D. (2007). *Ukrainian eBay scam turns down syndrome man into cash machine: How a mule made an ass of me*. Retrieved September 1, 2009, from The Register Web site: http://www.theregister.co.uk/2007/11/08/ebay_victims_track_their_mules/

Grimes, G. A. (2007, February). Compliance with the CAN-SPAM Act of 2003: Studying the application of the of the CAN-SPAM Act and its effect on controlling unsolicited email messages. *Communications of the ACM,* 50(2), 56-62.

Gross. G. (Jan 13, 2004). Is the CAN-SPAM law working? *PC World*.

Holt, T. J. & Bossler, A. M. (2009). Examining the applicability of lifestyle-routine activities theory for cybercrime victimization. *Deviant Behavior*, 30(1), 1-25.

Hulten, G., Penta, A., Seshadrinathan, G. & Mishra, M. (2004). Trends in spam products and methods. *Proceedings of the First Conference on Email and Anti-Spam, CEAS*.

Hutchings, A. & Hayes, H. (2009). Routine activity theory and phishing victimization: Who got caught in the 'net'?. *Current Issues in Criminal Justice*, 20(3), 432-451.

Internet World Stats: Usage and Population Statistics (2009). Retrieved July 20, 2009 from http://www.internetworldstats.com/stats.htm

IronPort Systems. (2006, June). *Spammers continue innovation: IronPort study shows image-based spam, hit & run, and increased volumes latest threat to your inbox.* Retrieved September 11, 2009 from CISCO IronPort Email and Web Security from Web site: http://www.ironport.com/company/ironport_pr_2006-06-28.html

Kanish, C., Kreibich, C., Levchenko, K., Enright, B., Voelker, G. M., Paxson, V., Savage, S. (2008). Spamalytics: An empirical analysis of spam marketing conversion. *Communications of the ACM.*

Kestenbaum, D. (August 8, 2006). *Spam goes literary*. Retrieved July 21, 2009 , from NPR Web site: http://www.npr.org/templates/story/story.php?storyId=5624749

Kleiner, K. (2008). *Happy spamiversary! Spam reaches 30*. Retrieved July 15, 2009, from New Scientist Web site: http://www.newscientist.com/article/dn13777-happy-spamiversary-spam-reaches-30.html?full=true

Krebs, B. (2007, October). Shadowy Russian firm seen as conduit for cybercrime. *Washington Post*.

Larkin, E. (2009, February). Economies of scale in the spam business. *PC World*, 47-48.

Lee, Y. (2005, June). The CAN-SPAM Act: A silver bullet solution? *Communications of the ACM*, 48(6), 131-132.

Lemos, R. (2006, November). New ways to nab spam. *PC Magazine*, 158.

MacFarlane, C., Harrington, E., Salsburg, D., & Goodman, M. (2003, April 29). FTC measures false claims inherent in random spam. *Federal Trade Commission*.

Majoras, D. P., Leary, T.B., Harbour, P.J., & Leibowitz, J. (2005, December). Effectiveness and enforcement of the CAN-SPAM Act: A report to Congress. *Federal Trade Commission.*

McCain, Sen. (2003, October 22). CAN SPAM Act of 2003. In Congressional Record 149, S13020.

McDowall, D., McCleary, R., Meidinger, E. E., & Hay, R. A. (1980). *Interrupted time series analysis.* Iowa City, IA: Sage Publications, Inc.

McMillan, R. (2009, May). 90 percent of e-mail is spam, Symantec says. PC World.

Mosher, D. (2007, March). Map: The world according to spam. *Discover Magazine*, 28(3), 23.

Mendes, S. M. (February 1, 2004). Certainty, severity, and their relative deterrent effects: Questioning the implications of the role of risk in criminal Deterrence policy. *Policy Studies Journal.*

Moustakas, E., Ranganthan, C., Duquenoy, P. (2005). Combating spam through legislation: a comparative analysis of US and European approaches. *Proceedings of Second Conference on Email and Anti-Spam, CEAS.*

Muris, T. J., Thompson, M. W., Swindle, O., Leary, T. B. & Harbour, P. J. (2004, June). National do not email registry: A report to Congress. *Federal Trade Commission*.

Nigerian advance fee fraud (1997, April). United States *Department of State Bureau of International Narcotics and Law Enforcement Affairs*, 10465, 1-33.

Paul, H. (2006). It's time to arrest cyber crime. *Business Week Online*, 17-17.

Register of known spam operations (ROkSO). (2009). Retrieved September 1, 2009, from Spamhaus Web site: http://www.spamhaus.org/rokso/

Sahami, M., Dumais, S., Heckerman, D., & Horvitz, E. (1998). A bayesian approach to filtering junk e-mail. *AAAI Workshop on Learning for Text Categorization.* Technical report WS-98-05.

Saltzman, M. (2009, March 20). *What happens when you buy "spam"?* Retrieved September 14, 2009, from Sympatico Web site: http://sync.sympatico.ca/How-To/ContentPosting_MS?newsitemid=34c50777-f4eb-4751-bb9f-82c9a6f07ab4&feedname=MARC-SALTZMAN&show=False&number=0&showbyline=True&subtitle=&detect=&abc=abc&date=True

Schiller, C. A., Binkley, J., Harley, D., Evron, G., Bradley, T., Willems, C. & Cross, M. (2007). *Botnets: The killer web app.* Syngress Publishing.

Siponen, M., Stucke, C. (2006). Effective anti-spam strategies in companies: An international study. *Hawaii International Conference on Systems Science*.

Sjouwerman, S., & Posluns, J. (2004). *Inside the spam cartel: Trade secrets from the dark side.* Syngress.

Smith, R. G., Holmes, M. N., & Kaufmann, P. (1999). Nigerian advance fee fraud. *Trends & Issues in Crime and Criminal Justice*, 121, 1-6.

