International Journal on New Computer Architectures and Their Applications (IJNCAA) 2(1): 81-90
The Society of Digital Information and Wireless Communications, 2012 (ISSN: 2220-9085)

Online Surveys System with Enhanced Authentication Intelligence: A Case of Online Course Evaluation System

Kosmas Kapis, Sijali Petro Korojelo

College of Information and Communication Technologies,
University of Dar es Salaam,
Tanzania.
Kapis@udsm.ac.tz, KKapis@gmail.com, sijalipeter@yahoo.com

ABSTRACT

Online survey offers cheapest data collection and processing cost, efficient service, easy data processing and wide coverage over traditional paper-and-pencil surveys, due to this its adoption recently has been inevitable. However, despite the number of advantages an online questionnaire can offer, there are still a number of problems and challenges related to authentication that need to be closely addressed. Multiple submissions, respondent authenticity/validity, and respondent's anonymity are among the issues that hinder the proliferation of online surveys.

Many recent online survey systems are critically vulnerable to the challenges aforementioned. Foe example, Zhang and Hussein [11] and [2] respectively presented systems which tried to tackle the above addressed problems, but they were not successfully in tackling respondent validity problem of which still the intended respondent were able to access, fill and submit the survey.

This paper addresses the aforementioned challenges and gives a solution that fills the left gap by improving previous named works and later presents an intelligent online survey system which behaves well in all aforementioned situations including an additional problem of identity theft.

KEYWORDS

Security, identity theft, authentication, confidentiality, multiple responses, system

1 INTRODUCTION

Due to the cheapest data collection and processing cost, efficient service, easy processing, and wide coverage that the online surveys offer over the traditional paper-and-pencil surveys, the adoption of the former has been inevitable in recent years [4], [3], [7].

Thomas [10] defines an on-line survey as the collection of data through a self-administered electronic set of questions by e-mail, the web, or a combination of both. Online surveys falls under two major categories which are Email-based survey and web-based survey.

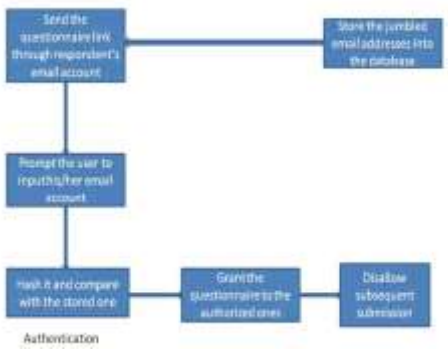
Email-based survey is the one in which the respondents are invited to participate in an online surveys through their email addresses. Usually the respondent of the survey receives an invitation email from the online survey administrator; the invitation email may contain a link address or URL which directs the respondent to the survey form. The other

type of online survey is the web-based survey in which the survey form is made available online and accessible by anyone. This type of online survey is normally intended to the generic respondents where the user authentication is hardly given priority.

Some surveys may be specifically targeted to a certain defined sample. In developing such a system several issues must be well addressed such as: i) How will the person administering the questionnaire guarantee that this survey response comes from the particular intended respondent and not someone un-intended?, ii) how will the person administering the survey make sure that all necessary survey questions are attempted by the respondent?, iii) How will multiple submissions from the same respondent be controlled so as not to cause biasness and "spoilage" of the reality of the survey results, iv) Lastly how to make sure that the submission of the survey form by the respondent is safe and that there is no easy tracking or identity theft by the third part (Hacker)? All of the above are the challenges and problems related to authentication which face the design and use of online surveys. Two research works have attempted to study and resolve some of the above named problems.

First work by Zhang [11], presented the Web-based online survey system which was used to collect data about the Library Information System (LIS). It was used to research how the use of internet has eased the work of researchers. The system was running in a Unix HP server. The web pages have been built using HTML, and the Computer Gateway Interface (CGI) program has been used to generate the case Identifications (IDs). The system worked by inviting each respondent using the respondents' email addresses, to participate in the survey by receiving an email from the system administrator. Using a CGI program in the server side, a 10 digit random case identification ID was produced and printed on the cover letter. Each intended respondent receives, through email which contains a hyperlink to the survey page, a cover letter with a case ID printed on it.

The system used the 10 digit case ID produced by the CGI server as the means of identifying the intended respondent. After receiving an email with the case ID in the cover letter, a respondent follows the hyperlink to the survey page. The respondent is then prompted to provide his/her case ID. The system then compares the given case ID with the list of those stored in a database. If the case ID is found in a database, then he/she is the legitimate user and thus authorized to access the survey page. The system used a database to store the case IDs, and other useful data about e-sources. Furthermore, it contains the "survey data" component which is used to record the submitted survey data so that respondents could review them later.



