Carleton University, Ottawa, Canada
School of Computer Science
Winter Term, 2017

COMP3008 Project 2
Quantitative Usability Evaluation

Specified: 2017-03-01 — Due: 2017-04-07

Introduction:

This project is to practice and explore the elements of Quantitative Usability Evaluation as presented and discussed in class and in the textbook. The application domain is knowledge-based authentication: password systems.

There are two parts to the project, one involving descriptive statics of authentication schemes based on sample data that will be provided to you, and one involving design and evaluation of a new authentication scheme, and inferential statistical to compare it with an established system. These two parts are described below, outlining the work you need to do, and how to prepare a report for assessment. Approximate page lengths for the sections of the report are suggested below, where a page is approx. 350 words.

This project is to be done working in small teams: 3–5 people. You are responsible for organizing and managing your teams. You may retain your team from project 1, or change your team. If you are changing your team, please use the forum on CULearn, and contact the Instructor or Teaching Assistants if assistance is required. On or before March 8th, please email your team details to `SanaMaqsood@cmail.carleton.ca`

Password Authentication

Authentication is the process of determining the truth of a claim, typically relating to identity or the right to access to services or resources. Authentication is typically determined on the basis of "something you have", such as a physical cards or tokens, "something you are", such as biometrics like fingerprints, or "something you know", such as passwords.

Despite the large number of options for authentication, text passwords remain the most common choice for several reasons. For example, they are easy and inexpensive to implement; are familiar to essentially all users; allow users to authenticate themselves while avoiding privacy issues that have been raised about biometrics; and have the advantage of portability without, for example, having to carry physical tokens. However, text passwords also suffer from both security and usability disadvantages. For example, passwords are typically difficult to remember, and are easy to guess if user-choice is allowed. This is sometimes called the "Password Problem".

In recent years, many novel approaches have been suggested to improve passwords to address the password problem. It is therefore important to carefully assess the security and usability of the new approaches. For more information, please see the following paper, available on the CULearn: R. Biddle, S. Chiasson, P.C. van Oorschot (2012). *Graphical Passwords: Learning from the First Twelve Years.* ACM Computing Surveys 44(4).

In this project, we focus on schemes that assure a certain level of security, but where we need to assess the usability. In particular, we focus on assigned random passwords with consistent password spaces. Our primary measures of usability will be the memorability of passwords, and the speed of password entry.

Part 1: Sample Data and Descriptive Statistics – 40%

This part of the project is to conduct usability analysis on an experiment that has already been conducted, and where sample data will be provided. Your work is to process the data and interpret the results.

The sample data is anonymous and comes from research to study various new password schemes done using the MVP framework, as discussed in class. Each data set involves approximately 10–25 users. Each user data involves uses of 3 websites, each with a different password, with repeated usage.

For more information on the MVP framework, please see the following paper, available on CULearn: S. Chiasson, C. Deschamps, E. Stobert, M. Hlywa, B. Freitas Machado, A. Forget, N. Wright, G. Chan, R. Biddle (2012) *The MVP Web-based Authentication Framework.* Financial Cryptography (FC), Spinger-Verlag LNCS.

1. Consider the schemes below. To explore each scheme, try the name of the scheme as a userid, and use the training page at: `http://mvp.soft.carleton.ca/trainer`

   Report: Suggest the advantages and disadvantages of each scheme, and provide screen images of you using each scheme from the training page — approx. 1–2 pages.

   Text28: Passwords of 6 random lower-case letters.
   Blankpt28: Passwords of 6 random tiles, chosen from a blank grid of 80 tiles.
   Imagept28: Passwords of 6 random tiles, chosen from a image made of 80 tiles.

2. Download the data from the CULearn folder for Project 2.

   The files contain log data in CSV format (comma separated values), and may be inspected in a text editor, imported to a spreadsheet such as Microsoft Excel or LibreOffice Spreadsheet, or read into R. Each row in the file has the fields below, and is in order by time.

   time: timestamp in the format `YYYY-MM-DD HH:MM:SS`
   user: userid
   site: website identifier
   scheme: password scheme/sub-scheme identifier
   mode: usage mode identifier
   event: event identifier
   data: data associated with the password event

   Design and implement software programs to process the log data for each scheme, producing a new file in CSV format containing results ready for analysis. For this processing you may use any language or environment you wish, but R is suitable. Think carefully about how to process the log. In particular, you may wish to re-order the data so that you can process the information by user and by site. The new file should contain at least the following information for each user of the scheme:

   - The userid.
   - The password scheme.
   - The number of logins, successful logins, and failed logins.

- The time taken to enter a password, recorded separately for successul logins, and for failed logins.

