Betrayed by the keyboard
How what you type can give you away

Matt Wixey
Research Lead, PwC UK Cyber Security
Disclaimer

• This content is presented for educational purposes only
• What this presentation isn’t...
Introduction

Matt Wixey

- Research Lead for the Cyber Security BU
- Work on the Ethical Hacking team
- PhD student at UCL
- Previously worked in LEA doing technical R&D
Why this talk?

- Based on some research I did at UCL
- Interest in side-channel attacks
- Humans have side-channels too
- Previous work on forensic linguistics
  - First degree = English Literature and Language
Agenda

- What is attribution?
- Problems
- Case Linkage Analysis
- Experimentation
- Results
- Implications
- Summary
What is attribution?

- Why would we want to do it?
- Benefits
- Types
- Approaches
What is attribution?

- Identifying an attacker’s location?
- Identify the country or organisation behind an attack?
  - Rid and Buchanan, 2014
- “Determining who is responsible for a hostile cyber act”?
  - Mejia, 2014
- “We must find a person, not a machine”
  - Clark and Landau, 2011
Benefits of attribution

- Deterring future attacks
- Improving defences
- Interrupting and disrupting attacks (Hunker et al, 2008)
- Does attribution actually lead to deterrence? (Guitton, 2012)
- Regardless, attribution is a desirable outcome (depending on which side you’re on!)
Types of attribution

• Hutchins et al, 2011:

Atomic
Computed
Behavioural
Problems with attribution

• Hiding atomic IOCs
• Issues with computed IOCs
• Lack of tangible benefits from behavioural IOCs
Hiding atomic IOCs

- These are the most effective identifiers
- Easy to resolve (usually)
- But also easiest to spoof/anonymise/obfuscate
Issues with computed IOCs

• Changes to malware make it harder
• Other methods:
  • Correlating activity with office hours in timezones (Rid & Buchanan, 2014; CloudHopper, 2017)
  • Deanonymising developers through artefacts (Caliskan et al, 2015)
  • Similar malware capabilities (Moran & Bennett, 2013; Symantec, 2011)
  • Distinguishing humans vs bots (Filippoupolitis et al, 2014)
Mo methods, mo problems

- Less focused on individuals
- Sufficient if aim is to identify a state/sponsor
  - Challenge is then legal/procedural
Behavioural profiling

- Less attribution
- More trying to understand *who* hacks, and *why*
  - Motivation, skills, attack behaviours (Landreth, 1985)
  - Attitudes and culture (Chiesa et al, 2008; Watters et al, 2012)
  - Psychological (Shaw et al, 1998)
Attack profiling

- Humans vs bots
  - Filippoupolitis et al, 2014: Skill, education, typing speed, mistakes, etc

- Skill level

- Attacker behaviour
  - Ramsbrock et al, 2007: Specific actions undertaken
The problem

- Profiling attackers is interesting
- Next logical step is *comparison*
  - To what extent is an attacker’s profile similar to another’s?
- Not really explored
Case Linkage Analysis

- The idea
- Discovering case linkage analysis
- Benefits of linking offences
- What case linkage analysis is (and isn’t)
- Methodology
- Example
- Exceptions
The idea

• I had an idea (rare occurrence - to be celebrated)
• Lurking in OSCP labs a few years ago
• Discussing attack techniques, commands, methodologies
  • Casual observation 1: everyone has their own way of doing things
  • Casual observation 2: this way of doing things rarely changes
Science!

- This seems obvious
- My first degree was English Lit
  - Could pretty much make it up as you went along
- Apparently, in science, you have to *prove* stuff
  - Can’t just write “this seems obvious”
  - Science is hard 🙃
Discovering case linkage analysis

• How could I empirically test this?
• Came across “Case Linkage Analysis”
• Methodology used in crime science literature
• Designed to link separate crimes to common offenders
• Based on behavioural aspects (Woodhams & Grant, 2006)
Benefits of linking offences

- Can attribute previously unsolved crimes
- Can investigate offences under one grouping – focused resources
- Useful evidentially
- Database of offences grows = better chance of success
- A minority of offenders commit the majority of crimes (?)
  - Not necessarily true of crime generally
  - But more accurate with specialist crimes
Benefits of linking offences

