Spring 2015

An Analytical Study Of Consumer Trust In Cloud Computing

III Albert Horvath
North Carolina Agricultural and Technical State University

An Analytical Study of Consumer Trust in Cloud Computing

Albert Steven Horvath III

North Carolina A&T State University

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Department: Computer Science

Major: Technology Management

Major Professor: Dr. Rajeev Agrawal

Greensboro, North Carolina

2015

The Graduate School
North Carolina Agricultural and Technical State University
This is to certify that the Master's Thesis of

Albert Steven Horvath III

has met the thesis requirements of
North Carolina Agricultural and Technical State University

Greensboro, North Carolina
2015

Approved by:

Dr. Rajeev Agrawal Major Professor	Dr. Cameron Seay Committee Member
Dr. Mahour Mellat-Parast Committee Member	Dr. Clay Gloster Department Chair

Dr. Sanjiv Sarin
Dean, The Graduate School

Biographical Sketch

Albert Steven Horvath III earned his undergraduate degree in English from Michigan State University in 1996. He was commissioned in the United States Army as an Infantry Officer upon graduation. In 2005 he attended the Information Systems Operations Leveler and Manager courses at Ft Gordon, Ga to earn the Military Occupational Specialty of a Functional Area 53 (Information Systems Manager.) In subsequent assignments he served as an Information Systems Manager at the 35th Signal Brigade and XVIII Airborne Corps G6 where he deployed twice to Iraq. Next he served as an Information Systems Manager and Information Assurance Manager at the 1st Sustainment Command where he deployed to Kuwait and Afghanistan.

His work in the Information Technology field spans practical experience in managing information services in austere environments that include: acquisitions, help desk operations, server room operations, tactical operations center operations, knowledge management, email, security policy, and Signal personnel management for a user base of over 40,000.

Upon completion of his Masters of Science in Technology Management from North Carolina A&T State he will join the Army Forces Central Command headquartered in South Carolina and Kuwait.

Dedication

This work is dedicated to my wife, Jodie. Our years together have been shaped by the Army, yet not defined by it one bit. It has given us all that we have while putting us through some very interesting times. Through all of it I have had no better champion, cheerleader, and partner than you.

A man's wife has more power over him than the state has.

-- Ralph Waldo Emerson

Acknowledgements

First and primarily, I want to thank Dr Rajeev Agrawal for his support throughout my stay at North Carolina A&T State. My entire graduate experience has been defined by him and, by extension, the other students he mentored with me. He knows what "right" looks like and through strong coaching will get any student to perform. You are an example to be followed. I also want to thank Col Theodore Lennon (US Army RET), A&T alumni and Dr. Eric Fretz, University of Michigan (US Navy LCDR RET) for their instrumental support in my acceptance to graduate school. Thank you to the US Army for the education, in school and out. Finally, thank you to the members of the US Army FA 53 community of technical leaders for their advice, encouragement, and technical support whenever I needed it.

Table of Contents

List of Figures

List of Tables

Abstract

The Internet has developed to such a point that many scholarly articles are calling it the 5th utility, behind water, power, sewage, and telephone. The usefulness of the fifth utility is undeniable and will certainly only grow. Soon, common internet users will be able to do more than just use it for entertainment and shopping. Emerging technologies have the ability to intelligently connect people to the data they need to improve their lives. For instance, heart rate monitors can be connected remotely to the internet and patients can live at home knowing that if there are any problems, help will be called even if they are alone and do not have the capacity to make the call themselves.

The largest obstacle holding the average person back from using the internet in more meaningful ways is trust. The patient with the heart monitor must understand that his personal data, phone number, identification numbers, address, and other personal information is safe from those who might take advantage of a sickly person. It is difficult for the average internet user to trust that their personally identifiable information (PII) is safe on the internet. Almost weekly the American news media reports new, devastating breaches of personal data in big business. Rarely do they publish how well some companies protect their users.

According to a 2010 survey conducted by the Fujitsu corporation, 88% of users, world-wide, are worried about who has access to their data and almost that much is worried about where their data is physically stored (Sato, 2010). We offer a survey and analysis to show that there is a consumer problem with trust and that there are ways for cloud service providers to gain that trust. The ultimate goal of the study is to educate users and CSPs of the problem that exists and suggest ways to overcome it.

CHAPTER 1

Introduction

The information age has matured to the point where most citizens of developed nations have access to computing resources. To comfortably exist in a highly developed nation the average citizen is nearly required to have access to computing resources in order to have access to other basic services such as banking and bill paying. The efficiency that computing brings to business makes paying for basic utilities (water, electricity, gas, and telephony, and soon to be health care) on the Internet the norm. The Internet is often referred to now as the fifth utility. The problem is user's trust in the services offered on Internet. If security is not handled properly, the entire area of cloud computing would fail since cloud computing mainly involves managing personal sensitive information in a public network. Data service providers of all types need innovative ways to be able to draw customers and learn from the data those customers they use. Service providers must become the biggest proponents of data security and privacy to gain the trust and business of the masses.

Trust is strongly tied to Internet security. One study shows that in a survey of scholarly papers on security concerns for cloud computing, few papers actually concentrated on the subject of security itself. Of the 200 articles reviewed, only 80 had anything to do with the subject of data security. This showed that data security should be the subject of more research. Of the 80 articles, 60 showed proposed solutions to data security problems so what little research has been done appears to be effective in finding solutions. (Gonzalez et al., 2011) The research being done is finding solutions but there simply isn't enough being done.

One challenge for the cloud storage industry is the importance of educating the consumer. Most consumers simply accept that their personal information is not safe anywhere on the

internet and they will refuse to put it there. The irony is that most consumers already use the vast majority, if not all of their PII in their online experience. People enjoy the freedom that the Internet provides to the point that some banking institutions are entirely online. There is no need to go to a branch office when the convenience of online banking and bill paying saves the trip. This, however, requires the use of PII online. A certain level of trust is given to banks. Social media is another area where a lot of PII is put online yet users are always distrustful of the ways social media sites use their information. Our goal is to come to a better understanding of why.

There are many challenges in trying to quantify and change something as vague as trust. It is even more unclear how trust relationships are established and maintained on the Internet. Even more interesting is that the problem with consumer trust in cloud computing has as much to do with education and behavior as it does with the underlying technology. Technology changes and improves over time but the more difficult issue is changing culture.

Our research shows that consumer trust in their cloud service providers (CSPs) is a significant issue and provides a proposed solution. We accomplish this by conducting a survey of Internet consumer opinions and combining that with interviews of CSP leaders. This will determine the extent of a problem with trust and what industry is currently doing about it. To date, only the survey of consumer opinion has been completed.

1.1 Background

In November and December of 2013 the retail store Target suffered one of the largest data breaches in history. While it was not the largest breach, it was one of the most covered by media. In the first fiscal quarter after the breach Target reported a 150 million dollar loss to investors. This loss came from breach related costs such as insurance and free credit scanning for affected customers. Over time the costs could reach up to a billion dollars. Lost in the

breach was credit card and personal information for over 100 million customers. (Abrams, 2014) It was not until the rising popularity of the Internet that companies were obligated to protect their customers from such dangers. Part of the responsibility lies with the customer to protect their own information. What is readily apparent is that this is a fairly new and ever developing problem. All of the top data breaches in Internet history have occurred since 2006. Most of them happened since 2013. (Palermo, 2015) Businesses that want to operate on the Internet are taking the steps they can but what has changed for the average user? How are they getting smarter on how to safely operate online within social media, shopping, or even medical records?

1.2 Problem Statement

Convenience and reliability have developed in recent years to the point that the average consumer should be able to rely on cloud storage providers for storage of personal data. Most consumers do not trust cloud storage providers for various reasons. There is, however, a news media culture that sensationalizes every lapse in security. The problem with reporting on data breaches is that when companies like Target do secure data responsibly, there is no positive news media report. The nature of the industry is such that there is no effective way for a storage provider to inform the public of how well it protects data other than advertising. To do so would probably invite hackers to accept the challenge. There is a distinct lack of good news in the media that only leads to paranoia and distrust by consumers. Additionally, the average consumer doesn't feel qualified to protect his or her own data as we will show in our survey.

1.3 Significance

The significance of this study is that if we discover the parameters of the trust issue between consumers and cloud storage providers will allow us to create software and education to solve the issue. The impact of the study will be an overall stimulation of the internet economy

by optimizing ads to consumers. This can be a boon to cloud service providers that need the revenue from more storage customers and the advertisers that would cater to them. Consumers benefit from confidence in the security of their Personally Identifiable Information and are exposed to advertising that actually appeals to them.

If consumers are given the tools they need and the education they want, both they and the corporate world will benefit. Consumers would have a safe and easy place to back up their PII while cloud storage providers would benefit from increased business.

1.4 Research Questions

- Is there a consumer lack of trust with cloud service providers?
- Is there a way for cloud storage providers to earn consumer trust?
- Will it be profitable for cloud service providers to work towards consumer trust?

1.5 Limitations

The survey conducted was limited to mostly residents of the United States. The survey had good responses but was limited only to distribution through social media and a group of United States Army Information Technology professionals. This survey was distributed online and therefore would not include the opinions of respondents who completely distrust the Internet or do not use it for economic reasons.

1.6 Definition of Terms

Personally Identifiable Information-(PII), as used in US privacy law and information security, is information that can be used on its own or with other information to identify, contact, or locate a single person, or to identify an individual in context.

Cloud Service Provider-(CSP), Any cloud computing company that, as a part of its business plan, provides data storage solutions for individual consumers.

Consumer- Any person who uses the internet for data processing needs. This includes any combination of email, purchasing, data storage, social networking, or browsing. A respondent is a consumer who completed the survey.

CHAPTER 2

Literature Review

2.1 Introduction

There has been much research on the business impacts of Internet commerce. The effect of social media on business is also an area that is well researched. Based on the difficulty we had on finding information, there is not enough scholarly focus on the trust relationship between a user and a business, especially Internet based businesses. The concepts that were difficult to find were those relating to trust establishment between consumers and Internet based businesses and the psychology of establishing and maintaining that relationship. The following sections outline the best and most recent research that we could find while trying to discover how trust is built without personal interaction. The review is divided into three sections. First we look at the psychology of trust whether or not it has anything to do with the Internet. Second, we look at the more popular technical methods that companies use to outline relationships with their customers. Finally, since there were not a lot of studies done directly relating to trust in cloud computing, we look at metadata studies of related issues. We did this to find out where current research is leading and how the subject of trust is trending.

2.2 Psychology of Trust

Two Patterns of Trust by Strub and Priest is not related to the technology field at all. Its worth comes from the fact that it is an early and respected study of trust in interpersonal relationships. The nature of relationships has changed due to developing technology but trust as a basic human emotion must be dealt with. The addition of technology to the act of establishing trust complicates the process. Personal interaction should occur for trust to be established more

easily. The paper has some insight on how to do that. Strub and Priest offer two methods of how trust can be established. They are through disclosure and extension. First, with the disclosure method, a person decides whether or not to trust another simply by observing how much information the second party is willing to give about a given subject. In this study, the subject was marijuana use. The second method is extension. I trust A and his judgment; A trusts B; I will trust B.

We can hypothesize about how these two methods would translate into an online establishment of trust. Through disclosure, a CSP can make all the information it can available to the prospective customer online. There are two problems with this method. First, the amount of information involved in protecting privacy is daunting. Privacy policies have to cover every product that a CSP offers and then go into detail about each one. This leads to multiple pages of information. Second, this information may be hard for the average consumer to digest. Privacy policies are written in a legal lexicon that can be hard to understand. Through the extension of trust technique, a company secures the trust of a few personalities or other companies with long standing reputations as good stewards of data; they can more easily establish trust with new consumers. This method also has drawbacks in that there are few well known personalities that are associated with trust and the Internet. The same can be said for companies. Security companies do not want to draw unnecessary attention as that can paint them as a target for hackers. (Strub & Priest, 1976)

2.3 Technological Approaches to Trust

In their paper, A Brief Survey on the Security Model of Cloud Computing, Xue and Zhang introduces the concept of security and different levels of security according to the desires of the user. They propose a fairly simple model: SACS includes Access Authorization, Security

API, cloud connection Security. The security model is somewhat simplistic. The flow chart of security doesn't stand on its own but it seems to be a good start. The authors accomplish essentially the same thing that we want to accomplish with more modern tools. They propose that all connections to the user are SSL and all communication is certificate based. We further propose that higher level classification items be protected by 1024 bit encryption. It is not likely that 1024 bit encryption can be broken in real time today. The important concept from this paper is that cloud security must be addressed at as many levels as possible to maximize security. The challenge, for companies that adopt this sort of model, is to educate the consumer on how they do business and why it is better than their competitors. The authors have a great experiment that maps system performance vs time while running their security measures and while the measures are turned off. This is one way to simplify security concepts for the average consumer. (Xue & Zhang, 2010)

Carlin and Curran use about half of their paper defining terms. There is good reason for that. It is common to see that there is no standardized lexicon for the emerging cloud computing field. Sometimes the lines are blurred between terms. A good example is Software as a Service (SaaS) versus Platform as a Service (PaaS.) SaaS refers to how applications are delivered. PaaS refers to the location of where applications are deployed. Google Docs can be viewed as either SaaS or PaaS depending on what equipment you are using it on, like a Chromebook, and what an individual's opinion is on the definition of the platform. Overall, the article is a good snapshot of the development of cloud systems in 2011. The interesting part is when they describe solutions to the trust/security problems in cloud environments. Only three years ago they suggested that encryption would be a large part of the solution to consumer trust in cloud service providers. They say that many believe that cloud authorization systems are not robust enough

with as little as a password and username to gain access to the system. The down side of the encryption solution is that even the slightest glitch in data can prevent a file from decrypting and it would be lost forever. The other down side is that encryption is processor intensive. This is not much of an issue on the user side, on the service provider side it could be a problem. Recently, a few cloud service providers started providing end to end encryption services. Files are encrypted before they leave the consumer's computer. This alleviates the problem of the provider using processing power. Carlin and Curran further suggest monitoring software and third party auditing. This may be worth developing into what the criteria for a cloud service provider auditing company should look like. Third party auditing is always a good way to get an impartial look at a business. In this case, it bolsters the case that consumers have good reason to trust the provider. (Carlin & Curran, 2011)

