I'm unique, just like you

Human side-channels and their implications for security and privacy

Matt Wixey

October 2019

pwc

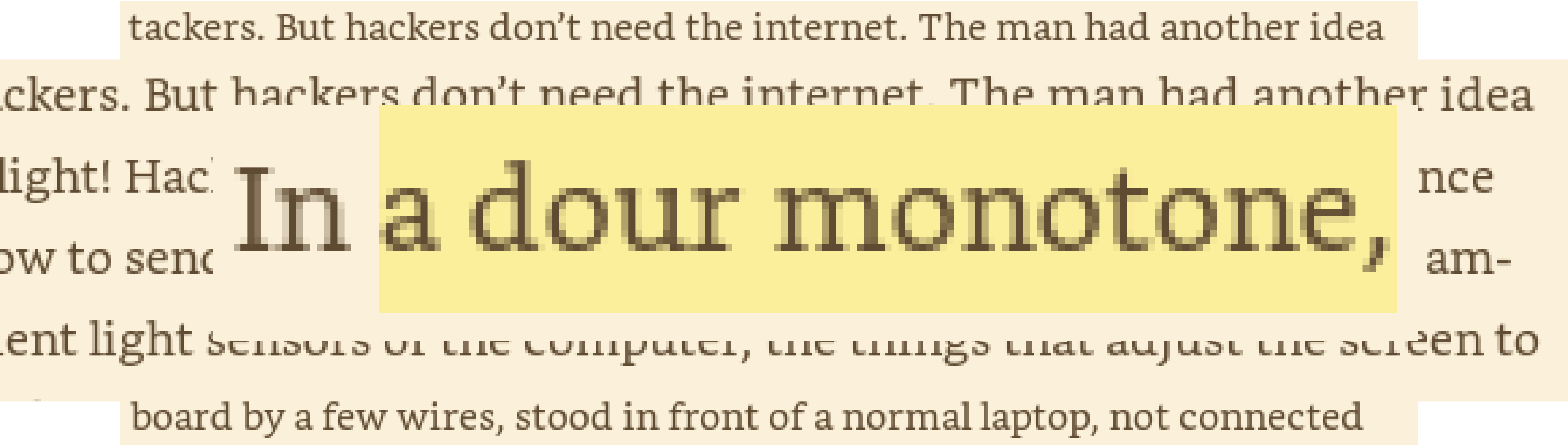
Matt Wixey

- Research Lead for the PwC UK Cyber Security practice
- PhD student at UCL
- Previously worked in LEA doing technical R&D
- Black Hat USA, DEF CON, ISF Congress, BruCon, 44Con, BSides, etc

Disclaimer (Miller, 2018)

8 ‘“What about computers not connected to the internet?”’ This was Matt Wixey, a security researcher at PwC UK. His talk was called ‘See no evil, hear no evil’. It was my personal favourite. https://

Disclaimer (Miller, 2018)



Aims

- Be aware of 3 human side-channels and how they work
- Practical takeaways for each side-channel, including tools
- Examine implications for security and privacy
- Know about possible countermeasures
- Explore future research ideas

Agenda

The John Christie case

https://www.radiotimes.com/news/2017-06-22/a-timeline-of-john-christies-crimes-and-their-discovery-and-the-bits-rillington-place-missed-out/

Background

How can we use identifiers to find an offender?

Background

- Various things we can look at in real-world crimes
  - Fingerprints, DNA, gait, irises, voice, etc
- What about digital offences?
  - IP and MAC addresses, domains, subscriber info, emails, usernames etc
  - New problem: easily obfuscated, spoofed, anonymised
  - Other methods take us further away from the individual
    - Activity correlated to timezones (Rid & Buchanan 2014)
    - TTPs (Symantec 2011)

A possible solution

Background

- Computers have “side-channels”
  – Unintentional leakage in primitive outputs, as a result of operations
- Is there a real-world equivalent?
  – Humans as bio-computers (Lilly, 1968) with outputs (writing, speech, etc)
  – Unintentional leakage (behavioural theory)
  – Distinctive and consistent (Shoda et al, 1994; Zayas et al, 2002)
    - Based on education, experience, training, environment, goals, etc
    - “Human side-channels”

Forensic linguistics

Me: Professor, I’d like to do my essay on the etymology of the word “f***”. I just wanted to check you’d be OK with that, or would it be inappropriate?

Professor: I don’t give a s***.

Theory of forensic linguistics

- Covers other aspects, but we're looking at one in particular:
- Authorship attribution via stylometry
  - Spelling and orthography
  - Grammar
  - Lexicon
  - Idiom
  - Identical expressions

Real-world use cases

Forensic linguistics

- Law enforcement investigations – ransom notes, texts, etc
- Plagiarism investigations
- Literature:
  - Shakespeare, The Federalist Papers, Primary Colors, JK Rowling
- Uncovering miscarriages of justice
  - e.g. police officers collaborating on statements

What forensic linguistics isn’t

- Detection of deception (cp. Van Der Zee et al, 2018; Wixey, 2018)
- Detection of intention
- Creating/comparing ‘textual fingerprints’
- Handwriting analysis
- Assessing context or content

Stylometry techniques

Complex

- Create corpus, extract features of interest
  - Parts of speech; word length; sentence length; pronouns; function words; hapax legomenon; dis legomenon; etc
- Statistical comparison of features
  - Support Vector Machines; Principal Component Analysis; Delta; etc

Basic

- Observing and noting unusual spellings/punctuation use
- Corpus/Google searching for these

Case studies (Olsson, 2009)

Forensic linguistics

http://news.bbc.co.uk/1/hi/england/south_yorkshire/4407944.stm
https://www.thetimes.co.uk/article/ice-cream-wars-feud-ended-before-death-of-thomas-campbell-cd2gwpwgk

Cyber-specific case studies

Forensic linguistics

- Academic research
  - Tweets (Sultana et al, 2017; Silva et al, 2011)
  - Sockpuppet detection (Solorio et al, 2013)
  - Forum posts (Abbasi & Chen, 2005)
  - Emails (Iqbal et al, 2010)
  - Source code (Caliskan-Islam et al, 2015; Frantzeskou et al, 2007)
  - Detecting authorship deception (Pearl & Steyvers, 2012)

Cyber-specific case studies

- Operation Tripoli (Check Point, 2019)
  - Large Facebook social engineering campaign
  - Searching for repeated spelling and grammatical errors
  - Revealed multiple profiles (over 30), appear to be by same actor
- Qualitative study of IRS phone scammers (Tabron, 2016)
  - Polar tag questions, narrative violation
  - "Strengthening the human link"
- Guccifer 2.0 (Argamon, 2016)

Other use cases

Forensic linguistics

- Spearphishing – different pretexts, same author
- Missives and manifestos posted online
- Ransomware instructions/notes
- Posts/Tweets claiming responsibility, coordinating attacks, etc
- Satoshi Nakamoto!

Scenario example

Forensic linguistics

- A new spearphishing email comes into your org
- You notice an unusual turn of phrase
- You Google it (using special operators)
- This leads you to a forum post with a username
- Law enforcement can attribute that username to an IP address, subscriber data, etc

Scenario example

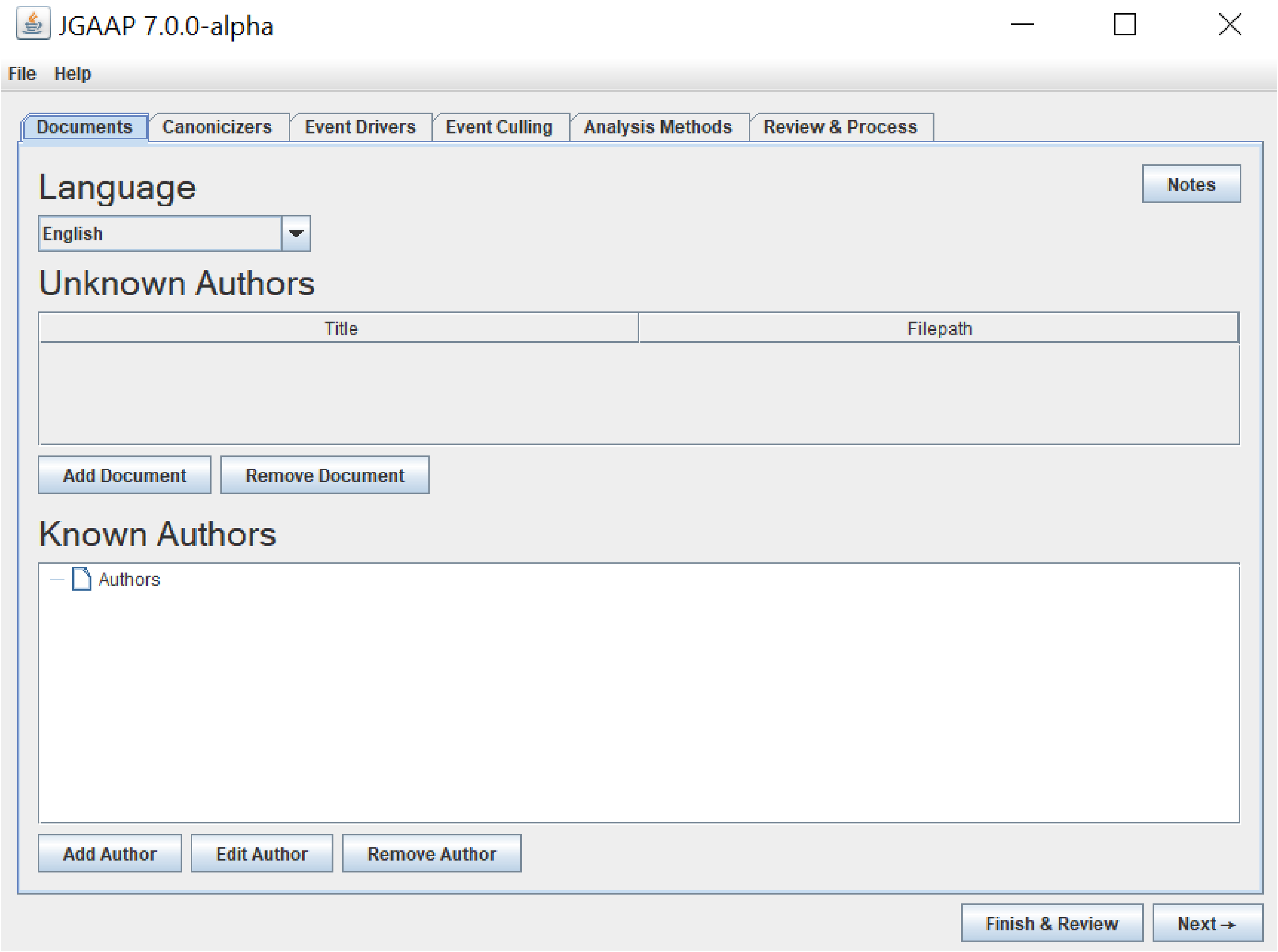
- You crawl forum posts of known threat actors and store them in a database (your corpus)
- Your org is hit by a DDoS attack using reflection/spoofing
- You notice the attack appears to be being coordinated on Twitter
- You search for other Tweets and compare them to your corpus
- You get a high match with posts by a particular user
- That user may be behind this attack