Sophos. (2006, April). *Sophos report reveals 'dirty dozen' spam relaying countries for January-March 2006: Asia named worst spam relaying continent.* Retrieved September 1, 2009, from Sophos Plc. Web site: http://www.sophos.com/pressoffice/news/articles/2006/04/dirtydozapr06.html

Stewart, J. (2006, August 15). *Mocbot Spam Analysis*. Retrieved January 28, 2010, from SecureWorks: http://www.secureworks.com/research/threats/mocbot-spam/

Stone, A. (2007). Evolution of a hacker. *Information Security*, 1-2.

Stone, B. & Levy, S. (2005). Grand theft identity. *Newsweek.*

Suler, J. (2003) The online disinhibition effect. *Cyberpsychology Behavior,* 7, 321-326.

Tive, C. (2006). 419 scam: Exploits for the Nigerian con man. iUniverse.

Types of spam. (2009). Retrieved September 1, 2009, from Virus List Web site: http://www.viruslist.com/en/spam/info?chapter=153350533

Weber, T. (2007). *Criminals may overwhelm the web*. Retrieved May 31, 2009 from BBC News: http://news.bbc.co.uk/2/hi/business/6298641.stm.

Weston, L. P. (Feb 5, 2010). *Four reasons we get ripped off.* Retrieved February 25, 2010, from MSN Money website: http://articles.moneycentral.msn.com/SavingandDebt/ConsumerActionGuide/weston-4-reasons-we-get-ripped-off.aspx?page=1

Why Am I Getting All This Spam? Unsolicited Commerical E-mail Research Six Month Report. (2003). Retrieved September 11, 2009 from Center for Democracy & Technology from Web site: http://www.cdt.org/speech/spam/030319spamreport.shtml

Wiedrick-Kozlowski, J. & Stinchombe, N, (2008). Software world. *Malicious Code Research Center (MCRC)*, 39(3), 16.

Wong, M. & Schlitt, W. (2006, April). Sender policy framework (SPF) for authorizing use of domains in e-mail, version 1. *Network Working Group*, 1.

Zeller, T. (Feb. 1, 2005). Law barring junk email allows a flood instead. *New York Times*, A1.

APPENDIX A: TIME SERIES CHARTS

Figure A.1 Line chart for spam messages received per month, 1998-2008

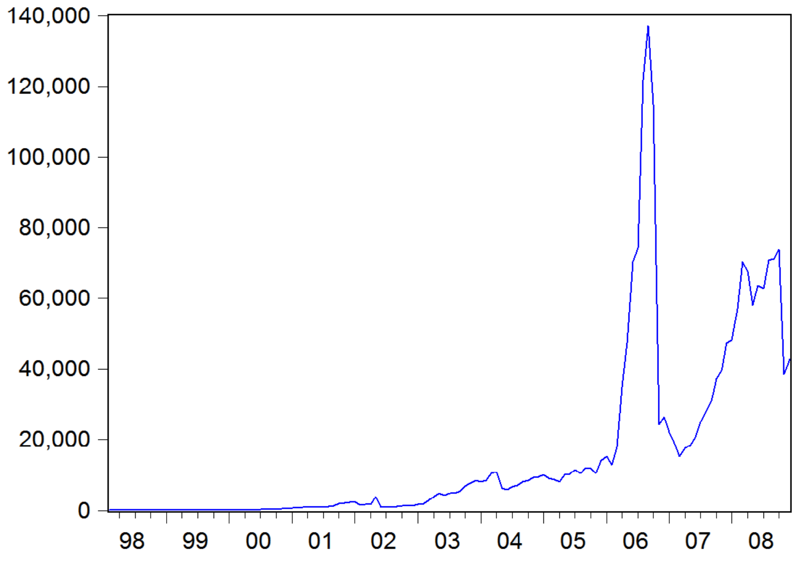


Figure A.2 Line chart for log transformed spam messages received per month, 1998-2008

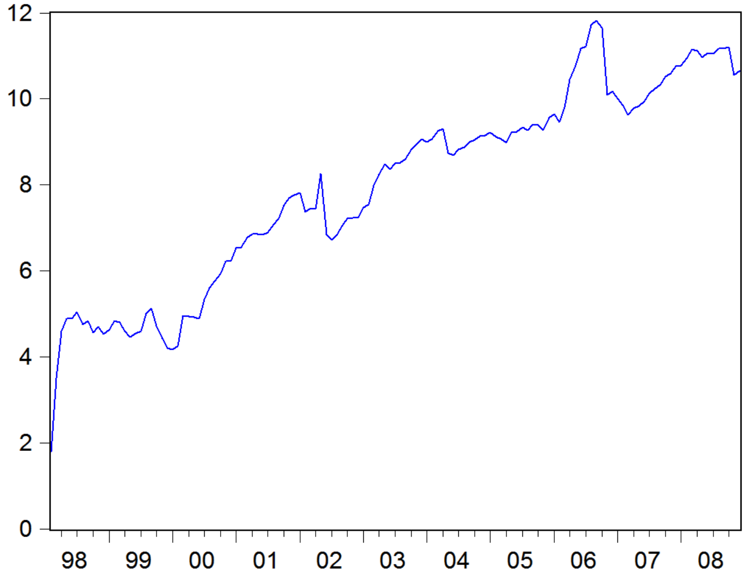


Figure A.3 Line chart for spam messages received per month collected by the FTC, 2000-2007

