Behavioural Profiling of Darknet Marketplace Vendors

Student: Sylvester Shan
UID: u6049249
Course code: COMP4560
Supervisor: Dr Ramesh Sankaranarayana
What is the Darknet?
Darknet

• Internet has three sections:
  – Clear net
  – Deep net
  – Dark net (subset of Deep net)

• Important properties
  – Fully anonymous and protects identity

• Why cannot shut it down?
  – Short answer: Network of hidden computers/servers
Avoid darknet at all cost?

Property of darknet benefits different groups of people:

– Journalist
– Platform for whistleblowers to spread information
Darknet Marketplace
What can you buy?

• Illegal drugs
  – Cocain
  – Heroin
  – Weed
  – COVID-19 vaccines

• Sensitive information (collected by hackers)
  – Zoom accounts and password
  – Credit card information

• Weapons
  – Guns
  – Explosives
Small recap
Darknet and marketplace

- Vendors are fully anonymous
- Illegal goods can be easily bought by anyone
- Darknet and marketplaces cannot be shut down easily
Statistics from other research

"Average life span of a darknet marketplace is 8 months"[1]

"Top 1% most successful vendors are responsible for 51.5% transactions"[2]

What happens to vendors when a marketplace shuts down?
MARKET A

Got the usual? YEP

MARKET A

MARKET B

SEIZED

MARKET B

Do I know you? Long time no...

MARKET B

*leaves*

BATHROOM

* sob *

* heart broken *
Ideally vendors need to allow returning customers to identify them on different marketplaces, but how?
Hypothesis

Assume vendors uses their username as a "brand" on different marketplaces
If hypothesis is true

• Simplify vendor correlation methods
• Help understand vendors movement between marketplaces
• Help provide evidence for law enforcement agencies
Procedure for testing hypothesis

1. Process raw data set
2. Correlate accounts in data set
3. Compute username similarity of correlated accounts
1. Process raw data set

- Data sets are from public data dump
- Used marketplace data set:
  - Evolution
    - 21-1-2014 ~ 17-03-2015
  - Silkroad2
    - 20-12-2013 ~ 06-11-2014
1. Process raw data set

Collect relative information

– Username
– PGP encryption key
– Dates
– Other potential useful attributes:
  • Eg listed item names, description for each item
2. Correlate accounts

• Method: Comparing PGP keys
  – PGP keys are long random string

<table>
<thead>
<tr>
<th>PGP</th>
<th>Username 1</th>
<th>Username 2</th>
<th>similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>p1</td>
<td>BossTweed</td>
<td>boss-tweed</td>
<td></td>
</tr>
<tr>
<td>p2</td>
<td>VendingSolutions</td>
<td>solutionsforvendors</td>
<td></td>
</tr>
<tr>
<td>p3</td>
<td>MeyerLansky2</td>
<td>arch</td>
<td></td>
</tr>
</tbody>
</table>
3. Username similarity of match accounts

• Username feature vector
  – Bi-grams
  – basic features
    • Length of username
    • Number of capital letters
    • Etc

<table>
<thead>
<tr>
<th>PGP</th>
<th>Username 1</th>
<th>Username 2</th>
<th>similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>p1</td>
<td>BossTweed</td>
<td>boss-tweed</td>
<td>1</td>
</tr>
<tr>
<td>p2</td>
<td>VendingSolutions</td>
<td>solutionsforvendors</td>
<td>0.689</td>
</tr>
<tr>
<td>p3</td>
<td>MeyerLansky2</td>
<td>arch</td>
<td>0.0</td>
</tr>
</tbody>
</table>
Experiment and Results
## Matching PGP statistics

