Forensic Analysis of Web Data

Erhard Dinhobl

Vienna University of Technology - Information and Software Engineering Group

5.10.2016

Introduction

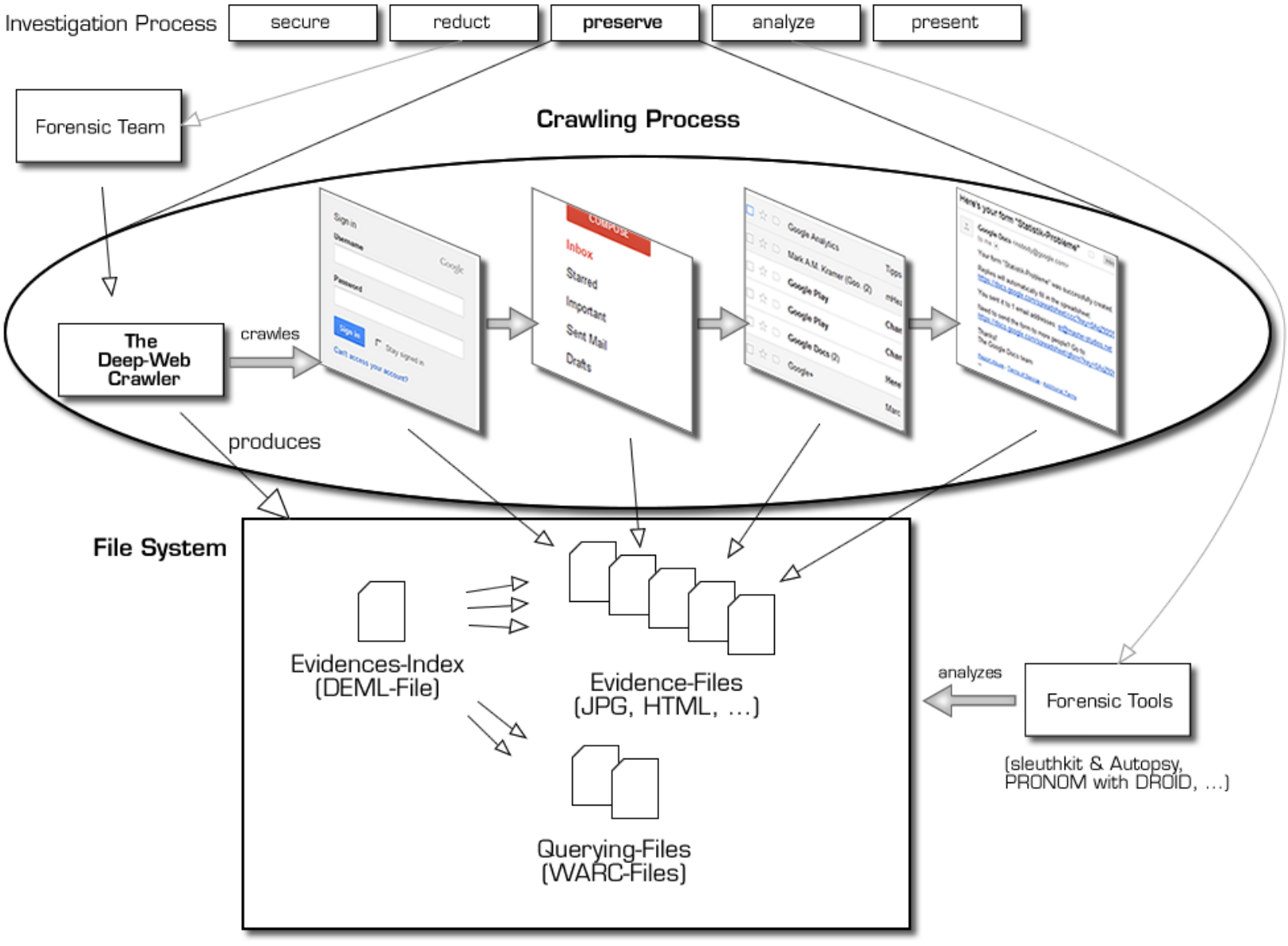
- using collected data in surveys
- influence to known investigative models
- appropriate investigation model
- digital preservation
- legal issues
- deep web crawling
- format of data
- designing a deep web crawler

Table of contents

The Investigation S(RP)AP

- secure
- **reduct**
- **preserve**
- analyze
- present

The Process


Investigation Process
secure
reduct
preserve
analyze
present
Forensic Team
Crawling Process
The Deep-Web Crawler
crawles
produces
File System
Evidences-Index (DEML-File)
Evidence-Files (JPG, HTML, ...)
Querying-Files (WARC-Files)
analyzes
Forensic Tools
(sleuthkit & Autopsy, PRONOM with DROID, ...)

The Process (2)

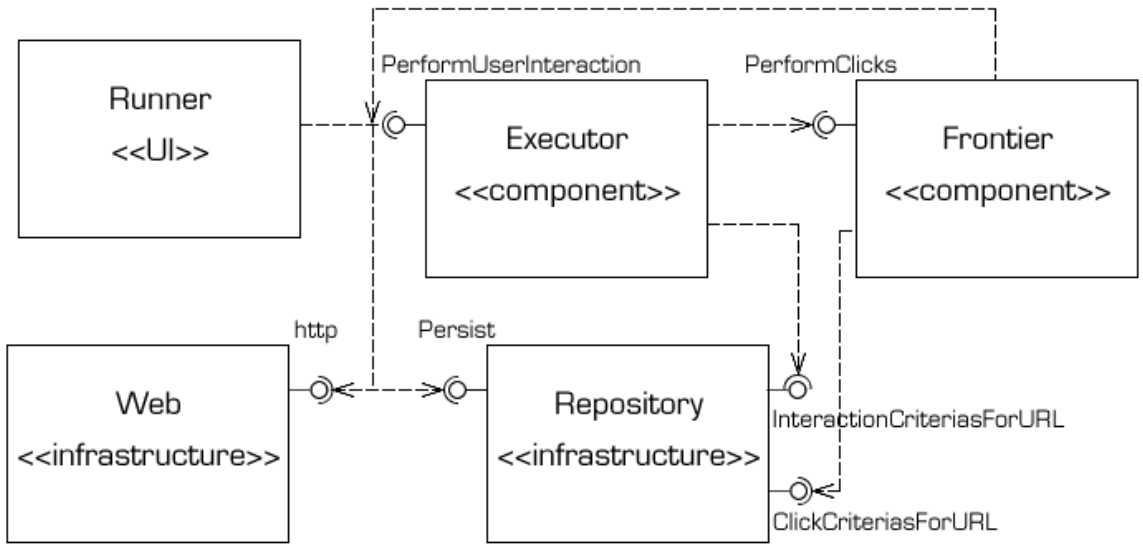
- Where was the evidence stored?
- Who had obtained the evidence?
- What has been done to the evidence?

The Deep Web Crawler

Challenges:

- the use of AJAX technology in web pages
- the simulation of user interaction with HTML elements (e.g. forms)
- the use of Captchas in web pages

The Architecture



Usage

- investigative
  - collecting bitcoin addresses (clear and dark net)
  - saving evidence files from the web (also via Tor)
- non-investigative
  - collecting event data
  - ... person data (incl. friendship)
  - ... news
  - ... locations
  - ... bitcoin addresses and names
  - many others still done