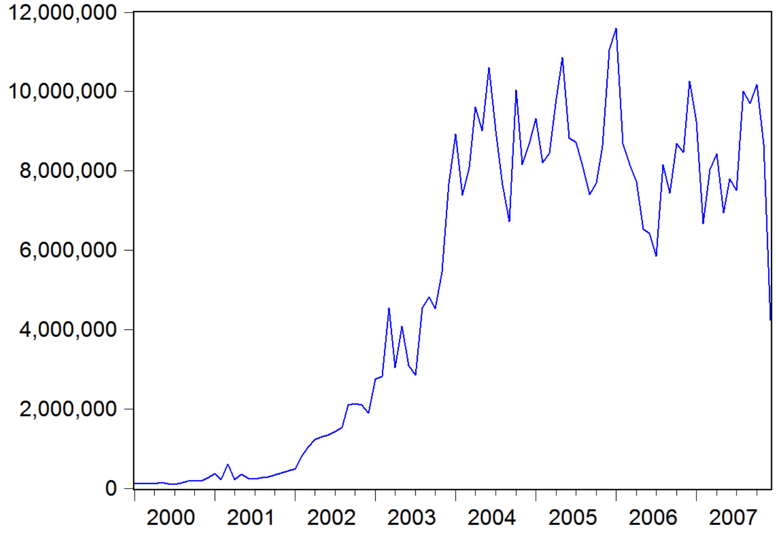


Figure A.4 Line chart for average compliance with CAN SPAM per month, 1998-2008

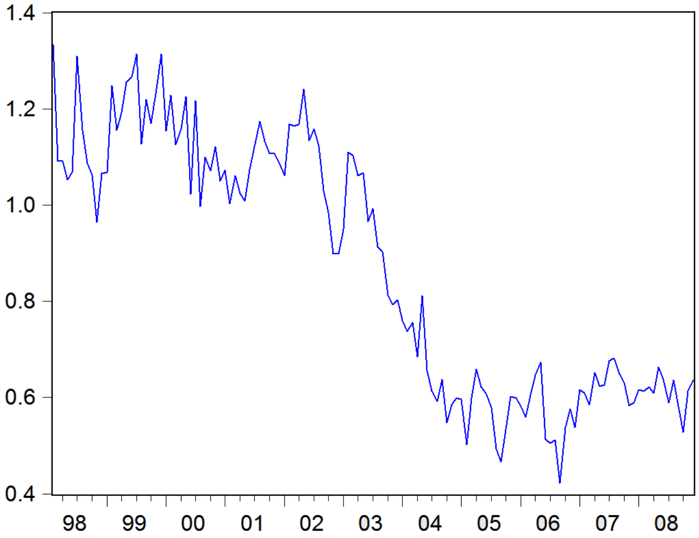


Figure A.5 Line chart of percentage compliance with accurate subject headings per month, 1998-2008

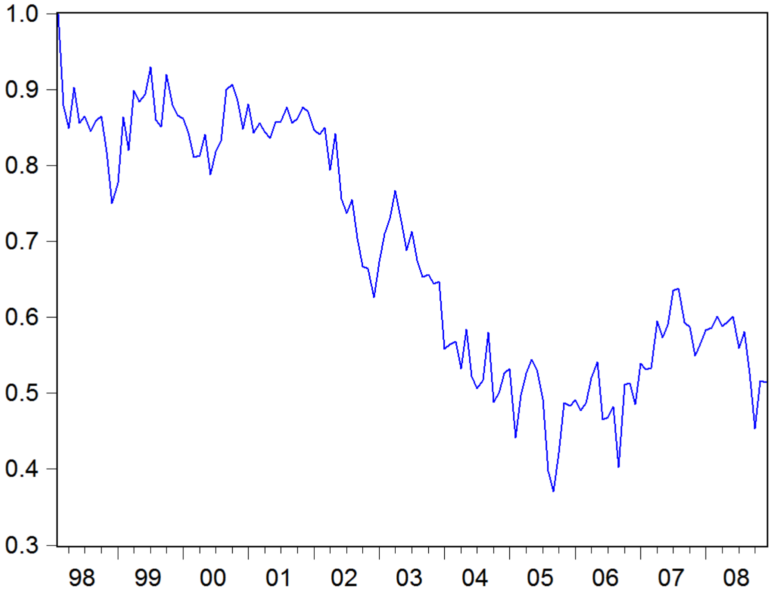


Figure A.6 Line chart of percentage compliance with physical address per month, 1998-2008

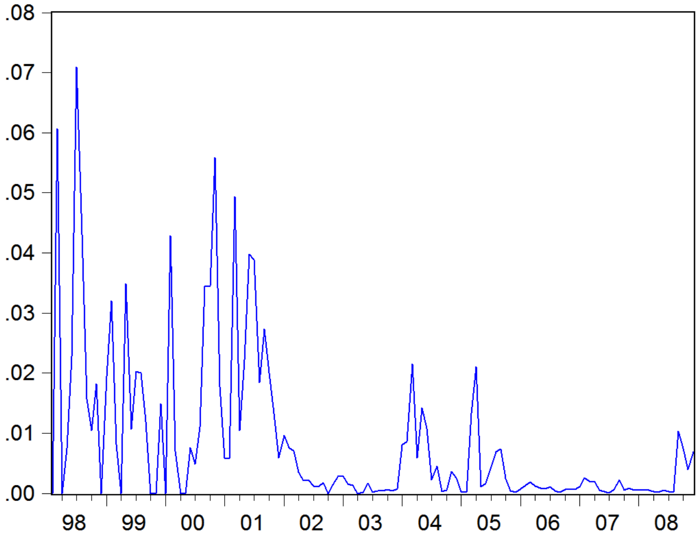


Figure A.7 Line chart for percentage compliance with ubsubscribe option per month, 1998-2008

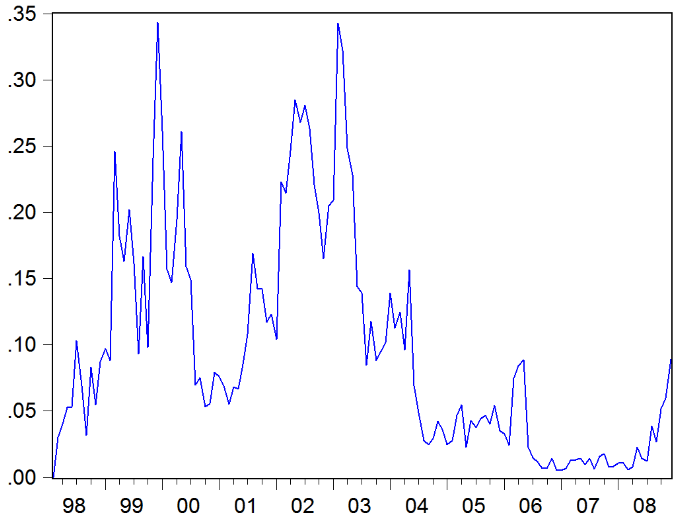


Figure A.8 Line chart for percentage compliance with notice of advertisement per month, 1998-2008

