Personal data in cloud computing: organizational information risk perception

Samir Husić

Ph.D. Candidate International University of Sarajevo

Abstract

There are different kind of private data collected by the Cloud service providers. Data privacy protection became increasingly important in the emerging Cloud information technology environment, where powerful data bases containing millions of people personal data of different kind became easy accessible online. The basic characteristic of the Cloud computing services is vulnerability for data disclosure. Researches show that tracking of the consumers is increasing in far more detailed, and in far more invasive way, than most customers realize. This problem spreads progressively, and so the public distrust in online database controllers, while the managers are faced with new challenges. Research study presented in this paper is independently conducted as part of PhD research in October and November 2013, and the results are presented in this paper. The goal of the research was to recognize what kind of organizational information is perceived to be under risk. Further on, we tried to identify privacy concerns about cloud computing and what cloud providers need to focus on. This paper presents an analysis of the key findings of our research, and will discuss considerations for business decision makers on privacy practices in the Cloud.

Key words:

Data privacy, Cloud computing, Risk perception

115

Sarajevo Journal of Social Sciences

Inquiry

Introduction

According to many researchers of online data privacy, one of the fastest growing internet businesses is spying on the internet users. It is not only limited to cookies, which are known to be recording users behavior and collecting personal information, but tracking technologies now increasingly use new tools for real-time scanning for instant access to personal information, like shopping interests. Some of these tracking tools are made to be persistent and continue to track and report, even after user try to stop it.

Such serious data privacy risk became increasingly concern for internet users. Our research was focused on this risk perception and identification of key factors which can help us to define data privacy concerns, causes of reduce of trust in Software as a service (“SaaS”), and to analyze potential solutions.

We anticipated our research and analysis from two standpoints. Firstly, we have to identify data privacy concerns for specific type of information in cloud computing, from customer’s perspective. In second part we have to assess how these privacy anxieties affect cloud computing business environment, with emphasis on SaaS model, from cloud service provider’s perspective.

Our analysis is based on survey, as the major method of research both in customer behavior observation and business strategies development. Target group have been selected from business developed urban area of Bosnia and Herzegovina, and they come from both public and private companies. Target group includes professionals, both users and providers of IT services, and where available, more specifically SaaS service providers. Survey assumes a method of gathering information from a sample via personal communication by phone or by e-mail.

Survey questions outlined some of the major issues that have emerged recently relating to data privacy in cloud computing. Survey evaluates communication privacy policy of the organizations management. Important questions include organization commitment to privacy policies, customers need for privacy choices for specific types of information, vigilance of cloud service users, users attention paid to the provider’s terms of service, and analyze of the difference between customer’s perceptions of cloud privacy threats. The primary purpose of statistical interpretation applied in this research is to gain insight into existing privacy concerns, and to assist

predictions of future changes. The ultimate purpose of this work is to offer concept or proposal to decision-making process for cloud computing companies.

Methodology

We focused on privacy concerns in this research, through gathering information for analysis and discussion about what business decision-makers in BiH small and medium organizations think about privacy when considering adoption of cloud computing services. We tried to produce data comparable and amenable to analysis. Most questions in our research survey are comparable to questions in relevant researches conducted recently. However, the problem with comparable research questions in cross-national surveys is that it cannot be taken for granted that results of survey in one social group can be compared with other groups, especially on international level. Those results may have partial or completely different meaning.

In order to avoid mistakes in our research results, the absence of bias needs to be demonstrated, not just to be assumed. Bias and functional equivalence between the target populations are regular terms used to consider this problem (Van de Vijver, 1997). Bias is expressed as difference between cultures or groups where one group has certain advantage of knowledge, for example geographical location of group when answering questions for which geographic knowledge is related to their place of residence. So the bias occurs when the differences in survey results on the question indicators do not correspond to trait or ability. Equivalence, in this context, is opposite term and means lack of bias.

