Paper Review

Database Forensic Analysis with DBCarver

Gurpreet Singh

February 11, 2018

CS4414

Jot notes

Abstract

- Forensics needs to be effective on databases that are configured incorrectly and have no protection techniques integrated in them. Sometimes they don't even have logging
- Paper talks about a tool called DBCarver.
- Reconstructs database context from a database image without using any log or system data.
- Uses page carving to reconstruct both query-able data and deleted data.
- Allows investigators to conduct new types of analysis on the data

Introduction

- Since most large applications use databases, large scale cyber crime almost always involves databases.
- A related concept is file carving, which retrieves files from storage on a file system even if they are deleted or corrupted
- Relational databases store data in pages
- We can reconstruct pages to retrieve data
- Targets extracting data maintained by the running database instead of recovering original
- Paper considers two scenarios, database is good or bad

Contribution and Related Work

- Explain Page carving
- Describe how to do sql analysis on data to solve real-world forensic problems
- Evaluate performance of the tool itself
- Other people have done work to prevent and detect database tampering but not much has been done to actively reconstruct data from a damaged database to find evidence
- Many other methods assume logs and detection mechanisms exist and leverage those to conduct an analysis which is not always possible in a real-world scenario
- This system leverages the fact that the metadata associated with the real data isn't present but the data is still in good condition.

Page Carving

- Follows national regulations set out by the National Institute of Justice
- Three tasks that occur in an investigation are evidence acquisition, evidence reconstruction, and evidence analysis.
- The investigator's goal is to preserve all evidence
- Assumptions are made about the running system and because it could be dynamically changed the data associated with runtime cannot be regarded valid for the investigation

Tool Overview

- Pages are the smallest unit of persistent storage in a relational database.
- DBCarver traverses pages within a desk image extracted from a relational database and tries to identify data within those pages to reconstruct evidence.
- Consists of two components: the parameter detector and the carver
- Parameter detector sets up a fake database using the specified Database Management System and loads in some data. Then it extracts the underlying identifying system configuration that indicates how the data was stored for this specific type of DBMS.
- The Parameter detector sets up the Carver using this configuration it has generated as the seed. Then the Carver uses this information on the real recovered data image to make sense of the data pages.
- Paper states the parameter detector requires "modest user intervention". Bad design redflag Requires making a wrapper class, running sql commands, schema changes, etc
- In most Database Management Systems the three major identifying components for each page are page header, row directory, and row data.
- The page header provides general information about the page and regarding how it relates to the database as a whole

- Row Directory helps index the data inside a single page. It can exist in different places within a page for each DBMS. This directory will have references to each of the separate Row Data entries.
- Each individual Row Data entry will have the raw data and the status in which the data was present. It will indicate if the data was deleted or not.
- The carver can then accept any type of data file of any size because it is solely comparing against the criteria extracted by the parameter detector
- Carver is able to process multiple databases with multiple extracted configurations concurrently
- Is a command line utility with simple input and output commands
- When input is a database image, a new database is created with the recovered data.
- When input is a RAM snapshot, the result is raw data present in the working memory in the form of cache pages.

Experiments

- Experiments were conducted with a well controlled environment meaning the results should be indicative of the performance of the tool
- Experiment 1 demonstrates creating a database and retrieving the contents via carving
- Experiment 2 shows how the performance of the tool scales up with pages in database. It shows that the larger the database the longer it will take to process per MB of data.

Summary

Strengths and Weaknesses