Facilitating Forensic Examinations of Multi-User Computer Environments through Session-to-Session Analysis of Internet History

(or, Who was sat at the Keyboard)

David W. Gresty, Diane Gan, George Loukas and Constantinos Ierotheou

C-SAFE Centre, Dept. of Computing and Information Systems University of Greenwich
Overview

• Aims of our research
• Sessions (two types investigated)
• Patterns
• Results
• Future work (and a sneak peek at our current work)
Aims of our research

• A specific problem we are investigating. In a possible multi user scenario, can we:
  • Identify patterns of habitual usage?
  • Can we identify the actual user that belongs to the patterns?
  • Can we display the above in a simple, forensically useful way?
Sessions

• A typical Internet history list of host
• Could come from a device or based upon the Internet Connection Record.
  • Fixed length session
  • Variable length session
Sessions (2)

• The ‘dog leg’ gives a god indicator where to break up the sessions for the variable-length sessions.

• With the fixed-length sessions there will be more sessions created.
Sessions (3)

• Display the sessions with a single box per host – binary at the moment, experimenting with greyscale.

• Ordered here by frequency L to R, could be ordered by Global Popularity etc. (see sneak peek)

• Visually this is quite striking and meets, at least in part, one of our initial goals.
Patterns

• Compare sessions-to-session pairwise.
• Set a threshold value.
• With a threshold of 0.5 in this example we see two patterns:
  • s4 and s5 perfectly match
  • S1, s2 and s3 which contains (C1 and C3) and (C4 or C5)
• In reality depending upon the number of components we are seeing thresholds above 0.75 being interesting.
Results

• Variable-length sessions work well for person identification.
• Thresholds of 600 seconds (10 mins) is reasonable for the Internet history we have experimented with.
• With fixed-length we see 20 mins as a reasonable amount of time to see habitual behaviour (or 2 minutes)
Results (2)

• We can map patterns onto a timeline and start to use it for reasoning within the investigation.

• There are gaps in the timeline. Not as much repetitive behaviour as we initially though, but we have now examined the ‘normality’ of the data.

• We tried removing ‘spoiler’ components and this did remove false positives, however it did this by removing a user who was incredibly repetitive.

- Identify patterns of habitual usage? [v]
- Can we identify the actual user that belongs to the patterns? [x]
- Can we display the above in a simple, forensically useful way? [v]
Future work

• A reasonable method of sessions size selection and a reasonable threshold selection will identify repetitive patterns. There is some error at this stage (90% correct identification).

• There is no examination of the components in this paper:
  • Examined the ‘spoilers’.
  • We have started looking at the ‘type’ of the components.
A sneak peek at our current work

- Normality of the data based upon the ‘Global Popularity’ of the component.
- Love ‘mrklingon.org’ but are surprised that it is ranked 3,637,715?
- The difference between local and global ranking indicates the idiosyncratic behaviours of the user.
- There is more going on inside a session than a pairwise session comparison can show.
Any Questions?

I'm from the government. We've been monitoring your internet activity.

Half of my department went blind and the other half needs counseling.

Sounds like not my problem.

We'd like to weaponize you.