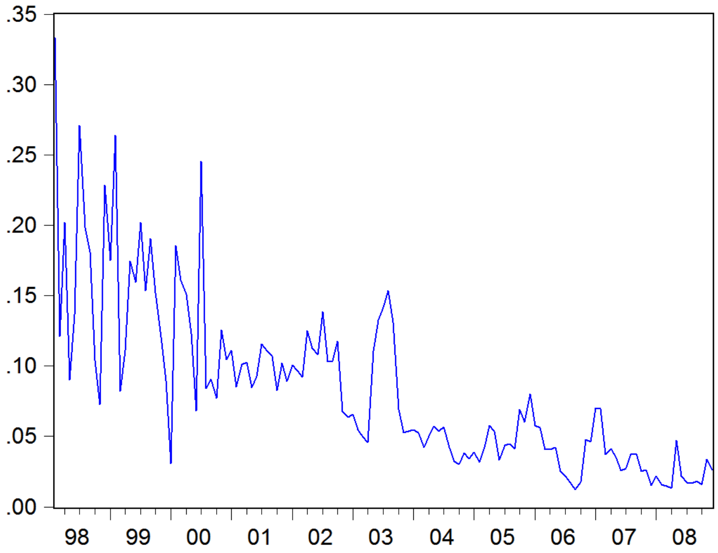
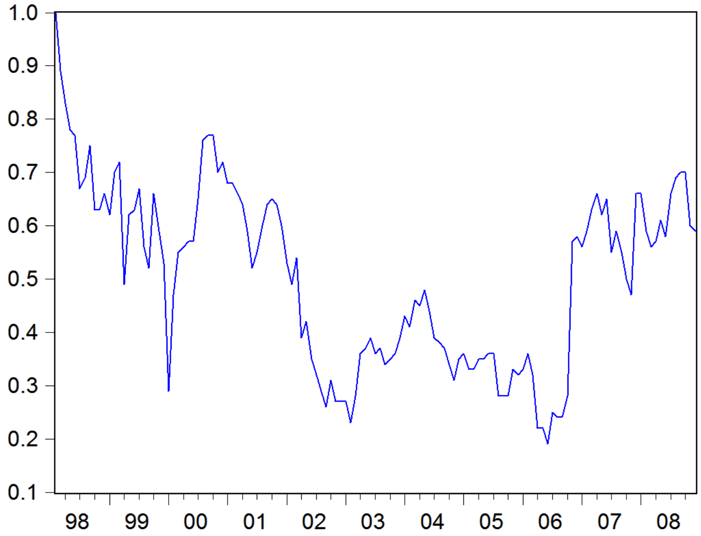
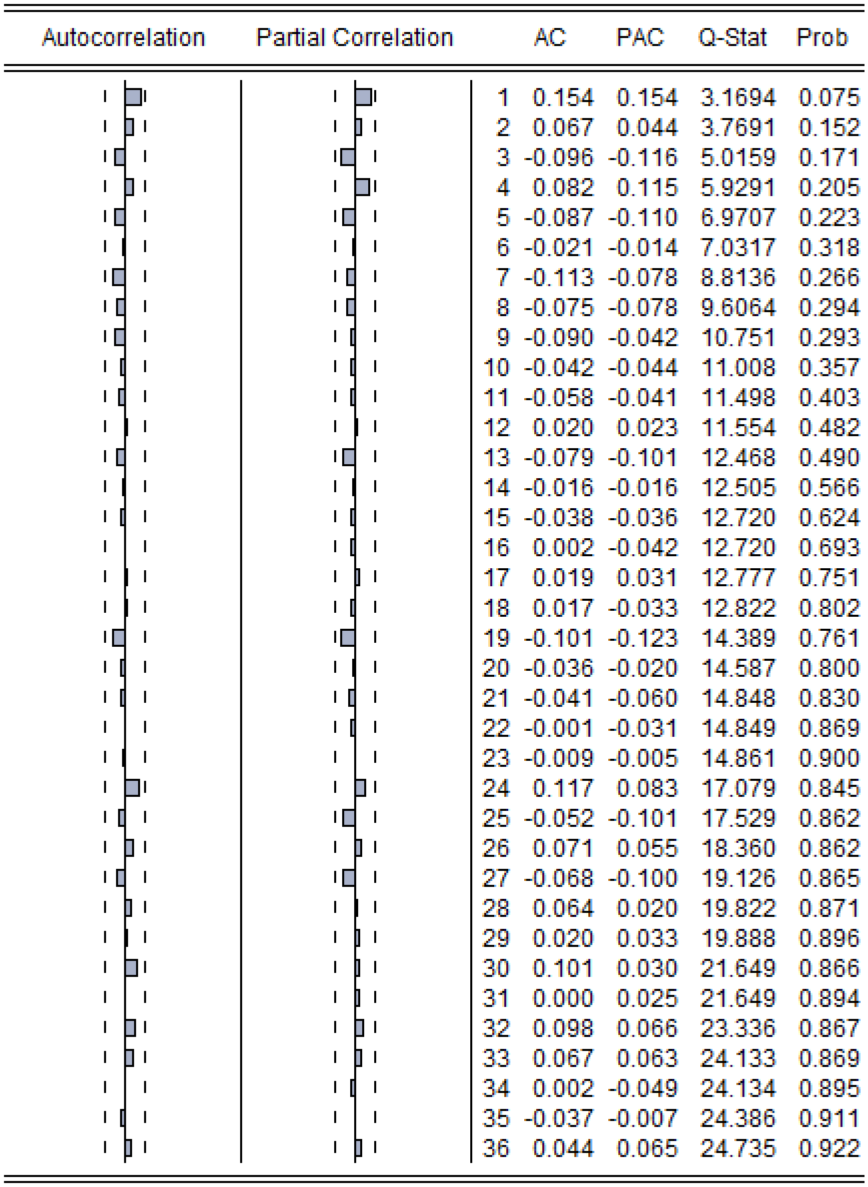


Figure A.9 Line chart for percentage of messages sent from the US per month, 1998-2008



