Mag. iur. Dr. techn. Michael Sonntag

Recovering web-browsing history

Institute for Information Processing and
Microprocessor Technology (FIM)
Johannes Kepler University Linz, Austria

E-Mail: sonntag@fim.uni-linz.ac.at
http://www.fim.uni-linz.ac.at/staff/sonntag.htm

Source files

- Source files: User profile of JDoe (extract only)
  - → JDoe\Favorites
    - » Bookmarks
  - → JDoe\Cookies
    - » Cookie directory
  - → JDoe\Local Settings\History
    - » Visited URLs
  - → JDoe\Local Settings\Temporary Internet Files
    - » Cache directory
  - → HKCU_Software_Microsoft_Internet Explorer.reg
    - » Registry (HKCU or HKU\<User-SID>; IE part only)
- Requirements:
  - → Software:
    - » Galleta
    - » Pasco
  - → Registry editor

Attention!

- As all web sites, the ones visited here change frequently
  - → This scenario was created in May 2008
- Cookies, webpages, URL format etc. might have changed significantly since then!
  - → But this is quite common in computer forensics too …
- Therefore:
  - → Compare with the current version by performing the same actions on a “clean” computer
  - → But don’t expect everything to be/work the same!
    - » This is **NOT** a hint for manipulations!

Favorites

- Your tasks:
  - → Investigate the bookmarks of the suspect
    - » Which sites did he visit?
    - » Did he add (which?) own bookmarks?
  - → Produce a list of all bookmarks
- Source: The JDoe\Favorites folder

- Please note: The file date/time cannot be original any more because it had to be copied for this course!

Favorites (1)

- The suspect did not add a single bookmark; everything to be found here is automatically created upon first start of IE!
  - → Desktop.ini: Windows file; no specific information
  - → These bookmarks all point to the "correct" location as well
- List:
  - → MSN.com.url: http://www.microsoft.com/isapi/redir.dll?prd=ie&pver=6&ar=IStart
  - → Radio Station Guide.url: http://www.microsoft.com/isapi/redir.dll?prd=windows&sbp=mediaplayer&plcid=&pver=6.1&os=&over=&olcid=&clcid=&ar=Media&sba=RadioBar&o1=&o2=&o3=
  - → RealOne Home Page.url: http://www.real.com/
  - → "Corel im Web" subfolder
    - » Buy.corel.com .url: http://product.corel.com/query.htm?lang=de&box=Coreldraw_10&topic=store&ver=10&src=bookmark
    - » Corel.com .url: http://product.corel.com/query.htm?lang=de&box=Coreldraw_10&topic=home&ver=10&src=bookmark
    - » CorelCity.com.url: http://product.corel.com/query.htm?lang=de&box=Coreldraw_10&topic=corelcity&ver=10&src=bookmark

Favorites (2)

- List (cont.):
  - → "Links" subfolder
    - » Customize Links.url: http://www.microsoft.com/isapi/redir.dll?prd=ie&pver=6&ar=CLinks
    - » Free Hotmail.url: http://www.microsoft.com/isapi/redir.dll?prd=ie&ar=hotmail
    - » RealOne Player.url: http://www.real.com/
    - » Windows Marketplace.url: http://go.microsoft.com/fwlink/?LinkId=30857&clcid=0x409
    - » Windows Media.url: http://www.microsoft.com/isapi/redir.dll?prd=ie&ar=windowsmedia
    - » Windows.url: http://www.microsoft.com/isapi/redir.dll?prd=ie&ar=windows

Favorites (2)

- Take a closer look on the Bookmark “MSN.com”
  - → Open it in a text editor: What’s additional in there?
  - → There is a “Modified” value!
    - » What could it be? A timestamp!
  - → Try converting it
    - » Result (Windows 64 Bit – Little Endian):
      Mo, 19 May 2008 12:41:14 UTC
      - – Note: We don’t know about the timezone (similar to FAT)
      - – This is also the file date “last change”
- Investigate “RealOne Home Page”
  - → Additional icon information → Not very interesting

Cookies

- Your tasks:
  - → Investigate the cookies of the suspect
    - » Which sites did he visit?
    - » Can we say which were visited intentionally?
      - – Advertisement banners!
  - → Produce a list of all sites with their visit time
- Sources:
  - → The JDoe\Cookies folder
  - → The index.dat file within
- Software:
  - → Use "Pasco" for the index.dat file and import the results into a spreadsheet for better investigation (“pasco –t ; index.dat”)
  - → Use "Galleta" for individual cookie files
- Please note: The file date/time cannot be original any more because it had to be copied for this course!