But how?

- How do you do all this?
- Isn’t forensic linguistics a really specialist discipline?
- Don’t I need at least an MSc in linguistics to do it?
- And don’t I need machine learning models, expensive statistics software, etc etc?
- Nope!

JGAAP (github.com/evllabs/JGAAP)

Forensic linguistics



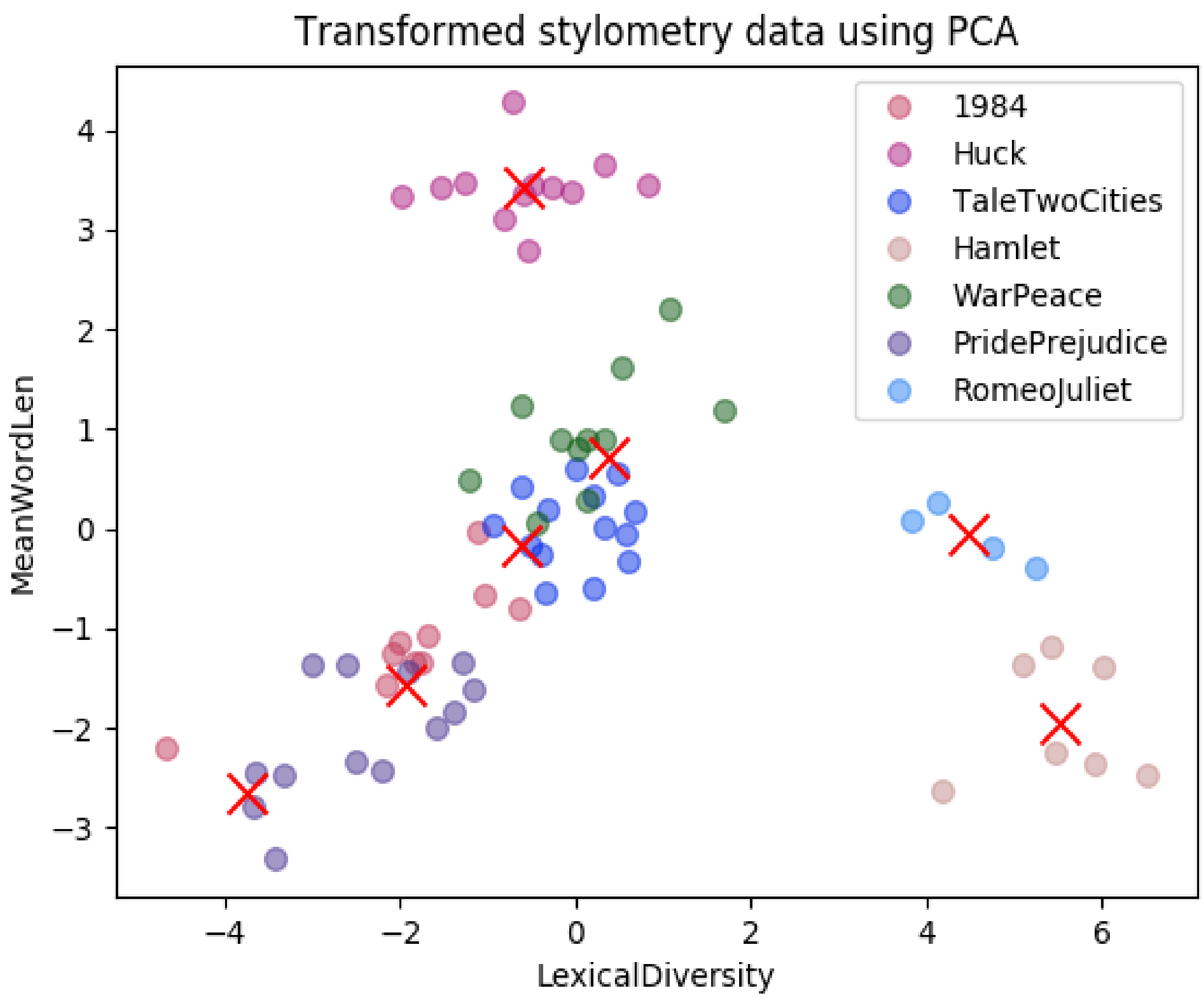
Delta spreadsheets (wp.nyu.edu/exceltextanalysis/deltaspreadsheets/)

	A	B	C	D	E	F	G	H	I
1	**Delta Calculation Worksheet 2019**				**Analysis Parameters**			**Instructions: View >**	
2	**© David L. Hoover**			Do It All	34	**Primary Samples**		20	**Secondary**
3	**Argamon's Delta: SUM(ABS((Test-Primary)/S.D.))**				Y	**Delete Personal Pronouns? If "Y", "personal pronouns"**			
4	**Analysis Area**				**70.00**	**Culling %--words for which a single text supplies more**			
5		MAX	394.85	D%chg 1-2	**2000**	**Words to Process--the number of MFW on which the ne**			
6		MIN	221.60						
7		MEAN	335.25	Dz%chg 1-	**4050**	**Word Count: the number of words in this sheet availabl**			
8		STDEV	35.41		**Test Sample**				
9	**Primary**	Stoker	2000.00	MFW	Stoker	**Primary Set**			
10	**Sample**	The Watter's Mou' (1)	**delta-score**	deltaz-score	The Watter'	**Std.Dev.**			
11	Jane Eyre (1)	Bronte, C_Jane Eyre (1)	311.96	-0.658037	7.8650246	0.854827636			
12	Shirley (1)	Bronte, C_Shirley (1)	332.09	-0.089428	3.8533914	0.609670005			
13	Vilette (1)	Bronte, C_Vilette (1)	296.80	-1.086214	2.5831835	0.348677134			
14	54HideSeek (1)	Collins_54HideSeek (1)	322.29	-0.366236	3.177658	0.263877382			
15	56 After Dark (1)	Collins_56 After Dark (	300.09	-0.993103	1.8176374	0.255871729			
16	57DeadSecr (1)	Collins_57DeadSecr (1)	356.16	0.5905793	1.8176374	0.186516296			
17	60WomanWh (1)	Collins_60WomanWh (1)	332.95	-0.065165	1.3728509	0.196812829			
18	62NoName (1)	Collins_62NoName (1)	359.02	0.6712535	0.8681892	0.149298976			
19	66Armadale (1)	Collins_66Armadale (1)	349.14	0.3922082	1.578137	0.183835864			
20	68Moonston (1)	Collins_68Moonston (1)	337.16	0.0539633	0.8125909	0.095727485			
21	70ManWife (1)	Collins_70ManWife (1)	354.43	0.5414657	0.9622787	0.117467696			
22	72PoorF (1)	Collins_72PoorF (1)	358.93	0.6686231	1.0093234	0.144218948			
23	73NewMagd (1)	Collins_73NewMagd (1)	382.86	1.3446892	0.5003849	0.101865662			
24	75LawLady (1)	Collins_75LawLady (1)	361.46	0.7401784	0.9622787	0.067778925			
25	76TwoDest (1)	Collins_76TwoDest (1)	338.09	0.0800497	0.975109	0.14678382			
26	79FallenL (1)	Collins_79FallenL (1)	375.22	1.1288705	0.3977419	0.157792639			
27	80Jezebel (1)	Collins_80Jezebel (1)	358.64	0.6604405	0.1539646	0.082783681			
28	81BlackR (1)	Collins_81BlackR (1)	369.01	0.9534379	0.5431529	0.165845115			
29	82HeartSci (1)	Collins_82HeartSci (1)	362.57	0.7714756	0.5688136	0.088644666			
30	84IsayNo (1)	Collins_84IsayNo (1)	394.85	1.6831709	0.1881789	0.180241452			

stylometry (github.com/jpotts18/stylometry)