<table>
<thead>
<tr>
<th></th>
<th>Evo + Evo</th>
<th>SR2 + SR2</th>
<th>Evo + SR2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Numer of unique accounts</strong></td>
<td>4366</td>
<td>1225</td>
<td>5591</td>
</tr>
<tr>
<td><strong>Numer of match accounts</strong></td>
<td>123</td>
<td>8</td>
<td>358</td>
</tr>
<tr>
<td><strong>Percentage of vendors with multiple accounts</strong></td>
<td>2.82 %</td>
<td>0.65%</td>
<td>6.40%</td>
</tr>
</tbody>
</table>

Vendors are more likely to create different accounts in different marketplaces
### Evolution and Evolution (PGP key match)

<table>
<thead>
<tr>
<th>Similarity score threshold</th>
<th>Percentage of match accounts above threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>13.82 %</td>
</tr>
<tr>
<td>0.95</td>
<td>13.82 %</td>
</tr>
<tr>
<td>0.90</td>
<td>13.82 %</td>
</tr>
<tr>
<td>0.85</td>
<td>14.63 %</td>
</tr>
<tr>
<td>0.80</td>
<td>16.26 %</td>
</tr>
<tr>
<td>0.75</td>
<td>17.07 %</td>
</tr>
<tr>
<td>0.70</td>
<td>17.07 %</td>
</tr>
</tbody>
</table>

Out of 123 vendors with multiple accounts, 17 used exact same usernames
SilkRoad2 and SilkRoad2 (PGP key match)

<table>
<thead>
<tr>
<th>Similarity score threshold</th>
<th>Percentage of match accounts above threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25.0 %</td>
</tr>
<tr>
<td>0.95</td>
<td>25.0 %</td>
</tr>
<tr>
<td>0.90</td>
<td>25.0 %</td>
</tr>
<tr>
<td>0.85</td>
<td>25.0 %</td>
</tr>
<tr>
<td>0.80</td>
<td>25.0 %</td>
</tr>
<tr>
<td>0.75</td>
<td>25.0 %</td>
</tr>
<tr>
<td>0.65</td>
<td>37.5 %</td>
</tr>
</tbody>
</table>

Out of 8 vendors with multiple accounts, 2 used exact same usernames
### Evolution and Silkroad2 (PGP key match)

<table>
<thead>
<tr>
<th>Similarity score threshold</th>
<th>Percentage of match accounts above threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>83.24 %</td>
</tr>
<tr>
<td>0.95</td>
<td>83.24 %</td>
</tr>
<tr>
<td>0.90</td>
<td>83.52 %</td>
</tr>
<tr>
<td>0.85</td>
<td>84.64 %</td>
</tr>
<tr>
<td>0.80</td>
<td>85.75 %</td>
</tr>
<tr>
<td>0.75</td>
<td>85.75 %</td>
</tr>
<tr>
<td>0.70</td>
<td>86.87 %</td>
</tr>
</tbody>
</table>

Out of 358 vendors with multiple accounts, 298 used exact same usernames.
Conclusion

- Vendors create new accounts on different marketplaces.
- If humans consider usernames with similarity score greater or equal to 0.85 are the same, 84.64% vendors will create a new account using the same username.
Limitation

- Increase number of matching usernames
- Correlation methods using other attributes
  - Compare output with ground truth
- Only used two marketplace data sets

<table>
<thead>
<tr>
<th>PGP</th>
<th>Username 1</th>
<th>Username 2</th>
<th>similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>p1</td>
<td>BossTweed</td>
<td>boss-tweed</td>
<td>1</td>
</tr>
<tr>
<td>p2</td>
<td>VendingSolutions</td>
<td>solutionsforvendors</td>
<td>0.689</td>
</tr>
<tr>
<td>p3</td>
<td>MeyerLansky2</td>
<td>arch</td>
<td>0.0</td>
</tr>
</tbody>
</table>
Future work

• Create a database, such that it's possible to update the database using newly scrapped webpages
• Time series analysis
• Consistent guide for web-scraper
• Improve correlation methods based on ground truth, which could be further used as training and testing data set for developing machine learning models
Special Thanks

Dr Ramesh Sankaranarayana
Thanks for Listening
Questions?