In attempt to avoid different types of bias (Van de Vijver, 1997) in inter-cultural survey questions comparison, we paid particular attention to the following variables. Geographical location of the target group did not affect the differences in compared results, as we distributed survey electronically irrespectively of location within country. “Method bias” may occur if samples are incomparable; thus, target groups must not differ in educational background. Our survey shows that only 3.4% of respondents have high school education, while 96.6% has university degree or some college. From the aspect of “instrument bias”, Cloud computing technology used and available to target groups was comparable. Target groups did not meet communication problems, since language used in ques-

tionnaire was English and wasn't become barrier or created difficulty for interviewee. Finally, the "item bias" shouldn't occur for different psychological meaning between different target groups, having in mind that we determined indicators in questions that do not differ between countries in comparison.

Considering that this research target sample is BiH professional population knowledgeable about Cloud computing, a special attention was paid to avoid translation bias. Translation of technical terms and psychological perceptions expressed in English may lead to mistaken answers if it is inappropriately just rewritten from Bosnian language. In any case, translation avoided hardly translated local idioms, and adaptation to local terminology did not affect measurement instruments.

Most relevant surveys in the field of measuring of trust in privacy and data security in recent years is conducted by Ponemon Institute (Ponemon, 2013). Ponemon Institute is considered as most excellent and most widely quoted independent research center in field of privacy and data protection. Its researches had influenced decision on trends of top corporations and government leaders about practices and potential threats regarding gathering, transfer and storage of sensitive and confidential data. Accordingly, surveys conducted by Ponemon Institute are used for comparison with results of our survey as most relevant for this study.

Importantly, we did not encourage privacy violations to prove that they are the cause for customer reluctance to use Cloud services. The subject of each question of our surveys within the capability of survey participants, requiring minimal technical knowledge of confidential information about their company. Revealing of sensitive and private information is avoided to the extent possible.

It was expected to find large divergence between shared aims of these stakeholders in cloud computing business, so attention will be on attempt to avoid imposing of predisposed concepts that might put at risk impartiality of research. This methodology included appropriate approach in asking questions and stating the results, while respecting ethical principles, privacy, company's liability, and cloud technology prosperity and effectiveness.

Finally, we applied confidence interval methodology for measuring

reliability of our data. Our desired confidence interval of 95% is set considering statistics typical standard for this kind of research.

Results

Identification of the organizational information risks took our analysis to first question: Which Web-based software services (SaaS, Cloud) does your organization use? Study results found that average usage of most popular Cloud computing services in Bosnia reaches 53.6%. Within these services, webmail is dominating with 83.5%. Around 68.4% of customers used communication services, while least commonly used services are business solutions like CRM with 38% share and security and identity management with 20.3%.

Most respondents are moderate or heavy users of at least some Cloud resources, where 91% of them use Cloud services professionally and even 96% who use them privately, and they expect their usage to increase in following years, despite existing fears of data privacy management in the Cloud. The furthermore analysis found that Cloud service purchasing decisions are highly dependent on privacy protection practices. About 26 % of respondents terminated relationship with Cloud service provider due to data privacy concerns.

Out of about 1000 relevant contacted target group potential respondents, we received 237 responses. Following chart indicates normal distribution of respondents by age, presenting real-valued random variables whose distribution was not known. This chart confirms that our respondents have been independently drawn and distributed approximately normally. The second chart on figure 2 shows numerical proportion of survey participants by job area, where dominating IT sector and management indicates decision makers profile in Cloud computing adoption processes.

Figure 1: Distribution of respondents by age

Figure 2: distribution of respondents by job area

Key results of our research are presenting perception of participants concern about privacy protection in the Cloud. Out of ten critical types of organizational information, financial data and accounting data are on the top, with 49% and 48% respectively considered extremely critical. On the other side, supplier records and meeting minutes are considered least critical with 14% and 6% labeled as extremely critical. From the aspect of the access to data, ultimate anxiety is who has the access to data, with 61% of extreme concern, followed by reliability of passwords and identification with 49%, sharing data to third party 48%, and question of who owns the data 41%. These key results will be discussed further on.