Forensic linguistics

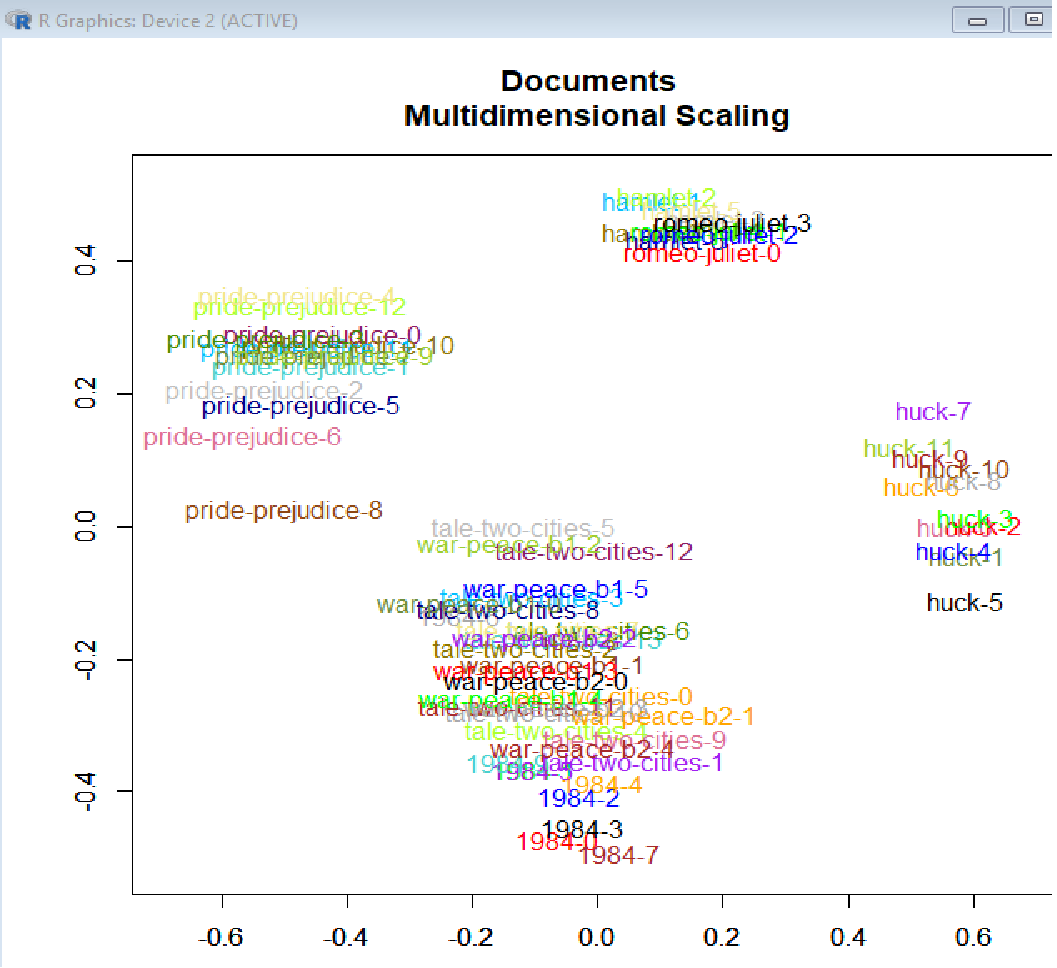
```
File  Edit  View  Search  Terminal  Help
ubuntu@ubuntu:~/stylometry$ python test-cluster.py
Reading corpus data...
Reading corpus data...
[6.48256219 3.75251274]
```



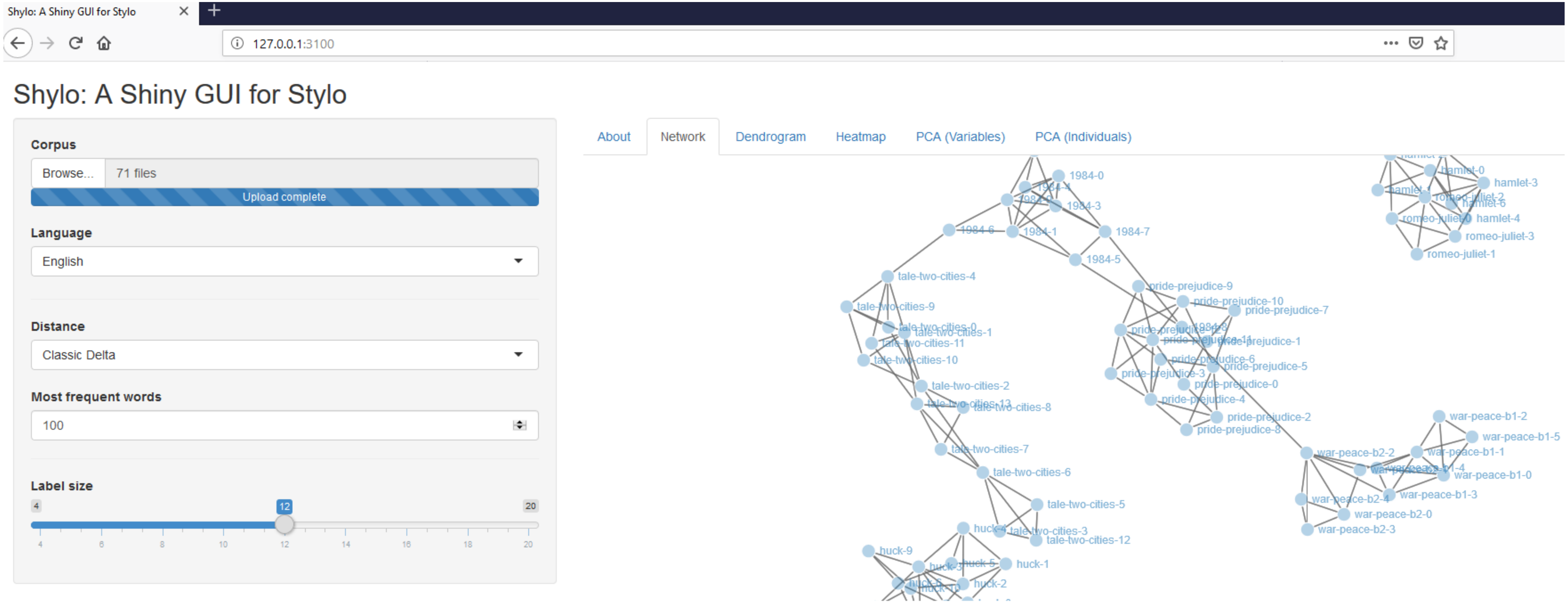
stylo (R library) - github.com/computationalstylistics/stylo

R Console

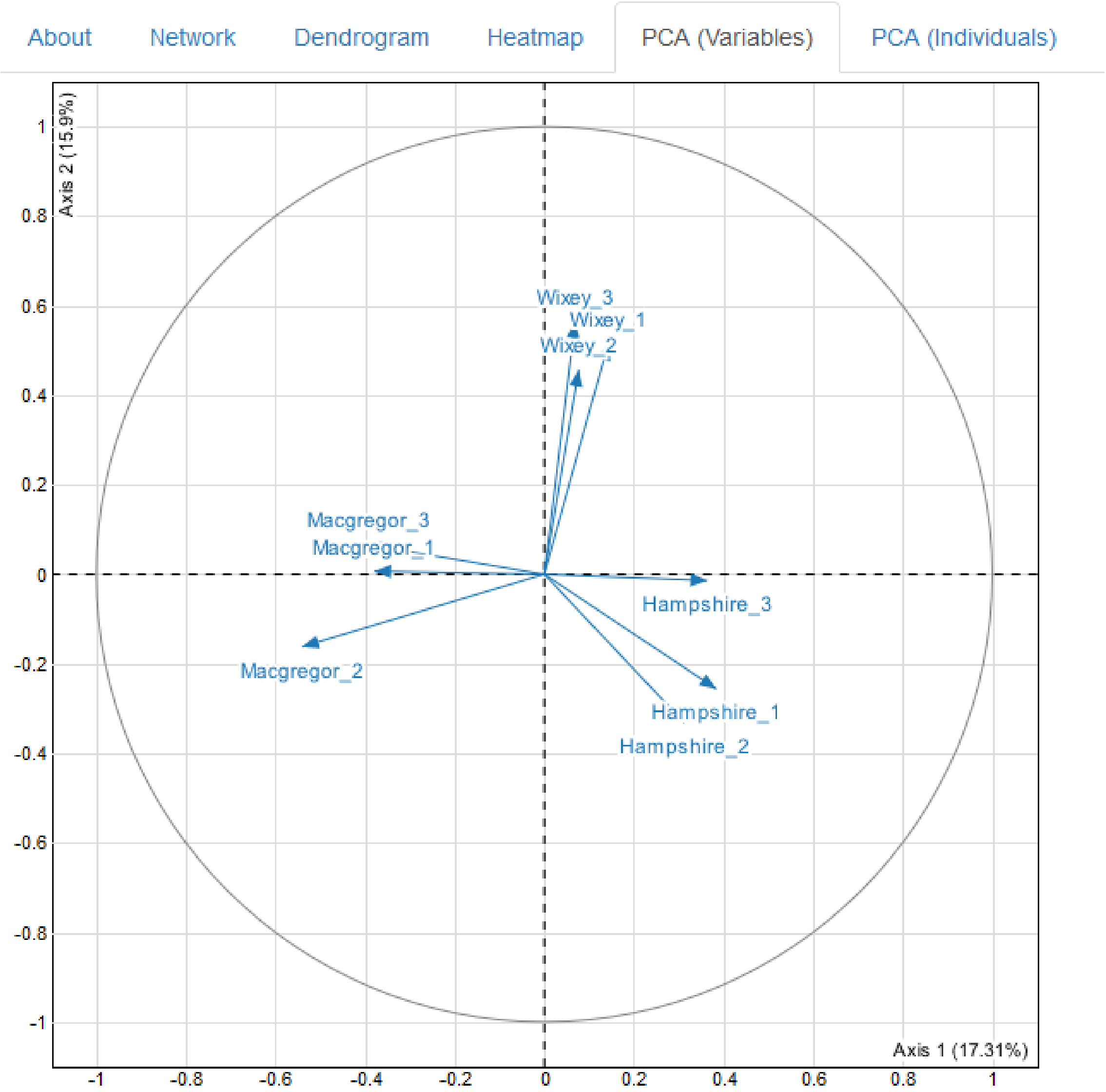
```
recent configuration, etc.

Advanced users: you can pipe the results to a variable, e.g.:
         hip.hip.hurrah = stylo()
this will create a class "hip.hip.hurrah" containing some presumably
interesting stuff. The class created, you can type, e.g.:
         summary(hip.hip.hurrah)
to see which variables are stored there and how to use them.


for suggestions how to cite this software, type: citation("stylo")


Warning messages:
1: In file(file, "r") : cannot open file 'Austen': Permission denied
2: In file(file, "r") : cannot open file 'Blogs': Permission denied
3: In file(file, "r") : cannot open file 'Dickens': Permission denied
4: In file(file, "r") : cannot open file 'Lit corpus': Permission denied
5: In file(file, "r") : cannot open file 'Orwell': Permission denied
6: In file(file, "r") : cannot open file 'Shakespeare': Permission denied
7: In file(file, "r") : cannot open file 'Shelley': Permission denied
8: In file(file, "r") : cannot open file 'Tolstoy': Permission denied
9: In file(file, "r") : cannot open file 'Twain': Permission denied
10: In file(file, "r") : cannot open file 'Tweet corpus': Permission denied
>
```



