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


JUST
JARED

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

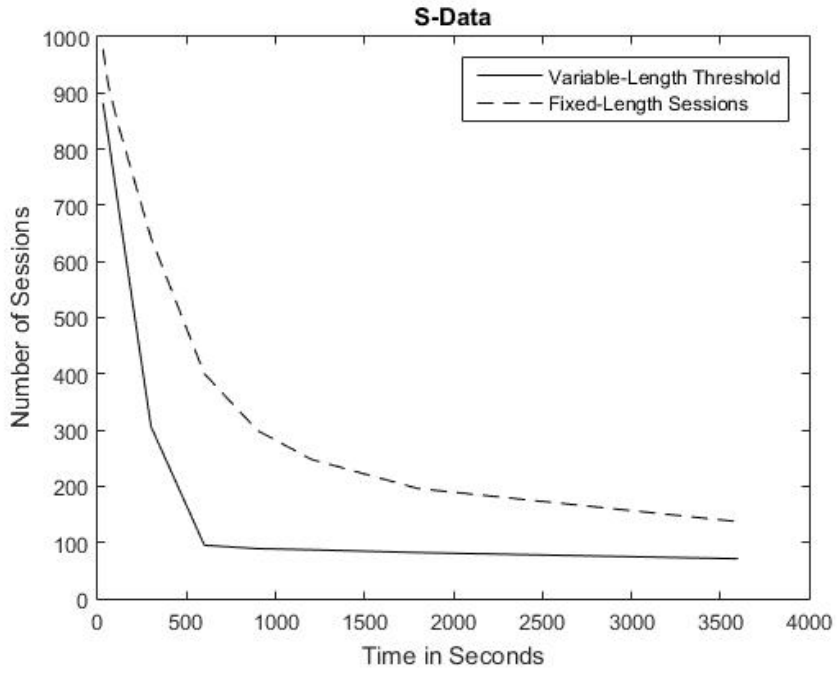
- A typical Internet history list of host
- Could come from a device or based upon the Internet Connection Record.
  - Fixed length session
  - Variable length session

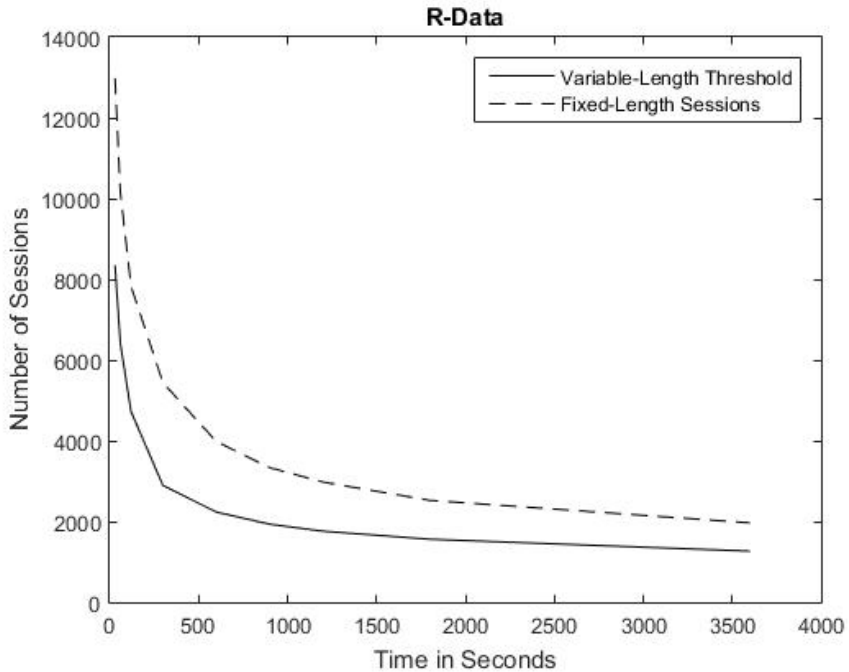
Time	Component Name
01/07/2009 22:50:37	wikipedia
01/07/2009 22:58:19	dell
01/07/2009 23:02:15	openoffice
01/07/2009 23:06:07	openoffice
01/07/2009 23:13:51	openoffice
01/07/2009 23:18:19	openoffice
01/07/2009 23:25:53	openoffice
01/07/2009 23:32:48	cccleaner
01/07/2009 23:34:53	gamblersanonymous
01/07/2009 23:35:30	gamblersanonymous
01/07/2009 23:37:44	gamblersanonymous
01/07/2009 23:43:50	gamblersanonymous
01/07/2009 23:44:14	scoresandodds
01/07/2009 23:44:18	scoresandodds
01/07/2009 23:45:07	scoresandodds
01/07/2009 23:45:11	scoresandodds
01/07/2009 23:52:25	scoresandodds
01/07/2009 23:52:29	scoresandodds
01/07/2009 23:54:31	scoresandodds
01/07/2009 23:56:14	mrklingon
02/07/2009 00:00:15	syfy
02/07/2009 00:06:02	syfy
02/07/2009 00:08:10	syfy

