Stylometry and Anonymity

Sadia Afroz
Privacy, Security and Automation Lab
Drexel University
Robert Galbraith

The Cuckoo's Calling

"The Cuckoo's Calling reminds me why I fell in love with crime fiction in the first place"
Val McDermid
Features used

Excerpt from 'The Cuckoo's Calling':

“The buzz in the street was like the humming of the flies”
Features used

Excerpt from 'The Cuckoo's Calling':

“The buzz in the street was like the humming of the flies”

Word length distribution
Features used

Excerpt from 'The Cuckoo's Calling':

“The buzz in the street was like the humming of the flies”

Word length distribution

Freq of 1000 most common words
Features used

Excerpt from 'The Cuckoo's Calling':

“The buzz in the street was like the humming of the flies”

- Word length distribution
- Freq of 1000 most common words
- Frequency of character 4-grams
JStylo

Project Lead: Ariel Stolerman

https://psal.cs.drexel.edu/
• Okay, it worked for 4 authors
• What if there were a lot of possible authors?
Authorship Recognition: State of the art

- 100 authors, Abbasi et al., 2008
- 10,000 authors, Koppel et al., 2010
- Narayanan et al., 2012
Where else this can be useful?
Where else this can be useful?

- De-anonymizing cybercriminals
Underground Forums

5 leaked forums

(Russian)
Antichat
(May 2002-Jun 2010)

(English)
BlackHat
(Oct 2005-Mar 2008)
BadHackerZ
(Nov 2003-May 2008)

(German)
Carders
(Feb 2009-Dec 2010)
L33tCrew
(May 2007-Nov 2009)
Challenges

- Forums are multilingual
- Texts include Leetspeak (l33t, pw0ned)
- Product information vs Conversation
# Exploit coded by “Alice” and “Bob"

```perl
use LWP::UserAgent;
    $ua = new LWP::UserAgent;
    $ua->agent("Mosiac 1.0" . $ua->agent);
if (!$ARGV[0]) {$ARGV[0] = "";}  
if (!$ARGV[3]) {$ARGV[3] = "";}  
my $path = $ARGV[0] . '/index.php?act=Login&CODE=autologin';
my $user = $ARGV[1]; # userid to jack 
my $iver = $ARGV[2]; # version 1 or 2  
....
if (!$ARGV[2])
    {print "The type of the file system is NTFS.\n\n"; 
    print "WARNING, ALL DATA ON NON-REMOVABLE DISK\n"; 
    ...
```
## More accounts

<table>
<thead>
<tr>
<th>Software</th>
<th>Windows Live Messenger</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protocol</td>
<td>MSN Messenger</td>
</tr>
<tr>
<td>User</td>
<td><a href="mailto:AliceSmith@ggg.de">AliceSmith@ggg.de</a></td>
</tr>
<tr>
<td>Password</td>
<td>Alice’s_password</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Software</th>
<th>ICQ Lite/2003</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protocol</td>
<td>ICQ</td>
</tr>
<tr>
<td>User</td>
<td>123454321</td>
</tr>
<tr>
<td>Password</td>
<td>Alice’s_password</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Name</th>
<th>AliceSmith</th>
</tr>
</thead>
<tbody>
<tr>
<td>Application</td>
<td>Hotmail/MSN</td>
</tr>
<tr>
<td>Email</td>
<td><a href="mailto:AliceSmith@ggg.de">AliceSmith@ggg.de</a></td>
</tr>
<tr>
<td>Server</td>
<td></td>
</tr>
<tr>
<td>Type</td>
<td>HTTP</td>
</tr>
<tr>
<td>User</td>
<td>AliceSmith</td>
</tr>
<tr>
<td>Password</td>
<td>Alice’s_password</td>
</tr>
<tr>
<td>Profile</td>
<td></td>
</tr>
</tbody>
</table>

---
Identify Products

1. Product information has repeated patterns
2. Conversation usually has verb

Product

Bankname: XX
CCNumber: XXXXXXXX
CCHolder: XX XXXX
CCExpire: X / XXXX
CVV2: XX
Vorname: XX
Nachame: YY
Adresse: XXXXX
Stadt: XXXX
PLZ: XXXX
Land: XX
Telefon: XXXXXX-XXXXX
E-mail: [email]victim@example.com[/email]
Geburtsdatum: XX / XX / XXXX

Conversation

i am selling my large targetted email list of more than 22.5 million emails
Identify Products