Shylo (stylo wrapper) - github.com/severinsimmler/shylo



- Shylo (stylo wrapper)



Summary of tools

Forensic linguistics

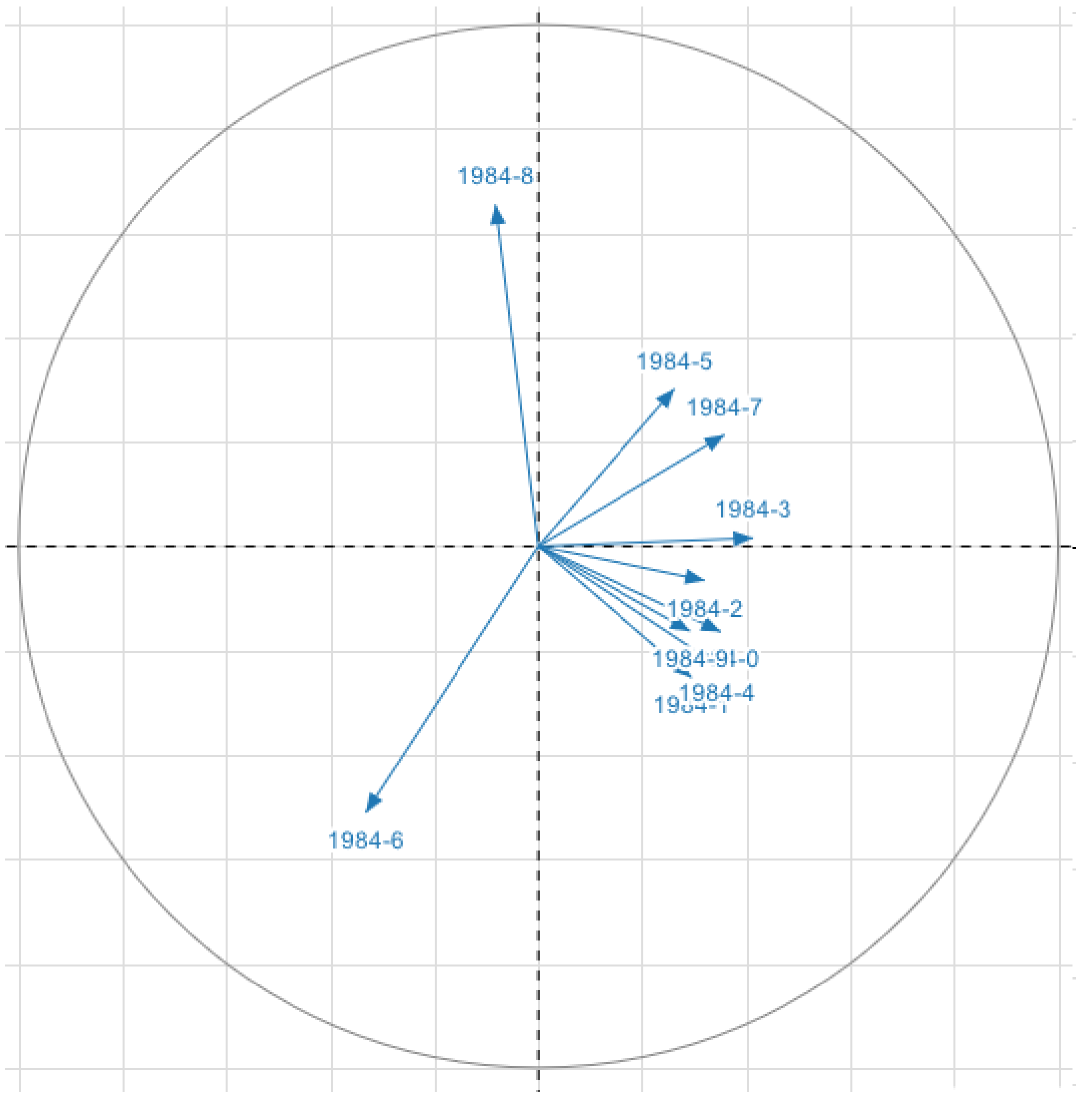
Tool	Free?	Ease of use	Method(s)	Outputs	Scalability
JGAAP	Yes	Hard	Multiple	Numeric	Possible
Delta sheets	Yes	Moderate	Delta	Numeric	Difficult
Stylometry	Yes	Easy	PCA	Graphs	Possible
Stylo (R)	Yes	Easy	Multiple	Graphs	Possible
Shylo	Yes	Easy	Multiple	Multiple	Possible

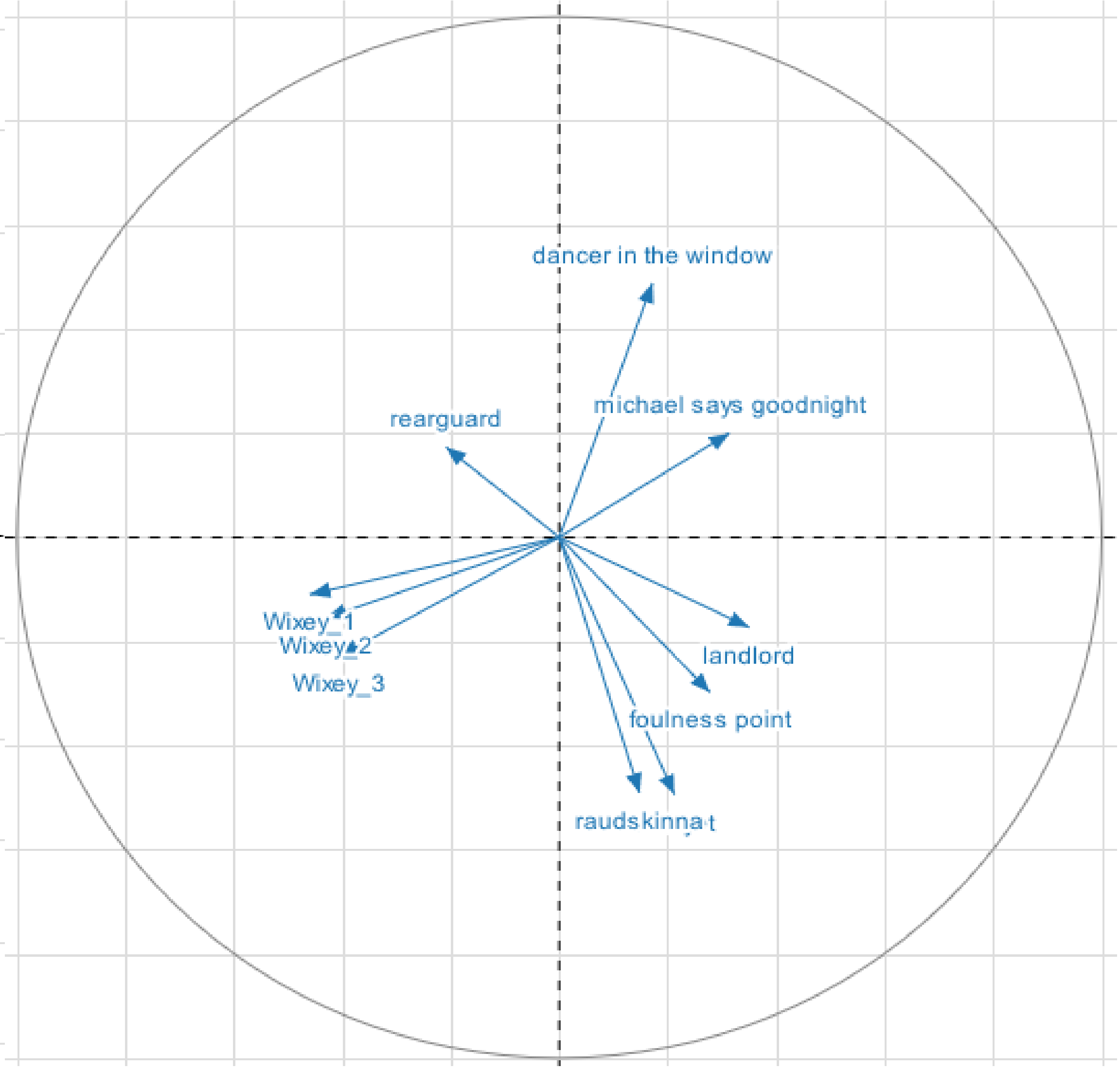
Caveats

- Register makes a big difference
- Need a baseline of text – sizeable samples
- Ground truth may also be required (depending on objective)
- Strategy will be decided by circumstances
- Time lapse may affect results
- Not fingerprints, no 100% accuracy – not a silver bullet