- Best method for linking: physical evidence (DNA, fingerprints, etc)
- Highly accurate, **but:**
  - May be absent or inconclusive (Grubin et al, 1997)
  - Does not really apply to cyber attacks
  - Closest approximation is forensic artefacts, but these are not always unique
  - Time-consuming and expensive (Craik and Patrick, 1994)
What case linkage analysis is

- Uses behavioural evidence
  - Things the offender does during the commission of an offence
- Classify granular crime behaviours into domains
- Create linked and unlinked pairs of offences
- Compare with behaviours in other offences
- Determine degree of similarity
What case linkage analysis isn’t

• It’s not offender profiling
• Offender profiling makes inferences about the offender
• Based on assumption of consistency between criminal and everyday behaviour (Canter, 2000)
  • Based on this behaviour, I infer that the perpetrator is a balding but charismatic researcher from the UK
What case linkage analysis isn’t

• CLA: statistical inferences about the similarity of 2 or more offences, based on common behaviours
  • Crime A, perpetrated by Matt “Charismatic But Balding” Wixey, has several features in common with Crime B
  • Therefore, Wixey may have also committed Crime B
Case linkage analysis in context

- Two key assumptions
  - **Behavioural consistency**
    - Offenders display similar offending behaviours across crimes
  - **Behavioural distinctiveness**
    - The way an offender commits crimes is characteristic of that offender
    - And distinguishable from the style of other offenders (Canter, 1995)
Case linkage analysis in context

• Both assumptions must be present
• Otherwise CLA is unlikely to be useful
• e.g. homicide: dumping a body in a remote location is consistent for many offenders
• But not distinctive
Case linkage analysis in context

- Individuals have stable, distinctive responses (Shoda et al, 1994)
- Cognitive-affective personality system (CAPS)
  - Mischel & Shoda, 1995; Mischel, 1999
  - System of goals, expectations, beliefs, plans, strategies, memories
- CAPS is consistent yet distinctive (Zayas et al, 2002)
Case linkage analysis in context

- Assumptions of stability/distinctiveness made in other fields
- Forensic linguistics
  - Word and sentence length; slang; typos; errors; syntax; idiolect; article frequency; syllable count; punctuation; hapax legomena; sentence length; stylistics
  - Language is socially acquired, continually – so may change
- Some biometrics
  - Typing speed; typos; typing habits
Case linkage analysis – does it work?

- Consensus: yes, in most cases
- Observed variance significantly smaller in linked crimes
  - Grubin et al, 1997; Mokros & Alison, 2002
- Significant evidence for cross-situational consistency
  - Both criminal and non-criminal behaviours (Tonkin et al, 2008)
Methodology

- Separate behaviours into domains
- Calculate similarity coefficient
- Input into logistic regression model
- Determine optimal combination of domains
- Receiver Operating Characteristic (ROC) curves
Methodology

• Lots of stats stuff
• I hate stats. I am bad at stats.
• Will try and explain this with a worked example
• None of that “left as an exercise for the reader” nonsense
Example

- Two burglaries, A and B
- We want to find out if the same offender did both
- Define a **dichotomous dependent variable**
  - This is a Y/N question, and we’re trying to ‘predict’ the answer
  - And find out what variables contribute more
  - “Are these two crimes linked?”
Example

- Take granular behaviours and put them into domains
  - e.g. *Entry behaviours* = method of entry; tools used; time of day; etc
  - *Property behaviours* = property taken; property damaged; and so on
- These are our **independent variables**
- Make these **dichotomous** by turning into yes/no questions
  - e.g. *Entry behaviours*: “was a screwdriver used? Was a crowbar used? Was a window open? Were the occupants home?” etc
Example

• Then apply a **similarity coefficient**
  • Index of similarity
  • **Jaccard’s** is coarse, but the measure of choice (Tonkin et al, 2008)
  • \( x = \) count of behaviours present in both
  • \( y = \) count of behaviours present in A but not in B
  • \( z = \) inverse of \( y \)

\[
J = \frac{x}{(x + y + z)}
\]
Example

- 1 = perfect similarity
- 0 = perfect dissimilarity
- 1 coefficient per domain
- Ignores joint non-occurrences
  - This is a concern when dealing with police data
  - Something may have been present, but not recorded
  - Less of a concern in this case
Example