A more recent paper serves as a good introduction to the issue of trust in a technological world. Trust in Cloud Services: Providing More Controls to Clients proposes that trust is not something that can be bought yet it will ultimately determine how successful cloud computing as a concept will be. It says that trust must be earned and currently isn't by cloud service providers. The paper doesn't focus on cloud storage providers but does provide valuable insight into the problem. The main focus, according to this article, is on control at many levels. If the user finds that he has more control over the cloud experience, it is the authors thought that more trust will be given to the provider. Specifically, they show five areas of control that could improve trust in cloud computing. The five areas are stored data, data during processing, software, regulatory compliance, and billing. In each area the authors give brief suggestions on how to improve trust with users. The problem is that the suggestions, while good, are not answered specifically enough. This concept hints at a possible way to educate users on the processes a cloud service

provider may use to protect their PII. Could this be used to find specific implementations for service providers to model? Could this information, when taught to the customers, be used to build trust? Ultimately, the article directly relates to our ideas in that the authors want to give control of sensitive information and computing assets to the users rather than asking the users to trust them. (Khan & Malluhi, 2013)

2.4 Metadata Studies of Trust

Personal Data in the Cloud: A global survey of consumer attitudes concentrates on the new and dynamic nature of cloud computing. It is the most significant influence on our work. This study supports the thought that as average consumers learn more about the cloud computing environment, the more they will feel they have control over their data and will trust large corporations with it. This work is important because in a search of scholarly articles, there are few with such detailed statistical analysis of consumer attitudes on cloud computing. Masaharu Sato's work shows that the average user, or consumer, is intrigued by the possibilities of cloud computing to add convenience to their lives. This implies that many goods and services can be marketed better if consumers use cloud computing, but they do not. The study shows that as much as 88 percent of consumers are worried about who has control of their data. It also shows that consumers across the world look to their governments to police the use of their data but do not trust their ability to do so. What emerges is the need for corporations to actively pursue gaining consumer trust. The paper has no suggestions on how to go about solving the problem but the study shows that no one size fits all approach will work. 91 percent of the consumers surveyed want a system which enables them to control how their data is used. Not every country or even every group in one country views personal data the same way. Whatever solution is found, it should be flexible and scalable. Further work in this area should concentrate on

methods that corporations use to gain consumer trust. Sato states that an analogy could be drawn with the evolution in banking from small, local, closed banks to today's globally integrated economy. The paper gives an overall hopefulness that consumer desire to enjoy the benefits of sharing data overrides the fear of losing it or it being used for criminal activity. (Sato, 2010)

A Quantitative Analysis of Current Security Concerns and Solutions for Cloud Computing is important because there are few quantitative studies of risk in cloud computing. The subject of security in cloud computing is difficult to measure except in the case that security related events have occurred in the past and now can be analyzed. This is difficult to do with cloud computing due to its relative infancy. The authors took the approach of quantifying the studies of certain areas of cloud computing. They looked for references in academia, organizations, and companies that covered a variety of security related topics within cloud computing. They took that raw data and produced pie charts to represent the variety of security concerns. They then used radar charts to illustrate the volume of solutions to the security concerns. By looking at what academia, groups, and organizations are studying, they are quantifying a somewhat qualitative topic. This is a great approach to understanding what risks in cloud computing security need the most attention. It is especially important for focusing future studies. The best example of how they do this was highlighted by showing that of the 200 articles reviewed; only 80 had anything to do with the subject of data security. This showed that data security should be the subject of more research. Of those 80 articles, 60 showed proposed solutions to data security problems so what little research has been done appears to be effective in finding solutions. The biggest gap came in the area of virtualization. This is not surprising because virtualization is a fairly newly developed concept. There are less than 20 articles proposing solutions to virtualization issues within the 200 articles surveyed. This shows a huge

gap in knowledge about virtualization security without actually diving into technical specifications. This study of studies is an overview of the many issues facing cloud computing. It is helpful to our research in that we can relate the subjects that most need attention to our thesis. It showed that virtualization, data control, and isolation are great concerns for further research without a lot of proposed solutions. This strikes at the heart of our research into the problems that consumers have in trusting cloud service providers with their data. Consumers have little understanding of the technical problems associated with virtualization, data control, and isolation. It is up to the service providers to find solutions and educate the consumers. The study very clearly shows the areas related to cloud computing that need attention. It is interesting that the areas that need the most work are not technical. Legal issues and regulatory compliance shine as the biggest concerns. Now there can be focus on the proper course of action for service providers to earn consumer trust. (Gonzalez et al., 2011)

The authors of Trust Management in Cloud Computing: A Critical Review spent considerable time on developing an exact definition of trust that is measureable. They propose twelve different trust calculation models and ultimately compare them in a chart. "If the security is not handled properly, the entire area of cloud computing would fail as cloud computing mainly involves managing personal sensitive information in a public network." The authors describe the models and the associated relationships of the key players without going into much detail. The key players are the user, the CSP, and third parties. Third parties can consist of storage providers, certificate authorities, auditors, trust verifiers, and others. The survey looks at mathematical methods that quantify the relationship between the parties. This, in spite of being a mathematical view, is a qualitative way of looking at the problem in measuring consumer trust in cloud computing. Ultimately the article poses that there should be a secure way to allow all of

the key players to be involved in establishing and maintaining a user's trust. The ultimate goal for the business is to make money on storing and mining the user's data, if permitted. The goal of the user is confidentiality. Somehow they should be able to agree on these mutually exclusive goals without sacrificing service. The authors propose that the answer can be found using the principles they discussed but none of the models discussed had been tested on cloud computing systems. This proposes many areas for future study such as implementing one or more of the trust models in a test environment. (Firdhous, Ghazali, & Hassan, 2012)

The paper Trust in Cloud Computing takes a bird's eye view of the consumer trust issue. It doesn't offer very specific solutions on a technical level to gain consumer trust. What the authors do is outline the trust issue and suggest, through a business practice example, how to gain consumer trust. The most important aspect of this article, though it is more qualitative than quantitative, is that it shows in 2013 the cloud service industry still struggles with consumer trust. Amazon launched its EC2 architecture in 2006. Microsoft followed suit with Azure in 2010. These two large service providers have had over 12 years of combined experience to earn trust yet the primary complaint of consumers is still trust. The paper suggests several avenues to gaining consumer trust. Service providers have to address not only trust but control of the data, ownership, prevention of attacks (rather than compensation after an attack), and security. Overall, the theme is prevention. In time, successful prevention leads to consumer trust. The paper identifies a top vulnerability to cloud computing. Virtualization software makes cloud computing possible by allowing the addition or subtraction of servers in a virtual environment. One physical machine can run many virtual machines (VMs). In the most prevalent virtualization software one VM must be created to manage all the others that are created. This one can control and view activity of the others and is vulnerability. One of the key solutions to

gaining trust and making cloud computing more secure is virtualization to accomplish data confidentiality for guest virtual machines. By applying this solution, even the infrastructure as service providers cannot access the private information of their customers. The paper then gives a hypothetical corporate structure to show how a trusted solution would work. In theory, if virtualization was indeed made completely confidential, it would not matter where the data is stored. There are some areas that this article should consider. For instance, if the data was called for within a confidential criminal investigation yet the storage provider was located in another country, there are differing laws on how to get to the data. Still, the storage provider could turn over all the data without compromising the trust of the customer by ensuring that the data is encrypted at all times. The data can be turned over as ordered but it would be up to the authorities to obtain the key or crack the encryption. Future work on this paper should be a more detailed look at how secure virtualization should be implemented. The authors also have given good examples of how outages have affected service providers. They should quantify they actual threat from criminals who exploit the top threats identified in the paper. Trust is not just a technical plan or implementation. It is a psychological issue that a technical solution can help, but not solve. Public relations and transparency of an organization are also areas that should be explored. (Bhosle & Kasurkar, 2013)

A Framework for Accountability and Trust in Cloud Computing strikes at the heart of the trust problem but it does not propose a very good solution to the problem. It is a great introduction to the subject of trust and is a good format for the terminology of the subject. It proposes a structure of data importance (classification levels) but does not expound on it. The article looks at cloud data from a provider side rather than a user perspective and it places emphasis on logging. The problem with the logging approach to trust is that it emphasizes the

provider's ability to view the data that is being stored rather than protecting the user's data within the law. Physical location of data in a virtual environment is important because of local laws. This should be addressed in a cloud service provider's promise to its customers. Additionally, a wholly detective approach to earning trust by emphasizing logging has two important problems. First, it doesn't address the psychology of trust. It doesn't allow for the initial gaining of trust with new customers. Logging is a tedious, but necessary task that is already mundane to system administrators. The concept of logging in order to add accountability to any particular cloud scheme would be completely lost on consumers. Auditing logs from an accountability standpoint is a good thing and is definitely a selling point to new customers; it is just not the best path to new customers. The second problem is that logging, while important, cannot be the entire depth of a cloud provider's defense. There are many layers, as the article suggests, but logging at each layer is only a small part of a good defense. (Ko et al., 2011)

Cloud Computing Privacy Concerns on Our Doorstep is a brief article that uses the example of conference organizing software to demonstrate the benefits and perils of cloud computing. It highlights the importance of clear, up-front, policies for the use of the data generated and collected. The article suggests that there are encryption systems that would allow CSPs to search uploaded encrypted data without decrypting it. This is not cost effective yet. There are times when a consumer wants to upload data that he only wants to share with certain people. The example in this paper is the case of an academic conference where it is necessary to share papers. The scenario could also apply in terms of sharing photos on a social media website. The scenario would require encryption with shared keys. This brings up an interesting concept in that trust could possibly be enhanced with social media with the use of user keys. The problem is that user education and cost are prohibitive. Even further, the use of tokens would

increase social media security greatly but again, is cost prohibitive and complicated. (Ryan, 2011)

Addressing Cloud Computing Security Issues does a great job of giving a quick history and definitions of the concepts in cloud computing. The authors propose, like many other papers, a more thorough integration of certificate use. They propose that a Trusted Third Party (TTP) is charged with security at different points in the cloud architecture. "Object reusability is an important characteristic of cloud infrastructures, but reusable objects must be carefully controlled lest they create a serious vulnerability." This is an excellent quote because cloud computing concepts allow for automatic growth based on consumer need, use reusable objects. They also talk about low and high levels of confidentiality. This implies differing levels of encryption and/or transport security for different classifications of data. This is something that perhaps, using a military classification scale; a user can have as an option when he or she sets up an account with the cloud service provider as an option. Education is important in this kind of model so the consumer knows what data is needs protected and what data is going to be mined. (Zissis & Lekkas, 2012)

The authors of Protection of Identity Information in Cloud Computing without Trusted Third Party explain why eliminating a Trusted Third Party (TTP) is beneficial. They show a few ways that Identity Management (IDM) is currently done in business today. Without a TTP the user must somehow be able to authenticate to a system without using PII. The user must also be able to use an untrusted host since there is not TTP to verify identity and trust. The authors propose a new method of securing/verifying traffic to ensure confidentiality but still use TTPs for some duties. For a practical standpoint, the TTPs are actually being replaced with several other parties. This doesn't do much to eliminate the need to communicate with another party, it

actually adds to it. The benefit is that the end result is much more secure. The idea is only in an experimental phase now and instead of using a TTP. It is not likely that this is revolutionary and that it will replace standard identification and security schemes already in place. (Ranchal et al., 2010)

Establishing Trust in Cloud Computing is a general paper that serves to outline the current state of consumer trust with cloud service providers. It shows that there are several specific reasons for consumers to be worried about their data and it breaks that information down into sections. First the authors define trust and the issues related to trust as seen in a cloud computing environment. Next, they give a generic example of the many complicated issues at play using an imaginary cloud service provider. They then make suggestions on how to solve the issues they first mentioned through things like service level agreements, software tools, and architecture standards. Several times throughout the paper the authors discuss different ways that virtualization should be addressed. This is particularly important as virtualization is the key to scalability in cloud implementations. There are several advantages to the approach the authors take to solving the problem of consumer trust in cloud computing. First, they understand and recommend how to keep data physically located in states and countries that have proper protective laws. They recommend that this information should be readily available to cloud users. Second, confidentiality should be treated with the latest in cryptographic technology and processed in such a manner that even the service provider cannot decrypt the data, only the user. Third, accessing the data should be audited and controlled by the user. Fourth, a certification process where industry best practices can be implemented and certified by independent companies. Finally, cloud architecture should be addressed on a physical level to ensure that the virtual level is secure. These are all great steps to follow to give assurances to

users that their data is secure. The down side of the paper is that it is not detailed enough. To the author's credit, addressing all of the above issues in one paper would be daunting. The first issue, user and company agreements could easily encompass an entire thesis to try and determine the best ways to implement. The same goes for the other issues. Additionally, the authors are focused solely on the company side of cloud services. They should take into account the training and understanding of the user before implementing changes. For instance, what good is a user agreement if the user doesn't understand what he is reading? Again, this relates to the level of detail in the paper. On a conceptual level this paper is great in outlining the problems in trust but on a practical level it doesn't scratch the surface. (Khan & Malluhi, 2010)

The authors of Taxonomy and Survey of Cloud Computing Services propose a standardized taxonomy for cloud services. They take current services and compare and contrast them. The valuable part of this paper is the realization that without a proper taxonomy, users are left confused and feel less secure because they do not understand. The same concept in cloud computing can be called a different name by different providers. The authors also present a detailed chart comparing the services and capabilities of a large number of CSPs. Interestingly, there are not many, except for the larger companies that use PKI and encrypted links. This is changing rapidly. (Rimal, Choi, & Lumb, 2009)

CHAPTER 3

Trust in Cloud Computing

3.1 Introduction

This chapter will frame the user experience for the purpose of highlighting the results of the survey we conducted. We start by looking at the many choices that the average Internet user is presented with in terms of cloud computing. We then take an interesting look at what our respondents had as concerns about their service providers. Finally, we look at another feature, the privacy policy, and its impact on the user experience.

3.2 User Perspective

In an effort to effectively explain the issues surrounding trust in cloud computing, we first have to explain the context. There are many different aspects of cloud computing to discuss but we will limit ours to the CSP to customer relationship. We also limit the discussion to user trust when it comes to putting personally identifiable information online in several different formats. These formats are social media, email, purchasing, and data storage services. We selected these to outline some of the most commonly used services. This was in an effort to create data that could be used in the future to develop techniques or tools for CSPs to use to increase user trust. This increase in trust should cause an increase in services use and revenue and should also benefit users.

