Author: rupp,scd
Project:ID Scanner
Date: 04/28/13

ID SCANNER SCREENS/ Specifications

Table of Contents

Illustration Index

1. Introduction

This document describes the screens and functional specifications for the Identity Scanner “ID Scanner”.

ID Scanner is a small utility that will scan and parse files on a PC to display softwares or files that holds sensitive informations (password,credit cards numbers etc...)

2. functionalities

2.1. Data Detection

ID scanner is multiplatform software and embeds a regular expression engine that allows to catch password and special data such as credit card data.

The regular expression engine will be provided with a set of rules to catch these personal data.

The regular expression engine could be DEELx or Qtregexp (http://en.wikipedia.org/wiki/Comparison_of_regular_expression_engines)

The rules will be hardcoded in the application (no rule xml definition file) and their final form will depend on the reexp enine selected.

It must be noted that there is no guarantee that the detected data will be a relevant sensitive data since the detection is purely based regular expression and as such it is heuristic.

The basic following rules should be applied:

a) Credit cards rules

- any string of the shape $$$$-$$$$-$$$$-$$$$ where $ is a number between 0 and 9

  – any 12 digits numbers with or without spaces, '-' between the 4 groups of 4 digits (for example '1234-567800000000' or '1234-5678-0000 0000'

  – any consecutive group of 12 digits that matches with PAN format / ISO/IEC 7812 (http://en.wikipedia.org/wiki/Bank_card_number)

b) Password rules

There is no *de facto* rules to identify a password in a stream of data.

However, if we find strings such as :

“password: XXXX”

“password= XXX”

“password is XXX”

we will consider the right string as a potential password.

Similar rules will be applied to detect paswords

c) Social security numbers rules

Rules created from: https://en.wikipedia.org/wiki/Social_Security_number

- a 9 digit number $$$$$$$$$ that may be sperated with spaces, “-” or other characters (US SSN)

d) Bank Accounts rules

rules created from :

- IBAN (http://en.wikipedia.org/wiki/International_Bank_Account_Number)

```
[a-zA-Z]{2}[0-9]{2}[a-zA-Z0-9]{4}[0-9]{7}([a-zA-Z0-9]?){0,16}
```

- US bank accounts (http://stackoverflow.com/questions/1540285/united-states-banking-institution-account-number-regular-expression)
  - routing/transit numbers
  - ACH transactions

e) Serial numbers

Rules to identify software (or other products) serial numbers

- groups of letters and numbers separated by dashes “-”
- similar rules

f) General Rules

In general if we detect a long non-english String inside a group of english language strings , and the entropy of the String is sufficiently important, we will process this string as “sensitive data, unknown type” (if we cannot classify it in the others categories above)

English language can be detected using Kappa tests, and the same test can ensure the entropy of a long String (for example “gdsgeyuej72y62hdhsbn” will be detected as long non-english String with important entropy)

2.2. Database

The engine will have a small db (liteSQL) that will store rules about the browsers, where they store information and how these information can be found.

The following tables will be defined:

- *Regexp_RULES: will store all the regular expression rules for senstive data detection*
- *Browser_Data: will store the pathes the browsers are using to store their data*
- *Email_Data: will store the pathes the emails clients are using to store emails*
- *Scans: will store the result of each scan*
- *Conf: will store the configuration*

2.3. Browser Scanner

The scan is done on the folders and files containing history, navigation data, autocompletion data etc … Some javascripts or browser -specific API calls will also be used

2.4. File Scanner

The engine will be able to scan entire folders and parse the content of each files to check for personal data.

2.5. Email Scanner

Also it will be able to check emails folders (knowing the place the emails clients store their mails) and also the registry.

2.6. Integration With PasswordMaster

A scan will contain data that password master is able to interpret (browser specific data etc...). There will be a button “Launch Password Master” available after a Scan which will transmit the scan data to PasswordMaster.

The way the data will be transferred could be through a temp file for example.