• Each coefficient into direct logistic regression model
• Predictive analysis
• “To what extent does a given factor contribute to an outcome?”
  • e.g. “to what extent does being a smoker contribute to the risk of having a heart attack?”
  • Or “does similarity in the entry behaviours domain predict whether or not the two burglaries are linked?”
Example

- Logistic regression tells us:
  - Whether a variable is positively or negatively correlated with the outcome
  - How well a given variable fits with the data
  - The amount of variance that a given variable explains
  - A p-value (probability of seeing this result if the null hypothesis is true)
- Run for each domain
Example

• Then forward stepwise logistic regression
  • Start with one domain
  • Add a domain at each step
  • If this contributes to the model’s predictive power, keep it
  • Else discard it
• Determines optimal combination of domains
Example

- Regression results into ROC curves
- Graphical representation
  - $x$ (probability of false positive) against $y$ (probability of true positive)
- More reliable measure of predictive accuracy
- Based on area under the curve (AUC)
Example

• Overcomes statistical issue of using pairs from same sample (Tonkin et al, 2008)
• No reliance on arbitrary thresholds (Santtila et al, 2005)
• Measure of overall predictive accuracy (Swets, 1988)
Example

- Diagonal: no better than chance
- The higher the AUC value, the greater the predictive accuracy
  - $0.5 - 0.7 = \text{low}$
  - $0.7 - 0.9 = \text{good}$
  - $0.9 - 1.0 = \text{high}$
- Swets, 1988
Exceptions

• Some offences are less suitable, e.g. homicide
  • Bateman & Salfati, 2007; Harbort & Mokros, 2001; Sorochinski & Salfati, 2010
• Some offenders show more distinctiveness than others
  • Bouhana et al, 2016
• Some behaviours less consistent, e.g. property stolen in burglaries
  • Bennell & Canter, 2002; Bennell & Jones, 2005
Exceptions

- MO is a learned behaviour, and offenders develop
  - Pervin, 2002; Douglas & Munn, 1992
- Offenders will change behaviours in response to events
  - Donald & Canter, 2002
- Behaviours under offender’s control more likely to be stable
  - Furr & Funder, 2004; Hettema & Hol, 1998
- So offences involving victim interaction may differ
  - e.g. whether victim fights back / runs / shouts for help, etc
Exceptions

- Most research only applied to solved crimes
  - Woodhams & Labuschagne, 2012
- Relatively small samples
- Only serial offences
  - Slater et al, 2015
Experimentation

- Concept
- Research design
- Hypothesis
- Analysis
- Results
Concept

• Could CLA be applied to network intrusions?
  • Specifically, where attacker has code execution
  • Has never been done before
• Take granular behaviours (keystrokes, commands, etc)
• Apply CLA methodology
Research design

- Common approach historically: use police reports
- Can be inaccurate and/or incomplete
- Victim accounts may be inaccurate
  - Alison et al, 2001; Canter & Alison, 2003
- Crimes are often traumatic
- Traumatic experiences can distort memories
  - Freyd, 1996; Halligan et al, 2003
Research design

- Crime reports unlikely to be granular enough
- Previous studies on attacker profiling used simulations
- Honeypot?
  - Needed ground truth, as CLA previously untested on this offence type
  - Same IP addresses do not guarantee same individual at keyboard
  - Need to also distinguish between bots and humans
  - Honeypots can be fingerprinted
  - Attackers may deliberately change approach
Research design

- Modified open source Python SSH keylogger (strace)
  - https://github.com/NetSPI/skl
- Two VMs, exposed on the internet (SSH)
- One account per user per box
- Deliberate privesc vulnerabilities
- Plus fake data to exfiltrate
Research design

• Obtained participants
  • 10x pentesters / students / amateur enthusiasts
• Asked to SSH into both machines and try to:
  • Get root
  • Steal data
  • Cover tracks
  • Poke around
• Meanwhile, I recorded all keystrokes on each VM
Hypothesis

Cyber attackers will exhibit consistent and distinctive behaviours whilst executing commands on compromised hosts, which will provide a statistically significant basis for distinguishing between linked and unlinked attack pairs.
Analysis