1. Product information has repeated patterns
2. Conversation usually has verb

```
Bankname: XX
CCNumber: XXXXXXXX
CCHolder: XX XXXX
CCExpire: X / XXXX
CVV2: XX
Vorname: XX
Nachame: YY
Addresse: XXXXX
Stadt: XXXX
PLZ: XXXX
Land: XX
Telefon: XXXXX-XXXXX
E-mail: [email]victim@example.com[/email]
Geburtsdatum: XX / XX / XXXX
```

Product

```
i am selling my large targetted email list of more than 22.5 million emails
```

Conversation
Features used

Freq. Leetspeak

Example from Carders

| l33t down? **Neh, die Lösung!* |
| Ne klappt nit, denke mal eher das sie mal wieder DNS probleme haben |

Freq. of n-grams

Freq. of punctuations

Freq. of special characters

Function words

Parts of speech
## Results

<table>
<thead>
<tr>
<th></th>
<th>Public</th>
<th>Private</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Antichat</strong></td>
<td>44.4% (1459)</td>
<td>84% (25)</td>
</tr>
<tr>
<td><strong>Blackhat</strong></td>
<td>72% (81)</td>
<td>80.7% (35)</td>
</tr>
<tr>
<td><strong>BadhackerZ</strong></td>
<td>28.7% (69)</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Carders</strong></td>
<td>60.3% (346)</td>
<td>82.8% (210)</td>
</tr>
<tr>
<td><strong>L33tcrew</strong></td>
<td>68.8% (1215)</td>
<td>77.2% (479)</td>
</tr>
</tbody>
</table>
What about Anonymity?
Evading Authorship Recognition?

• Can we evade authorship recognition?

• Can we detect if someone is trying to evade authorship recognition?

• How to anonymize writing style?
Assumption of Authorship recognition

• Writing style is invariant.
  – It’s like a fingerprint, you can’t really change it.
WikiLeaks, a Postscript

By BILL KELLER
Published: February 19, 2012

THIS is apparently the revenge of Julian Assange: everyone who runs afoul of the rock-star leaker is condemned to spend eternity discussing the cosmic meaning of WikiLeaks. As the editor of The Times during our publication of many articles based on that treasury of military and diplomatic secrets, and as the lucky man the WikiLeaks founder singled out as his Least Favorite Journalist, I have participated in half a dozen panel discussions, and turned down at least that many. I can’t complain about the one in Madrid, where, after holding forth in a packed auditorium, the American, British, German, French and Spanish editors who broke news based on WikiLeaks commemorated the collaboration with an after-hours prowl through the Prado Museum and a 27-course meal cooked by master chef Ferran Adrià. (If Europe is dying, Spain is where I plan to go for the wake.) Unforgettable in a different way was the retrospective in Berkeley, where Assange himself, then as now awaiting an extradition ruling in England, was Skyped in on a giant screen, like the mighty Oz, to pontificate on Western media’s failure to turn the files into a kind of Nuremberg trial of American imperialism. About half the audience seemed on the verge of tossing their underwear at the screen.

Add to that the three or four documentaries on the WikiLeaks adventure, the dozen books — including, weirdly, Assange’s unauthorized autobiography — and a couple speculative Hollywood projects, in which I have a twofold interest. (1. The very slight possibility that I might make some money for my small piece of the story. 2. The exceedingly remote chance that a director will take up my wife’s brilliant idea that Assange be played by Tilda Swinton.)

It’s amazing they keep inviting me to these things, since I’m a bit of a spoilsport. My consistent answer to the
OP-ED COLUMNIST

WikiLeaks, A Post Postscript

By BILL KELLER
Published: July 29, 2012

As rumors build about the potential financial blockade against the New York Times by Visa, Mastercard, and American Express for hosting U.S. government cables published by WikiLeaks, I find myself in the awkward position of having to defend WikiLeaks. During the House Judiciary Subcommittee hearing on July 11th, several Republicans made it clear they also want New York Times journalists charged under the Espionage Act for their recent stories on President Obama’s ‘Kill List’ and secret US cyber attacks against Iran.

As those of you who have followed my turbulent relationship with WikiLeaks and its Guru-In-Chief Julian Assange know, I am first in line when it comes to distancing myself from his brand of transparency without government checks and balances. You don’t have to embrace Assange as a kindred spirit to believe that what he did in publishing those cables falls under the protection of the First Amendment. The backroom pressures by the Obama Administration’s State Department to expand its financial blockade targeting WikiLeaks to include news organizations that host information from their trove of pilfered documents goes too far.