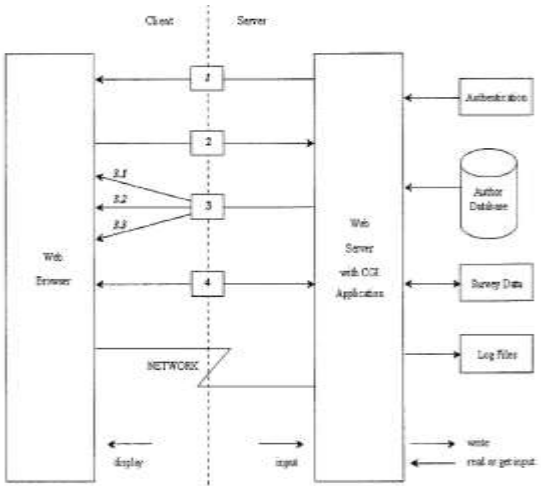


Figure 1. Web-based survey system architecture (source: [9]

Second, [6] on the other hand studied the work of Zhang he came with an alternative system better than the Zhang's system. In his system, he used email based survey. The system is presented in figure 2.

Figure 2. Overview of Online Survey System (source: [7]

His system worked by the system administrator inviting the respondents to participate in an online survey by emailing them. The respondents are then linked to the survey home page where they are prompted to provide their email addresses. Using the one-way MD5 hash function the system compares the entered email address with that in the database. If the two hash codes match, then the respondent is an intended one and is thus privileged to access the questionnaire page. Otherwise the respondent is an un-intended one and thus he/she is denied the access to the questionnaire. After the first submission, the respondent is then prevented from any further subsequent responses.

Zhang's system did well in the area of tackling the problems of multiple responses of which he managed to control it but not for the cases of identity theft and respondent validity or authentication. His system failed in the case the survey invitation email is accidentally delivered to an un-intended respondent where that un-intended responded was able to access and fill the survey page since the case ID was printed on the invitation email. [6], on the other hand improved the work done by Zhang by tackling additional problem of respondent anonymity. He did it by hashing of respondent details before transmitting them hence someone could not tell the real identity of the person since the respondent details were hashed and could not be decrypted. However Hussein's work failed to do well in the case of respondent validity. It was seen that when the intended respondent received the survey link page from the system administrator and forwarded that link to another person who is not meant to receive the survey link before that particular intended respondent responds to the survey page, in that case this person receiving the survey link page could still access, fill and submit the survey page without any problems using the email address of its sender usually printed on the "from" field of the email.

This work reviews previous works of other people as far as the issue of authentication in online surveys is concerned, see what further authentication challenges, problems and weaknesses they face, proposes suitable alternatives for combating the problems and lastly presents the prototype of an advanced system which tackles the

issues mentioned by combating the observed weaknesses. The authors of this work hypothesize that all the addressed problems above are associated by the use of poor authentication methods/techniques of which this work seek to find solution to them.

2 APPROACH

In order to resolve the prior named problems the authors used password authentication techniques/method. Deb [5] argues that most of people are familiar with password authentication, to log onto a computer or network, you enter a user account name and the password assigned to your account of which this password is checked against a database that contains all authorized users and their passwords. Password authentication method is usually considered weak, but in this work MD5 hash function was used to strengthen its performance. Using this function respondent details containing both email addresses and passwords were hashed before being stored or transmitted, this helps to tackle problems of respondent anonymity and respondent tracking by a third party user or hacker. The approach used to solve the problems is depicted in figure 3.

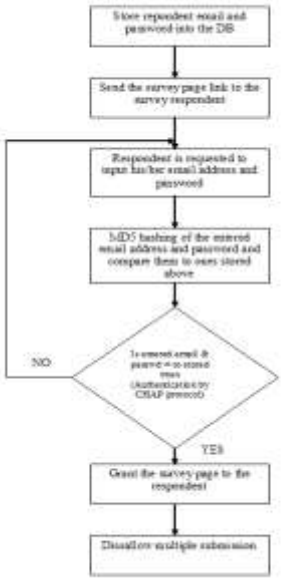


Figure 3. Proposed design of online student course evaluation survey system

The use of both the email address and password eliminated the respondent validity problem which was unresolved by Hussein [6]. In the problem, when the intended respondent received the survey link page from the system administrator and forwarded that link to another person who is not meant to receive the survey link before that particular intended respondent responds to the survey page, it was seen that this person receiving the survey link page could still access, fill and submit the survey page without any problems using the email address of its sender usually printed on the "from" field of the email.

In this new system it is difficulty for any un-intended respondent to fill and submit the survey page unless he knows both email address and password of its sender (intended respondent) who has sent him/her the questionnaire, so even if he gets the email from the "from" field of the email received from the intended respondent he can't know the password of that particular respondent unless the intended respondent decides to give that person the password.