3.2.1 Cloud Service Providers

In order to understand what the context is for measuring trust in cloud computing by the average user we have to first understand where the users operate. The average user, as our survey results will later show, uses cloud services in a few different ways. They may not be aware that what they are using is classified as a cloud service but they do use them.

The top CSPs offer a variety of services so comparing them directly for the entirety of their products is problematic. There are many different services with many different prices and features offered. While pricing is competitive, users have different needs. Some need a simple product that works in an intuitive way. Other users want very specific products with certain specifications. For instance, an older user that is not particularly Internet savvy may want only a service to archive their photos. That person's child may want a service that offers the same thing but they want advanced sharing and even editing software added. Some providers listed in table 1 offer exactly this. Some do not.

Table 1. *Cloud service provider comparison*. (table as of March, 2014)

Provider	**Cost/month**	**Size**	**Encrypt data in motion**	**Encrypt data at rest**
Google Docs	$4.99	100GB	Yes	No
MS OneDrive	$4.16	107GB	Yes	No
Carbonite	$5	Unl.	Yes	1024 bit
Dropbox	$10	100GB	Yes	256-bit AES
SugarSync	$9.99	100GB	Yes	256-bit AES
Amazon Cloud Drive	$50	100GB	Yes	No
Box	$5	100GB	Yes	No
Bitcasa	$10	1TB	Yes	Hash
SpiderOak	$10	100GB	Yes	2048 RSA

All ten of the top CSPs mentioned in table 1 have some free storage to lure new customers. Storage is cheap and getting cheaper so in the competition for new customers, more free storage is offered over time. In February 2015, MS OneDrive and Google both offered 15GB of free storage across their services. Less than one year ago $4.99 bought 100GB of storage space. Now it is only $1.99. In another year the price will likely be lower or the amount of free space offered will increase. Some larger CSPs are working towards not charging for storage at all. The storage price war itself indicates that there are vast corporate interests in

having average Internet users store their data in the cloud. They see the value of the data the store in a couple of different ways. First, they can sell services that help users manipulate the data in ever more innovative ways. For instance, box.com's CEO Aaron Levie was quoted in an interview saying of larger CSPs, "They're not doing it because there's any economics of the storage of information; the economics are that information is now currency." He went on to say that in his business model, "The way to think about it is, we buy storage, and sell the service on top of that." (Lynley, 2014)

3.3 User Concerns on CSP Security

The very essence of user sentiment on cloud computing security can be summed up in one word: education. One survey of more than 1,000 American adults conducted in August of 2012 by Wakefield Research shows that cloud computing is still misunderstood. Some of the highlights of the survey include:

- 95% of those who think they're not using the cloud, actually are
- 3 in 5 (59%) believe the "workplace of the future" will exist entirely in the cloud
- 40% believe accessing work information at home in their "birthday suit" would be an advantage
- More than 1/3 agree that the cloud allows them to share information with people they'd rather not be interacting with in person
- After being provided with the definition of the cloud, 68% recognized its economic benefits
- 14% have pretended to know what the cloud is during a job interview

(Citrix, 2012)

The average Internet user does not know enough about cloud use or cloud security to form a proper opinion. The foundation of trust is a belief in reliability. If Internet users do not even understand CSPs and their products, how can they be expected to trust them with their most personal and valuable data? This is the CSP's problem. If they want users to upload more data and trust them, they have to educate them.

3.3.1 Word Cloud

A good way to visualize the difference between the security sentiments of CSPs and their users is through a word cloud. A word cloud is a simple tool to help visualize what is important in a given text. The bolder and larger a word is, the more it appears in the original text. This is a way of quickly gaining an understanding of what is important to the author without reading long passages of tedious text. In figure 1 we see a word cloud assembled from the text of Facebook's privacy policy.

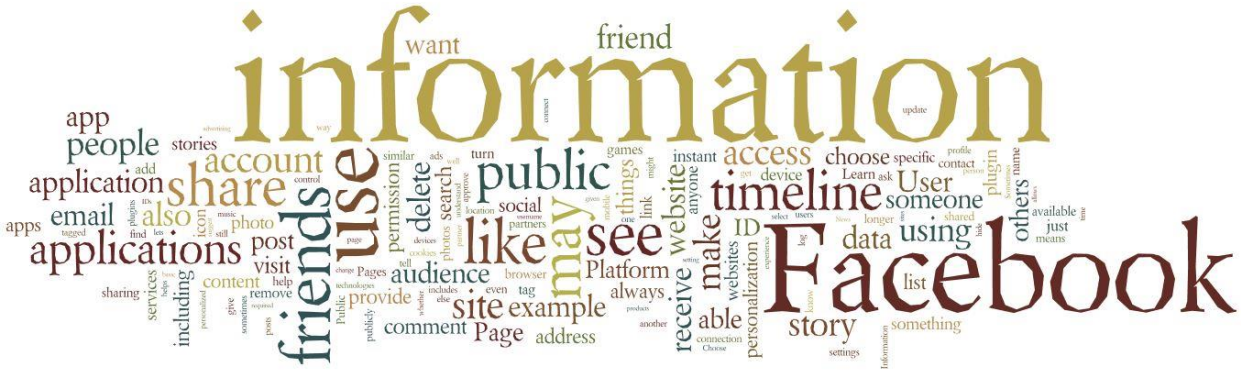

Figure 1. Facebook privacy policy word cloud.

In this word cloud we see many innocuous words and a lot of nouns. If you did not know where the original text came from you could guess that it came from some kind of manual that has to do with information and Facebook.

Figure 2 shows the word cloud produced from the open ended questions from our survey. These questions asked why respondents did not trust service providers and how those providers could gain their trust. We can see with a quick look that the respondents have data and information security in the foremost of their thoughts when they are considering the issue of trust in their CSP. What is particularly interesting is that the word "information" is roughly the same in importance in the Facebook privacy policy and in what respondents found important. Trust is also an often used term in their responses. This adds validity to the idea that Facebook and the respondents are thinking about the same thing.

Figure 2. Respondent security concerns word cloud.

If you didn't know where this cloud came from, you notice there are a lot of emotionally charged words on the lower levels that do not appear in Facebook's word cloud. We see words like care, just, risk, want, vulnerable, yes, and need. None of these words appear in the Facebook

word cloud because to the CSP, user security simply is not an emotionally charged subject. There is a noticeable increase in the number of adjectives as well. This can be attributed to the respondents describing what they really want as opposed to Facebook describing what they are willing to offer.

Open ended questions are difficult to quantify. In an effort to consolidate personal sentiments and project an overall feeling of the open ended questions another technique that stems from the word cloud is the word count. We counted the words and how often they appeared in the text. The more the word appeared in the text, the greater the importance of the word. We used this technique on the Facebook privacy policy and on the open ended questions that our respondents gave us that concerned why they do not trust CSPs and what can CSPs do to earn their trust. Figure 3 shows the top ten words used by the respondents. The words data, security, and trust occupy the three top positions. From this we can clearly see that in an open response discussion, data security and trust relationships are the top subjects on our respondents' minds.

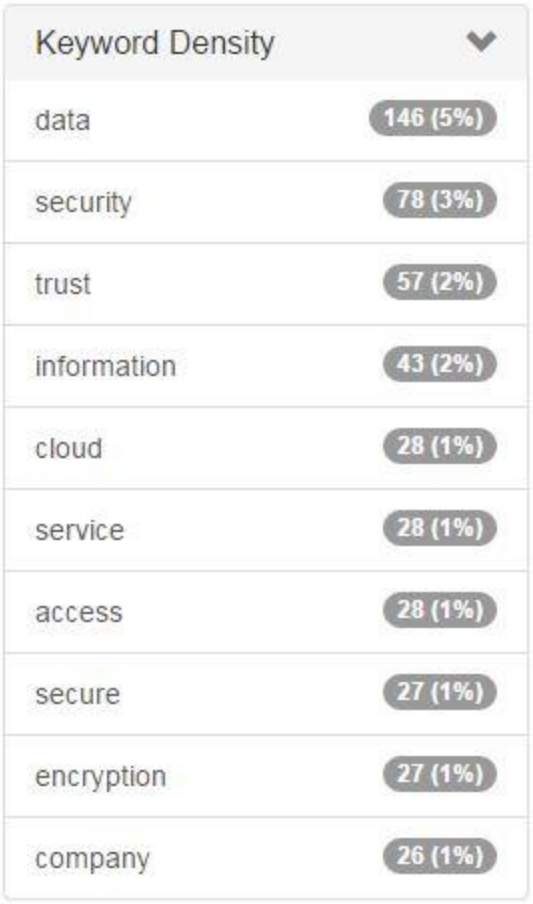

Figure 3. Respondent concerns word count.

The same technique applied to the Facebook privacy policy in figure 4 shows that the top three results are similar to what the respondents' concerns were. The fourth word shows a sharp departure from respondents' desires. That word is share. The most important items to the respondents are data security and trust but Facebook is clearly more concerned with its users sharing information. We can see a clear contradiction of interests between Facebook and potential Facebook users. The cause of that difference can easily be summed up by understanding that it is rooted in the education of the respondents in terms of cloud computing. Facebook has the need and desire for respondents to share information. This is a direct

contradiction to the respondents' desire to secure their information and trust those who have access to it.

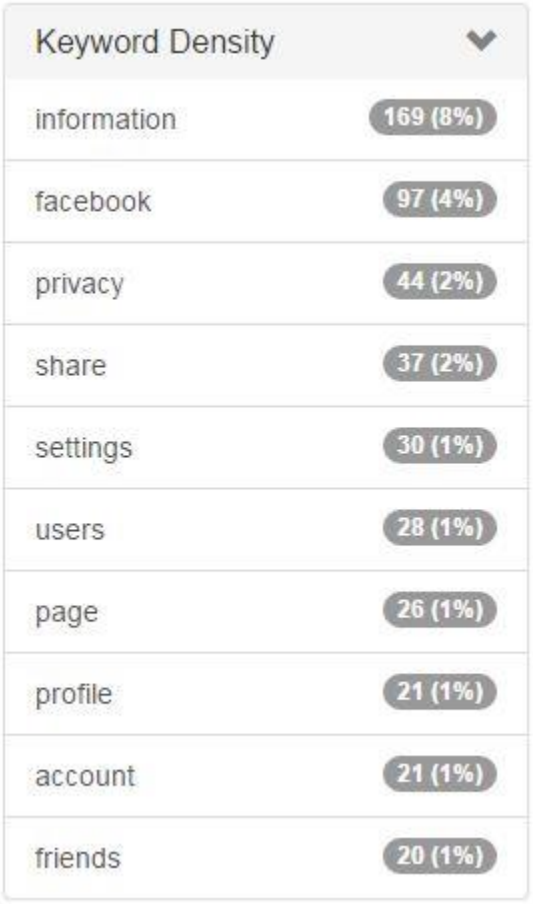

Figure 4. Facebook privacy policy word count.

3.4 Privacy Policies

This section will attempt to clarify the trouble with CSP privacy policies and how they relate to the education of the user. Ease of use is very important to the end user as will later be shown in chapter 4. Privacy policies are as varied as the services that CSPs offer. First we show three different privacy policies from similar CSPs. We compare them in terms of presentation, complexity, and utility. After the observations are complete we will have a better idea of what

the average Internet user experiences when trying to learn about privacy and how his or her data is handled.

We first look at Google's privacy policy in figure 5. The current policy can be found at http://www.google.com/policies/privacy/. It is downloadable into a .pdf format but much like the online document, it is full of links to follow. It is laid out in order focusing on the information they collect and how they use that information. In general, it is difficult to zero in on a specific part of the policy. For example, if a user wanted to know more about privacy in terms of Google's email service, Gmail, there is no quick way to get to it. The term Gmail doesn't appear on the privacy home page at all yet it is the most used service besides the search engine. (Google, 2015)

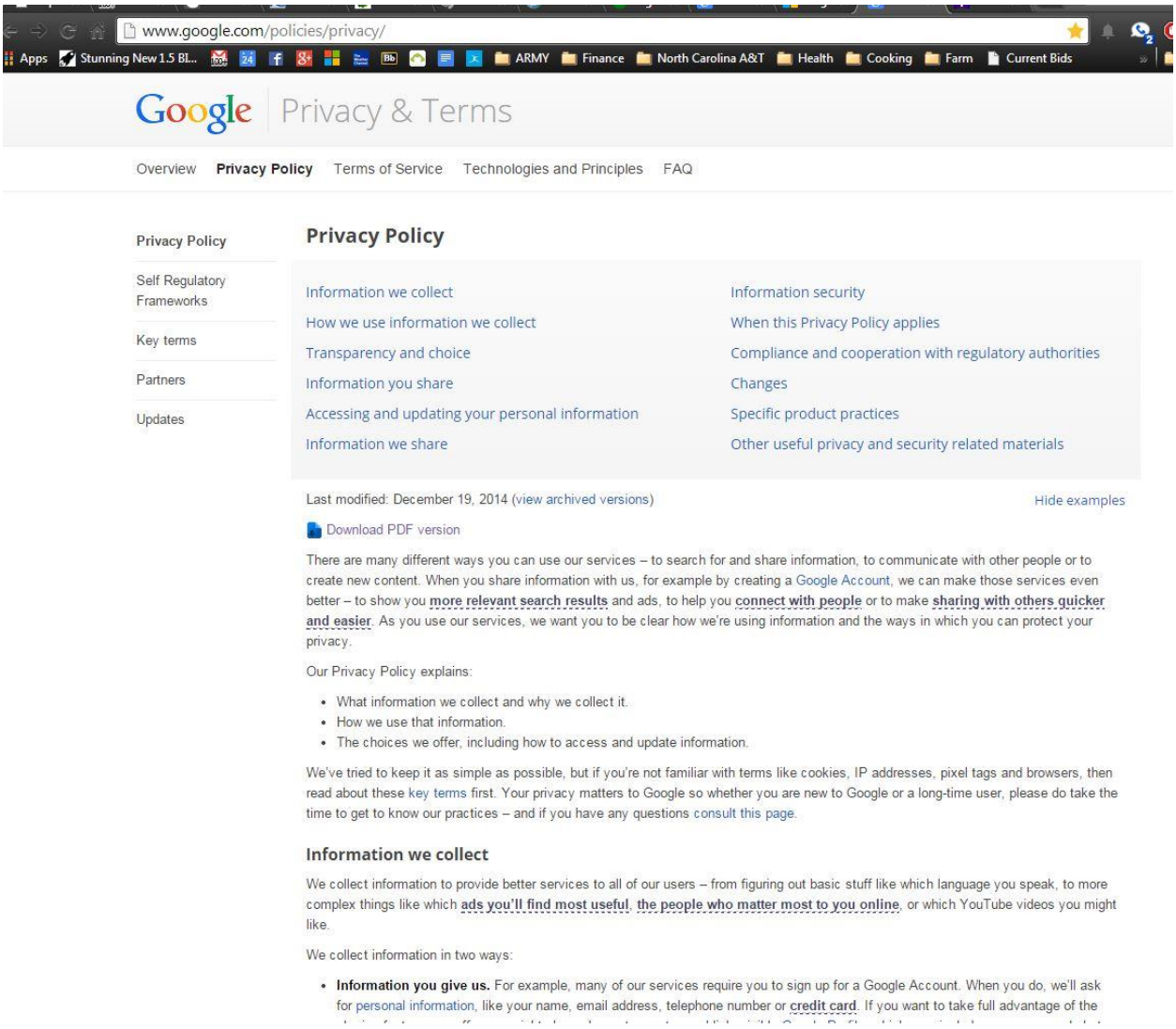

Figure 5. Google's privacy policy.