Discussion:

Statement of hypothesis

Customers in Bosnia and Herzegovina are reluctant to use Cloud services, in particular Software as a service, because they are concerned about data privacy. Type of information customers are concerned for are related to industry and job area. On the other hand, whether the information under risk is private or professional, is not much influencing customer's level of concern. Customers are making a difference between data privacy and data security, and they are clearly more worried about data privacy. Cloud service users and decision makers who are adopting Cloud computing are careful in moving

their data to the Cloud, even when they consider Cloud as reliable. Finally, there might be correlations between users' behavior and decisions they make while they care for data privacy, and those correlations might be used for prediction of customers' behavior.

Concerns regarding types of critical data

What kind of information will be collected depends, first of all, from the Cloud services utilized and the type of Cloud deployment model. Apparently, public based SaaS will have a lot more involvement in collecting private data than Platform as a service.

The highest interest of SaaS companies of users is behavioral advertising. Through monitoring software they usually utilize customize advertisements based on users browsing preferences and demographic data.

Cloud computing is expanded in many developing countries, since it is very convenient, with on-demand network access to a shared pool of configurable computing resources, with minimal investment. For instance, Cloud computing in India is particularly in development, where estimations predict it's grow from a \$50 million industry in 2009 to a \$15 billion industry by 2013. (Greengard, 2010a).

Such wave of cloud computing expansion is expected in Bosnia and Herzegovina as well, since Bosnia is on the list of typical emerging and developing economies according to the International Monetary Fund's World Economic Outlook Report, April 2010. Getting intense increase of SaaS application in Bosnia rises preceding question of needs and reasons for use of SaaS. Thus, our survey started with question for which reason do you use Software as a service, taking SaaS as a basic model of Cloud computing. Respondents could choose more than one answer. The answers are presented on Figure 3 below.

Figure 3: For which reasons (may be many) do you use SaaS?

Most respondents in our survey (86.1%) consider SaaS solutions practical. Remote access is next leading incentive for adopting SaaS with 64.6% of responses. However, only 19% consider SaaS reliable, making reliability the last reason for adoption of SaaS, which supports our hypothesis of lack of trust in Cloud based services. Remarkable answer is that considerably low number of responses finds SaaS as cheaper solution. We can compare our results with results of Information Week Analytics 2010 SaaS survey (Wang, 2010), which analyzed incentives to move to the SaaS. Information Week Analytics results presents savings on capital and operational expenses as second most important incentive for adopting SaaS, with 3.5 out of 4 score. Thus cost effectiveness is the only indicator not corresponding to expected results, and our research directed us to possible explanations. One explanation might be high level of software piracy, which makes on the premise solutions incomparably cheaper. Illegal use of software is likely to disrupt the cost balance between Cloud based and local solutions.

Having reasons why customers in Bosnia are adopting Cloud services, next question is what kind of typical Software as a service they use. Multiple answers were available. Dominating service is webmail and web hosting taking 83.5% respondent's choices. Communication types of services derived from decade preceding Cloud computing transformation is obviously still dominating in Bosnia and Herzegovina, which is typical for transitional country. Second most common use of communication software services with 68.4% of answers, such as Skype, is confirming previous conclusion. Following group are banking and financial services, and social relation (and again communication) services with 60.8% and 57% respectively. Storage services, recruitment web services, and online office software are in group just above 50% of uses. Contemporary business solutions, expectedly, are close to the bottom of the list. Finally, security and identity management services are last with 20.3% of use by participants.

Figure 4: type of Cloud software services used

Reliability of the online software services for the Bosnian professionals is generally not an issues. About 70% of users claim that SaaS used in their organization is very of extremely reliable. While 27% consider it moderately reliable, only 3% is getting concerned.

Figure 5: How reliable are online software services (SaaS) for your organization?