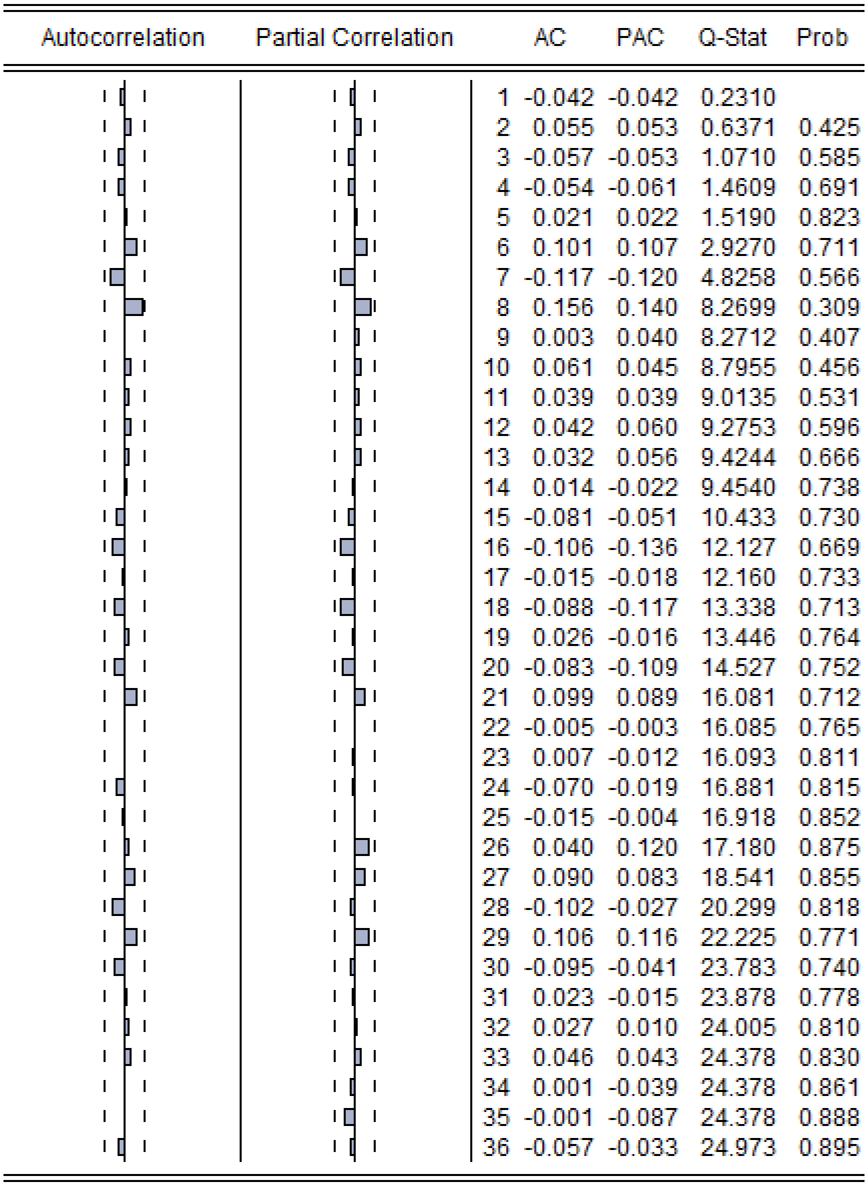
APPENDIX B: RESIDUAL CORRELOGRAMS

Figure B.1 Correlogram of residuals for log transformed spam rate model



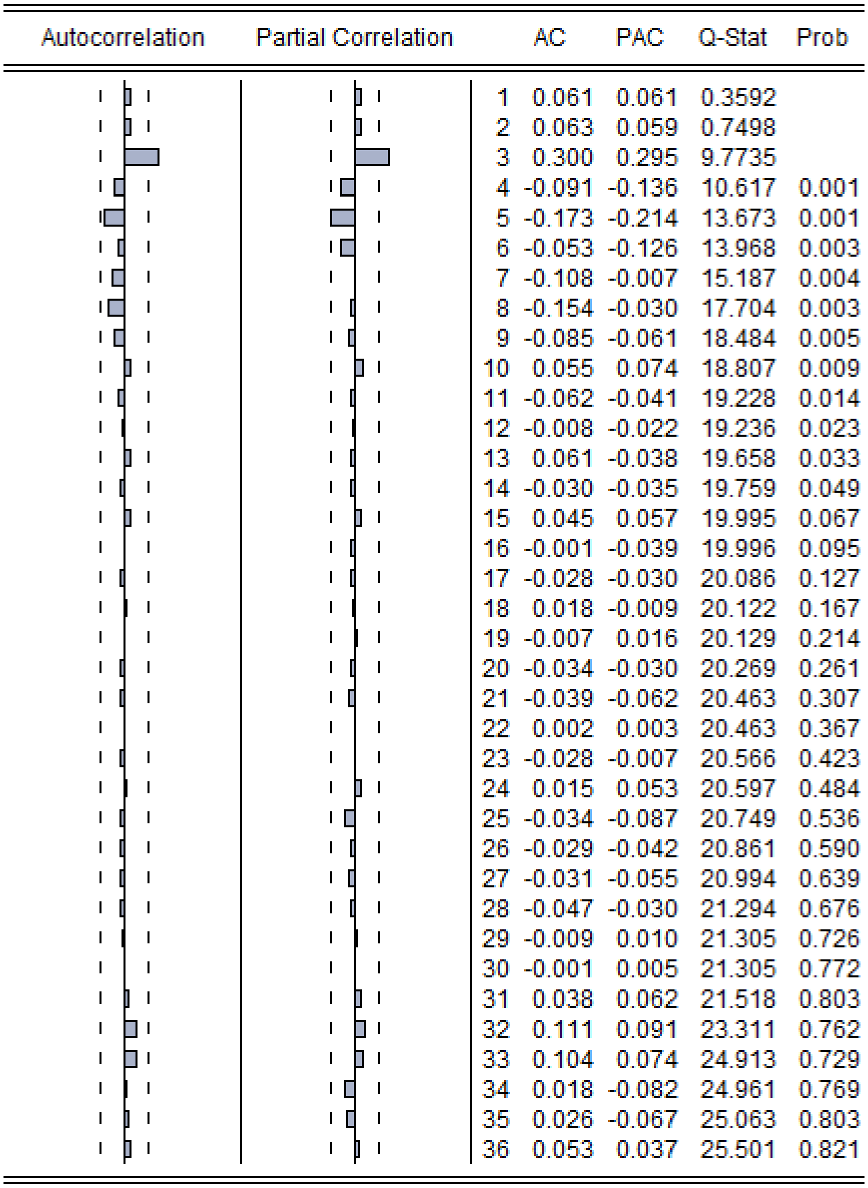
Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.154	0.154	3.1694	0.075
		2	0.067	0.044	3.7691	0.152
		3	-0.096	-0.116	5.0159	0.171
		4	0.082	0.115	5.9291	0.205
		5	-0.087	-0.110	6.9707	0.223
		6	-0.021	-0.014	7.0317	0.318
		7	-0.113	-0.078	8.8136	0.266
		8	-0.075	-0.078	9.6064	0.294
		9	-0.090	-0.042	10.751	0.293
		10	-0.042	-0.044	11.008	0.357
		11	-0.058	-0.041	11.498	0.403
		12	0.020	0.023	11.554	0.482
		13	-0.079	-0.101	12.468	0.490
		14	-0.016	-0.016	12.505	0.566
		15	-0.038	-0.036	12.720	0.624
		16	0.002	-0.042	12.720	0.693
		17	0.019	0.031	12.777	0.751
		18	0.017	-0.033	12.822	0.802
		19	-0.101	-0.123	14.389	0.761
		20	-0.036	-0.020	14.587	0.800
		21	-0.041	-0.060	14.848	0.830
		22	-0.001	-0.031	14.849	0.869
		23	-0.009	-0.005	14.861	0.900
		24	0.117	0.083	17.079	0.845
		25	-0.052	-0.101	17.529	0.862
		26	0.071	0.055	18.360	0.862
		27	-0.068	-0.100	19.126	0.865
		28	0.064	0.020	19.822	0.871
		29	0.020	0.033	19.888	0.896
		30	0.101	0.030	21.649	0.866
		31	0.000	0.025	21.649	0.894
		32	0.098	0.066	23.336	0.867
		33	0.067	0.063	24.133	0.869
		34	0.002	-0.049	24.134	0.895
		35	-0.037	-0.007	24.386	0.911
		36	0.044	0.065	24.735	0.922