Another top CSP is Microsoft. Microsoft is closest to Google in terms of services offered, like email, data storage, and online document creation. It would stand to reason that the privacy policy would be similar in layout and readability. This is not the case as can be seen on their consolidated privacy website at http://www.microsoft.com/privacystatement/en-us/core/default.aspx. Figure 6 shows that Microsoft's privacy policy is laid out according to service. There are tabs for each service that Microsoft offers and they are easy to find. We used the same example to find the email service, Outlook's privacy policy, after three clicks; it is still

not apparent where the policy is actually located. On the surface it appears to be well organized but drilling down to the precise information needed for Outlook email is problematic. (Microsoft, 2014)

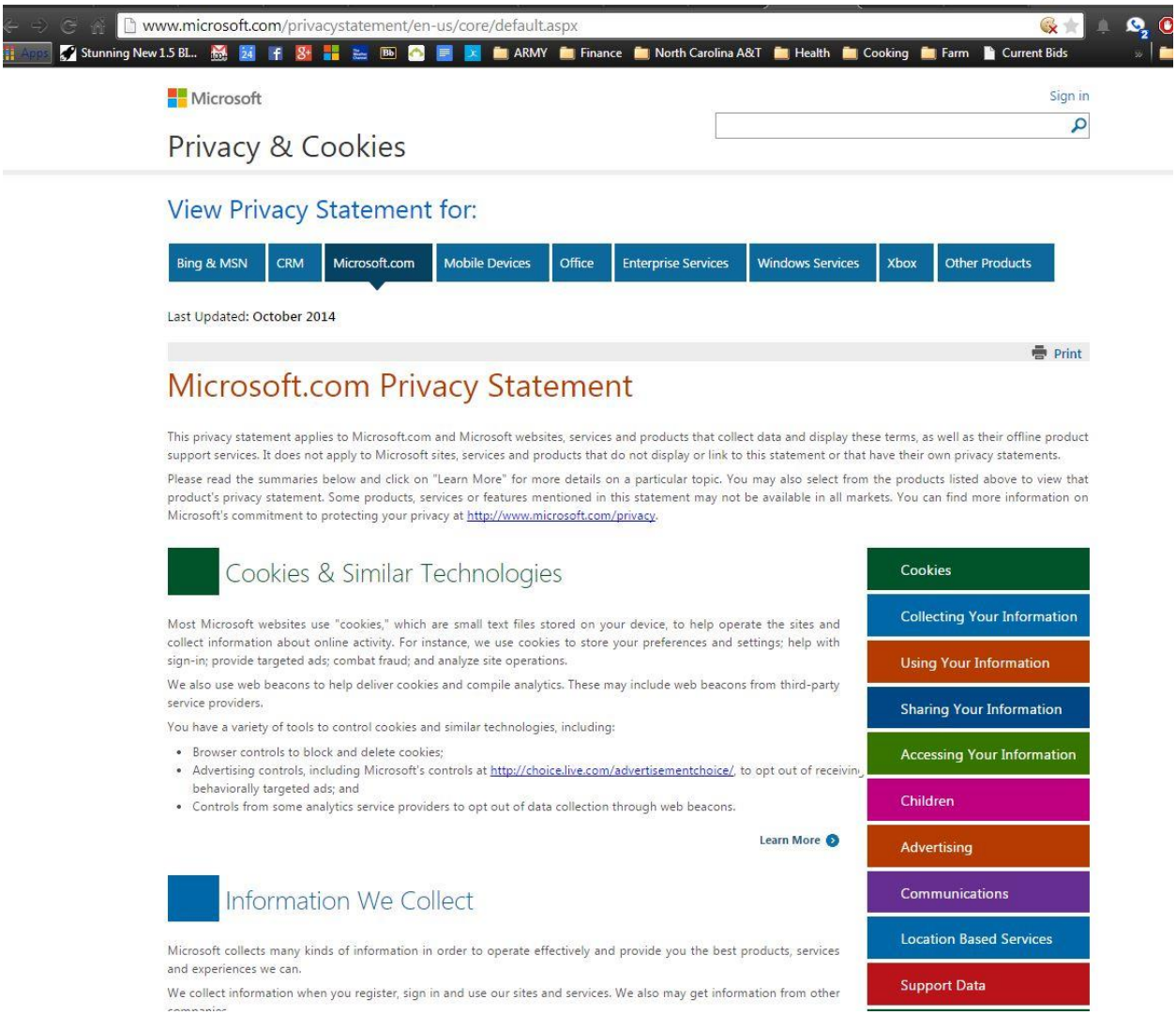

Figure 6. Microsoft's privacy policy.

Finally we look at yahoo.com's privacy policy at http://info.yahoo.com/privacy/us/yahoo/products.html in Figure 7. We again look to see how difficult it is to find information on yahoo's email privacy policy. In one click it was readily

available. Like the others, there were many other links off the email privacy policy page that serves to dissuade users rather than help them. In this case, the information was intuitive and easily accessed. In one click the information pertaining to privacy as it related to Yahoo mail was available. There were many links from the policy out to more specific information but Yahoo clearly made efforts to simplify the understanding of the policy. (Yahoo, 2015)

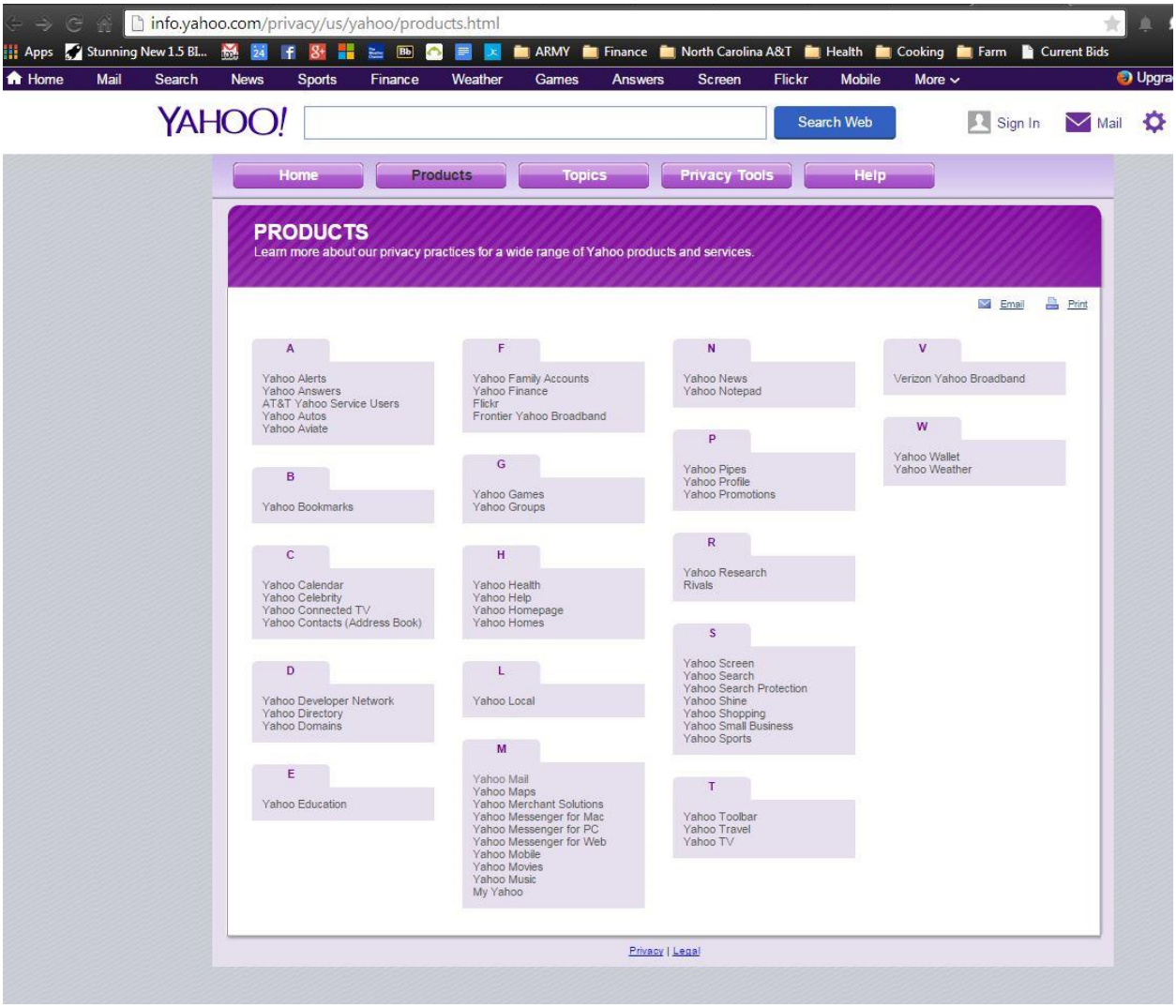

Figure 7. Yahoo's privacy policy.

In summary, all three policy pages are fairly inadequate. They are complicated, full of technology jargon, and massive amounts of information to be digested. Some of the jargon is

necessary for legal reasons and some concepts simply cannot be explained without it. We will show in chapter four that users inherently do not trust CSPs. Here we see that CSPs have not made a strong effort to reach out to users and try to gain their trust. Perhaps they do not understand the problem properly

A suggestion for bridging the gap between users and CSPs is a fairly simple concept. This possible improvement for delivering information can be found with Google. Even though their policy hierarchy was difficult to navigate, they did have videos that outlined the basic concepts that each page had to offer. Videos are an easy, fast way to share a lot of information by combining spoken text with graphs, charts, and images to facilitate better understanding. Yahoo was clearly the best at breaking down information into subject areas. A combined Yahoo/Google approach to presenting policy information would be an improvement. A simplified policy page with links to videos explaining those policies would take a lot of mystery out of them. It would also help in terms of trust because it puts a face to a product. Rather than read a black and white printed page about policies, a user could see a person who explains it.

In summary, it is easy to see how regular Internet users can be inundated with information. They have to choose which cloud service provider they would like to use based on features they want. No two are the same and some have cellular telephone integration. This makes the choice even harder. They also have to decide if the level of security offered by that company is sufficient for their needs. Some companies encrypt data at rest, others do not. Individual company privacy policies are long and difficult to navigate and understand. Each company handles policies in different ways so the provider that is chosen should be well understood. There is also price to consider. Most companies charge roughly the same but they vary by service and size of storage. Ultimately the real deciding factor is going to be trust. Trust

implies education. Users have to know enough about the company to be happy with the level of security they want. Not all users have the technical ability to differentiate between encryption levels so it is up to the CSP to figure out how to teach the user.

CHAPTER 4

Analysis of Cloud Security Survey

4.1 Introduction

This chapter is dedicated to the analysis of the survey conducted for this thesis. We will do this by outlining the objectives of the survey, the methodology used in conducting the survey, and then a detailed analysis of the survey. Each question in the survey will be analyzed within the context of some of the other relevant questions. We will do this to highlight contrasts in opinions and practices of the respondents. Finally we show some statistical analysis of the responses.

4.2 Survey Objectives

Our survey was designed to answer three underlying research questions. First, is there a consumer lack of trust with cloud service providers? Second, is there a way for cloud storage providers to earn consumer trust? Finally, will it be profitable for cloud service providers to work towards consumer trust? The questions have three areas of emphasis, demographics, opinions on security, and finally users' current online habits. They also vary in the data type collected. The survey we conducted is a mix of true/false, Likert scale, and open-ended questions. The section of open ended questions was given but the data from those questions is currently not processed.

4.3 Survey Methodology

The survey was distributed online through several different venues. There were 236 total respondents. The two primary distribution points were Facebook and a listserver for a group of United States Army Information Systems Managers. Facebook proved to be a powerful collection tool in that once respondents took the survey; many also shared it with their friends.

This led to a more diverse population. The initial link was shared on two user's pages but quickly spread through re-sharing with many users that were completely unknown to the original posters. The US Army listserver also produced good results. On the days that the request was posted, survey participation spiked. This was good for participation but due to the education level of the respondents on those days, it may have slightly skewed some of the results as will be seen.

Prior to starting the survey, the respondents were given the definition of Personally Identifiable Information (PII) for purposes of the survey. PII, as used in US privacy law and information security, is information that can be used on its own or with other information to identify, contact, or locate a single person, or to identify an individual in context. (Phone number, address, SSN, banking information, driver's license information) This definition is important to standardize responses. The following section will outline each question and what the preliminary results are.

4.4 Analysis of Survey Results

The survey was divided into three sections.

- Demographic
- User habits/Internet use
- User opinions

An additional section was open ended questions that covered the first two subjects but the results were only used to highlight the statistical responses of the quantifiable responses.

4.4.1 Demographic Questions.

The emphasis on the survey was intended to be about the nature of user opinions rather than comparing the opinions of differing demographics. In hindsight, more demographic

information may have been helpful in answering the second research question in terms of finding certain ways for cloud service providers to earn the trust of consumers. This can be more useful for a more global survey. The focus of the survey, however, was not in contrasting user identities; rather, it was intended to find user opinion regardless of their demographics.