3. Screens

3.1. Form Overview

The software is a simple windows form interface that displays scan results:

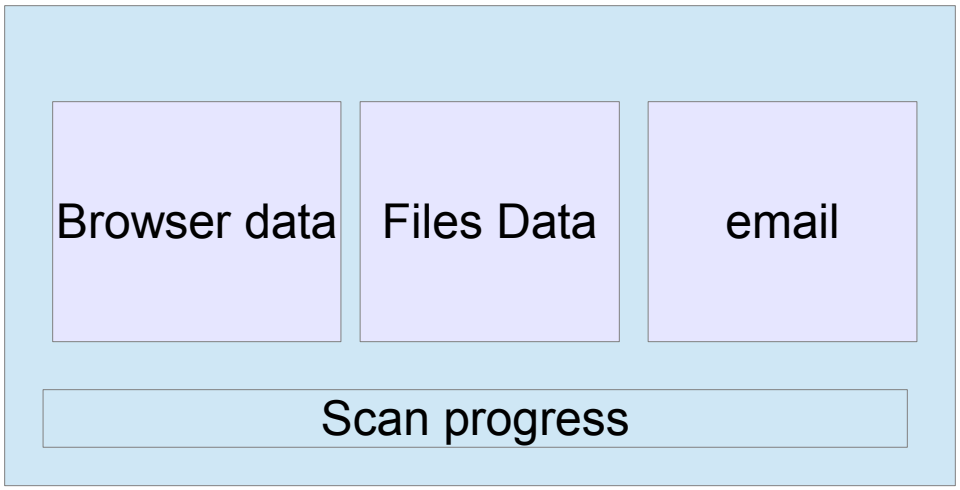

Illustration 1: Form Overview

The scan results are displayed in one of the three columns:

- ***browser data***
- ***files***
- ***email***
- ***registry***

Each of the columns have scrolling bars.

Additionally there are progress bars on the bottom to indicate the status of a current identity scan

a) Browser data:

- Show the registered/found internet browsers and offer the ability to specify a path to a browser
- will scan the temporary folders, permanent folders of each browser and scan the content to retrieve identity data
- Can scan web database (webDB), most common web storage (cookies, InnerDb, Web Sqlite, HTML5 storage etc...)
- Can detect evercookies, zombie cookies, etc...

b) Files

- Offer the ability to specify paths (drives/folders/etc...) to scan
- Can scan content of zip files
- Can scan registry

c) Email

- Display registered/found email clients
- Offer the ability to specify custom paths to emails clients
- will scan email folders to retrieve identity data