I’ve said repeatedly, in print and in a variety of public appearances, that if the welcoming of these same U.S. cables in The New York Times is seen as controversial, it is only because the U.S. press is already facing a financial blockade from Visa, Mastercard and American Express. What worries me more is why there is still no indication of any financial threat to the New York Times or any other news organization and why Visa, Mastercard and American Express have not threatened to freeze the accounts of any of the many other news organizations publishing these cables.

And it is not only the financial blockade that raises serious questions. The New York Times is not the only news organization providing a public forum for the U.S. government cables, none of which had been published before. The site The Intercept, which has been a leader in publishing these cables, is also facing a financial blockade.

The financial blockade alone can’t explain what is happening here. It is not a war on financial institutions but an attempt to chill freedom of the press and the U.S. political process. This is a dangerous and transparent attempt to limit access to information at a time when the American people want more and more government transparency.

So I do not have the final word on this, but it is clear to me that this is an attempt to limit access to information at a time when the American people want more and more government transparency.
AS rumors build about the potential financial blockade against the New York Times by Visa, Mastercard, and American Express for hosting U.S. government cables published by WikiLeaks, I find myself in the awkward position of having to defend WikiLeaks. During the House Judiciary Subcommittee hearing on July 11th, several Republicans made it clear they also want New York Times journalists charged under the Espionage Act for their recent stories on President Obama’s ‘Kill List’ and secret US cyber attacks against Iran.

As those of you who have followed my turbulent relationship with WikiLeaks and its Guru-In-Chief Julian Assange know, I am first in line when it comes to distancing myself from his brand of transparency without government checks and balances. You don’t have to embrace Assange as a kindred spirit to believe that what he did in publishing those cables falls under the protection of the First Amendment. The backroom pressures by the Obama Administration’s State Department to expand its financial blockade targeting WikiLeaks to include news organizations that host information from their trove of pilfered documents goes too far.

I’ve said repeatedly, in print and in a variety of public appearances, that WikiLeaks is a threat to the interests of the United States. But I don’t believe that it is a threat to personal privacy or the freedom of the press. I believe it is a threat to national security and the integrity of the democratic process. I believe it is a threat to the rule of law and the balance of power between the executive and legislative branches of government.

I believe that WikiLeaks is a threat to the safety and security of the United States. I believe that WikiLeaks is a threat to the well-being of the American people. I believe that WikiLeaks is a threat to the future of our country. And I believe that WikiLeaks is a threat to the future of the world.

I believe that WikiLeaks is a threat to the future of the world. I believe that WikiLeaks is a threat to the future of the world. I believe that WikiLeaks is a threat to the future of the world.

I believe that WikiLeaks is a threat to the future of the world. I believe that WikiLeaks is a threat to the future of the world. I believe that WikiLeaks is a threat to the future of the world.

I believe that WikiLeaks is a threat to the future of the world. I believe that WikiLeaks is a threat to the future of the world. I believe that WikiLeaks is a threat to the future of the world.
Who actually did it?

WikiLeaks @wikileaks

Yes. We admit it. WikiLeaks (Assange & co) and our great supporters where behind the successful NYTimes banking blockade hoax on @nytkeller.
Authorship Recognition

- 2 random authors + Bill
Authorship Recognition

• 2 random authors + Bill

The hoax is attributed to Bill
Authorship Recognition

• 2 random authors + Bill
  The hoax is attributed to Bill

• 2 random authors + Julian’s blog + Bill
Authorship Recognition

• 2 random authors + Bill
  The hoax is attributed to Bill

• 2 random authors + Julian’s blog + Bill
  The hoax is attributed to Bill
Wrong Assumption!

• Imitation or framing attack
  – Where one author imitates another author
• Obfuscation attack
  – Where an author hides his regular style
Imitating Cormac McCarthy

“On the far side of the river valley the road passed through a stark black burn. Charred and limbless trunks of trees stretching away on every side. Ash moving over the road and the sagging hands of blind wire strung from the blackened lightpoles whining thinly in the wind.”
Obfuscating writing style

Your goal is to fool the computer into thinking that your passage was **NOT** written by you.
Authorship Recognition
Accuracy on Regular Documents

- 9-Feature (NN)
- Synonym-Based
- Writeprints Baseline (SVM)
- Random
Authorship Recognition
Accuracy on Regular Documents

More than 80% accurate in detecting authorship with 40 authors in regular document.
Authorship Recognition Accuracy on Obfuscated Documents

- 9-Feature (NN)
- Synonym-Based
- Writeprints Baseline (SVM)
- Random
Authorship Recognition Accuracy on Obfuscated Documents

Accuracy is less than random chance in obfuscated documents
Authorship Recognition Accuracy on Imitated Documents

- 9-Feature (NN)
- Synonym-Based
- Writeprints Baseline (SVM)
- Random

Number of Authors
Authorship Recognition Accuracy on Imitated Documents

Accuracy is less than random chance in imitated documents
Evading Authorship Recognition?