Figure B.2 Correllogram of residuals for unsubscribe compliance percentage model



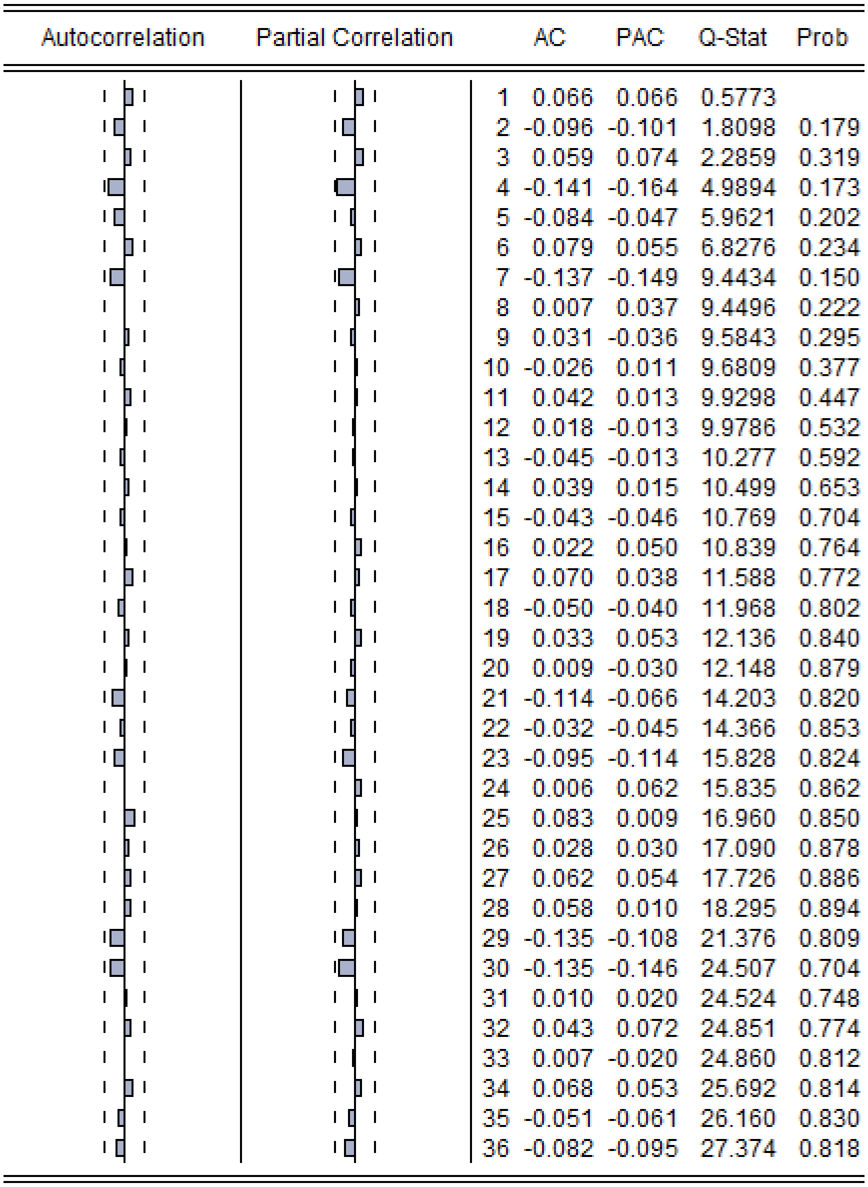
Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.042	-0.042	0.2310	
		2	0.055	0.053	0.6371	0.425
		3	-0.057	-0.053	1.0710	0.585
		4	-0.054	-0.061	1.4609	0.691
		5	0.021	0.022	1.5190	0.823
		6	0.101	0.107	2.9270	0.711
		7	-0.117	-0.120	4.8258	0.566
		8	0.156	0.140	8.2699	0.309
		9	0.003	0.040	8.2712	0.407
		10	0.061	0.045	8.7955	0.456
		11	0.039	0.039	9.0135	0.531
		12	0.042	0.060	9.2753	0.596
		13	0.032	0.056	9.4244	0.666
		14	0.014	-0.022	9.4540	0.738
		15	-0.081	-0.051	10.433	0.730
		16	-0.106	-0.136	12.127	0.669
		17	-0.015	-0.018	12.160	0.733
		18	-0.088	-0.117	13.338	0.713
		19	0.026	-0.016	13.446	0.764
		20	-0.083	-0.109	14.527	0.752
		21	0.099	0.089	16.081	0.712
		22	-0.005	-0.003	16.085	0.765
		23	0.007	-0.012	16.093	0.811
		24	-0.070	-0.019	16.881	0.815
		25	-0.015	-0.004	16.918	0.852
		26	0.040	0.120	17.180	0.875
		27	0.090	0.083	18.541	0.855
		28	-0.102	-0.027	20.299	0.818
		29	0.106	0.116	22.225	0.771
		30	-0.095	-0.041	23.783	0.740
		31	0.023	-0.015	23.878	0.778
		32	0.027	0.010	24.005	0.810
		33	0.046	0.043	24.378	0.830
		34	0.001	-0.039	24.378	0.861
		35	-0.001	-0.087	24.378	0.888
		36	-0.057	-0.033	24.973	0.895