UNIVERSITY of GREENWICH

Sessions (2)

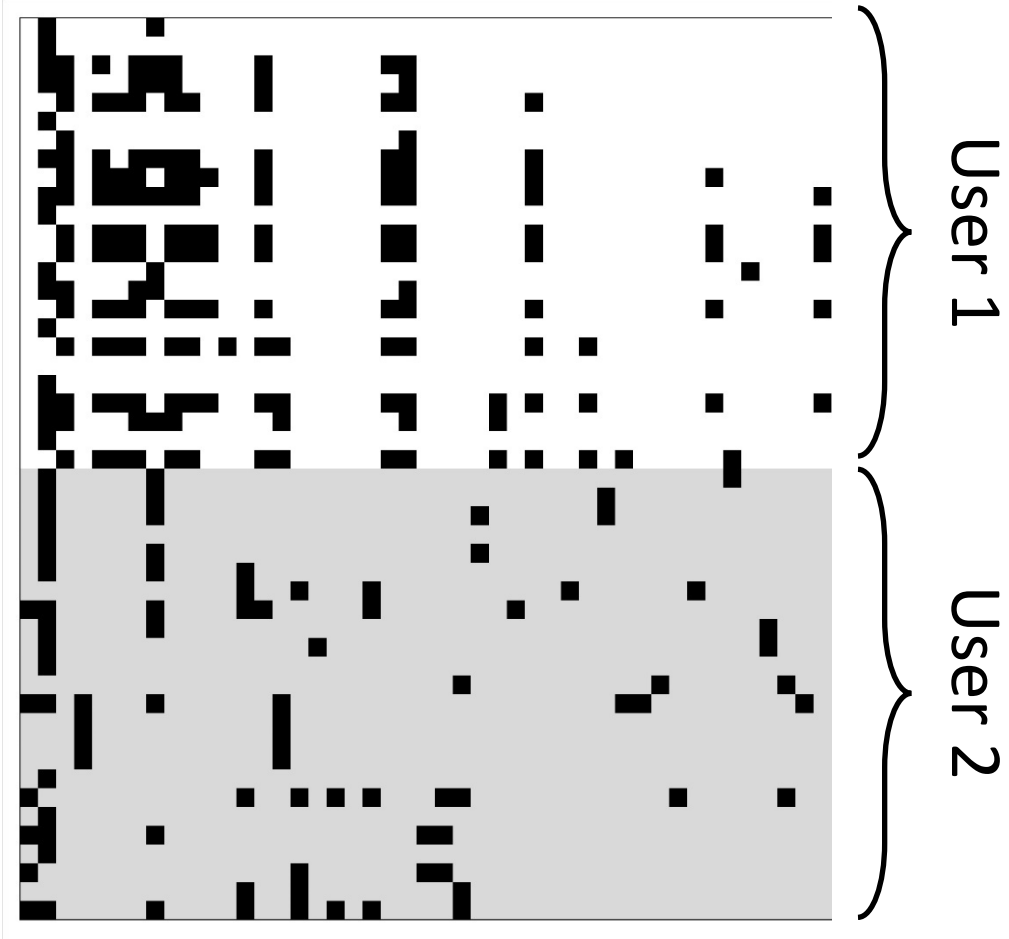
- The ‘dog leg’ gives a god indicator where to break up the sessions for the variable-length sessions.
- With the fixed-length sessions there will be more sessions created.





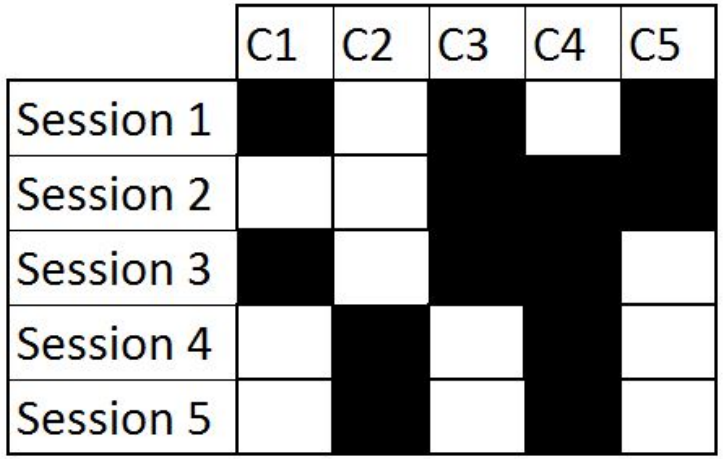
Sessions (3)