- Split into behavioural domains, 40 behaviours each:
  - Navigation – moving through filesystem
  - Enumeration
  - Exploitation – privesc and exfil attempts
- Also coded for 3 metadata variables:
  - Number of ms between each keystroke
  - Number of ms between each command
  - Number of backspaces (as percentage of all keystrokes)
**Metadata variables**

- Non-dichotomous
- Used in other CLA work, in addition to behavioural domains
  - Intercrime distance (Bennell & Canter, 2002)
  - Temporal proximity (Tonkin et al, 2008)
- Filippoupolitis et al, 2014: commands typed per second
  - Problematic: length of command, time to complete, and time spent interpreting or manipulating output
Example behaviours

- `pwd`
- `cd /`
- `cd ..../..
- `cd ../`
- `ls /remote/dir`
- `ls -al`
- `cd /..
- `rm`
- `cd /remote/dir`
- `ls -la`
- `ls -la |grep Ctrl+D`
- `ls with wildcards`
- `date`
- `exclamation commands`
- `ls -al /remote/dir`
- `man`
- `which`
- `tab autocomplete`
- `pipe errors to dev null`
- `cd../[dir]
- `cat/etc/passwd`
- `cd .`
- `cd/[dir]
- `cd..
- `ls -a`
- `ls [dir] -al`
- `clear`
- `cd -R`
- `cd -r`
- `ls -R`
- `ls -ahl`
- `cat /remote/file`
- `find exec`
- `find /usr`
- `grep -i`
- `locate`
- `find / -name`
- `less`
- `python escape shell
  Uses FTP for exfil
  kill`
- `sync`
- `Used exploit suggester
  /bin/sh`
- `telnet`
- `su [username]
  sudo`
- `su`
- `sudo [command]
  sudo [username]
  sudo -n`
- `su root`
- `su - [username]
  sudo -s`
- `sudo su`
- `gcc file.c -o file`
- `CVE exploits`
- `wget`
- `wget 127.0.0.1
  Python script
  cp to home dir
  mkdir in tmp
  mv file
  cp file to tmp
  scp file`
- `touch`
- `chmod 755
  chmod 777
  chmod +x`
- `chmod +x [dir]
  vi`
- `nano`
- `cat /etc/sudoers`
- `sudo -s`
- `sudo -l`
- `bash`
- `looks for ssh authorized keys`
- `mount`
Analysis

- Average attack time per host: **133.34 minutes**
- Average commands per host: **243**
- 2 participants got root on Host A
- 1 participant got root on Host B
Similarity coefficients

- 10 attackers, 2 machines = 100 crime pairs
- Compare each attack against Host A to each attack against Host B
- 10 linked pairs, 90 unlinked pairs

Wrote application to calculate the similarity coefficient:
- For each pair for the 3 behavioural domains
- And differences between the 3 metadata variables

Ended up with CSV file:
- ID, paired (y/n), coefficients for each domain, differences for each metadata variable
## Similarity coefficients - behaviours

<table>
<thead>
<tr>
<th>Variables</th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
<th>Variance</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Navigation(linked)</td>
<td>0.756</td>
<td>0.756</td>
<td>0.166</td>
<td>0.28</td>
<td>0.5</td>
</tr>
<tr>
<td>Navigation (unlinked)</td>
<td>0.163</td>
<td>0.125</td>
<td>0.134</td>
<td>0.018</td>
<td>0.75</td>
</tr>
<tr>
<td>Enumeration (linked)</td>
<td>0.641</td>
<td>0.708</td>
<td>0.259</td>
<td>0.067</td>
<td>0.857</td>
</tr>
<tr>
<td>Enumeration (unlinked)</td>
<td>0.108</td>
<td>0.087</td>
<td>0.122</td>
<td>0.015</td>
<td>0.567</td>
</tr>
<tr>
<td>Exploitation (linked)</td>
<td>0.58</td>
<td>0.555</td>
<td>0.281</td>
<td>0.079</td>
<td>0.875</td>
</tr>
<tr>
<td>Exploitation (unlinked)</td>
<td>0.091</td>
<td>0.077</td>
<td>0.097</td>
<td>0.009</td>
<td>0.455</td>
</tr>
</tbody>
</table>

*Table 1 showing mean, median, standard deviation, variance and range scores for the three behavioural domains for linked and unlinked subsets (Jaccard’s coefficient)*
### Similarity coefficients - metadata