Register

Forensic linguistics





Privacy implications

- Attribution of texts written under a separate identity
- Diminish anonymity

Countermeasures

- Linguistic style is often unconscious
- Awareness of it can facilitate disguising it
- Imitating another's style, either during or after writing
  - Writing in another 'voice' (cp. *1984*)
- Google Translate
- Combining with other authors
- Running forensic linguistic tools – Anonymouth (Brennan et al, 2012; McDonald et al, 2012)

What can I do now?

- Test tools out
  - Text from previous attacks & open source data
  - Start building corpus
  - Have a play, let me know what you think!
- Explore how useful/applicable it would be for your use cases
- Think about other scenario/contexts it could be used in

Behavioural signatures

“I got an AUC of 0.99 but that’s basically 1” – Jay-Z (a ROC fella)

Background

Behaviour

- Active area of research in attribution: who hacks, and why
  - Motivation, skills, attack behaviours (Landreth, 1985; Salles-Loustau et al, 2011)
  - Attitudes and culture (Chiesa et al, 2008; Watters et al, 2012)
  - Psychological elements (Shaw et al, 1998)
  - Specific actions undertaken (Ramsbrock et al, 2007)

Background

Behaviour

- What hasn't been done: comparing profiles of attackers
- Case Linkage Analysis (CLA)
  - Linking crimes together based on common features
  - Note: this is **not** offender profiling!
  - Offender profiling: *After analysing this crime, I think the offender is a charismatic security researcher with a fast-disappearing hairline*
  - CLA: *After analysing crimes A and B, they have features XYZ in common. I know charismatic balding researcher Matt Wixey committed crime A, so he may have also committed crime B*

Background

Behaviour

- Statistical comparison of crime scene behaviours (Woodhams & Grant, 2006)
- Some success in academic literature, with real-world crimes
  – Homicide, burglary, robbery, sexual assault, arson, etc
  – But not cyber attacks (until this research!)
- Grubin et al, 1997; Mokros & Alison, 2002; Tonkin et al, 2008
- Based on same principles of distinctiveness and consistency

Why? What's the point?

Behaviour

- If we can conclude that two crimes are linked, we can:
  – Save time and resources by investigating them together
  – Build up a body of evidence against an offender
  – Potentially identify weaknesses/flaws in offender's strategies
  – Attribute multiple crimes to one offender if/when they're identified
  – Decision-making aid

Example – Crime A

Behaviour

https://www.businessinsider.com/why-banksy-has-nothing-to-do-with-real-graffiti-culture-2013-10?r=US&IR=T

Example – Crime B

Behaviour

https://www.etsy.com/au/listing/245982767/banksy-graffiti-art-super-mario-various

Linking crimes

Behaviour

- We know these crimes are probably linked
- But how do we prove it?
- What features of the crime might we look at?

Step 1: Identify behaviours

Behaviour

- Create behavioural domains – broad categories of the crime, e.g. “equipment used”, “property targeted”, etc
- For each domain, look at very granular behaviours and turn them into yes/no questions
- E.g. for “equipment used”: did attacker use stencil? Did they use colour? Did they sign the image? Did they use X paint? Or Y paint?
- Repeat this for all behavioural domains – the more granular, the better!

Step 2: Similarity coefficient

Behaviour

$$J = \frac{x}{(x + y + z)}$$

- Jaccard's Coefficient (Tonkin et al, 2008)
- 1 per domain
- X = count of behaviours present in both attacks
- Y = count of behaviours present in Crime A, but not B
- Z = present in Crime B, but not A
- 1 = perfect similarity, 0 = perfect dissimilarity

Step 3: Logistic regression

- Can we predict whether the crimes are paired (e.g. committed by the same person)?
- Logistic regression lets us test this out
- Statistical way of finding out which domain contributes more
  - e.g. is “equipment used” more effective than “property targeted”?
- And, combined, how well they can be used to predict linkage?
  - SPSS, R, etc – loads of tutorials online

Step 3: Logistic regression

Behaviour

- Run for each behavioural domain to get:
  - Positive or negative correlation
  - A p-value (statistical significance)
  - Amount of variance that a variable explains
- Repeat with forward stepwise logistic regression
  - Will automatically start with one domain, and add one at each step
  - If it contributes to predictive power, keep it, else discard from the model
  - Determines optimal combination of domains

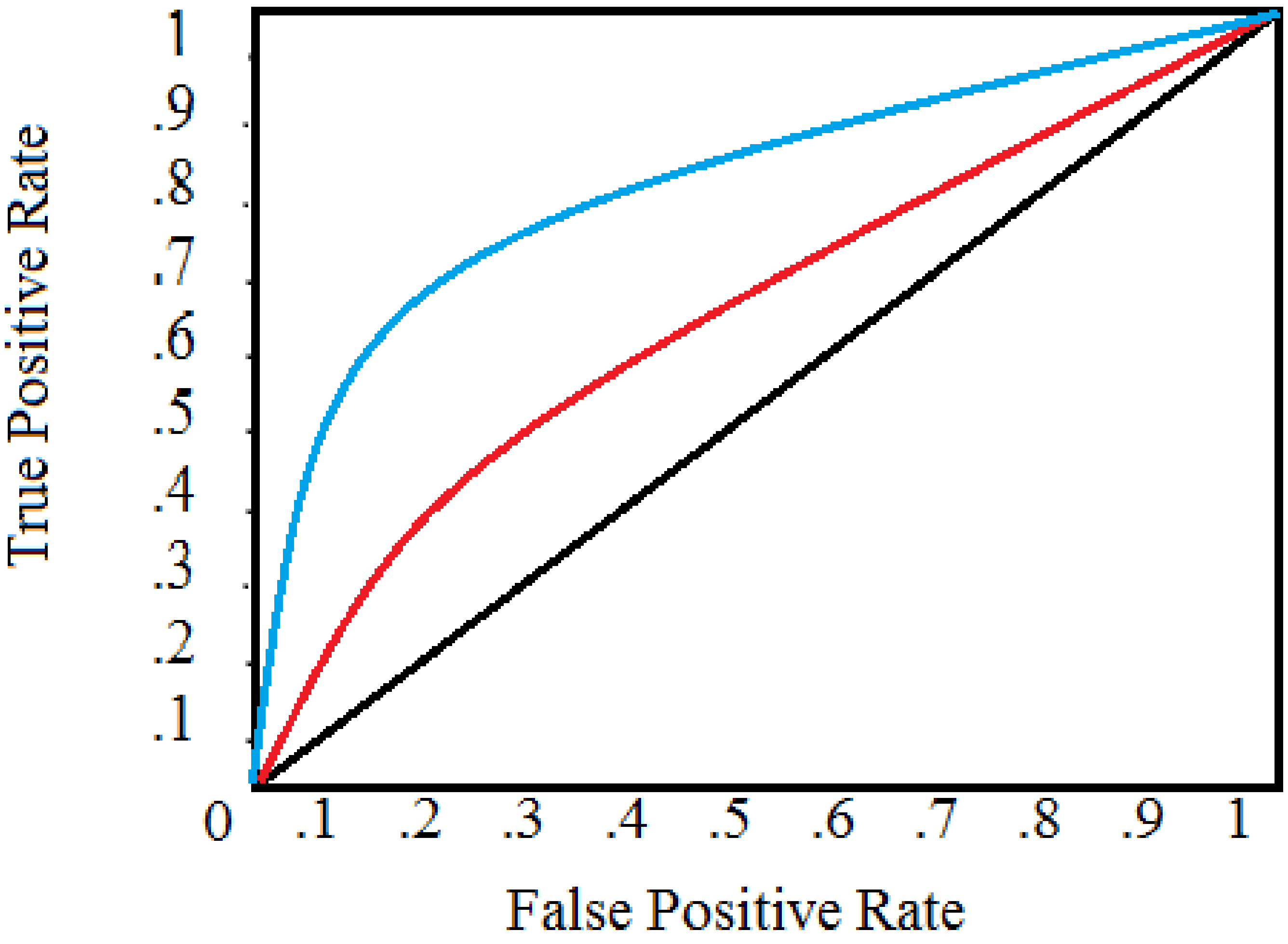
Step 4: ROC Curves

Behaviour

- Put regression results into ROC curves
- Graphical representation of performance
- Commonly used to look at predictive accuracy of machine learning
  - Plots x (prob of false positive) against y (prob of true positive)
  - More reliable measure of predictive accuracy (Tonkin et al, 2008; Swets, 1988)
  - You'll get 'area under the curve' (AUC) values

Step 4: ROC Curves

Behaviour



- Diagonal: no better than chance
- The higher the AUC value, the greater the predictive accuracy
- **0.5 – 0.7 = low**
- **0.7 – 0.9 = good**
- **0.9 – 1.0 = high**
- Swets, 1988

https://www.statisticshowto.datasciencecentral.com/receiver-operating-characteristic-roc-curve/

Why apply it to cyber attacks?

