Progress Report Template and Rubric

The purpose of the progress report is to plot a snapshot of the work already accomplished, the work in progress and the remaining work by the end of the project. Everyone on the team should work on the report; teamwork is essential here.

1. Project description relevant at the reporting time (*3 points*)
   Describe the current project scope together with the current limitations or changes with respect to the initially proposed project description.
2. Work accomplished (*2 points*)
   In addition to a description of the work completed, include a brief description of the overall team effort and which team members worked on each task.
3. Work remaining and timeline indicating both the major project steps and the current progress report (*3 points*)
   Update the project timeline from the proposal to include all the recent changes to project steps and milestones (if there are any), and the time of the progress report. In addition to a description of the work remaining, describe which team members are expected to work on which parts.
4. Updated bibliography; if no updates, say so (*1 point*)
5. Any project difficulties experienced up to the reporting phase and respective mitigation plan; if none, say so (*1 point*)
   Indicate any difficulties experienced with the project such as technical problems, missing expertise, logistics problems or internal team problems. Come up with a plan to mitigate these problems and explicitly note if the the team needs an additional assistance of any kind.

Accompany the draft report with a presentation that communicates the essential points of the progress report concisely and clearly. Each group will give a 10–12 minute progress report using the slides.

Attached are two example progress reports from a previous class (at Purdue). Slides for the second are also included.

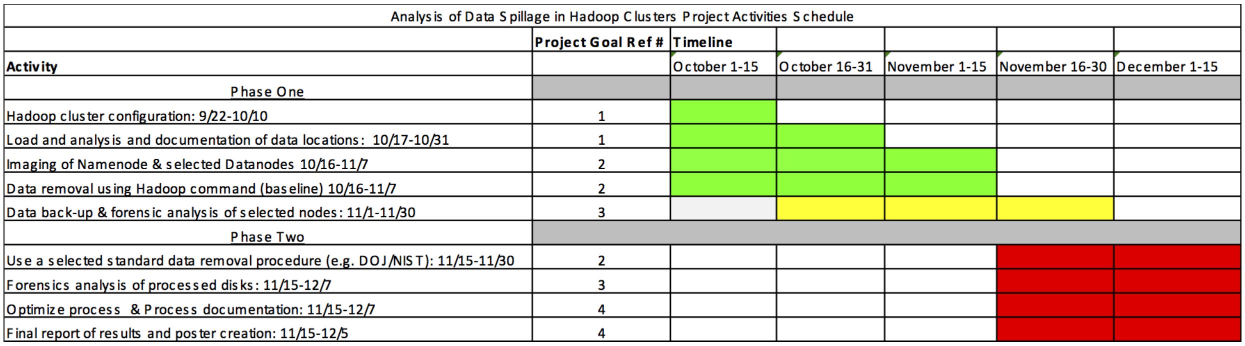
Example: Data Spillage in Hadoop Clouds

redacted

Problem Statement

Data spillage is defined as security threat that occurs when sensitive information is introduced into a non-authorized platform. This project focuses on a scenario that involves user introduction of sensitive information into a non-authorized Hadoop distributed file system (HDFS) cluster. Using a forensic analysis approach of the cluster's Name node and Data nodes, the goal of this project is to create a procedure to remove all sensitive material from the HDFS cluster with minor impact to cluster availability. In support of this work, we defined four project goals:

1. Locate and retrieve, the Name node meta-data and retrieve the locations of all the nodes on which the sensitive data resides following the upload of a document designated by the project team as sensitive.
2. Determine and select a set of appropriate removal methods and remove the file meta-data and data blocks from the HDFS cluster. Our optimal removal processes will be governed by NSA specifications and process maximization of system availability.
3. Forensically examine the impacted nodes following the removal process in order to confirm the inability of potential attackers to find residual sensitive information.
4. Document the forensics findings and procedures carried out for each step to allow for reproducibility on more complex data sets/types.



Analysis of Data Spillage in Hadoop Clusters Project Activities Schedule						
	Project Goal Ref #	**Timeline**				
Activity		October 1-15	October 16-31	November 1-15	November 16-30	December 1-15
Phase One						
Hadoop cluster configuration: 9/22-10/10	1	🟩				
Load and analysis and documentation of data locations: 10/17-10/31	1	🟩	🟩			
Imaging of Namenode & selected Datanodes 10/16-11/7	2	🟩	🟩	🟩		
Data removal using Hadoop command (baseline) 10/16-11/7	2	🟩	🟩	🟩		
Data back-up & forensic analysis of selected nodes: 11/1-11/30	3		🟨	🟨	🟨	
Phase Two						
Use a selected standard data removal procedure (e.g. DOJ/NIST): 11/15-11/30	2				🟥	🟥
Forensics analysis of processed disks: 11/15-12/7	3				🟥	🟥
Optimize process & Process documentation: 11/15-12/7	4				🟥	🟥
Final report of results and poster creation: 11/15-12/5	4				🟥	🟥

Timeline. Green cells indicate ongoing work. Red cells indicate completion of work.