4.4.2 Internet Use Questions.

The following questions focused on opinions of the average consumer on the security of their data. The intent was to first try and figure out how respondents currently used the Internet in terms of PII security and then later ask what they expected in terms of security. These questions brought some surprising results in that it initially appears that respondents actually do put a lot of PII on the Internet already. In fact, we found that almost every conceivable form of PII is already placed on the Internet by respondents, just not all in the same place at the same time. Question 3 asked, "How often do you use the Internet for storing data of any type?" 70.08 percent of respondents said they store data daily or a few times a week. This question shows the significance of data storage for the average user.

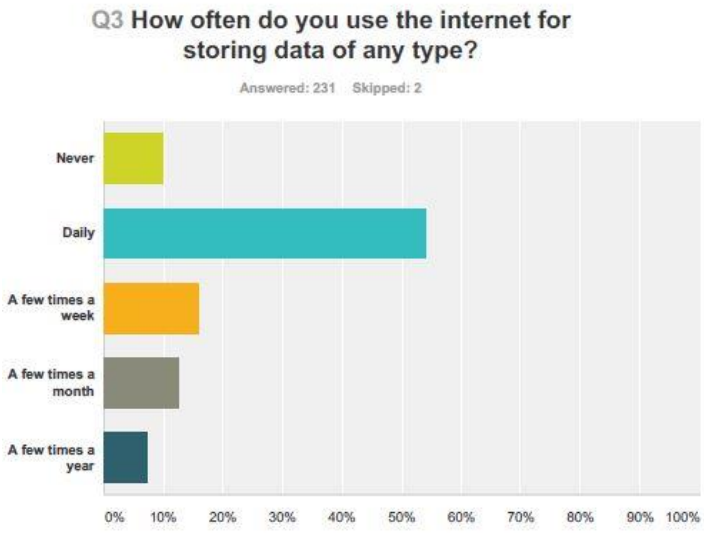

Figure 8. Question 3.

Questions 4 through 6 asked, on an increasing scale of threat, how much PII respondents have online at the time of the survey. Question 4 asked, "Do you currently use credit cards to make purchases on the Internet?" The intent of the question was a little crafty as it was meant to show that even though average users are skeptical and even unsatisfied with the state of information security on the Internet, they still choose to use the Internet to make credit card purchases. The results demonstrate exactly this point. Almost 97 percent of respondents answered that they do use credit cards on the Internet. This is a direct contrast to question 13 which helps to establish that respondents don't trust the entities that currently maintain their medical records. It also shows a contrast with question 14 where respondents overwhelmingly state they have concerns with who has access to their data online.

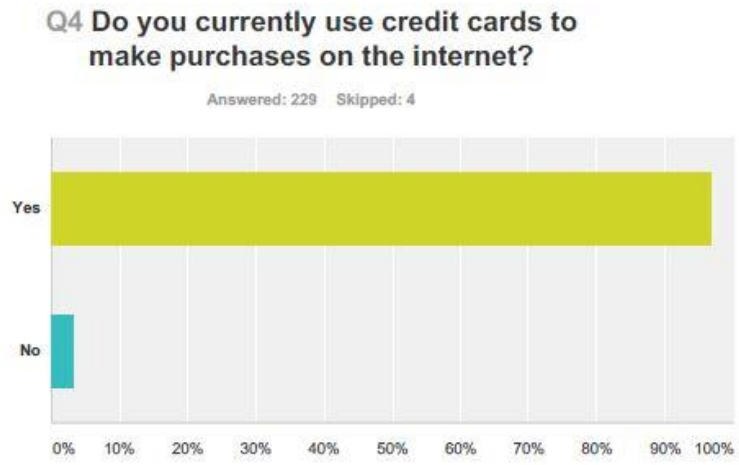

Figure 9. Question 4.

Question 5 was much like question 4 in that it was used to gauge respondent's current trust in online security. It asked, "Have you done your US Tax forms on the Internet? This is a slightly elevated level of trust in that there is more PII on tax forms than a credit card purchase. Also, if intercepted, that PII can cause more damage because it contains banking information, social security numbers, and other information that is more useful to criminals. There was a

corresponding decrease of trust to this increased threat in that only 65 percent of respondents did trust their tax information on the Internet. Later in the survey we show that respondents do not trust their government to enforce the laws that protect their data.

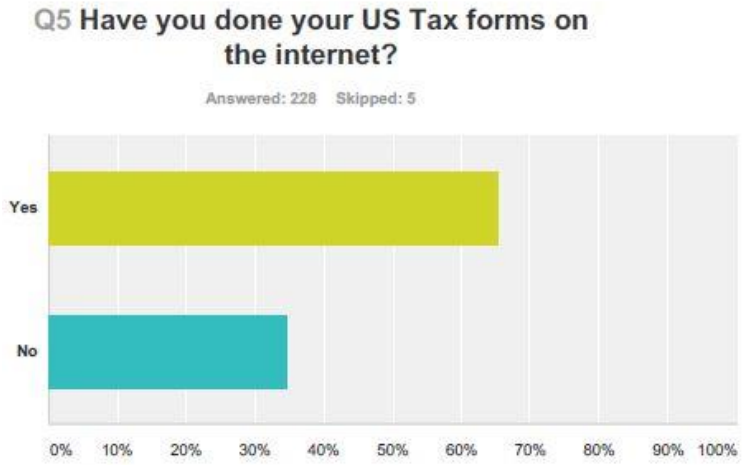

Figure 10. Question 5.

Question 6 openly asks, "Do you currently store Personally Identifiable Information in some type of cloud storage?" This question is a little crafty in that the previous two questions already alluded to the fact that respondents, whether they knew it or not, were already storing PII in the cloud. Again, there is a drop in affirmative responses as this question poses the highest risk of all questions asked thus far. 51 percent said they do consciously store PII online. This is interesting in that when directly asked, 51 percent of respondents admit to doing what they consider to be the wrong thing. It is so wrong, that it is central to the reason they do not trust cloud service providers. This is proven in question 16 where respondents show they do not trust their government.

Figure 11. Question 6.

Question 7 attempts, like question 3, to gauge online activity. Specifically it asks, "Do you currently use cloud based email (Gmail, Hotmail, etc.?") The reason for this was to get a better feel for respondent's level of use of popular email services. The most popular services also offer online storage services that will also likely be used in conjunction with the email service. Almost 90 percent of respondents do use cloud based email. The interesting part about email is that as a service it becomes central to the use of other products like social media and photo sharing. If a respondent uses Gmail, for instance, it is likely that they use Google + for social media and photo sharing. It is also convenient to use their document services as well. The nearly 90 percent response to the question makes it safe to assume that respondents also use services other than email from that provider. Further questioning is necessary for proof but it suggests that of those who use the Internet, 90 percent are exposed to email threats to PII.

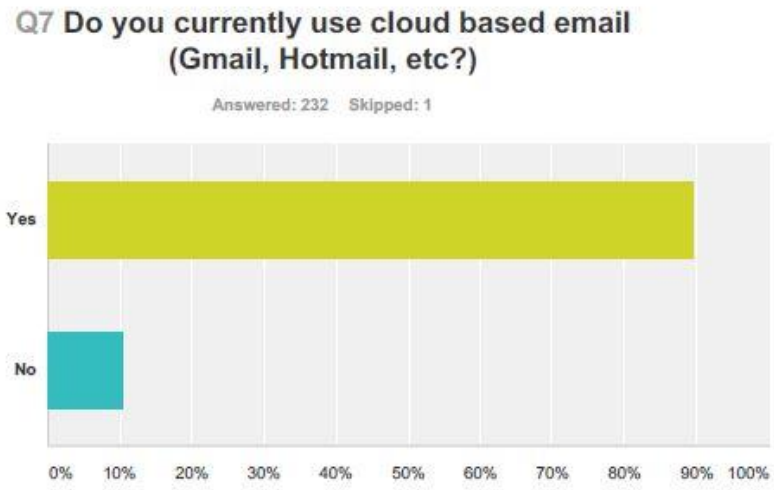

Figure 12. Question 7.

Question 8 was asked to try and begin making a comparison between respondent's current online habits and their current concerns about security. This may eventually lead to investigation into how to better; more simply, inform users of their rights and protections during online purchases. Almost 70 percent of respondents don't read or strongly disagree that they read the online terms and conditions.

Our theory is that online terms and conditions are too large and structured with legal language that makes them a burden to read. As we have seen in chapter 3, there are complicated issues at play between the CSP and the respondents. Through further questioning, we will see that privacy and data security of PII are among the top concerns of respondents. However, if respondents are so concerned with privacy and data protection, why don't they read the privacy policy? A good follow up question would ask why they do or don't read terms and conditions. Terms, conditions, and policies are a good subject for a thorough examination all by themselves. While this is worthy of much more examination, it is a good example of expanded work that falls outside the parameters of this study.

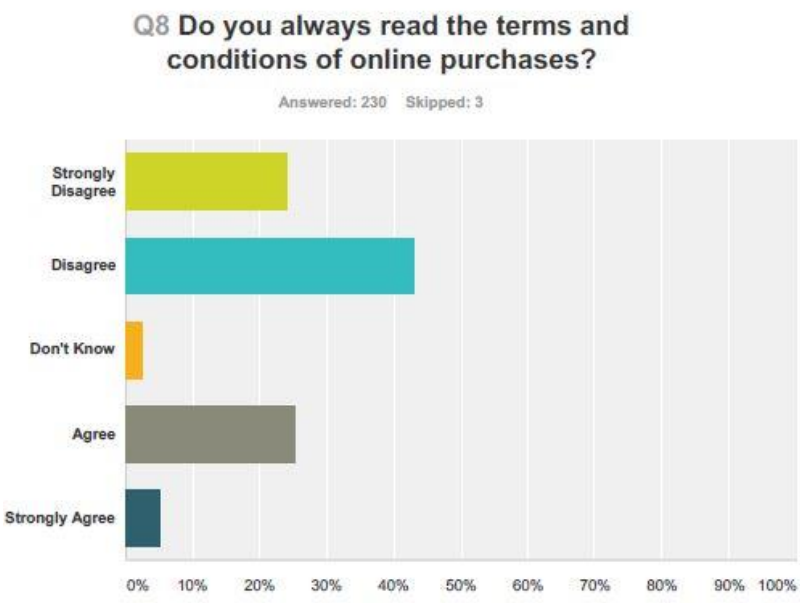

Figure 13. Question 8.

Question 9 also attempted to establish online use habits by respondents. Social media is an area that is ripe for criminal activity. It is dangerous because people put PII there without realizing it. Criminals gain valuable insight into a target's life through the process of compiling information about them. Through social media, criminals gain valuable PII like birthdates, mother's maiden name, street addresses, and family pet names. All of these items can be used to access banking or other online accounts by answering security questions. This is an example, certainly not the only way social media can be used. Over 80 percent of respondents use social media. From the previous question, 70 percent of those respondents have not read the online terms and conditions that contain important warnings about PII security. A recent white paper from a company that provides a support system for law enforcement agencies shows that social networking is a potential gold mine for criminals. In various ways, criminals leverage users' personal details into financial opportunity [7].

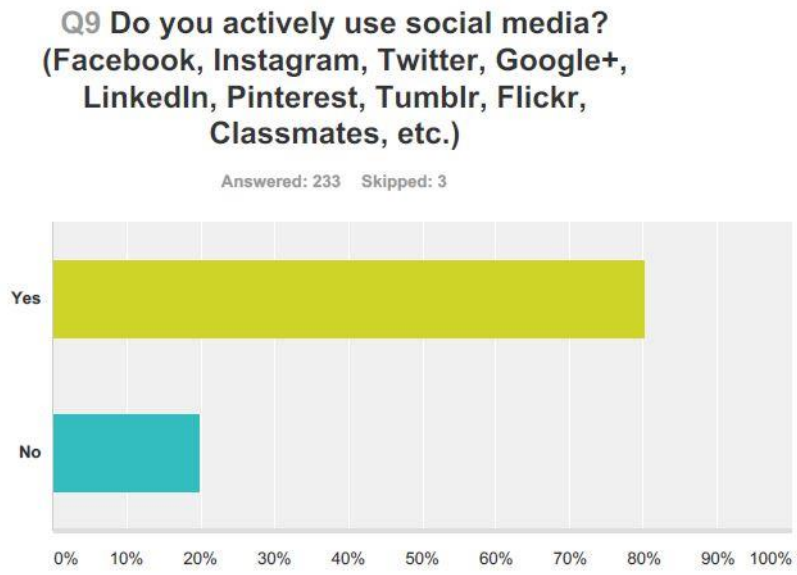

Figure 14. Question 9.

Question 10 allows the respondents to reveal if they actively knew they were putting PII on social media. They should know that doing that is a poor decision. It remains to be seen if the reason they did it anyways was a lack of knowledge because they didn't read the terms and conditions or that they simply chose to ignore the risk. Again, this one question would be a good starting point for another complete investigation. The results of the question show, related to the other risk measuring types of questions, that the average user fully realizes there is a risk. Less than 40 percent admitted that they posted PII to social media. 40 percent is a very high number when we consider that in later questions we establish that data security and privacy are a primary concern.

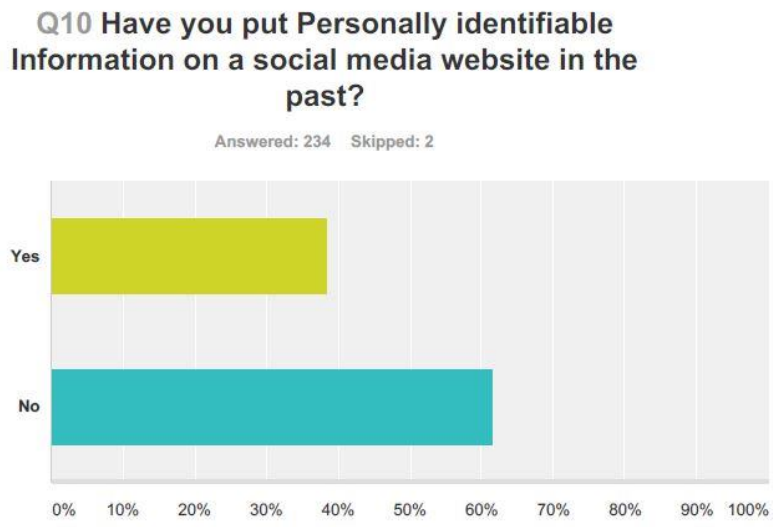

Figure 15. Question 10.