Behaviour

- In principle, same concepts will apply
- Never been done before
- OSCP, 2014 –idea
- New contribution to CLA body of literature

Cyber attacks - scenario

Behaviour

- In 2017, Business Corp is attacked
- The attacker infects the network with a malicious macro doc
- And then pokes around the filesystem
- Sets up a permanent backdoor
- And starts exfiltrating data
- In 2019, Business Corp is attacked again
- The methodology looks similar – but how do we know it's the same threat actor?

Experiment – cyber attacks

Behaviour

- Modified open source Python SSH keylogger (strace)
  – https://github.com/NetSPI/skl
- Two VMs, exposed on internet over SSH (like honeypots)
- One account per user per box
- Deliberate privesc vulnerabilities, plus fake data to exfiltrate
- 10x pentesters/students asked to SSH in (2 attacks each)
  – And get root, steal data, cover tracks, poke around

Classification

- Define behavioural domains e.g. ‘navigation’, ‘enumeration’, etc
- Classify keystrokes as commands (‘behaviours’)
  - Turn into ‘yes/no’ questions
  - “Did attacker try to wget malware from a remote site after compromise?”
  - Assign 1 if yes, 0 if no
  - End up with binary string for each offence in each domain

Experiment

- Keystrokes collated per user, split into behavioural domains
  - Navigation, enumeration, exploitation
  - 40 individual behaviours per domain

sudo	chmod 755
su	chmod 777
sudo [command]	chmod +x
sudo [username]	chmod +x [dir]
sudo -n	vi
su root	nano
su - [username]	cat /etc/sudoers
sudo -s	sudo -s
sudo su	sudo -l
gcc file.c -o file	bash
CVE exploits	looks for ssh authorized keys
wget	mount

Experiment

Behaviour

- Automated calculation of Jaccard values

Variables	Mean	Median	SD
Navigation(linked)	0.756	0.756	0.166
Navigation (unlinked)	0.163	0.125	0.134
Enumeration (linked)	0.641	0.708	0.259
Enumeration (unlinked)	0.108	0.087	0.122
Exploitation (linked)	0.58	0.555	0.281
Exploitation (unlinked)	0.091	0.077	0.097

Experiment

Behaviour

- Imported results into SPSS
- Performed logistic regression (direct and forward stepwise)
- Also used SPSS for ROC curves

Variable	AUC	Sig.	SE	95 %CI
Navigation	0.992	$p < 0.001$	0.007	0.978 - 1.0
Enumeration	0.912	$p < 0.001$	0.081	0.753 - 1.0
Exploitation	0.964	$p < 0.001$	0.028	0.91 - 1.0
Keystroke Interval	0.572	NS	0.102	0.373 - 0.771
Command Interval	0.58	NS	0.113	0.358 - 0.802
Backspaces	0.702	$p < 0.05$	0.094	0.519 - 0.886
Optimal	1	$p < 0.001$	0	1.0 - 1.0

Applicability and approaches

Behaviour

- Honeypots
- Build up a corpus of attackers
- Could also identify attackers who’ve trained together

Caveats

- Some offenders show more distinctiveness than others
  – Bouhana et al, 2016
- Some behaviours less consistent
  – Bennell & Canter, 2002; Bennell & Jones, 2005
- MO is a learned behaviour, and offenders develop
  – Pervin, 2002; Douglas & Munn, 1992
- Offenders will change behaviours in response to events
  – Donald & Canter, 2002

Caveats

Behaviour

- This experiment:
  - Small sample, only commands
  - Only one OS/scenario
  - Not ‘real’ attackers – knew they wouldn’t suffer consequences
  - Not all attackers will have the same motivations, could affect results
  - Not 100% accurate

Privacy implications

Behaviour

- People can be linked to separate hosts/identities
- Based on approaches, syntax, and commands
- Regardless of anonymising measures
- Regardless of good OPSEC elsewhere
- Could be linked to historical or future activity

Countermeasures

Behaviour

- Similar to defeating authorship identification
- Make a conscious decision to disguise your style
  - CLA different – e.g. alias command would not work
  - Hard to automate – can't predict commands in advance
  - Could semi-automate, using scripts
  - Randomising ordering of command switches
  - Switching up tools e.g. wget instead of curl; vi instead of nano, etc

What can I do now?

Behaviour

- Give it a go!
  - – Keylogger on CTF machines (make sure participants are aware, take appropriate ethical measures)
  - – Classification and calculate Jaccard score – pretty simple
  - – Calculate logistic regression scores – again, pretty simple
  - – ROC curve analysis (same tools)
  - – Have a go at automating! R/Python probably best place to start
  - –Other behavioural domains, e.g. evasion techniques
  - –Whitepaper available (contact me!) or see DEF CON 2018 talk

Cultural CAPTCHAs

“Of course I remember Crinkley Bottom”

Background

- “Is this account a human or a bot?”
  - Lots of academic and practical research (Filippoupolitis et al, 2014)
  - Botometer, Twitteraudit, Botcheck, Botsentinel
- Certain behaviours/features can be “tells”
- Harder question: “Is this account owner really X nationality?”
- Context: hostile accounts influencing conversations or consensus
  - We think they’re probably human
  - But how do we prove they’re *authentic*?

Background

- Enter “cultural CAPTCHAs”
- Cultural artefacts which haven’t spread beyond origin
- In many cases this can be popular culture, but also:
  - Language
  - Cultural norms and expectations
  - Food
  - Music
  - Traditions, etc

Experiment

- Let's try an example - who are these two men?
- **RAISE YOUR HAND** if you know

Experiment

- Let’s try another

https://www.independent.co.uk/arts-entertainment/tv/news/barry-chuckle-dead-brothers-latest-cause-comedy-death-manager-a8477966.html

Experiment

Cultural CAPTCHAs

vicandbob.jpg × reeves mortimer

All | Images | Maps | Shopping | More | Settings | Tool

About 2,890,000,000 results (1.00 seconds)

Image size:
770 × 375

Find other sizes of this image:
All sizes - Small - Medium

Possible related search: ***reeves mortimer***

Vic and Bob - Wikipedia

https://en.wikipedia.org/wiki/Vic_and_Bob ▾

Vic and Bob, also known as **Reeves** and **Mortimer**, are a British comedy double act consisting of Vic **Reeves** and Bob **Mortimer** (born 23 May 1959). They have ...

Experiment

Picture1.jpg × official

All | Images | Maps | Shopping | More | Settings

About 2 results (1.61 seconds)

Image size:
612 × 376

No other sizes of this image found.

Possible related search: ***official***

Zedd, Katy Perry - 365 (Official) - YouTube
https://www.youtube.com/watch?v=YrbgUtCfnC0
14 Feb 2019 - Zedd & Katy Perry - 365 (**Official** Music Video) Katy Perry Complete Collection on Spotify: http://katy.to/SpotifyCompleteYD Katy Perry Essentials ...

Official | Definition of Official by Merriam-Webster
https://www.merriam-webster.com/dictionary/official
3 days ago - **Official** definition is - one who holds or is invested with an office : officer. How to use

Experiment

- One for any Americans ☺
- Who’s this, and where is he from?

https://www.qthemusic.com/articles/the-latest-q/vic-bob-the-real-morrissey-hated-morrissey-the-consumer-monkey-q349-preview

Another example

https://knowyourmeme.com/memes/jake-from-state-farm
https://www.reddit.com/r/MovieDetails/comments/7vt5wh/inglourious_basterds_2009_you_can_clearly_see_the/

Other possible examples

Cultural CAPTCHAs

- Food
- Music
- Cultural norms and quirks
- Popular culture
- Education

https://www.youtube.com/watch?v=2cgRd2WJXpo

Case studies

BotSentinel.com

45%

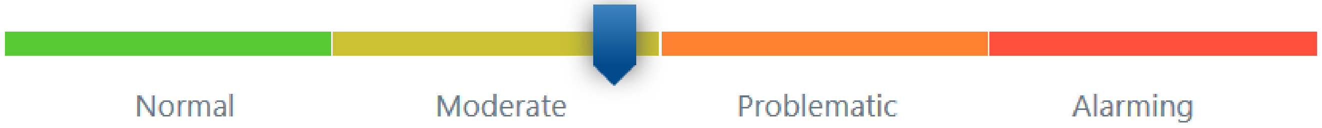
Donate

Block List

Trollbot Rating: Moderate

This report was created for [redacted]