Report: Documentation for your log data processing software, including high-level explanation and pseudocode for your approach, and the documented source code. Also provide the resulting data in CSV format.

3. Compare the usability of the schemes, by developing programs using R to calculate descriptive statistics and produce graphs. The descriptive statistics should include: the mean, standard deviation, and median of number of logins per user, total, successful, and unsuccessful; and the mean, standard deviation, and median of the login time per user, successful, and unsuccessful. The graphs should include: histograms for the number of per user, total, successful, and unsuccessful; and histograms and boxplots for the login time per user, successful, and unsuccessful. Interpret these results, and discuss which scheme is best. *Report: your descriptive statistics, your graphs, and your interpretation of them, discussing which password scheme has the better usability. Provide the documented R source code that produces your statistics and your graphs.*

Part 2: Design, Implementation, Statistical Inference – 60%

This part of the project is for you to design a new authentication system and conduct quantitive usability testing.

1. Design a new knowledge-based authentication scheme. It should assign passwords randomly, rather than let users choose them, and the password space should be approximately 28 bits. That is, there should be approximately 2^{28} possible passwords. The scheme may use characters, words, images, or any combination, or something novel. Try to leverage what you know about human memory. Even simple schemes will earn a passing mark, but extra marks will be reserved for innovative schemes.

   Report: Design rationale explaining your scheme, why you think it might have good usability, and your calculation of the password space — approx. 1–2 pages.

2. Implement your password scheme. You may use any programming language and environment you wish, as long as it will be possible to test your scheme with participants.

   Report: Documentation including screen images of your program in use, both for password creation and password entry, together with notes explaining the images; and the documented source code for your software for us to review (see Assignment Submission, below) — approx. 2–3 pages.

3. Create a simple framework to assist quantitative testing of your new password scheme. Your system should assign and test 3 passwords. For each password, the system should say what it is for (e.g. "Email", or "Bank", or "Facebook"), show the user the password, and have the user enter the password to confirm they know it. When this has been done for all 3 passwords, the system should request the user enter the passwords again, in a random order. It should then repeat this process again. The user may be allowed up to 3 attempts to enter each password before declaring failure. Instrument your software to record details useful for assessing usability, such as login times, success, failures, etc. You should record the event details in a log file, to allow unanticipated analysis later. Make sure you save the log files!

Report: Documentation including screen images for your system, together with notes explaining the images; and the documented source code for your software — approx 2–3 pages.

4. Create a simple questionnaire to investigate the user's perception of the new password scheme, in comparison to normal user-chosen text passwords. It should use approx. 10–20 Likert scale questions to assess relevant perception. You should create your questionnaire using the COMP3008 survey system at: `https://hotsoft.carleton.ca/comp3008limesurvey` Your team will be given a userid on request.

   Report: your survey questions in pdf (generated by the survey software), and a link to your actual questionnaire on the survey website.

5. Conduct testing of your password scheme, using your system and your questionnaire. Your participants should be other students at Carleton, and can include your own team members and other students in COMP3008. You should recruit at least 20 participants.

   You must again obtain informed consent by participants. You should use the consent and debriefing forms for Project 2 provided on CULearn. Completed consent forms must be handed in to COMP3008 box in HP 3115 before your project will be graded.

6. Compare the results of your usability testing with the results obtained by the TEXT28 scheme. You should produce descriptive statistics and graphs for success and failure, and also for login time, as in Part 1. You should use inferential statistics to determine whether the results are statistically significant. You should use the results of your survey to compare perception of your scheme with a normal password scheme, and produce descriptive and inferential statistics. Use R for all graphs and statistics.

   Interpret the results of your testing and survey, and assess your findings. Do your best to answer the questions: Is your new scheme better than TEXT28? Would users prefer it to a normal password scheme?

   Report: presentation and interpretation of your study results, and your interpretation and discussion — approx. 2–5 pages.

Assignment Submission:

Throughout the period of the project, the teaching assistants will be available during their office hours for advice and feedback as you work through the project steps. For example, you may wish to seek feedback on your password design, understanding of the experiment data, or interpretation of results.

For the project report, please write up the steps of your study as described above. Please ensure that the entire report is in a single PDF document. Upload the file following the submission instructions on CULearn.

All program source code should be helpfully documented internally. In particular, every file must begin with a block comment briefly explaining all the classes or functions within it, and their purpose. All data files should be named to make the contents clear. Your set of files should also have a file "readme.txt" that explains each of the other files and their purpose. The documentation should include sufficient detail for someone else to compile or run the code themselves.

The project is due at 5PM EDT (as indicated on CULearn) on April 7th. Projects submitted or updated late will be penalized by a deduction of 10 marks (out of a possible 100) per 24 hour period, or part thereof.