d) Settings

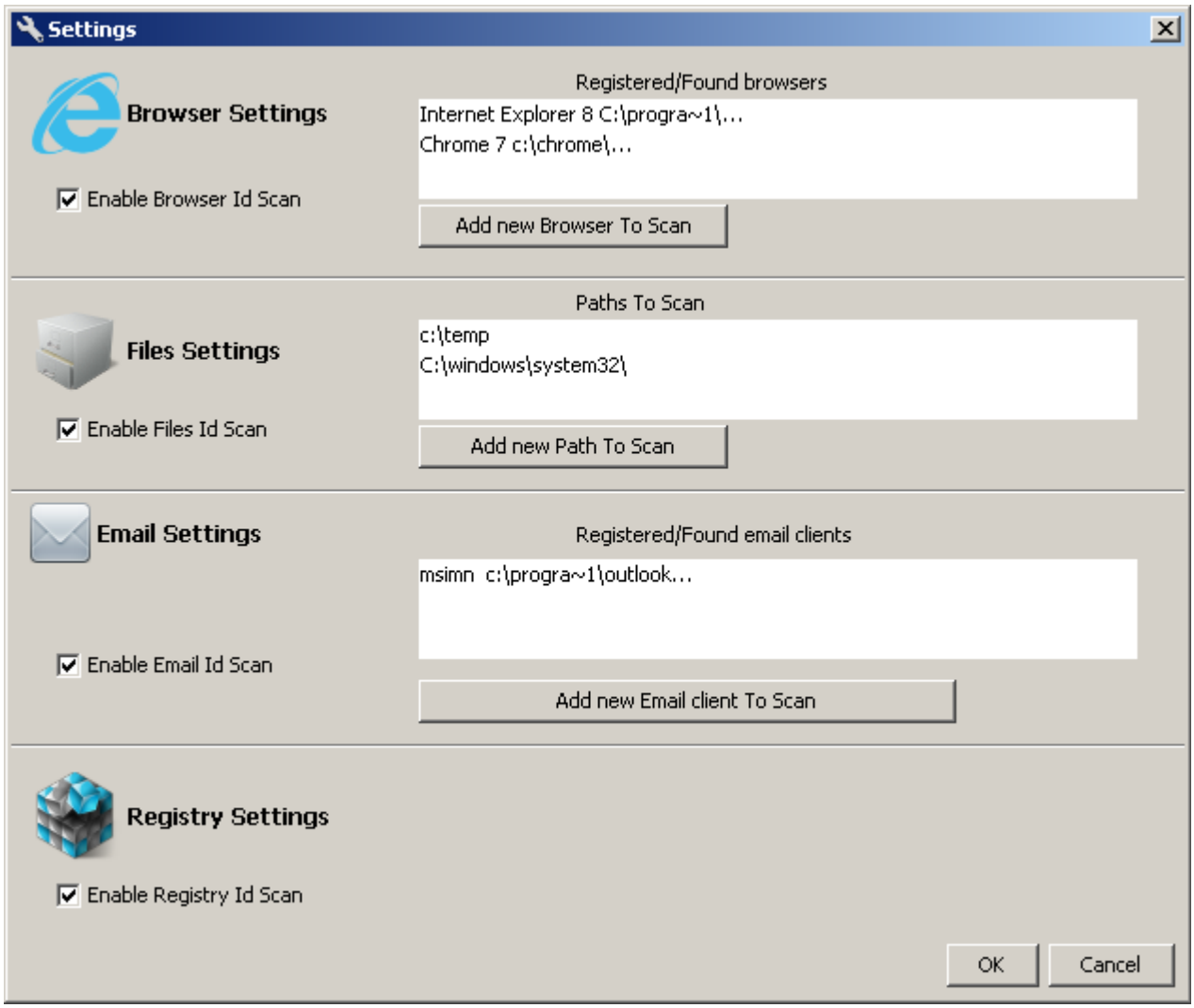

Illustration 2: Settings for Id Scan

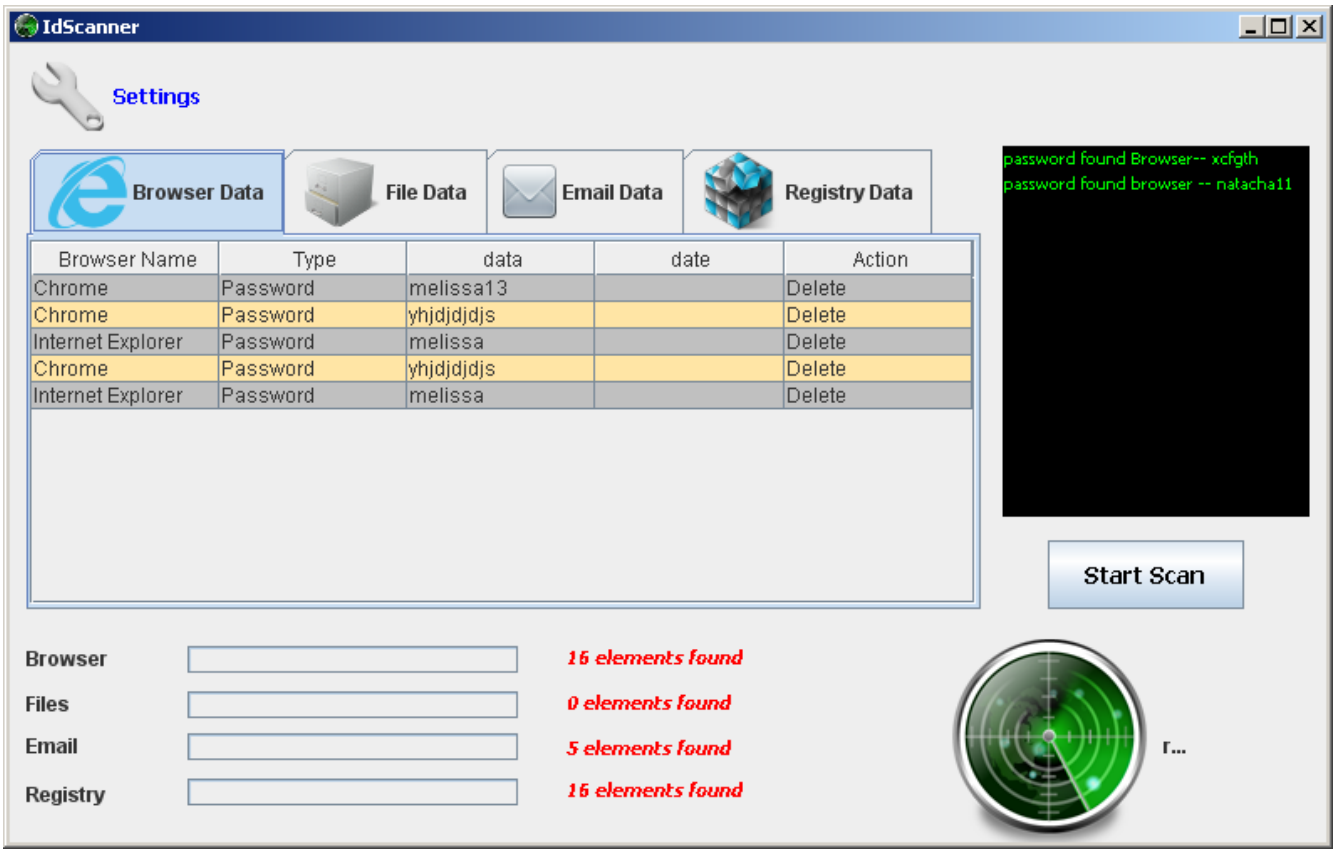

Illustration 3: IdScanner GUI

3.2. Detailed info

Any element in the column is clickable and when clicked, a