When asking participants about types of critical information they are most worried about, we received different responses depending on the nature of information. Out of ten critical types of organizational information, financial data and accounting data are on the top, with 49% and 48% respectively. We can associate this result with moderate use of banking and financial Cloud based services above mentioned. Employee records, customer records and intellectual property are expectedly shaping second most important data privacy critical group. Strategic plans are sixth on the list and still considerably critical with 51% of responses stating it extremely of very critical. Following are, surprisingly, health records, which are close to the bottom of data importance list. Finally, supplier records and meeting minutes are considered least critical with only 14% and 6% labeled as extremely critical.

Figure 6: How risky and critical is the following organizational information?

Concerns regarding data access

Worries about privilege access to the private information stored in computer networks has been recognized in the United States in early reports used as a basis for the Electronic Communication Privacy Act enacted in 1986. Even long time before the Patriot Act, US legislation established system which allows government to easily obtain privilege access to citizens electronically stored private data. Law enforcement agencies, particularly US authorities as Cloud service providers are dominantly US-based, perceive trans-national law-conflicting environment in cloud computing. They justified concern that Cloud services are possibly used for criminal activities, and they certainly want to obtain fully controlled access to Cloud based activities, stored data and logs, for the investigation and forensic purposes.

Ponemon institute research in customers' consideration of most significant privacy-related threats (Ponemon, 2012) shows that government surveillance is second most perceived privacy threat among Cloud service customers with 56% of responses. The first one, out of ten threats, is identity theft with 61% of responses. Spam is below average with 32%, and employer (management) intrusions with 30% of responses.

We discovered slightly different ratio of responses in our research, as presented in chart below. Concern for government surveillance is on fifth place, while spam and management interference is at the bottom of customer's concern list. Identity and password theft is second on top, similar to Ponemon results. Authorizations for data access and data sharing to third parties are among top concerns. From the aspect of the access to data, crucial anxiety is who has the access to data, with 61% of extreme concern, followed by reliability of passwords and identification with 49%, sharing data to third party 48%, and question of who owns the data 41%.

Figure 7: Which of the following question is a concern for your company?

Service Level Agreement (“SLA”) is the key document defining relationship between CSP and Cloud service user. Laws and by laws are not sufficient to protect user’s privacy in the Cloud, and because these are too broad and provide many exceptions. Thus, SLA is basis for the customer’s trust in the CSP.

However, most Internet services users are not reading “terms of use” of the service provider. It is psychological phenomenon specific for the internet, and it has particular impact for Cloud users. There have been many researches proving this fact. One experiment conducted in funny manner, proven that users missed free 5 GBP bonus just because they didn’t read terms and conditions of the Game Station site (Smith, 2010). They didn’t even scroll down the Agreement they sign, to see the bonus option. Only 12% users noticed this option, meaning that they at least take a glance look down the Agreement text.

Similar test is made by PC Pitstop Company in their limited offer for anyone who read their End User License Agreement (“EULA”). After four months and more than 3,000 downloads of this software, only one person contacted company for reward. That person got a reward of \$1,000 (Magid, 2012).

Our survey shows on following figure 8 that considerable number of users is actually reading SLA. Almost 25% are reading, and additional 56.5% at least take a glance down the agreement text. Only 18.8% clearly stated that they are not reading SLA.

Figure 8: Do you read the agreement when you register on online services?

Common misconception we researched is that Cloud service provider is responsible for protecting the data. In an attempt to fully understand all terms of use contained in Service Level Agreement, which typically contain responsibility for data protection, consumers need to weight conditions and services offered towards their organization needs. It is interesting in most legal systems that burden of responsibility for safeguarding the data is transferred to the customer. This rule remains even when the data is processed by third parties, as long as the customer has control over the data. On following chart in our research we discovered different perception of customers. Very few respondents, only 10.1%, accept sole responsibility for the data privacy protection in the Cloud environment. However, most expected and rational answer “shared responsibility” is dominating with 59.4% responses.

Figure 9: Who is most responsible for protecting data?