<table>
<thead>
<tr>
<th>Variables</th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
<th>Variance</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keystroke Interval (linked)</td>
<td>1726.642</td>
<td>351.723</td>
<td>4011.047</td>
<td>16088496.15</td>
<td>13017.14</td>
</tr>
<tr>
<td>Keystroke Interval (unlinked)</td>
<td>1827.393</td>
<td>488.678</td>
<td>3731.947</td>
<td>13927431.81</td>
<td>13425.62</td>
</tr>
<tr>
<td>Command Interval (linked)</td>
<td>46677.82</td>
<td>17058.83</td>
<td>61843.71</td>
<td>3824644150</td>
<td>194211</td>
</tr>
<tr>
<td>Command Interval (unlinked)</td>
<td>56354.3</td>
<td>29744.82</td>
<td>71853.04</td>
<td>5162858878</td>
<td>369751.9</td>
</tr>
<tr>
<td>Backspaces (linked)</td>
<td>5.471</td>
<td>2.416</td>
<td>7.53</td>
<td>56.695</td>
<td>24.355</td>
</tr>
<tr>
<td>Backspaces (unlinked)</td>
<td>9.574</td>
<td>6.941</td>
<td>8.715</td>
<td>75.944</td>
<td>38.249</td>
</tr>
</tbody>
</table>

Table 2 showing mean, median, standard deviation, variance and range values for the three timing and error variables for linked and unlinked subsets (keystroke and command interval in ms; backspaces as a percentage of total keystrokes).
Logistic regression

- Imported CSV file into SPSS
- Strenuous Package for Sad Students 😞
- Significant Probability of Statistics-related Stress 😞
- Direct logistic regression for each predictor variable
- Then forward stepwise logistic regression
- Six models in total, for each domain
- Plus an optimal combination/order of all domains
Results

Here comes the slide you’ve all been waiting for...
<table>
<thead>
<tr>
<th>Variables</th>
<th>Statistics</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
<th>Model 6</th>
<th>Model 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>b</td>
<td>-8.782</td>
<td>-5.773</td>
<td>-5.815</td>
<td>-2.184</td>
<td>-2.083</td>
<td>-1.553</td>
<td>-104.458</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>2.741</td>
<td>1.318</td>
<td>1.297</td>
<td>0.37</td>
<td>0.421</td>
<td>0.502</td>
<td>8648.491</td>
</tr>
<tr>
<td></td>
<td>OR</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.003</td>
<td>0.003</td>
<td>0.113</td>
<td>0.125</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.212</td>
</tr>
<tr>
<td>Navigation**</td>
<td>b</td>
<td>14.786</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>119.572</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>4.802</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10021.408</td>
</tr>
<tr>
<td></td>
<td>OR</td>
<td>2639086.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8.50E+51</td>
</tr>
<tr>
<td>Enumeration*</td>
<td>b</td>
<td>10.701</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>136.097</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>2.716</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>31066.523</td>
</tr>
<tr>
<td></td>
<td>OR</td>
<td>44415.895</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.28E+59</td>
</tr>
<tr>
<td>Exploitation*</td>
<td>b</td>
<td>14.012</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td>3.772</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>OR</td>
<td>1217542.54</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KS Interval+</td>
<td>b</td>
<td></td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td></td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>OR</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cmd Interval+</td>
<td>b</td>
<td></td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td></td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>OR</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Backspaces+</td>
<td>b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.9</td>
</tr>
<tr>
<td></td>
<td>SE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.065</td>
</tr>
<tr>
<td></td>
<td>OR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.914</td>
</tr>
</tbody>
</table>