Other methods such as smartcard and biometrics are available but they are expensive and sometimes they require specialized hardware and software which are expensive, this makes them uneconomic for a person to use for the reason of collecting online data. The system is also built such that multiple submissions are rejected after first submission is successful.

In case of the authentication protocol, this work used CHAP (Challenge

Handshake Authentication Protocol). Dax Networks [4], Argues that CHAP is a type of authentication in which the authentication agent (typically a network server) sends the client program a key to be used to encrypt the username and password. This enables the username and password to be transmitted in an encrypted form to protect them against eavesdroppers. [4] Further argues that Challenge Handshake Authentication Protocol (CHAP) is used to periodically verify the identity of the peer using a 3-way handshake (challenge, response, failure or success message). CHAP is the best option for the work for several reasons, First CHAP verifies that the two devices have the same "shared secret" but doesn't require that the secret be sent over the link which is not good as far as security is concerned, second CHAP provides protection against replay attacks, where an unauthorized user captures a message and tries to send it again later on. This is done by changing an identifier in each message and varying the challenge text. Third using CHAP, the server controls the authentication process, not the client that is initiating the link. Lastly, [8] argues that CHAP is an only internet standard that uses MD5 (Message Digest (5th version)) hashing algorithm of which was also used in this work.

In this work a new system controlled by the system administrator, first stores respondent information including email accounts and passwords to access the survey page into the database, he/she then invites the respondents to participate in an online survey by emailing them and sending them the questionnaire link (see figure 4). The respondents are then linked to the survey home page where they are prompted to provide their identities which are the emails addresses and password known to them prior (either sent to them by a separate mail or by phone). Then using the one-way MD5 hash function the system compares the entered email address and password with that in the database by applying CHAP authentication protocol. If the two hash codes match, then the respondent is an intended one and is thus privileged to access. She or he then fills and submits the questionnaire page. Otherwise the respondent is an un-intended one and thus he/she is denied the access to the questionnaire. Multiple submissions are rejected after first submission is successful.

On its completion, the new system was tested in all challenging cases addressed prior. First in the case of respondent validity (either the respondent is intended or un-intended, see figure 5), second in the case of restricting multiple submission, and third for the case of avoiding respondent tracking/ tracing by the third part user also known as the system hacker and ensuring that respondent anonymity is preserved.

Twenty three users were involved in the system testing. Out of them, ten were asked to act as un-intended respondents, eleven as intended respondents, and two as system hackers who were asked to track/trace the system. The email was adopted as the means of communication between the survey users and the system administrator. It was used by both the system to invite the survey respondents to participate in the system testing and by the respondents to send back the findings to the administrator. All

respondents whether intended or un-intended were emailed and asked to try logging into the system and access the survey page and provide the feedback of the findings to the system administrator.
On the other hand during the system evaluation against multiple responses, the intended respondents were emailed asked to attempt submitting the questionnaire several times as well as trying to forward the invitation email to someone else to respond on their behalf and finally they were asked to give the findings back to the system administrator. Lastly 2 users were emailed to act as the hacker and asked to try if they could trace the respondents by knowing which questionnaire has been submitted by whom.

Figure 4. Online student course evaluation survey page

Figure 5. Respondents Login page

3 RESULTS

In this work, the online course evaluation survey system was implemented with 23 respondents being involved during system testing. The cases that the system was tested to see its performance are: the case of respondent validity (either the respondent is intended or un-intended), the case of multiple submission to see if the system could restrict that and the case of respondent tracking/ tracing by the third part user also known as the system hacker and the case of preserving respondent anonymity. Below are the results obtained during that system testing.

3.1 Respondent Validity/ Authentication

Ten un-intended respondents were forwarded a survey invitation email, among ten un-intended respondents invited only 5 responded by attempting to access the questionnaire page. Each time they tried to login into the system they were not able access the survey page, their logging in details (email addresses and passwords) were captured and stored in the "non_respondents" table so that it can be used as evidence during the system evaluation. As shown in the figure 6 below five out of five un-intended respondents who tried to log into the system were blocked.

Furthermore, eleven intended respondents were forwarded a survey invitation email, among them, only seven filled and submitted the

questionnaire successful, two respondents among ten did not attempt the survey but they were asked to try to forward the invitation email to someone else so that he/she can access the questionnaire page, results showed that those people forwarded the survey invitation email did not succeed to access and fill the survey and their details were blocked too.

Figure 6. List of un-intended respondent

3.2 Anonymity/Confidentiality and Respondent Tracking

Among the two users acting as system hackers who were asked to trace the system, only one tried to access the system's database where he was asked to trace which questionnaire has been submitted by whom. The hacker then reported that he couldn't do that because the submitted questionnaires and respondents' identities are stored in the separate tables within the database, see figure 7 and figure 8. Furthermore, even though he was privileged to access the system database such as "respondents' table which contains the respondents' hashed email addresses and passwords, stealing or identifying them was not possible at all and hence the anonymity of the respondents was preserved. This is due to the fact that the stored email-addresses and passwords were hashed using an MD5 hash function in which the encrypted values couldn't be decrypted back to their original format.