Through the results we have collected so far we can see a trend that the respondents definitely know there is a risk to any PII on the Internet. This does not necessarily mean that they know how to classify their data into more and less important categories. If they can, however, they may be able to operate more intelligently. Question 11 shows that almost half of the respondents can make the distinction between more and less important data and are willing to share that with companies in order to make their browsing experience better. This has potential for CSPs in that they can make offers of more storage space or access to special features to users. In exchange for these benefits, CSP's can potentially have access to almost 50 percent more data to mine for targeted advertising, CSP development, or other big data analytics. The more pessimistic view of question 11 is that more than half of the respondents don't want any company looking at their data in any way.

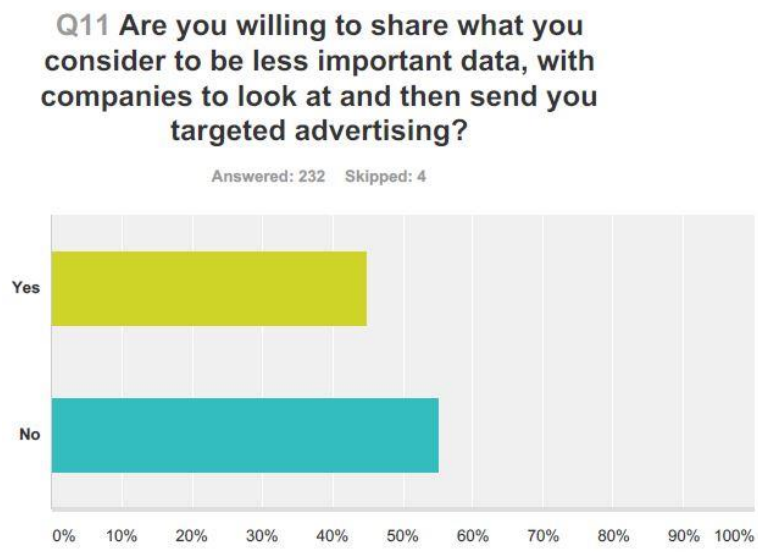

Figure 16. Question 11.

Question 12 follows up by asking if respondents would be inclined to share their data if companies made a substantial offer of free storage space. Many companies already offer free space but only an amount that hooks a user into using their services. Once the services are learned and used, more storage is available at a price. Only about 32 percent of respondents thought that would be a good idea. This combined with the results from question 11 points to the possibility that there isn't much a company can offer in terms of free goods and services to allow more data mining. The important lesson for CSPs in this question is that even the 32 percent of respondents that would not allow access to their data might possibly be persuaded with a generous offer.

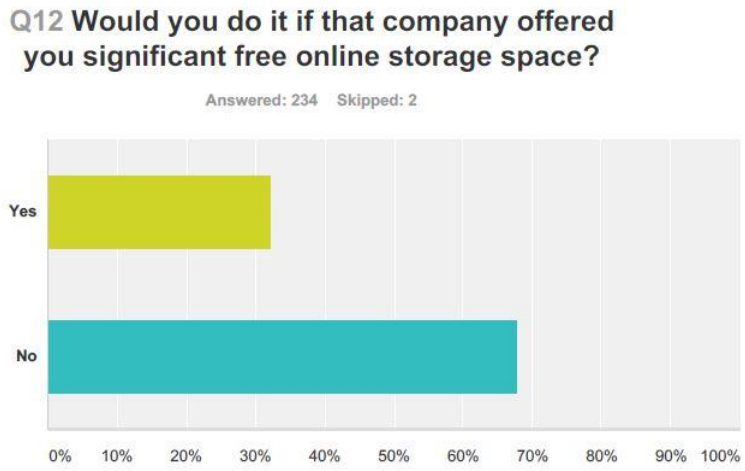

Figure 17. Question 12.

Question 13 was a standalone question related to the research question: Is there a lack of trust with cloud storage providers? Medical records offer another source of criminal mischief. They provide all pertinent PII along with a host of financial information and opportunity to insert false billings that would not be caught due to the volume of information they often contain. The respondents again showed that they know the value of the data in the records in that over 80 percent of them would rather protect their data themselves.

Outside the scope of this paper, in the open ended question section, we asked respondents how they currently back up their data. There were a variety of responses but very few chose the cloud backup method. Using the cloud to back up files is currently the safest, long term storage solution. One study by the RAND Corporation highlights the fact that physical media such as CDs, digital tape, and magnetic disks may have a practical lifetime of tens to hundreds of years. The problem is that the average time until the technology to read them becomes obsolete is only about five years. This begs the question, is the average respondent knowledgeable and diligent enough to maintain their own records by keeping current copies and updating the technology at least every five years? For now, the cloud is the safest place in terms of data integrity in spite of

the overwhelming response to question 13. Respondents had data security in mind more than integrity when they answered the question.

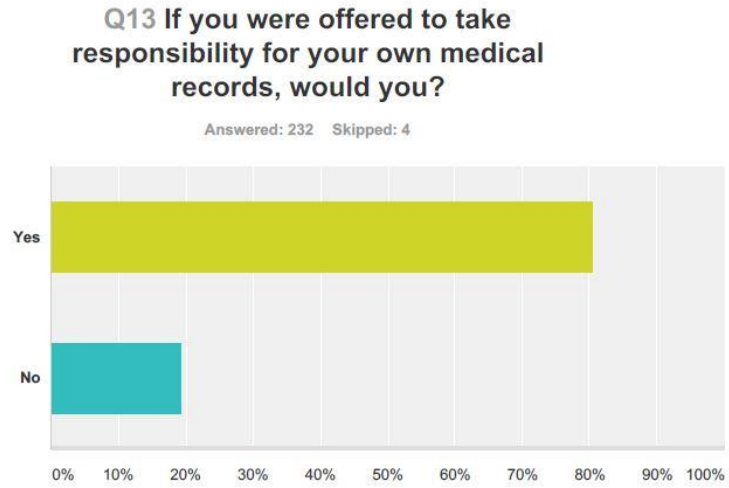

Figure 18. Question 13.

Question 14 directly addresses the concern that respondents had with why they don't have trust in cloud service providers. The strongest concern is not that they fear their data will be destroyed in a cloud service. Their fear is that they have concern that others may gain access to it. This doesn't bode well for a storage model that would allow for mining. More than 93 percent of respondents have concern over who has access to their data online. This indicates a basic level of trust that seems insurmountable.

The response to Question 14 highlights the massive level of mistrust between respondents and CSPs. It also highlights the hypocrisy of the average respondent. Through the questions thus far, we have established that respondents do not take basic precautions to protect their data. In large part, respondents don't read the terms and conditions, they admit to putting PII social media, and they use the Internet for filing taxes and making credit card purchases. They knowingly admit to using PII in risky ways yet they do it anyways.

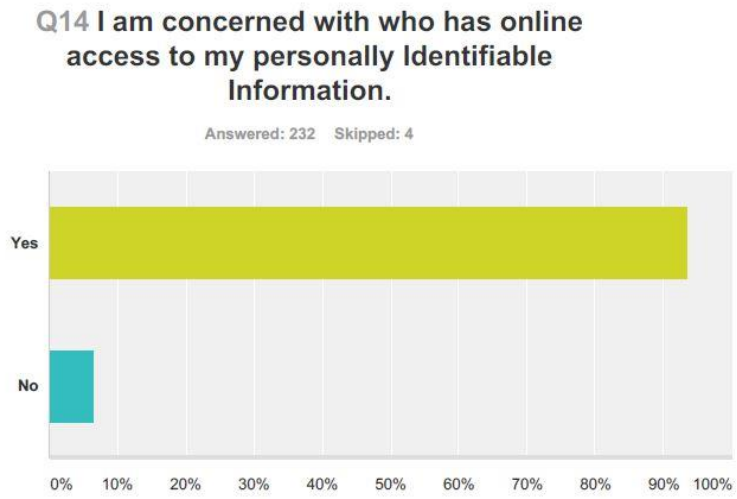

Figure 19. Question 14.

Question 15 is a follow on question to the trust question. By asking if the respondent knows where his or her data is actually stored online, we attempt to discover if, perhaps, the trust issue is related to the knowledge level of the respondent. The more someone knows a stranger, the more likely a person is to trust them. Friends are trusted because we know how they act in a given situation. Online, it is hard to know exactly how a company is going to handle your data. It is even harder to find out how CSPs handle your data. Users who do not read the terms and conditions of online storage users are not likely to trust them or find out how their data is handled. When the terms and conditions are so bulky and hard to understand, it is unlikely that the respondent will ever try. Perhaps if the terms and conditions were easier or simpler to understand then more trust would be given.

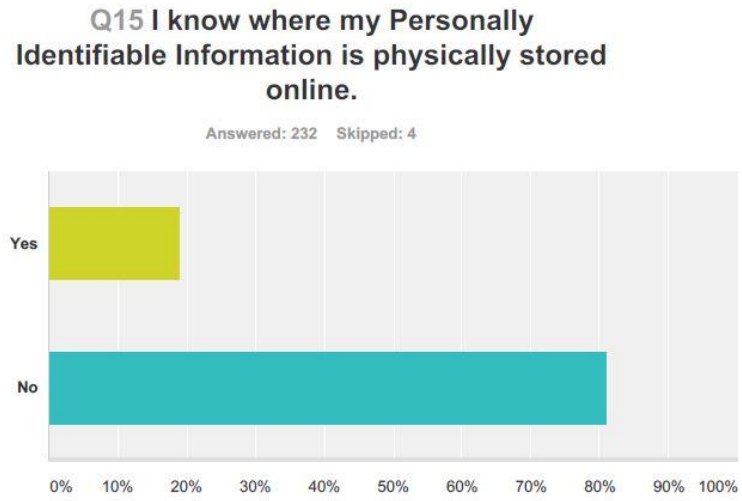

Figure 20. Question 15.

Question 16 takes a different angle on trust. Rather than focusing on the cloud service provider, it asks if users would trust their government to protect their data with laws and regulations governing online activity. Over 74 percent of respondents do not trust that their government has the ability to protect their data. One survey respondent echoed this sentiment well, "Right now, the risk is clearly worth the reward for cyber criminals." Enforcement is the issue, not protection. The Internet is a global community and national laws are nearly impossible to enforce. A felony offense between two parties in different countries has little chance of resolution.

What users want is solid protection before a crime ever happens. This can be seen in the open ended questions when respondents replied to the question, "Why don't you trust cloud service providers with your personally identifiable information." Respondents were most worried about data breaches. This is the basis for the problem CSPs face. In information technology, you are only as good as your last successful email, transaction, upload, etc. Additionally, success is expected so when things go wrong, it is a big deal. The problem is that when you are successful as a company and you are good at what you do, for instance email or

web hosting, it is transparent to the user. It is expected. When things go wrong and data is lost, that is where the difference is made in a prospective user's decision to use that service. How does a CSP prove that is protects user data better than its competitors? One respondent summed it up very well, "It's not about trust. It is about acceptable risk. I accept risk where millions subscribe. (There is) Power in numbers and more likelihood to law enforcement and Federal response." This is a theme we will see again in our analysis.

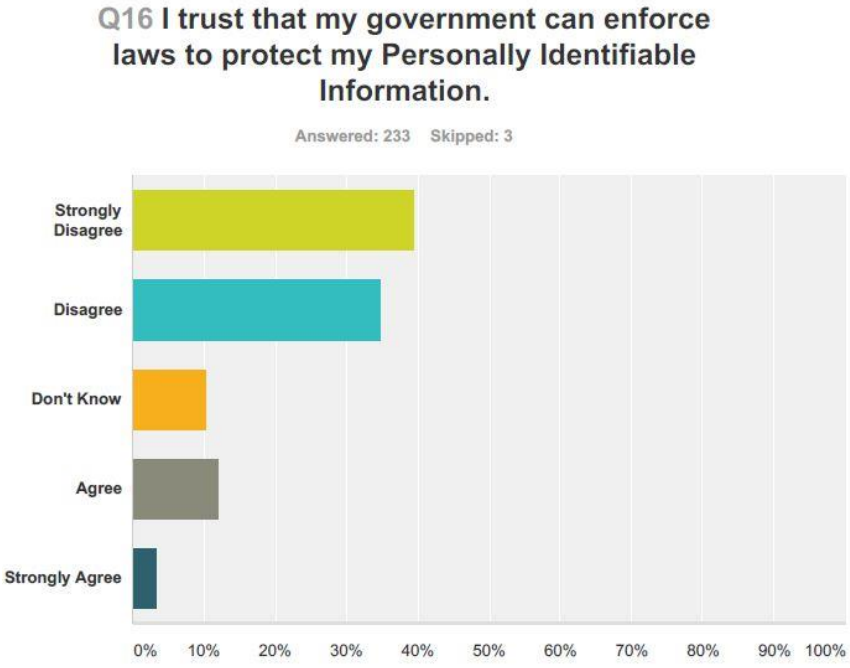

Figure 21. Question 16.

4.4.3 User Opinion Questions.

Through the survey data so far, it is safe to say that education is important to trust. Respondents must understand what is in the terms and conditions of online transactions. It is also important that respondents understand where and how their data is stored. The more they know, the more they trust. For the cloud service provider it is, therefore, important to find out how much time people are willing to spend to learn more. This directly equates to respondent's

willingness to try and trust a service provider. It can be interpreted from the results of the following questions that the respondents clearly know there is a problem but they have not done much to correct it. They haven't read more about PII security and they see that it may be too complicated for them to solve. The good news is that they are willing to do something. They just need to be shown the way.

The results of question 17 deserve a more thorough statistical analysis as it appears there isn't an even distribution of responses. There is a grouping of responses in question 17 that say that 30 minutes to an hour of time would be sufficient to learn about protecting their data. Another large grouping shows that more than 4 hours is a reasonable amount of time. What is important about this question is that respondents are clearly interested in learning. Some are willing to spend a lot of time doing so. It is in the cloud service provider's best interest to provide a learning environment to take advantage of that. They should make it easy for their users to learn their products and become deeply convinced their data is safe.