Figure B.3 Correllogram of residuals for physical address compliance percentage model



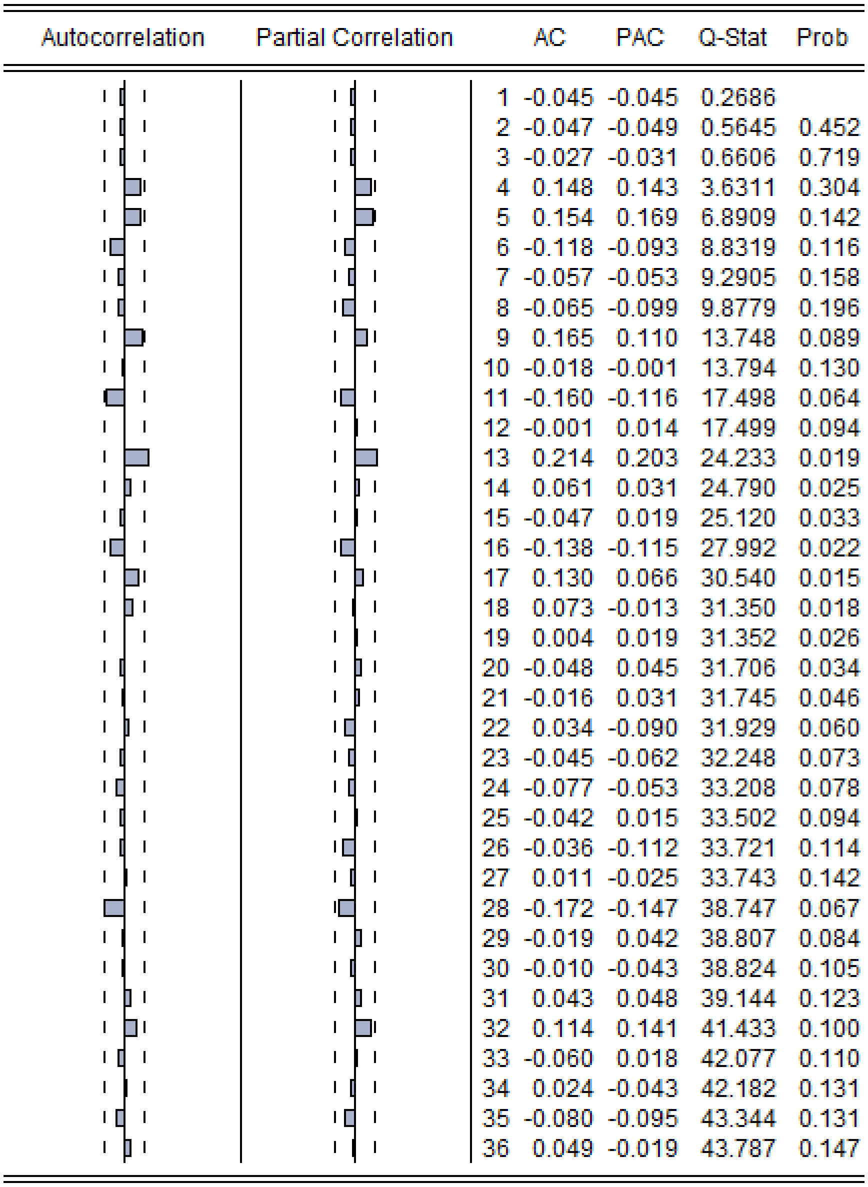
Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.061	0.061	0.3592	
		2	0.063	0.059	0.7498	
		3	0.300	0.295	9.7735	
		4	-0.091	-0.136	10.617	0.001
		5	-0.173	-0.214	13.673	0.001
		6	-0.053	-0.126	13.968	0.003
		7	-0.108	-0.007	15.187	0.004
		8	-0.154	-0.030	17.704	0.003
		9	-0.085	-0.061	18.484	0.005
		10	0.055	0.074	18.807	0.009
		11	-0.062	-0.041	19.228	0.014
		12	-0.008	-0.022	19.236	0.023
		13	0.061	-0.038	19.658	0.033
		14	-0.030	-0.035	19.759	0.049
		15	0.045	0.057	19.995	0.067
		16	-0.001	-0.039	19.996	0.095
		17	-0.028	-0.030	20.086	0.127
		18	0.018	-0.009	20.122	0.167
		19	-0.007	0.016	20.129	0.214
		20	-0.034	-0.030	20.269	0.261
		21	-0.039	-0.062	20.463	0.307
		22	0.002	0.003	20.463	0.367
		23	-0.028	-0.007	20.566	0.423
		24	0.015	0.053	20.597	0.484
		25	-0.034	-0.087	20.749	0.536
		26	-0.029	-0.042	20.861	0.590
		27	-0.031	-0.055	20.994	0.639
		28	-0.047	-0.030	21.294	0.676
		29	-0.009	0.010	21.305	0.726
		30	-0.001	0.005	21.305	0.772
		31	0.038	0.062	21.518	0.803
		32	0.111	0.091	23.311	0.762
		33	0.104	0.074	24.913	0.729
		34	0.018	-0.082	24.961	0.769
		35	0.026	-0.067	25.063	0.803
		36	0.053	0.037	25.501	0.821