3.3 Multiple Submissions/Responses

Results showed that the submitted questionnaires were 7 even though each intended respondent was asked to attempt submitting them several times. This shows that after the respondent has responded once, then the rest of the subsequent responses were denied. However the system was tested by instructing one of the intended respondent who already filled and submitted the survey form to try to re-submit a new filled form once again. He argued that the attempt was denied by giving the error message "Sorry you have automatically restricted from submitting more than one questionnaire".

Figure 7. Submitted questionnaires

Figure 8. The database of hashed respondent identities

4 DISCUSSION

Respondent validity, anonymity, identity theft, tracking, and multiple responses are some of the critical problems that an online survey suffers from. Some of the suggested approaches such as that of [11] and [6] are still not sophisticated enough to provide complete solution to all the named problems.

The online course evaluation survey system that has been implemented in this work behaves well all of the challenging situations, first in the situation where the invitation email is accidentally sent to un-intended recipient by the administrator. In this case having received an invitation email, that particular un-intended respondent has to be authenticated using the valid email address and password of which he/she will not have because his/her details are not stored into the system database, second in the situation where the invitation email gets forwarded accidentally to un-intended respondent by the intended respondent having received the invitation email from the system administrator and before that particular intended respondents fills and submits the survey. The one who receives has to be authenticated like in the previous case, in this case even if the unintended respondent knows the sender email address from the "from" field of an email, he/she has to know also the sender's password (used to access the survey page) so to access the survey page, hence that particular un-intended respondent can't access the survey unless his/her sender decides to tell him/her the password of which this will be a purposive forwarding of the invitation email.

Based on the results obtained and discussed in the results section, the system proves to be more stable against the submission of multiple responses from the same respondent. The system

accepted one and only one response from each respondent.

The system also offers the reasonable solution against both identity theft and respondent tracking, the results show the submitted questionnaires and the hashed respondent's identities are stored in the separate database tables. This indicates that even though someone accesses the system database, he/she can hardly tell who submitted a particular questionnaire. Storing the hashed identities indicates that even if one gets into the database and steals information, decrypting them will be difficulty since the encryption schemes used was only one way, i.e. you can only encrypt data but decryption isn't possible. Also hashing during authentication plays an important role against identity theft on transit.

Despite all of the above achievement of this work, there are still a number of challenges that particularly need to be dealt with. First low response rate is the issues which need much effort in order to establish a suitable online survey system. During the system test, out of 23 respondents invited to participate in the online questionnaire, only 13 responded (7 intended respondents, 5 un-intended respondents, 1 hacker). Second results show that only two respondent sent feedback during the system testing. This is because there wasn't a log file in which users could type the comments and send back to the system administrator instead of emailing him. This finding agrees with the "feedback" limitation as reported by other researchers.

5 REFERENCES

1. Abdullah, A. Protecting your good name: Identity Theft and its Prevention. On the proceedings of the 1st annual conference on Information security curriculum. Kennesaw State University, Georgia: USA, (2004).
2. Anita, K. Password Authentication with Insecure Communication. Department of Computer Science, University of Waterloo, Waterloo, Ontario:Canada, (1981).
3. Articlesbase. Advantages and Disadvantages of Using Online Surveys. http://www.articlesbase.com/business-articles/advantages-and-disadvantages-of-using-online-surveys-1377003.html (2009).
4. Dax Networks. CHAP (Challenge Handshake Authentication Protocol. www.daxnetworks.com /Techno logy /Tech Dost/TD-06 14 0 6. pdf (n.d)
5. Deb, S. Understanding and selecting authentication methods. http://articles.techrepublic.com.com/5100-10878_11-1060518.html (2001)
6. Hussein, A. Authentication in Online Surveys. University of Southampton, UK (2008)
7. Peter, M. Advantages & Limitations of Forms in Web Pages. http://www.ehow.com/list_6331726_advantages-limitations-forms-pages.html (2010)
8. Shinder, D. Understanding and selecting authentication methods. http://articles.techrepublic. com.com/5100-10878_11-1060518.html (2008)
9. ScienceDaily. Online Access With a Fingerprint. http://www.sciencedaily.com/releases/2010/12 /101 21 5082942.htm (2010)
10. Thomas, M. Points for Consideration On-Line Surveys. Ohio State University: USA (2003)

11. **Zhang, Y. Using the internet for Survey Research: A case Study. Journal of the American society for information science, John Wiley & Sons, Inc, .57-68** .http://www3.interscience.wiley.com/cgi-bin/ fullte xt/6 95009 1 9/PDF START **(2000)**