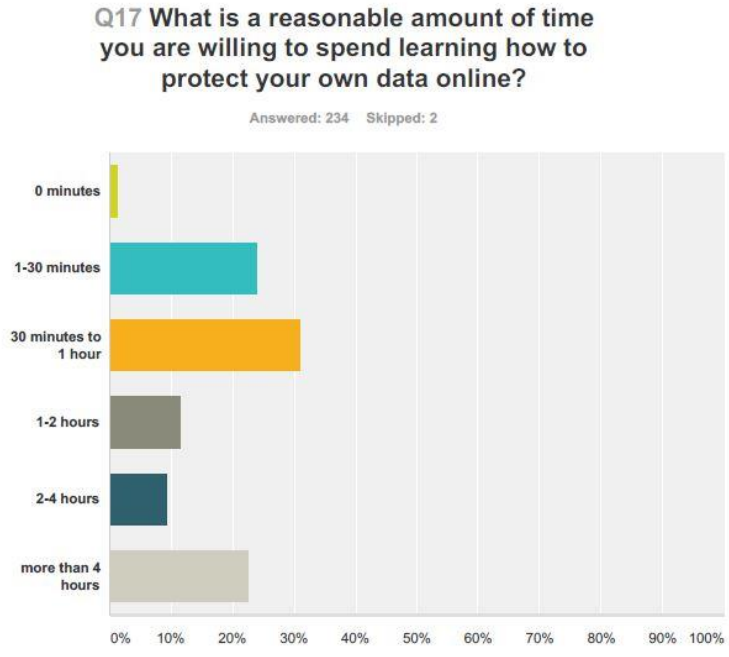

Figure 22. Question 17.

Question 18 is the converse of 17. Nearly half of the respondents have little or no knowledge of how their data is protected online. Of the respondents that have some knowledge, very few are confident they know exactly how their data is protected. Based on the open ended questions at the end of the survey, some of those respondents have end to end knowledge of their data protection procedures because they protect it entirely on their own and do not use cloud services. This is likely a little skewed from a normal distribution in that the survey audience, in this case, is likely more technically competent. With consideration to that observation, this question shows that most people simply don't know how their data is protected. Again, this should be a motivator for cloud service providers to make protection information a priority and simple to understand.

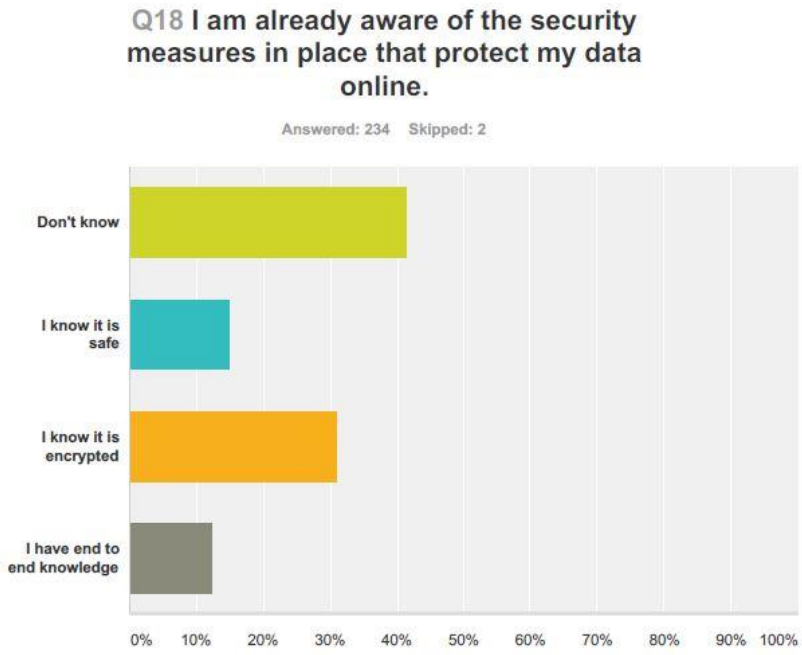

Figure 23. Question 18.

Question 19 has a similar distribution as question 17. Both distributions may be accounted for by the population mentioned in the introduction. Some of the respondents were technology savvy and this may account for the population that has spent more time reading about

PII safety. Future studies may be more accurate with a more diverse population. We can, however, learn from this question that respondents may find that up to an hour of reading (from question 18) may not be sufficient to give them the trust they should have to use cloud storage. It is important to note that more than 30 percent of respondents spent only 1 to 30 minutes reading about how to safely store their PII online. That is a very short period of time considering what is at stake if their PII is compromised. If a person becomes a victim of identity theft, it can take a lifetime to get corrected.

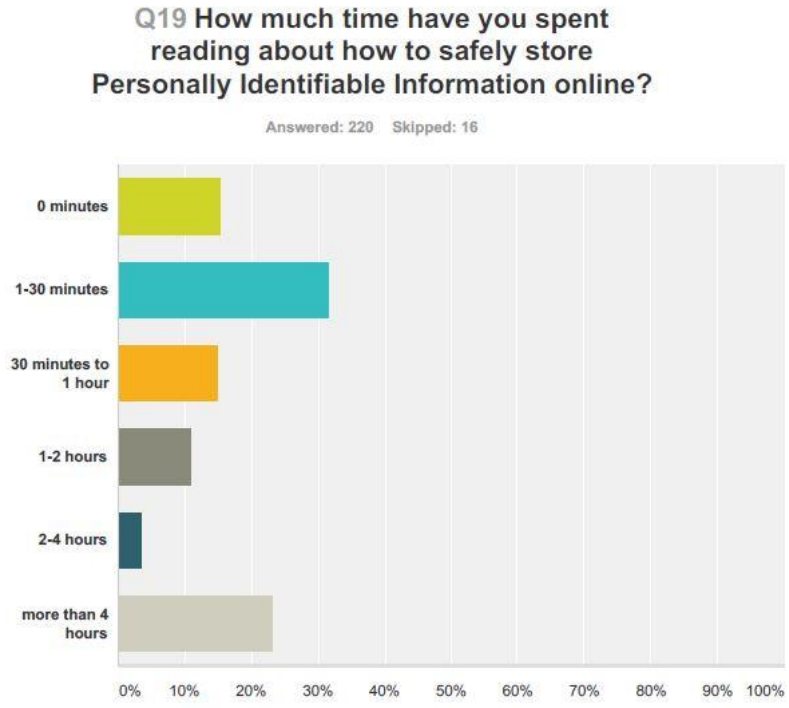

Figure 24. Question 19.

Question 20 is a departure from previous questions. It is designed to determine if the traditional folder structure is a pathway to helping people make smarter security decisions about their data. In the attempt to educate respondents, CSPs should take into account what they already know and what they already use. A folder structure to store files is intuitive and had been integrated into the earliest operating systems. If a CSP took action to educate users and

created a tool to help them classify their data in order of importance, this would be a wise starting point. Question 20 shows that over 90 percent of respondents already use a folder structure. CSPs could offer differing levels of protection based on different folders that people put their documents in. Respondents already classify their friends on social media based on how much access they desire them to have. This is a similar concept using files and folders.

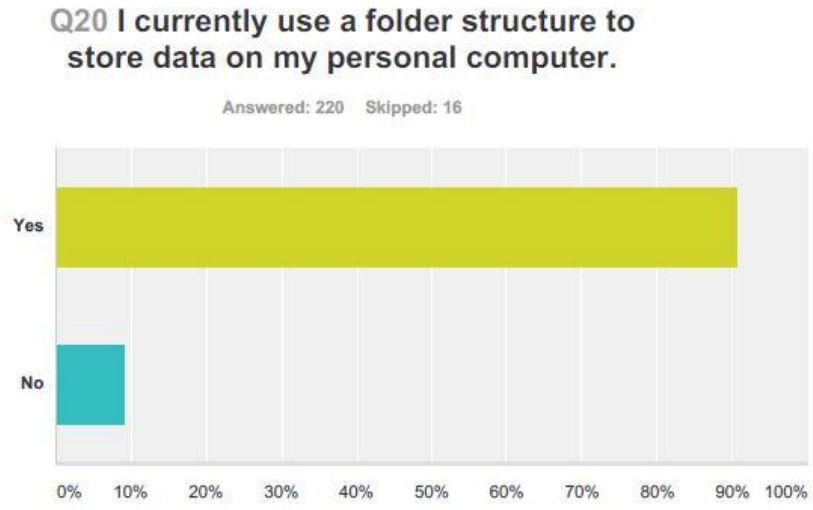

Figure 25. Question 20.

Question 21 is related to question 14 in that it seeks to understand the lack of trust respondents have. This question attempts to quantify that lack of trust. By doing so, this is one of the more important questions for CSPs to pay attention to. Over 73 percent of respondents agree or strongly agree that security is the most important issue with using cloud storage. This is, again, a strong indicator that cloud service providers should find innovative ways to gain the trust of users. Other questions in the survey point to education being a key pathway to that trust.

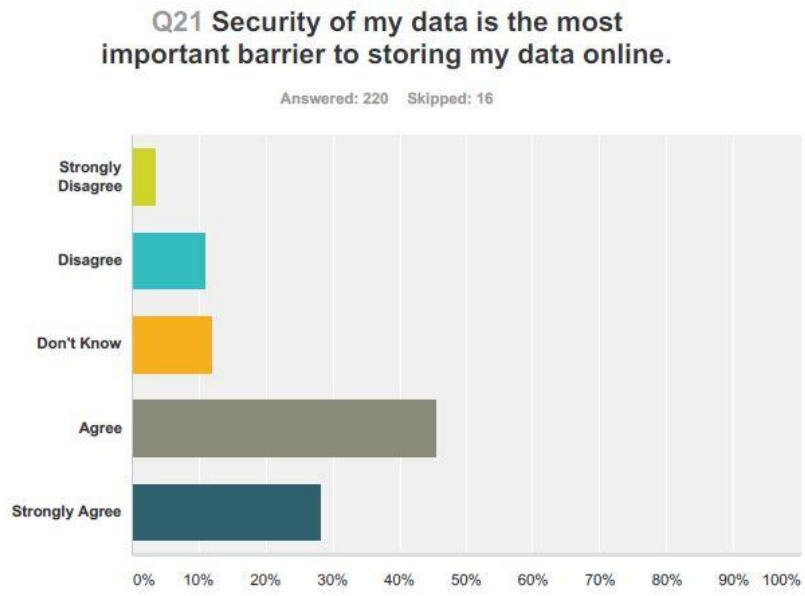

Figure 26. Question 21.

Question 22 essentially asks that if money were a factor, would it make a difference. The responses were evenly divided. Again, the technology savvy group may have had an effect on this answer. Technology workers understand that when data is given to them, it is their duty to protect it. Non-technology workers may not have this understanding. This question was an attempt to see if it would be profitable for cloud service providers to offer pay plans for increased protection. Based on the preliminary results it seems this would not be true. The problem is that the question may not have been clear or could be taken differently by different people. It may have been clearer if the question asked if respondents were willing to pay extra to keep their information secure. Throughout the survey it can be inferred that the respondents generally feel that it is the responsibility of the CSP to secure their data because they don't know where it is stored or how (question 15) yet they already store at least some PII online.

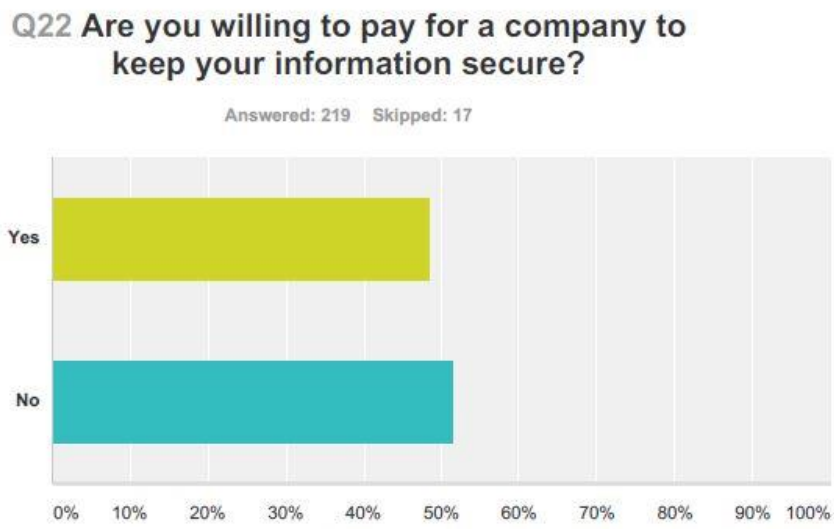

Figure 27. Question 22.

Near the end of the survey we have determined a few trends. First, there is definitely a problem with online trust. Respondents do not trust any CSP completely. Second, money cannot buy that trust. Finally, education seems to be the key to online trust. It would be helpful to companies to know that the average user recognized all of this and is willing to do something about it. Question 23 answers that soundly. More than 88 percent of respondents are willing to try something new. They are not happy with the way they currently operate online. They need someone to teach them.

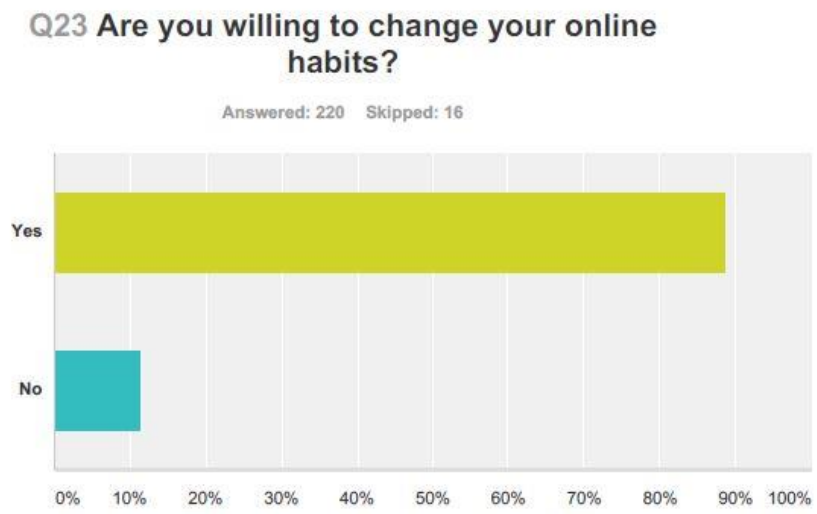

Figure 28. Question 23.

Given the answers to the rest of the survey, it can be safely assumed that respondents are not entirely happy with the current state of CSP services. About 31 percent of respondents do not use cloud services. The primary reason according to question 21 is that they do not trust the service. It is encouraging to note that 70 percent of respondents are active users of CSP services. This confirms the earlier assumption from question 7 that if a respondent used cloud based email then they are likely to use other cloud services from the email provider. In spite of the very valid reasons they have given throughout the survey, they still use the service. This can be attributed to the fact that respondents value the convenience and utility of cloud services more than the risk they take by using them.