Work accomplished

As indicated by the green cells in the time line shown above, the project is slightly ahead of schedule. We have created the our test environment to, load sample data sets, locate the sensitive data within the cluster using information provided by the name node, and removed that data using HDFS commands. We have also completed the retrieval the memory images of the Name node and "infected" Data nodes related to our first sample data set.

Work remaining

The team is currently working on the following:

1. Analyzing the forensic images of all nodes in the cluster, as denoted by the yellow cell in the chart above, in order to: ensure that all sensitive data has been located, test procedures for removal of the sensitive data, and supplement our data archives which enable future research.
2. Ongoing literature review is also currently focused on details of HDFS data storage and removal. This work has created a basis for creation and analysis of the proposed data removal procedure, and for future work relating to data storage in HDFS.
3. Completing an analysis of our procedure. We expect that we will find remnants of the sensitive data as we analyze the node images taken after the removal of that data using HDFS commands. As a result, we expect that our removal procedure will include steps to remove these remnants in a way that maximizes the availability of the cluster.
4. Documentation of our procedure and analysis results is the main deliverable of the project.

Updates to bibliography

None.

Problems encountered

None.

Example: Coping Mechanisms in Password Selection

redacted

Problem Statement

Stringent password policies include requirements such as not including a dictionary word, or including a capital letter, special character, or number. To analyze how the requirements of these policies effect users, passwords will be collected from individuals within three different password policies over seven required password modifications, and will be reviewed for coping mechanisms such as using a capital letter first, using a number or special character last, repeating the same character or word multiple times within the password, repeating the same root password over multiple changes, or incrementing numbers or special characters over multiple changes. The NIST entropy calculation will be modified for any of these coping mechanisms observed and with the results showing numerically how coping mechanisms decrease the actual entropy of a specific policy.

Work accomplished

The group rewrote the proposal, edited and modified the literature review, and wrote the IRB Application. In addition, we have completed the required consent form and wrote the survey questions. Based upon timing concerns, the group decided to pursue the Mechanical Turk collection process initially, and have submitted the required IRB Application. After receiving feedback from the IRB, the Application has been modified as required and will be resubmitted for review on Friday October 31, 2014.

We have met with Statistical Consulting at Purdue University as well as with Lisa Zilinski to discuss our data management plan. Additionally, we have collected fake passwords from classmates and professors at Purdue and calculated both the actual NIST entropy and a post coping mechanism entropy for each password. This has allowed us to determine the mean of the NIST entropy and post coping mechanism entropy, the standard deviation, and confidence interval. In addition, through the analysis of this fake data the group has narrowed down and determined the specific coping mechanisms to be evaluated, and assigned a value for each one. We have used this information to create the presentation to be given in class on October 31, 2014.

Work remaining

The most time sensitive item remaining is to finish creating the website. If a website designer is not located, the team will work to create a website on our own with the assistance of other individuals in the CERIAS program.

We have calculated the actual NIST entropy and post coping mechanism entropy of the fake passwords by hand, and we need to determine whether we will continue to do this calculation by hand with the actual data, or whether there is a program to assist with this task. Additionally, the team has used SAS software for calculation of the mean entropy, standard deviation, and confidence interval, and this will also be reviewed for determination of whether this is the most appropriate software for use in our analysis. These decisions will be made once the full analysis of the fake data is complete, which we hope to accomplish before November 14, 2014.

We have received initial feedback from the Institutional Review Board (IRB) and modified the Application as requested. Upon receiving final approval from the team will launch the Mechanical Turk HITS and the website. Within two days after the first Mechanical Turk HIT is posted, we will review whether the each participant correctly completed the survey. If yes, then the payments to the Mechanical Turk participants will be approved, and each participant will be invited back to participate in the second iteration of the password creation. If a participant does not complete the task, then we will notify Mechanical Turk of this and that participant will not receive compensation nor an initiation to participate in a subsequent round.

Each instance of the data collection will be downloaded and analyzed as it is received. This will allow the team to present limited data findings by the end of the semester as well keep the workload spread out over the data collection, instead of attempting to calculate the entropy of all passwords at one time. The team will present a final presentation and paper on the data that has been collected to date on December 5, 2014. If the data collection is not completed by that date, *redacted* will continue on after that date with the final iterations of password collection.

The team is still interested in collecting data from Purdue students, and will be submitting an IRB Application for that portion of the project. The goal is to have that IRB Application turned in on or before December 1, 2014, to allow sufficient time for review and any modifications that may be needed prior to the spring semester starting on January 12, 2014.

Updates to bibliography

None.

Problems encountered

None.