Report created: 2019-04-23 09:50:18



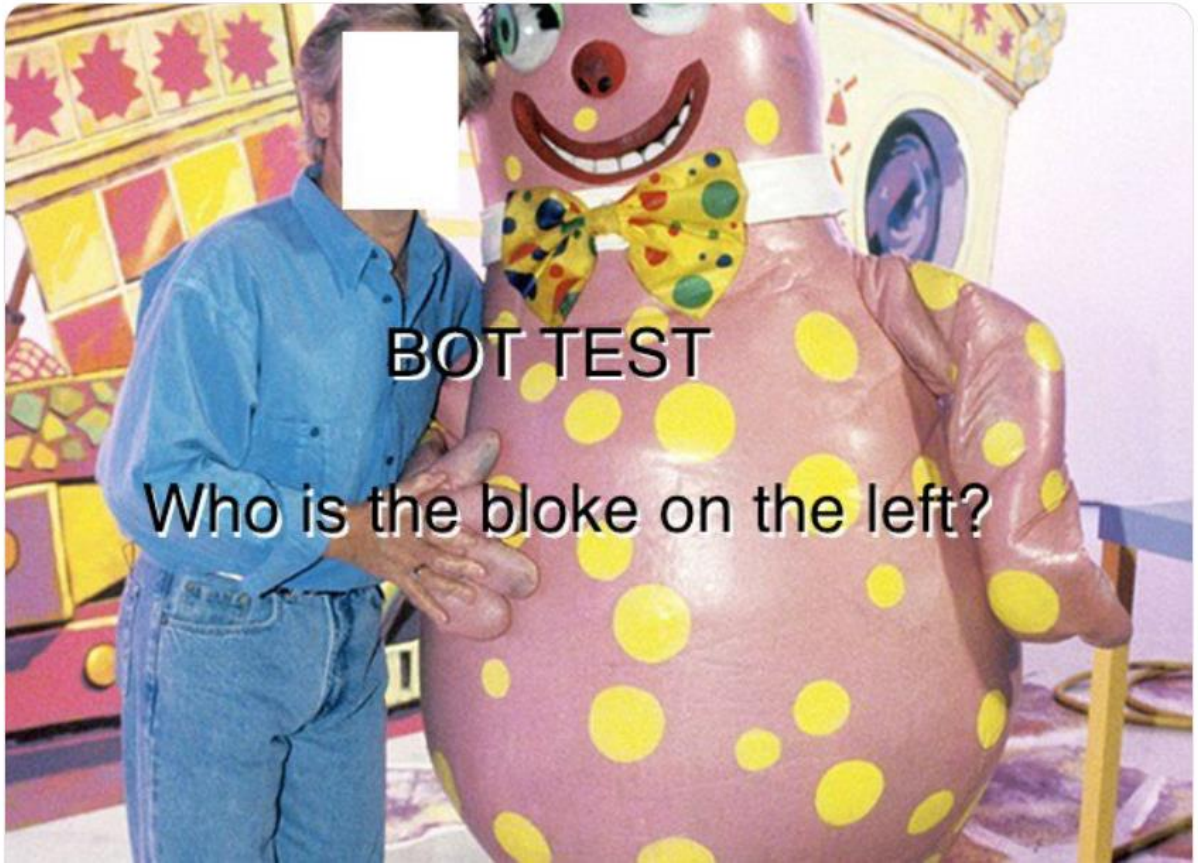
- Normal: 0% - 24%
- Moderate: 25% - 49%
- Problematic: 50% - 74%
- Alarming: 75% - 100%

Our analysis has concluded [redacted] exhibits moderate tweet activity and is not a trollbot account.

Case studies

Cultural CAPTCHAs

I have administered this test multiple times now, on multiple pro-Brexit accounts with multiple linked patterns of posting. Never gets a reply. They can't answer it.



Case studies

Cultural CAPTCHAs

[redacted] 18h
Replying to [redacted] and 3 others
You still can't, can you? Pathetic. You have no idea. You're not what you say you are, at all. It's all a lie.
1 1 3

[redacted] 18h
Replying to [redacted] and 3 others
Still can't name them! Which farm do you work in, then? How much do they pay you to fake being a Brexiter?
1 8

[redacted] 18h
I think I've found my first "click-farm" worker on Twitter. Interesting. Still can't answer the image, but very touchy about it.
1 5

[redacted] 18h
Replying to [redacted] and 3 others
That's still not the answer, is it?
2 1

[redacted] 16h
British people challenge liars when they meet them. It's part of our national identity.
1 2

[redacted] 16h
Provide irrefutable evidence that I'm not British or that I'm a bot
2

[redacted] 16h
Answer the question
1 2

[redacted] 16h
How about you just fuck off you annoying little worm
2

[redacted] 15h
Answer it.
1 1

[redacted] Follow

Replying to @iamsimonyoung @AmusingName0 and 4 others

You fuck off too

Case studies

Cultural CAPTCHAs

[redacted] · 18h
Replying to [redacted] and 3 others
Fuck off
1

[redacted] · 18h
Replying to [redacted] and 3 others
Fuck off

[redacted] · 18h
Replying to [redacted] and 3 others
Why should I tell you
1

[redacted] · 18h
Replying to [redacted] and 3 others
I know who they both are thanks

[redacted] · 18h
Replying to [redacted] and 4 others
Piss off
1

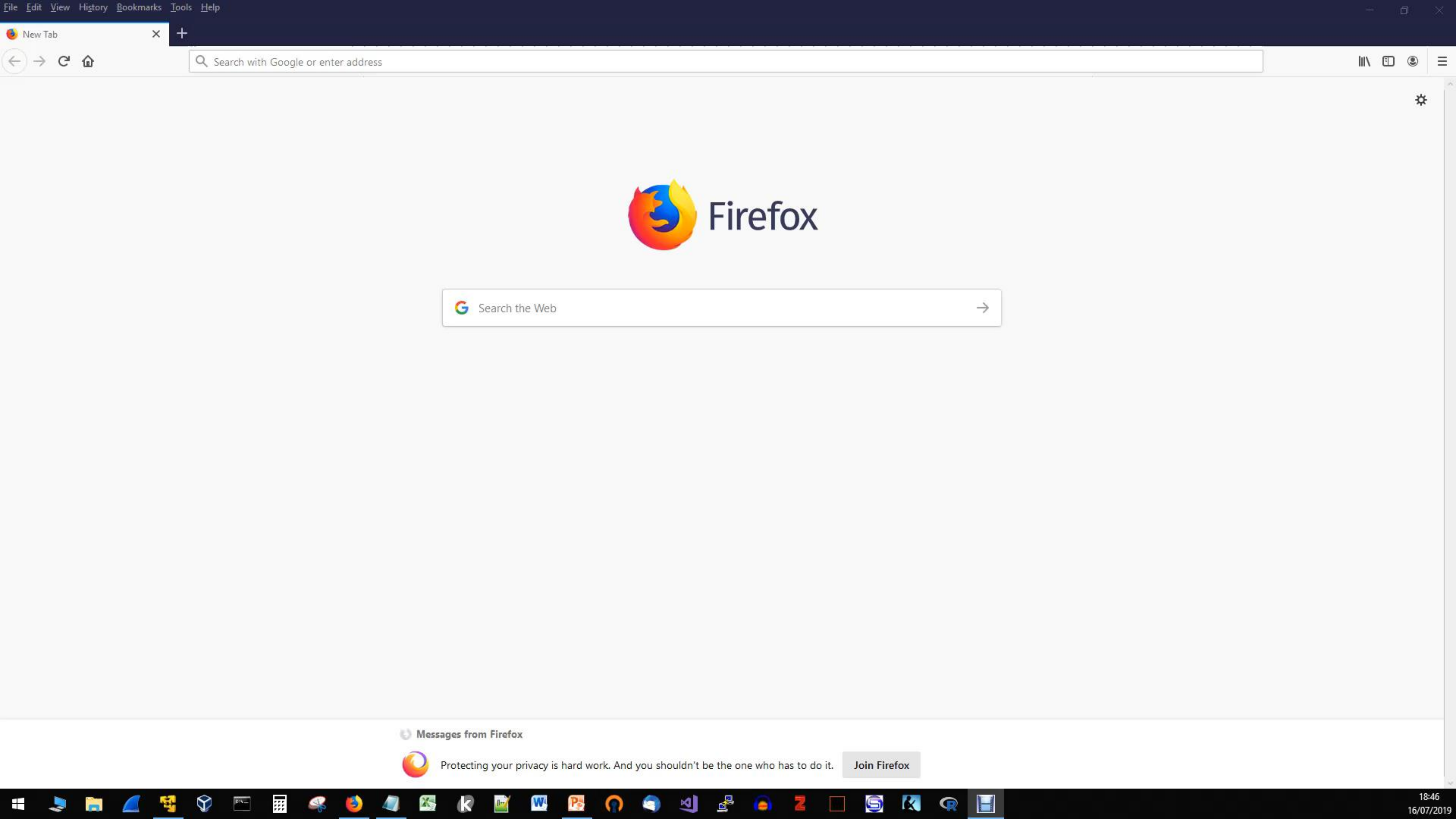
[redacted] · 18h
Replying to [redacted] and 4 others
Fuck off you little prick
3

[redacted] 18h
Replying to [redacted] and 4 others
I'm so fucking pleased for you
2

[redacted] · 18h
Replying to [redacted] and 3 others
Oh go fuck yourself
1

Applicability and approaches

- ‘CAPTCHA’-style verification system
- For accounts reported as possibly false/hostile?
- Give users option of selecting a different CAPTCHA
  – They genuinely may not know the answer!


File Edit View History Bookmarks Tools Help
New Tab
Search with Google or enter address
Firefox
Search the Web
Messages from Firefox
Protecting your privacy is hard work. And you shouldn't be the one who has to do it.
Join Firefox
18:46
16/07/2019

Caveats

- Reliant on specific cultural knowledge
  - Some may be age-dependent
  - May become increasingly hard to find examples
  - Users may genuinely not know the answer
  - cp. genuine CAPTCHAs
- Images cannot be searchable online
  - Manipulation/generation to avoid TinEye, reverse image search, etc

What can I do now?