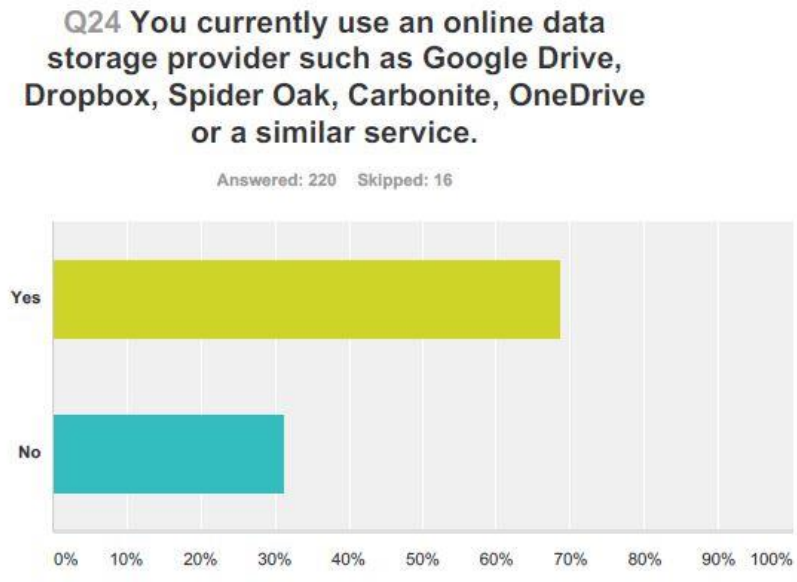

Figure 29. Question 24

4.5 Statistical Analysis

For analysis we used the Levene's test for equality of variances and the Pearson Coefficient. The Levene's test is an inferential statistic used to assess the equality of variances for a variable calculated for two or more groups. In this case it is male and female results. We hypothesized that men and women handle trust issues differently. The Pearson Coefficient is a correlation that represents the relationship between two variables that are measured on the same interval or ratio scale. In this case it is used to assess the relationship between questions that used the Likert scale. The Likert scale is used to grade questions based on levels of agreement. This helps to quantify personal attitudes on various subjects.

Table 2 shows the correlation of question 21 to question 23. We did this correlation to see if respondents who are concerned primarily with security are willing to learn more about it. The Pearson correlation coefficient is 0.145. This indicates that there is a positive correlation between question 21 and question 23. We can draw the conclusion that indeed, there is a

relationship between the user's awareness of the security risks to their data and their willingness to do something about it.

Table 2. *Correlation of Questions 21 and 23.*

		Security of my data is the most important barrier to storing my data online	Are you willing to change your online habits?
Security of my data is the most important barrier to storing my data online.	Pearson Correlation	1	.145*
	Sig. (2-tailed)		.032
	N	220	219
Are you willing to change your online habits?	Pearson Correlation	.145*	1
	Sig. (2-tailed)	.032	
	N	219	220

*. Correlation is significant at the 0.05 level (2-tailed).

In table 3 we correlate questions 21 and 19 to find that if security is important to the respondent, they spend more time reading about it. Here we see that the correlation is .133. This is statistically significant at the .05 level. Respondents are willing, able and are actually reading more about security. Why is it still their top concern? They have taken the necessary steps to protect themselves yet still do not trust CSPs. The one word answer is trust.

Table 3. *Correlation of Questions 21 and 19.*

		Security of my data is the most important barrier to storing my data online.	How much time have you spent reading about how to safely store Personally Identifiable Information online?
Security of my data is the most important barrier to storing my data online.	Pearson Correlation	1	.133*
	Sig. (2-tailed)		.049
	N	220	220
How much time have you spent reading about how to safely store Personally Identifiable Information online?	Pearson Correlation	.133*	1
	Sig. (2-tailed)	.049	
	N	220	220

*. Correlation is significant at the 0.05 level (2-tailed).

Another interesting fact is that women and men have differing needs when it comes to establishing trust. Table 4 shows that women do not use the Internet for storing data as much as men. This suggests that there is a significant difference between male and female on how often they use internet for strong data. The two tailed value is .005. This means that men store data on the Internet more than women do.

Table 4. *Independent Samples test Men vs Women use of CSP storage.*

Group Statistics

	Are you Male or female?	N	Mean	Std. Deviation	Std. Error Mean
How often do you use the internet for storing data of any type?	0	74	1.85	1.235	.144
	1	157	1.39	.958	.076

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
How often do you use the internet for storing data of any type?	Equal variances assumed	12.960	.000	3.112	229	.002	.463	.149	.170	.756
	Equal variances not assumed			2.845	115.946	.005	.463	.163	.141	.785

CHAPTER 5

Conclusion and Future Research

5.1 Lessons Learned

The investigation into trust in cloud computing has shown several interesting trends.

- Users do not trust current systems
- The government cannot fix trust issues
- Education is key to trust
- Trust cannot be purchased

Respondents clearly do not trust current systems in place but begrudgingly use them anyways. A little more than 50 percent of respondents say they consciously put PII on the Internet but further questioning shows the reality. 90 percent use cloud based email and 80 percent use social media. Within those two applications there is a high likelihood that some PII will be introduced to the Internet through email correspondence or social postings. Over 74 percent of respondents do not trust that the US Government can protect their data. The reasons are many but the fact that respondents already trust industry with email and social media is a positive endorsement. Industry must find better ways to earn the trust of their users.

Education of the user is very important to establishing trust. Education for the purpose of this survey is defined as level of knowledge on the subject of cloud computing security. We are not referring to what educational degrees a respondent may or may not have, rather, we talk about the education of respondents in terms of safely using PII online. Respondents are clearly willing to spend some time learning how to safely protect their data online, but that time is limited. It is also clear that people are unwilling to read and understand the legal language that

most terms and conditions use. People want a concise explanation of what is going to happen to their data. They want to know, in simple terms, how the CSP is going to secure their data. It is up to the CSP to be the teacher to earn that trust. A clear, concise way of explaining the security of user's PII would be a great start. That explanation should include easy ways to get more detail if the user wants it. Currently, of the top CSPs, it is difficult to find privacy policies and navigate through them.

Trust cannot be bought. It is the product of a relationship over time. This is evident in a recent paper that shows in 2013 the cloud service industry still struggles with consumer trust. Amazon launched its EC2 architecture in 2006. Microsoft followed suit with Azure in 2010. These two large service providers have had 12 years of combined experience to earn trust yet the primary complaint of consumers is still trust. (Bhosle & Kasurkar, 2013)

5.2 Conclusion

In this study we conducted a survey to measure consumer trust in cloud computing. What was found was that consumers trust cloud computing more than they admit to even themselves. They trust only to the extent that the risk is perceived to be low and the convenience payoff for them is big. There still is a problem with consumer trust and it is beneficial for consumers and industry to come to an agreement where the Internet becomes more useful to the consumer and the consumer becomes more profitable for industry.

More meaningful results will be obtained through the use of regression modeling. This modeling will more precisely determine the relationships between the questions and concepts we have presented. With stronger evidence to the relative importance of these concepts, it is easier to determine the most effective path towards earning trust. The ultimate goal is to turn this data into action.

5.3 Future Work

This work can be expanded by executing another, more detailed survey. For example, the results of this survey may have been skewed by the education level of the respondents in question 22. The question involves more factors than are readily apparent. Are users willing to pay for CSPs to secure their data? According to our survey, half of the users are. It is important to CSPs to understand why. More detailed questions are the key to finding the reasons people act the way they do when it comes to personal data security. Another factor in expanding this work is screening a larger audience for the survey. Our survey covered mostly Americans; however, the problems presented in this work are not unique to the United States. A more thorough, global survey would incorporate two improvements. First, the questions should be asked with more statistical analysis in mind. For example, questions will gather more demographic information. Second, we would make efforts to maximize targeting users who are not Internet consumers to try and understand their fears. Finally, the open ended questions would be better focused and used as more than highlights for the quantifiable responses. Ultimately, the information gathered would produce a model or best practices for Internet businesses to use for improving sales. This model would ideally also improve security for consumers.

References

Abrams, R. (2014). Target Puts Data Breach Costs at $148 Million, and Forecasts Profit Drop. Retrieved March 3, 2014, from http://www.nytimes.com/2014/08/06/business/target-puts-data-breach-costs-at-148-million.html?_r=0

Bhosle, P. I., & Kasurkar, S. A. (2013). Trust in Cloud Computing. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 2*(4), pp: 1541-1548.

Carlin, S., & Curran, K. (2011). Cloud computing security. *International Journal of Ambient Computing and Intelligence (IJACI), 3*(1), 14-19.

Citrix. (2012). Most Americans Confused by Cloud Computing According to National Survey. Retrieved February 19, 2015, from http://www.citrix.com/news/announcements/oct-2012/cloud-confusion-survey.html

Firdhous, M., Ghazali, O., & Hassan, S. (2012). Trust management in cloud computing: a critical review. *arXiv preprint arXiv:1211.3979*.

Gonzalez, N., Miers, C., Redigolo, F., Carvalho, T., Simplicio, M., de Sousa, G. T., & Pourzandi, M. (2011, Nov. 29 2011-Dec. 1 2011). *A Quantitative Analysis of Current Security Concerns and Solutions for Cloud Computing.* Paper presented at the Cloud Computing Technology and Science (CloudCom), 2011 IEEE Third International Conference on.

Google. (2015). Privacy Policy. Retrieved March 1, 2015, from http://www.google.com/policies/privacy

Khan, K. M., & Malluhi, Q. (2010). Establishing Trust in Cloud Computing. *IT Professional, 12*(5), 20-27. doi: 10.1109/MITP.2010.128

Khan, K. M., & Malluhi, Q. (2013). Trust in Cloud Services: Providing More Controls to Clients. *Computer, 46*(7), 94-96. doi: 10.1109/MC.2013.254

Ko, R. K. L., Jagadpramana, P., Mowbray, M., Pearson, S., Kirchberg, M., Qianhui, L., & Bu Sung, L. (2011, 4-9 July 2011). *TrustCloud: A Framework for Accountability and Trust in Cloud Computing.* Paper presented at the Services (SERVICES), 2011 IEEE World Congress on.

Lynley, M. (2014). Cloud Storage Wars. Retrieved February 19, 2015, from http://www.buzzfeed.com/mattlynley/cloud-storage-wars#.ce5ZodvXDv

Microsoft. (2014). Privacy Statement. Retrieved March 1, 2015, from http://www.microsoft.com/privacystatement/en-us/core/default.aspx

Palermo, E. (2015). 10 Worst Data Breaches of All Time. Retrieved March 3, 2015, from http://www.tomsguide.com/us/biggest-data-breaches,news-19083.html

Ranchal, R., Bhargava, B., Othmane, L. B., Lilien, L., Kim, A., Kang, M., & Linderman, M. (2010). *Protection of identity information in cloud computing without trusted third party.* Paper presented at the Reliable Distributed Systems, 2010 29th IEEE Symposium on.

Rimal, B. P., Choi, E., & Lumb, I. (2009). *A taxonomy and survey of cloud computing systems.* Paper presented at the INC, IMS and IDC, 2009. NCM'09. Fifth International Joint Conference on.

Ryan, M. D. (2011). Cloud computing privacy concerns on our doorstep. *Communications of the ACM, 54*(1).

Sato, M. (2010). Personal data in the cloud: A global survey of consumer attitudes. Minato-ku, Tokyo 105-7123, JAPAN.

Strub, P. J., & Priest, T. B. (1976). Two Patterns of Establishing Trust: The Marijuana User. *Sociological Focus, 9*(4), 399-411.

Xue, J., & Zhang, J.-j. (2010, 10-12 Aug. 2010). *A Brief Survey on the Security Model of Cloud Computing.* Paper presented at the Distributed Computing and Applications to Business Engineering and Science (DCABES), 2010 Ninth International Symposium on.

Yahoo. (2015). Products. Retrieved March 1, 2015, from http://info.yahoo.com/privacy/us/yahoo/products.html

Zissis, D., & Lekkas, D. (2012). Addressing cloud computing security issues. *Future Generation Computer Systems, 28*(3), 583-592. doi: http://dx.doi.org/10.1016/j.future.2010.12.006

Appendix: Consumer Survey Questions

1. What is your age?

2. Are you male or female?

3. How often do you use the internet for storing data of any type?

4. Do you currently use credit cards to make purchases on the internet?

5. Have you done your US Tax forms on the internet?

6. Do you currently store Personally Identifiable information in some type of cloud storage?

7. Do you currently use cloud based email (Gmail, Hotmail, etc?)

8. Do you always read the terms and conditions of online purchases?

9. Do you actively use social media? (Facebook, Instagram, Twitter, Google+, LinkedIn, Pinterest, Tumblr, Flickr, Classmates, etc.)

10. Have you put Personally identifiable Information on a social media website in the past?

11. Are you willing to share what you consider to be less important data, with companies to look at and then send you targeted advertising?

12. Would you do it if that company offered you significant free online storage space?

13. If you were offered to take responsibility for your own medical records, would you?

14. I am concerned with who has online access to my personally Identifiable Information.

15. I know where my Personally Identifiable Information is physically stored online.

16. I trust that my government can enforce laws to protect my Personally Identifiable Information.

17. What is a reasonable amount of time you are willing to spend learning how to protect your own data online?

18. I am already aware of the security measures in place that protect my data online.

19. How much time have you spent reading about how to safely store Personally Identifiable Information online?

20. I currently use a folder structure to store data on my personal computer.

21. Security of my data is the most important barrier to storing my data online.

22. Are you willing to pay for a company to keep your information secure?

23. Are you willing to change your online habits?

24. You currently use an online data storage provider such as Google Drive, Dropbox, Spider Oak, Carbonite, OneDrive or a similar service.

25. How do you currently back up your personal files? (Photos, documents with personal information) - Open-Ended Response

26. If a company showed you how they protected their data would you be more inclined to use their services? Why? - Open-Ended Response

27. Why don't you trust cloud service providers with your Personally Identifiable Information? (Google Docs, Dropbox, OneDrive) - Open-Ended Response

28. If you had a Google or Dropbox CEO in your office, what would you tell him to do to earn your trust? Be as specific as you can.

29. Please provide any additional information or opinions about this subject or the survey that you think are relevant.

30. If you would like the results of this survey, please enter your email address.