Establishing Causal Links

The final focus of our scientific research is reflected in the hypothesis, through the search for the causal links within customers' loss of trust in Cloud computing services. First steps in understanding of concept of data privacy and its connection to customers trust in

cloud services may include speculations. Further on we will discuss the arguments and their causal relationships to the effects to the customer's trust, and we will try to find a way to explain these relationships.

During the research, we are circumspect and aware that causation is not a matter of absolute true or false. Certain phenomenon may produce different effect under different circumstances, but also same cause may have different outcomes. The key challenge of our causal methodology is how to prove that particular cause that generated the outcome on a certain small group, or in a different environment, will have the same or similar outcome on a relatively large group. The additional challenge is occurrence of multiple causal factors. Effects are rarely caused by a single factor, thus it becomes difficult to be sure that effects are not influenced by other causes, or that different causes will create different effects on a larger population. To counter the bias in concluding the causal links, and rule out the chance, our sample need to be random and to represent the larger population.

In following linear regression model we present scatter graph, as one of many we produced to represent values of two variables out of 242 correlations tested in this research. The graph shows how one variable (who is accountable for the data) is affected by the other (who owns the data). The data is presented as group of dots, positioned to present how explanatory variable on x-axis is affected by response variable on y-axis.

Coefficient of determination in our model is considerably strong with 0.55 value, thus indicates that our example explains the variability of the response data.

Figure 10: Linear regression model: explains established causal links

Conclusion

This paper presented privacy concerns in Bosnian business Cloud computing environment, and gives an overview of necessity for building privacy protection system rapidly. Cloud computing services are excellent stepping stone for small and medium businesses in transitional or unstable economies, like Bosnian. However, our research shows that Cloud based software services are seems distrustful for customers in Bosnia and Herzegovina, thus they are hesitate to use that stepping stone.

Type of vulnerable organizational information is definitely influencing customer's perception of risk. This perception of risk reflects on reliability of security utilities applied by Cloud service provider. Reliability is based on many other modes, and contractual agreements are among most trustworthy for the customers. Therefore it is critical that Cloud service users read, understand and deliberately accept the terms of use enclosed in CSP's agreement, before signing its conditions.

Forthcoming rapid spread of Cloud services will continue in emerging markets like Bosnian, regardless of lack of reliable data privacy protection assurance. But service providers might lose their pace if they neglect founded anxiety for the data privacy, which is alerted more than ever.

References:

AEA Guiding Principles.(2004). American Evaluation Association Guiding Principles for Evaluators, July 2004.

Angwin, J. (2010). The Web's New Gold Mine: Your Secrets, Wall Street Journal. (July 30, 2010)

Greengard, S. (2010 a). Cloud Computing and Developing Nations. Communications of the ACM, may 2010, vol. 53, no. 5

International Monetary Fund's World Economic Outlook Report, April 2010. Retrieved from <http://www.imf.org/external/pubs/ft/weo/2010/01/>

Magid, Larry (2012). It Pays To Read License Agreements. Retrieved from: <http://www.pcpitstop.com/spycheck/eula.asp> 21.04.2014.

Ponemon Institute (2013). About PonemonInstitute,retrieved from<http://www.ponemon.org/>, last accessed 03.05.2014.

Ponemon institute research: 2012 most trusted companies for privacy (2013). Research Report, January 28, 2013.

Smith, Catharine (2010). 7,500 Online Shoppers Accidentally Sold Their Souls

To Gamestation. Retrieved from: http://www.huffingtonpost.com/2010/04/17/gamestation-grabs-souls-o_n_541549.html 21.04.2014.

Van de Vijver, F., & Tanzer, N. K. (1997). Bias and equivalence in cross-cultural assessment: An overview. European Review of Applied Psychology, 47(4), 263-280.

Wang, R. (2010).Monday's Musings: The Hidden Value In SaaS Deployments.

January 25, 2010. Available at: <http://blog.softwareinsider.org/2010/01/25/mondays-musing-the-hidden-value-in-saas-deployments/>