Cookies

- Note: Some cookies are from the "initial" folder, which are copied when a new account is created
  - → These can be distinguished by the date (2002), and the URL (administrator@...)
- JDoe visited the following sites
  - » And a few others, these are presumable ads
  - → MSN
  - → Beate Uhse
  - → Amazon
  - → ORF
  - → GMX
  - → EBay
  - → Google
  - → 66.29.38.208 → "Net Access Corporation“ (Now defunct)

Cookies

- All visits took place between 14:54:50 and 15:17:13
  - → A very brief timeslot!
- Only few cookie have a (possibly) interesting content:
  - → jdoe@shop.beate-uhse[1].txt
  - → Agecheck
    18
    shop.beate-uhse.at/
    1024
    2960605440
    29933360
    2316118208
    29931952
    *
  - → There seems to have been an agecheck which was answered with "18" or "18 and above"
- About "intentionality" we cannot say anything really
- See Cookies.csv and Cookies.xls for the complete list!

Cookies

- Amazon cookies:
  - → session-id-time: Unix numeric timestamp?
    - » Monday, 26 May 2008 00:00:00 +0200
    - » File timestamp: 19.05.2008 15:10 (= valid for 1 week)
  - → The rest are internal values
    - » session-id: Unique number identifying the session/user
    - » session-token: Probably a good random value against CSRF
    - » ubid-acbde: Unknown; probably also session related
      - – Amazon.de; amazon.com (ubid-main) and amazon.co.uk (ubid-acbuk) have different names! Perhaps a hint to the server “section” serving these requests
- GMX cookies (4 files!):
  - → GUD: Probably session identifier
  - → POPUPCHECK: Unix TS: Tue, 20 May 2008 15:15:46 +0200 (=1 day)
  - → Visits: 3 (probably the visit count)

Cookies

- Galleta output:
  - → Tab-delimited file

Cookie File: jdoe@shop.beate-uhse[1].txt

SITE	VARIABLE	VALUE	CREATION TIME	EXPIRE TIME	FLAGS
shop.beate-uhse.at/	agecheck	18	05/19/2008 15:02:06	05/26/2008 15:02:02	1024
shop.beate-uhse.at/	ANZ_ARTIKEL	7	05/19/2008 15:03:14	05/20/2008 15:03:11	1024

Visited URLs

- Your tasks:
  - → Investigate the URLs the suspect visited
    - » Which sites did he visit?
    - » What can we find out about the actions on these sites?
  - → Produce a list of all URLs with their visit time
- Sources:
  - → The JDoe\Local Settings\History folder
  - → The History.IE5\index.dat file within
- Software: Use "Pasco" for this and import the result to a spreadsheet for better investigation
- Please note: The file date/time cannot be original any more because it had to be copied for this course!

Visited URLs

- Again some remnants from the administrator are present
  - → Ignored in further investigation
- Beate Uhse: Shop locations were searched ("filiale_ort.php")
  - → Shopping for bed sheets
  - → Basket was shown and edited
- Amazon.com: Looking for memory cards
  - → Signed in as a new user
  - → Wishlist accessed → Item was added to it
- ORF visited
  - → Individual stories could be checked!
  - → They occur at various times interspersed
    - » Probably in a separate window and browsed concurrently

Visited URLs

- GMX was accessed
  - → First login failed
  - → We have a customer number: 47680471!
  - → We have a date/time of last login: 2008-05-19 14:45:10!
  - → The suspect checked the inbox and the spam folder
  - → He viewed a message in the spam folder
  - → He replied to a message and sent it
  - → He changed the password
  - → Finally he logged off
  - → He logged in again
    - » Checking the new password works…

Visited URLs

- Google was used for searching
  - → Search terms: "how to hide my data from the police"
  - → Clicked presumably on some result links
    - » Research.ibm.com, Windowsitpro → Check content yourself!
  - → Searched again: "hiding data"
  - → Searched again: "free data hiding software"
  - → Clicked on a result link
    - » www.freedownloadscenter.com
    - » www.e-cronies.com
  - → Actual download of such software is **not** present!

Cache

- Your tasks:
  - → Investigate the cache for the E-Mail messages of the suspect
    - » Can we find out some content?
  - → Investigate the shopping behaviour (Beate Uhse, Amazon)
- Sources:
  - → The JDoe\Local Settings\Temporary Internet Files folder
  - → The Content.IE5\index.dat file within
- Software: Use "Pasco" for this and import the result to a spreadsheet for better investigation
- Please note: The file dates/times cannot be original any more because it had to be copied for this course!