small windows appears and displays detailed information about the elements.

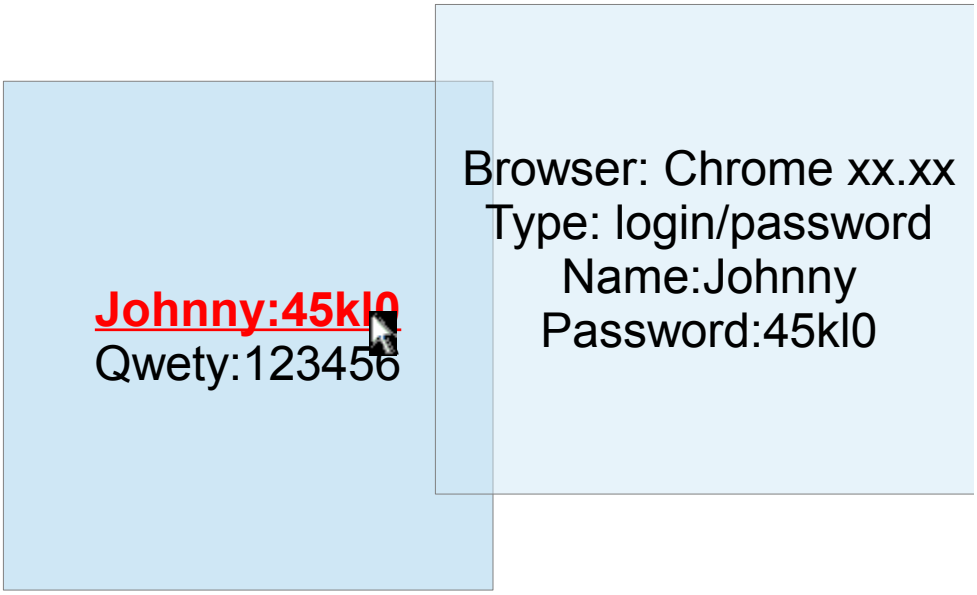

Illustration 4: Detailed Info

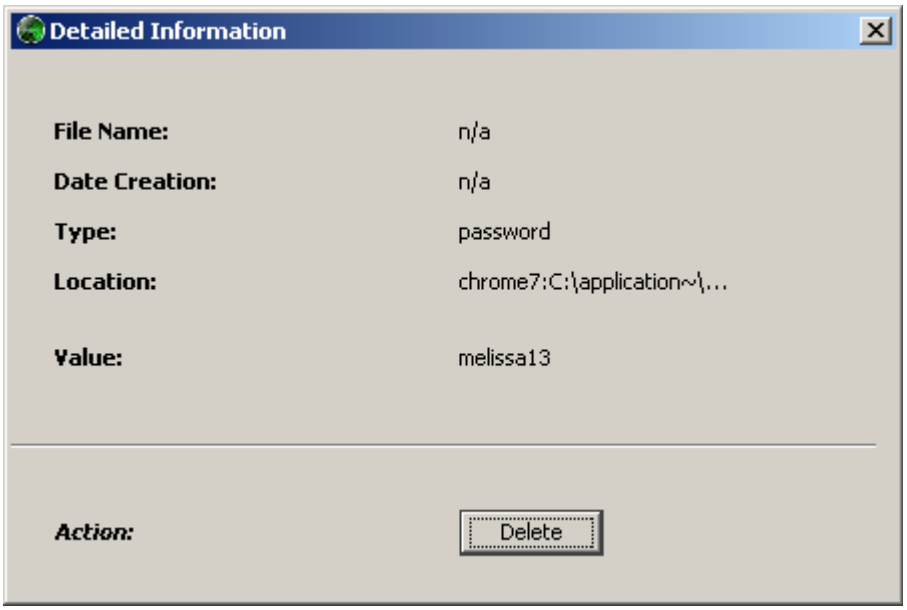

Illustration 5: Detailed Info when clicked on an element