Digital Network Information & Traces Extraction (Dignité)

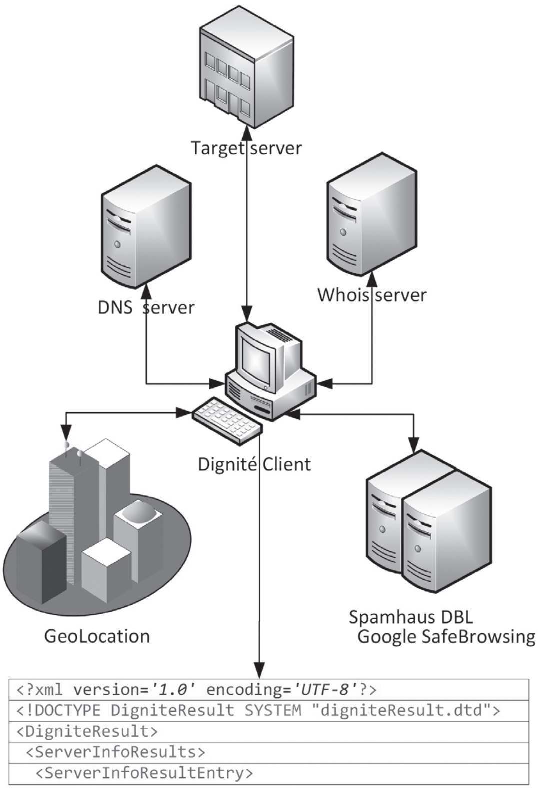
Subject: IT Security
Thesis advisors: Claude Fuhrer, Dr. Olivier Biberstein
Expert: Dr. Joachim Wolfgang Kaltz (PostFinance AG)

Cybercrime often involves websites, hosted by web servers all-over the world. Those servers contain pieces of information, often called digital traces, which reveal insights on the type of illegal activity or on the persons behind. The goal of our project is to develop an extensible API and library to acquire automatically a large range of digital traces. For a convenient demonstration of our library, we created a simple graphical application using Scala Swing components.

Digital Traces

To retrieve traces we need Internet access, which allows us to get website information and query other external tools, like the Whois database or a geo-location service. But not all traces can be fetched directly. To handle this, we developed our own tools using already developed libraries and our algorithms using the Scala programming language.

To gather information about the target website we use a selection of traces. Those traces allow to retrieve server version and geographical location of the physical server, the chain of SSL certificates or the DNS and Whois information stored in the public databases. Furthermore we examine the content of a website and its different pages to find interesting information like e-mail addresses or phone numbers as well as HTML metadata like the author or character encoding. To include Social Network information, we also look for Facebook or Twitter code snippets and account IDs from traffic analyzing services, for example Google Analytics. As a further source of information we contact established blacklist providers like Google Safe-Browsing or Spamhaus.



```
<?xml version='1.0' encoding='UTF-8'?>
<!DOCTYPE DigniteResult SYSTEM "digniteResult.dtd">
<DigniteResult>
 <ServerInfoResults>
  <ServerInfoResultEntry>
```

The Dignité Client uses different services to gather information

Testing and Interoperability

Testing the developed library was an important task to verify the correct functioning of our work. To avoid network queries for every test run we developed mock objects to return typical service responses which are parsed by our trace strategies.

Although the library and API is developed in Scala, it can be used with Scala or Java applications and using the XML export functionality the results can be analysed with other programming languages. This allows to use the retrieved information in Big Data projects establishing connections between different websites.

Legal aspects

Being Swiss citizens and operating out of Switzerland we have to respect the current Swiss law. This constrains us to using only publicly accessible information that has not been protected to restrict access. We are also not allowed to conduct any invasive operation or to take offensive action to disrupt the operation of the target website or any intermediary device. Nevertheless we are able to extract a lot of information.

New possibilities

With the finished product, a user can now easily collect publicly available information about a given target website using only one tool instead of many different command-line or graphical utilities.

Furthermore all information retrieved is stored in a structured format using well-established XML standards.

Thomas Marcel Ender
thomas.ender@the-online.ch

Patrick Vananti
patrick@vananti.ch