- 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

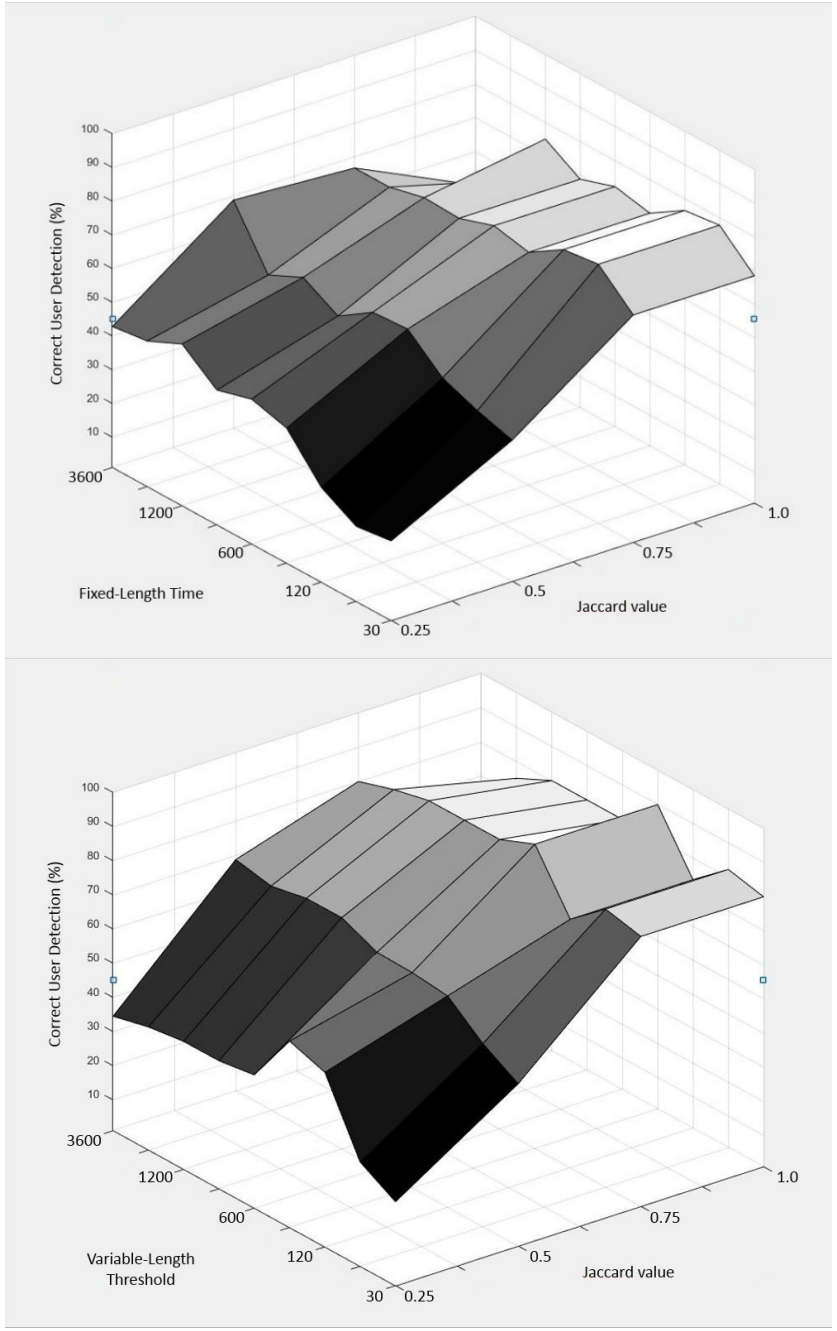
- Compare sessions-to-session pairwise.
- Set a threshold value.
- With a threshold of 0.5 in this example we see two patterns:
  - s4 and s5 perfectly match
  - S1, s2 and s3 which contains (C1 and C3) and (C4 or C5)
- In reality depending upon the number of components we are seeing thresholds above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



	s1	s2	s3	s4	s5
s1	1	0.5	0.5	0	0
s2	0.5	1	0.5	0.25	0.25
s3	0.5	0.5	1	0.25	0.25
s4	0	0.25	0.25	1	1
s5	0	0.25	0.25	1	1

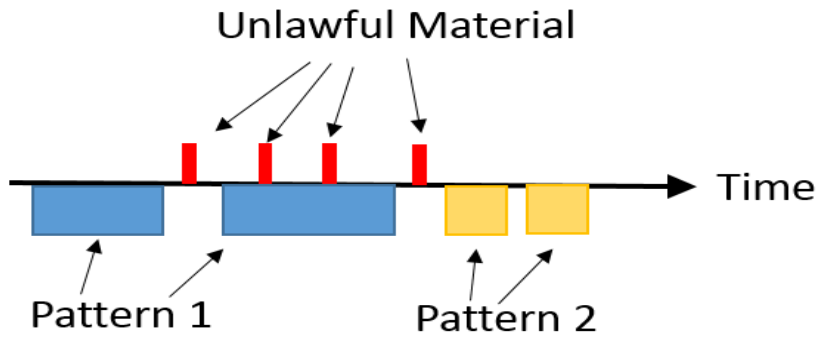
Results

- 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]

UNIVERSITY of GREENWICH

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.

UNIVERSITY of GREENWICH

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.

UNIVERSITY of GREENWICH

Any Questions?

UNIVERSITY of GREENWICH