**Model 1-6: Binary logistic regression**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Model $X^2$</th>
<th>$X^2$ sig.</th>
<th>Nagelkerke $R^2$</th>
<th>HL $X^2$</th>
<th>HL $X^2$ sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>50.16</td>
<td>p &lt; 0.05</td>
<td>0.825</td>
<td>0.211</td>
<td>p &gt; 0.05</td>
</tr>
<tr>
<td></td>
<td>41.673</td>
<td>p &lt; 0.05</td>
<td>0.713</td>
<td>11.26</td>
<td>p &gt; 0.05</td>
</tr>
<tr>
<td></td>
<td>42.526</td>
<td>p &lt; 0.05</td>
<td>0.725</td>
<td>5.627</td>
<td>p &gt; 0.05</td>
</tr>
<tr>
<td></td>
<td>0.007</td>
<td>Not sig.</td>
<td>0</td>
<td>8.808</td>
<td>p &gt; 0.05</td>
</tr>
<tr>
<td></td>
<td>0.185</td>
<td>Not sig.</td>
<td>0.004</td>
<td>10.876</td>
<td>p &gt; 0.05</td>
</tr>
<tr>
<td></td>
<td>2.747</td>
<td>Not sig.</td>
<td>0.057</td>
<td>18.605</td>
<td>p &gt; 0.05</td>
</tr>
<tr>
<td></td>
<td>65.017</td>
<td>p &lt; 0.001</td>
<td>1</td>
<td>0</td>
<td>p &gt; 0.05</td>
</tr>
</tbody>
</table>

Table 3 showing summary of logistic regression models for individual predictor variables.}

b: logit coefficient; SE: standard error; OR = odds ratio

*: p < 0.05
**: p < 0.005
*: Not statistically significant

HL: Hosmer and Lemeshow goodness-of-fit test

Model 7: Forward stepwise logistic regression (inclusion criteria: p < 0.05)
Dependent variable: linked attack pair (1), unlinked attack pair (0)
You’re too kind

(waits for applause to die down)
What does this tell us?

- Three behavioural domains can classify linked/unlinked offences
- High level of accuracy
- Navigation: most effective predictor
  - Followed by exploitation, then enumeration
- Strong positive correlation to dependent variable
- Keystroke and command interval variables not reliable predictors
- Backspace: weak negative correlation to linkage
- Results statistically significant for behavioural domains
- But not for any metadata variables
## ROC curves

- Results used to build ROC curves

<table>
<thead>
<tr>
<th>Variable</th>
<th>AUC</th>
<th>Sig.</th>
<th>SE</th>
<th>95 %CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Navigation</td>
<td>0.992</td>
<td>p &lt; 0.001</td>
<td>0.007</td>
<td>0.978 - 1.0</td>
</tr>
<tr>
<td>Enumeration</td>
<td>0.912</td>
<td>p &lt; 0.001</td>
<td>0.081</td>
<td>0.753 - 1.0</td>
</tr>
<tr>
<td>Exploitation</td>
<td>0.964</td>
<td>p &lt; 0.001</td>
<td>0.028</td>
<td>0.91 - 1.0</td>
</tr>
<tr>
<td>Keystroke Interval</td>
<td>0.572</td>
<td>NS</td>
<td>0.102</td>
<td>0.373 - 0.771</td>
</tr>
<tr>
<td>Command Interval</td>
<td>0.58</td>
<td>NS</td>
<td>0.113</td>
<td>0.358 - 0.802</td>
</tr>
<tr>
<td>Backspaces</td>
<td>0.702</td>
<td>p &lt; 0.05</td>
<td>0.094</td>
<td>0.519 - 0.886</td>
</tr>
<tr>
<td>Optimal</td>
<td>1</td>
<td>p &lt; 0.001</td>
<td>0</td>
<td>1.0 - 1.0</td>
</tr>
</tbody>
</table>

Table 4 showing area under the curve (AUC) values for each of the six predictors as well as the optimal model. 
SE: Standard error
ROC curves

I got 0.992 AUC, but it just ain’t 1

Jay-Z
(A ROC fella)
**ROC curve results**

- Navigation = 0.992
- Enumeration = 0.912
- Exploitation = 0.964
- Keystroke internal = 0.572
- Command interval = 0.58
- Backspace variable = 0.702
- **Optimal model (navigation & enumeration) = 1.0**
Implications

- Observations & comparisons
- Investigation implications
- Privacy implications
- Defeating CLA
- Threats to validity
Observations & comparisons

• High levels of consistency and distinctiveness
• Navigation and enumeration combined
  • No need for exploitation (in this study)
• Why was navigation specifically so prominent?
  • Something everyone does, every day
  • Enumeration & exploitation only done during attacks
  • Navigation behaviours may be more ingrained
Observations & comparisons