Figure B.4 Correllogram of residuals for accurate subject compliance percentage model



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.066	0.066	0.5773	
		2	-0.096	-0.101	1.8098	0.179
		3	0.059	0.074	2.2859	0.319
		4	-0.141	-0.164	4.9894	0.173
		5	-0.084	-0.047	5.9621	0.202
		6	0.079	0.055	6.8276	0.234
		7	-0.137	-0.149	9.4434	0.150
		8	0.007	0.037	9.4496	0.222
		9	0.031	-0.036	9.5843	0.295
		10	-0.026	0.011	9.6809	0.377
		11	0.042	0.013	9.9298	0.447
		12	0.018	-0.013	9.9786	0.532
		13	-0.045	-0.013	10.277	0.592
		14	0.039	0.015	10.499	0.653
		15	-0.043	-0.046	10.769	0.704
		16	0.022	0.050	10.839	0.764
		17	0.070	0.038	11.588	0.772
		18	-0.050	-0.040	11.968	0.802
		19	0.033	0.053	12.136	0.840
		20	0.009	-0.030	12.148	0.879
		21	-0.114	-0.066	14.203	0.820
		22	-0.032	-0.045	14.366	0.853
		23	-0.095	-0.114	15.828	0.824
		24	0.006	0.062	15.835	0.862
		25	0.083	0.009	16.960	0.850
		26	0.028	0.030	17.090	0.878
		27	0.062	0.054	17.726	0.886
		28	0.058	0.010	18.295	0.894
		29	-0.135	-0.108	21.376	0.809
		30	-0.135	-0.146	24.507	0.704
		31	0.010	0.020	24.524	0.748
		32	0.043	0.072	24.851	0.774
		33	0.007	-0.020	24.860	0.812
		34	0.068	0.053	25.692	0.814
		35	-0.051	-0.061	26.160	0.830
		36	-0.082	-0.095	27.374	0.818

Figure B.5 Correlogram of residuals for percentage of spam within the US model



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.045	-0.045	0.2686	
		2	-0.047	-0.049	0.5645	0.452
		3	-0.027	-0.031	0.6606	0.719
		4	0.148	0.143	3.6311	0.304
		5	0.154	0.169	6.8909	0.142
		6	-0.118	-0.093	8.8319	0.116
		7	-0.057	-0.053	9.2905	0.158
		8	-0.065	-0.099	9.8779	0.196
		9	0.165	0.110	13.748	0.089
		10	-0.018	-0.001	13.794	0.130
		11	-0.160	-0.116	17.498	0.064
		12	-0.001	0.014	17.499	0.094
		13	0.214	0.203	24.233	0.019
		14	0.061	0.031	24.790	0.025
		15	-0.047	0.019	25.120	0.033
		16	-0.138	-0.115	27.992	0.022
		17	0.130	0.066	30.540	0.015
		18	0.073	-0.013	31.350	0.018
		19	0.004	0.019	31.352	0.026
		20	-0.048	0.045	31.706	0.034
		21	-0.016	0.031	31.745	0.046
		22	0.034	-0.090	31.929	0.060
		23	-0.045	-0.062	32.248	0.073
		24	-0.077	-0.053	33.208	0.078
		25	-0.042	0.015	33.502	0.094
		26	-0.036	-0.112	33.721	0.114
		27	0.011	-0.025	33.743	0.142
		28	-0.172	-0.147	38.747	0.067
		29	-0.019	0.042	38.807	0.084
		30	-0.010	-0.043	38.824	0.105
		31	0.043	0.048	39.144	0.123
		32	0.114	0.141	41.433	0.100
		33	-0.060	0.018	42.077	0.110
		34	0.024	-0.043	42.182	0.131
		35	-0.080	-0.095	43.344	0.131
		36	0.049	-0.019	43.787	0.147

APPENDIX C: IMPACT ASSESSMENT TABLES

Table C.1 Linear regression model for log transformed spam rates, 1998-2008

Variable	Coefficient	Standard error	*t*	*p*-value
Intercept	.069	.029	2.36	.02
Intervention	-.132	.335	-.395	.694

R^2 = .001

Table C.2 Linear regression model for accurate subject compliance percentage, 1998-2008

Variable	Coefficient	Standard error	*t*	*p*-value
MA(1)	-.312	.083	-3.77	< .001
Intervention	-.087	.038	-2.325	.022

R^2 = .09

Table C.3 Linear regression model for physical address compliance percentage, 1998-2008

Variable	Coefficient	Standard error	*t*	*p*-value
AR(1)	.326	.082	3.983	< .001
AR(4)	.318	.073	4.343	< .001
AR(36)	.125	.044	2.833	.006
Intervention	.002	.002	.94	.35

R^2 = .56

Table C.4 Linear regression model for unsubscribe compliance percentage, 1998-2008

Variable	Coefficient	Standard error	*t*	*p*-value
Intercept	.127	.024	5.373	< .001
AR(1)	.809	.053	15.207	< .001
Intervention	-.054	.03	-1.8	.075

R^2 = .77

Table C.5 Linear regression model for percentage of spam sent from the US, 1998-2008

Variable	Coefficient	Standard error	*t*	*p*-value
Intercept	.46	.064	7.158	< .001
AR(1)	.895	.034	26.691	< .001
Intervention	.026	.059	.431	.667

$R^2 = .849$