- Come up with your own examples and implementations
- Test on social media
- Research on effectiveness at scale
- How resilient are cultural CAPTCHAs?
  – Not an area I know much about, but with click-farm workers, catfish, etc – how much research do they do into culture and language?
  – Interesting area for future work

Conclusion

Key takeaways

Conclusion

- Human side-channels offer under-explored, unconventional, and often cost-effective, opportunities for attribution and defence

- These are often specialist areas – but barrier to entry isn't as high as you might think!

- Tools and resources are available now, often open-source, to test these things out

Next steps and future research

Conclusion

- Expanding PoCs, applying techniques to more scenarios
- Other side-channels
- Further research into nature and scope of cultural CAPTCHAs
- Further research into applicability and effectiveness of forensic linguistics and behavioural signatures as investigative tools
- Automate some of this stuff, especially FL and CLA
- Get in touch! Let’s discuss ☺
- matt.wixey@pwc.com, @darkartlab

Aims - review

Conclusion

- Be aware of 3 human side-channels and how they work
- Practical takeaways for each side-channel, including tools
- Examine implications for security and privacy
- Know about possible countermeasures
- Explore future research ideas

www.pwc.co.uk

Thank you!

@darkartlab
matt.wixey@pwc.com

References

Abbasi, A., & Chen, H., 2005. Applying authorship analysis to extremist-group web forum messages. IEEE Intelligent Systems, 20(5), 67-75.

Argamon, S.E., 2016. Guccifer 2.0: Russian, not Romanian. https://multaverba.blogspot.com/2016/07/guccifer-20-russian-not-romanian.html

Bennell, C. and Canter, D.V., 2002. Linking commercial burglaries by modus operandi: Tests using regression and ROC analysis. Science & Justice, 42(3), 153-164.

Bennell, C. and Jones, N.J., 2005. Between a ROC and a hard place: A method for linking serial burglaries by modus operandi. Journal of Investigative Psychology and Offender Profiling, 2(1), 23-41.

Bouhana, N., Johnson, S.D. and Porter, M., 2014. Consistency and specificity in burglars who commit prolific residential burglary: Testing the core assumptions underpinning behavioural crime linkage. Legal and Criminological Psychology, 21(1), 77-94.

Brennan, M., Afroz, S., & Greenstadt, R., 2012. Adversarial stylometry: Circumventing authorship recognition to preserve privacy and anonymity. ACM Transactions on Information and System Security (TISSEC), 15(3), 12.

Caliskan-Islam, A., Harang, R., Liu, A., Narayanan, A., Voss, C., Yamaguchi, F., & Greenstadt, R., 2015. De-anonymizing programmers via code stylometry. In 24th {USENIX} Security Symposium ({USENIX} Security 15) (pp. 255-270).

Caliskan-Islam, A., Yamaguchi, F., Dauber, E., Harang, R., Rieck, K., Greenstadt, R. and Narayanan, A., 2015. When Coding Style Survives Compilation: Deanonymizing Programmers from Executable Binaries. arXiv preprint arXiv:1512.08546.

Check Point, 2019. Operation Tripoli. https://research.checkpoint.com/operation-tripoli/

Chiesa, R., Ducci, S. and Ciappi, S., 2008. Profiling hackers: the science of criminal profiling as applied to the world of hacking (Vol. 49). CRC Press.

Donald, I. and Canter, D., 1992. Intentionality and fatality during the King's Cross underground fire. European journal of social psychology, 22(3), 203-218.

Douglas, J.E. and Munn, C., 1992. Violent crime scene analysis: Modus operandi, signature and staging. FBI Law Enforcement Bulletin, 61(2).

Eder, M., Kestemont, M., & Rybicki, J., 2013. Stylometry with R: a suite of tools. In DH (pp. 487-488).

Filippoupolitis, A., Loukas, G. and Kapetanakis, S., 2014. Towards real-time profiling of human attackers and bot detection. http://gala.gre.ac.uk/14947/1/14947_Loukas_Towards%20real%20time%20profiling%20(AAM)%202014..pdf, accessed 05/07/2018.

Frantzeskou, G., Stamatatos, E., Gritzalis, S., Chaski, C. E., & Howald, B. S., 2007. Identifying authorship by byte-level n-grams: The source code author profile (scap) method. International Journal of Digital Evidence, 6(1), 1-18.

github.com/computationalstylistics/stylo

github.com/evllabs/JGAAP

github.com/jpotts18/stylometry

github.com/NetSPI/skl

github.com/psal/anonymouth

github.com/severinsimmler/shylo

Grubin, D., Kelly, P. and Brunsdon, C., 2001. Linking serious sexual assaults through behaviour (Vol. 215). Home Office, Research, Development and Statistics Directorate.

Hoover, D. L., 2007. Updating delta and delta prime. Graduate School of Library and Information Science, University of Illinois, 79-80.

ibm.com/uk-en/analytics/spss-statistics-software

Iqbal, F., Binsalleeh, H., Fung, B. C., & Debbabi, M., 2010. Mining writeprints from anonymous e-mails for forensic investigation. digital investigation, 7(1-2), 56-64.

Juola, P., 2009. JGAAP: A system for comparative evaluation of authorship attribution. In Journal of the Chicago Colloquium on Digital Humanities and Computer Science (Vol. 1, No. 1).

Landreth, B., 1985. Out of the inner circle: A hacker guide to computer security. Microsoft Press.

Lilly, J. C., 1972. Programming and metaprogramming in the human biocomputer. Julian P.

McDonald, A. W., Ulman, J., Barrowclift, M., & Greenstadt, R., 2013. Anonymouth revamped: Getting closer to stylometric anonymity. In PETools: Workshop on Privacy Enhancing Tools (Vol. 20).

Mokros, A. and Alison, L.J., 2002. Is offender profiling possible? Testing the predicted homology of crime scene actions and background characteristics in a sample of rapists. Legal and Criminological Psychology, 7(1), 25-43.

Olsson, J., 2009. Wordcrime: Solving crime through forensic linguistics. A&C Black.

Pearl, L., & Steyvers, M., 2012. Detecting authorship deception: a supervised machine learning approach using author writeprints. Literary and linguistic computing, 27(2), 183-196.

Pervin, L.A., 2002. Current controversies and issues in personality. 3rd ed. John Wiley & Sons.

Ramsbrock, D., Berthier, R. and Cukier, M., 2007, June. Profiling attacker behavior following SSH compromises. In 37th Annual IEEE/IFIP international conference on dependable systems and networks (DSN'07) 119-124

Rid, T. and Buchanan, B., 2015. Attributing cyber attacks. Journal of Strategic Studies, 38(1-2), 4-37

Salles-Loustau, G., Berthier, R., Collange, E., Sobesto, B. and Cukier, M., 2011, December. Characterizing attackers and attacks: An empirical study. In Dependable Computing (PRDC), 2011 IEEE 17th Pacific Rim International Symposium on Dependable Computing 174-183

Shaw, E., Ruby, K. and Post, J., 1998. The insider threat to information systems: The psychology of the dangerous insider. Security Awareness Bulletin, 2(98), 1-10.

Shoda, Y., Mischel, W. and Wright, J.C., 1994. Intraindividual stability in the organization and patterning of behavior: incorporating psychological situations into the idiographic analysis of personality. Journal of personality and social psychology, 67(4)

Silva, R. S., Laboreiro, G., Sarmento, L., Grant, T., Oliveira, E., & Maia, B., 2011. 'twazn me!!!;('automatic authorship analysis of micro-blogging messages. In International Conference on Application of Natural Language to Information Systems (pp. 161-168). Springer, Berlin, Heidelberg.

Solorio, T., Hasan, R., & Mizan, M., 2013. A case study of sockpuppet detection in wikipedia. In Proceedings of the Workshop on Language Analysis in Social Media (pp. 59-68).

Sultana, M., Polash, P., & Gavrilova, M., 2017. Authorship recognition of tweets: A comparison between social behavior and linguistic profiles. In 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC) (pp. 471-476). IEEE.

Swets, J.A., 1988. Measuring the accuracy of diagnostic systems. Science, 240(4857), 1285-1293.

Symantec, 2011. W32.Duqu: The precursor to the next Stuxnet. Symantec Corporation, California, USA. Available from https://www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/w32_duqu_the_precursor_to_the_next_stuxnet.pdf

Tabron, J. L., 2016. Linguistic features of phone scams: A qualitative survey. In 11th Annual Symposium on Information Assurance (ASIA'16).

Tonkin, M., Grant, T. and Bond, J.W., 2008. To link or not to link: A test of the case linkage principles using serial car theft data. Journal of Investigative Psychology and Offender Profiling, 5(1-2), 59-77.

Van Der Zee, S., Poppe, R., Havrileck, A., & Baillon, A., 2018. A personal model of trumpery: Deception detection in a real-world high-stakes setting. arXiv preprint arXiv:1811.01938.

Watters, P.A., McCombie, S., Layton, R. and Pieprzyk, J., 2012. Characterising and predicting cyber attacks using the Cyber Attacker Model Profile (CAMP). Journal of Money Laundering Control, 15(4), 430-441.

Wixey, M.S., 2018. Every ROSE has its thorn: The dark art of Remote Online Social Engineering. Black Hat USA 2018.

Woodhams, J. and Grant, T., 2006. Developing a categorization system for rapists' speech. Psychology, Crime & Law, 12(3), 245-260.

Zayas, V., Shoda, Y. and Ayduk, O.N., 2002. Personality in context: An interpersonal systems perspective. Journal of personality, 70(6), 851-900.