- Can we evade authorship recognition?

- Can we detect if someone is trying to evade authorship recognition?

- How to anonymize writing style?
Detecting Deception

1. Create models of deceptive and regular documents
2. Identify deceptive documents
Detecting Deception

1. Create models of deceptive and regular documents
2. Identify deceptive documents
Detecting Deception

1. Create models of deceptive and regular documents
2. Identify deceptive documents
Detecting Deception

1. Create models of deceptive and regular documents
2. Identify deceptive documents
Detecting Deception

1. Create models of deceptive and regular documents
2. Identify deceptive documents
Dataset

- Imitation and Obfuscation dataset
- Imitating Hemingway and Faulkner
- Hoax blog: A Gay Girl In Damascus
Result

- Imitation and Obfuscation dataset: > 85%
- Imitating Hemingway and Faulkner: > 88%
- Hoax blog: A Gay Girl In Damascus: Failed
Information Gain vs Feature Change

- **Left Graph:** Information gain ratio vs Change in n-grams
  - X-axis: Change in n-grams
  - Y-axis: Information gain ratio

- **Right Graph:** Information gain ratio vs Change in function words
  - X-axis: Change in function words
  - Y-axis: Information gain ratio
Information Gain vs Feature Change

Graphs showing the relationship between change in n-grams and information gain ratio, as well as change in function words and information gain ratio.
Evading Authorship Recognition?

• Can we evade authorship recognition?

• Can we detect if someone is trying to evade authorship recognition?

• How to anonymize writing style?
Original text:
"Lunch Time With Smosh."
...35 translations later, Bing gives us:
"Many cats"

Original text:
"9gag - just for fun."
...35 translations later, Bing gives us:
"Only the pool entertainment (9)"

Original text:
"9gag - just for fun."
...35 translations later, Bing gives us:
"In addition, it is interesting."

Original text:
"Now I wonder if those guys couldn't make a bigger mess."
Why not machine translation?

They passed through the city at noon of the day following.

(German)

(Japanese)
Why not machine translation?

They passed through the city at noon of the day following.

(German)

(Japanese)

They passed the city at noon the following day.
Why not machine translation?

Just remember that the things you put into your head are there forever, he said.

(German)

(Japanese)
Why not machine translation?

Just remember that the things you put into your head are there forever, he said.

(German)

(Japanese)

You are dead, that there always is set, please do not forget what he said.
Why not machine translation?

Translation alone cannot anonymize text
Anonymouth

Project lead: Andrew McDonald

https://psal.cs.drexel.edu/
How Anonymouth works

1. Background corpus
2. Document
How Anonymouth works

1. Background corpus
2. Document
3. Is the document anonymized? (Jstylo)
How Anonymouth works

1. Background corpus
2. Document
3. Is the document anonymized? (Jstylo)

If yes, the process ends. If no, proceed to:

Which feature to change (Rank features)
How Anonymouth works

1. Background corpus
2. Document
3. Is the document anonymized? (Jstylo)
4. Which feature to change (Rank features)
5. How to change feature (Feature clustering)

YES!

Document
How Anonymouth works

1. Background corpus
2. Is the document anonymized? (Jstylo)
3. Which feature to change (Rank features)
4. How to change feature (Feature clustering)
5. Shows suggestion
How Anonymouth works

1. Background corpus
2. Document
3. Is the document anonymized? (Jstylo)
   - Yes! Document
   - No
4. Which feature to change (Rank features)
5. Shows suggestion
6. How to change feature (Feature clustering)
Summary

- Writing style can be used to recognize author of a document
- Writing style can be changed by imitating other people’s writing style
- JStylo and Anonymouth can help in changing writing style
My co-authors:

Rachel Greenstadt  
Mike Brennan  
Ariel Stolerman  
Aylin Caliskan Islam  
Andrew McDonald  
Damon McCoy