Cache: GMX

- GMX information starts in the solution file in line 1003
  - → We are out of luck, there are lots of images, but no page after the login page
    - » Investigate the reason by logging in on GMX
    - » Every "personal" page has the following headers:
      - – <meta http-equiv="Cache-Control" content="no-store, no-cache" />
      - – <meta http-equiv="Pragma" content="no-cache" />
      - – <meta http-equiv="Expires" content="-1" />
    - » This means, these pages will not be cached or stored in disk!
  - → We could not find anything about the actual content!

Cache: Beate Uhse

- Beate Uhse starts in line 174
  - → First page → Age verification!
    - » HCU5LTVF\beate-uhse[1].htm
  - → Branch office search (2 times)
    - » D6WZ0ELD\filiale_ort[1].htm
    - » D6WZ0ELD\filiale_ort[2].htm
  - → Adding to basket
    - » 6XPZPFS4\index[1].htm
  - → Forms to fill in customer data
    - » 6XPZPFS4\index[2].htm
    - » R2KRGKYU\index[1].htm
  - → Complete basket
    - » HCU5LTVF\index[2].htm

Cache: Amazon

- Amazon starts at line 345
  - » Actually we are at Amazon.de, redirected from Amazon.at
- → Memory subcategory
  - » 6XPZPFS4\ref=amb_link_42071665_3[1].htm
- → Restricted to only the cheap ones (10-15 €)
  - » D6WZ0ELD\ref=sr_nr_p_36_2[1].htm
- → Individual product viewed: MiniSD card
  - » R2KRGKYU\ref=sr_1_5[1].htm
- → New registration:
  - » 6XPZPFS4\sign-in[1].htm
- → The login page cannot be seen → only garbage?
  - » R2KRGKYU\select[1].html
    - – Investigate headers: "Content-Encoding: gzip
    - – Copy it somewhere else, add ".gz" extension, decompress
    - – Now it can be viewed **(contains an E-Mail address → GMX!)**
- → Added to wishlist (see at the right)!
  - » HCU5LTVF\new[1].htm

Cache: IBM

- IBM journal article access: Line 1474
  - → Try to find it online: Is it still available?
    - » Note: We do know the URL!
    - » But: Moved to IEEE Explore and new URL unknown (or: journal issue, … to search for it!)
  - → But: Take a look at the cache!
    - » Subdirectory “HCU5LTVF”, File “bender[1].htm”
  - → Now we ca (still) find out, what the user was interested in!

Typed URLs

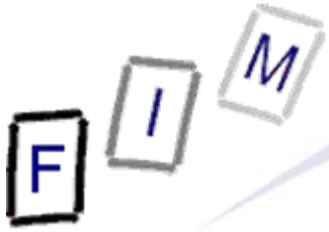
- Your tasks:
  - → Check what URLs were entered manually
    - » These should match the sites visited, as we have found no bookmarks for them!
    - » These sites were obviously visited intentionally
      - – No pop-ups, banners etc.!
- Sources:
  - → HKCU_Software_Microsoft_Internet Explorer.reg
- Software: Use a text editor
  - → Import to registry is not that ideal, as it is incomplete and would be added locally
    - » Would work better if it was a complete hive; could be "mounted"
- Please note: This is an export of a subtree. It only contains the data, but not associated information, like the last access date/time of keys!

Typed URLs
Text editor

- Open the .reg file with a text editor
  - → Unicode → Can be hard to read, depending on editor!
    - » Try Wordpad!
- Typed URLs can be found in the key "HKEY_CURRENT_USER\Software\Microsoft\Internet Explorer\TypedURLs"
- Here we have only five URLs (reverse order of visit):
  - → http://www.google.at/ Last typed URL
  - → http://www.gmx.at/
  - → http://www.orf.at/
  - → http://www.amazon.at/
  - → http://www.beate-uhse.at/ First typed URL

Conclusions

- We can find out quite a lot about the suspect, especially through the cache
  - → What he shopped for, his interests (wishlist)
  - → His E-Mail address used at Amazon for registration
    - » Can be found again on GMX
- The registry provides information on explicit user actions
  - → Allows removing the "automatic" displays → Ads
- Visited URLs: What the suspect searched for
  - → Google search URLs
- Cookies did not provide a lot of useful information
  - → Some IDs might help in combination with data from the websites, but these are not accessible for us here!



Questions?

Thank you for your attention!