- Higher accuracy than other crime types
- Behaviours less subject to influence may be more stable
  - Nature of offence: offenders less likely to be influences
  - Broader approach may change
  - But possibly not granular command choice
  - Especially navigation
Observations & comparisons

• Metadata variables significantly weaker
• *What* you type has greater linking power than *how* you type
• Latency may have affected some of the results
• But mistakes/typos show some promise
• Needs further exploration
**Implications for investigators**

- Can link separate offences to common offenders
- With no atomic or computed IOCs
- But need a lot of information
  - Previous CLA/attribution work: limited, specific info required
  - Bennell & Canter, 2002; Hutchins et al, 2010; Clark & Landau, 2011
  - Here, need as much as possible
  - As granular as possible
Implications for investigators

• Need to be in a position to capture commands/keystrokes
  • High-interaction honeypots
  • Verbose and detailed logging
  • Backdoored CTFs or vulnerable VMs
Implications for investigators

• Could also link attackers who trained together
• Or who have all done a certain certification
  • Sample commands and code
  • Dilutes CLA assumption of distinctiveness
  • But could still assist with attribution
Privacy implications

• People can be linked to separate hosts/identities
  • Based on approaches, syntax, and commands
• Regardless of anonymising measures
• Regardless of good OPSEC elsewhere
Privacy implications

• Like forensic linguistics, exploits stable behavioural traits
• Won’t be 100% accurate obviously
• And affects less of the population, cp. forensic linguistics
  • e.g. ~86% of the population is literate*
  • Less people than that can operate a command-line

* https://data.worldbank.org/indicator/SE.ADT.LITR.ZS, 27/06/18
Privacy implications

• This study only focused on commands
• May also apply to:
  • Typos, and the way you correct them
  • How you form capitals
  • Using PgDn/PgUp
  • Using arrow keys rather than the mouse
  • Tabs/spaces
  • Keyboard shortcuts
  • Use of, and preference for, bracket types
Privacy implications

• If someone can log your keystrokes, you have issues anyway
  • But this is less about *identification*
  • If someone can log your keystrokes, it’s not hard to find out who you are
• This is more about *attribution via linkage*
• Could be used to link you to historical/future activity
• Used to build up repository of command profiles
Defeating CLA

- Similar to defeating authorship identification
- Make a conscious decision to disguise your style
  - Forensic linguistics: solutions range from crude (Google Translate) to sophisticated (automated changes to sentence construction, synonym substitution, etc)
  - CLA different – e.g. alias command would not work
  - Hard to automate – can’t predict commands in advance
  - Could semi-automate, using scripts
Turns: People who succeed have momentum. The more they succeed, the more they want to succeed, and the more they find a way to succeed. Similarly, when someone is failing, the tendency is to get on a downward spiral that can even become a self-fulfilling prophecy.
Note on Google Translate

- English -> Norwegian
- Norwegian -> French
- French -> Afrikaans
- Afrikaans -> Romanian
- Romanian -> Japanese
- Japanese -> English

U wot m8

She is trying to explain my style through various translations such as demonstration.
Defeating CLA

- Conscious changes are probably the best way to do it
- Randomising ordering of command switches
- Switching up tools used e.g. `wget` instead of `curl`; `vi` instead of `nano`; `less` instead of `cat`
Threats to validity

- Very small sample
- Not real-world data
- Attackers were willing volunteers
  - Knew they had permission, with no risk of reprisal
- Linux only
- One scenario (low-priv shell)
- Attackers may not always want/need to escalate
Summary

- Topics for future research
- Collaboration
- Conclusion
- References
Future research

• Explore effects of expertise and temporal proximity
• Further research into metadata variables for mistakes
• Real-world data
• Stochastic analysis
• Greater environmental and scenario diversity
• Real-time or near real-time automation
Collaboration

- Get in touch if you want to discuss
- @darkartlab
- matt.wixey@pwc.com
Conclusion

- Small, novel study
- Some promising results
- Significant implications for defenders/investigators
- As well as implications for privacy
- Needs further investigation
References


data.worldbank.org/indicator/SE.ADT.LLTR.ZS, accessed 27/06/2018
References


github.com/NetSPI/skl, accessed 27/06/2018


References


References


Thoughts, questions, feedback:

@darkartlab
matt.wixey@pwc.com