What is an “Insider Threat”? 

- Malicious Insider 
  - a current or former employee, contractor, or business partner who meets the following criteria: 
    - has or had authorized access to an organization’s network, system, or data 
    - has intentionally exceeded or intentionally used that access in a manner that negatively affected the confidentiality, integrity, or availability of the organization’s information or information systems 

- Can also be inadvertent (non-malicious) 

Source: The CERT Insider Threat Center
Conversations around Insider Threat

- Why look at public conversation? Unlikely to find any insider threats...
- ...but, there may be actors trying to shape the conversation to their own ends – corporations, nation states, etc.
- Understanding the conversation will lead to informed research
- **Research question**: Can dynamic network analysis be used to discover the nature of public conversations around insider threat and related organizational threats?

<table>
<thead>
<tr>
<th>Category</th>
<th>Hashtags</th>
</tr>
</thead>
<tbody>
<tr>
<td>General</td>
<td>#insiderthreat</td>
</tr>
<tr>
<td></td>
<td>#insiderattack</td>
</tr>
<tr>
<td></td>
<td>#cyberespionage</td>
</tr>
<tr>
<td></td>
<td>#dataloss</td>
</tr>
<tr>
<td>Corporate</td>
<td>#industrialespionage</td>
</tr>
<tr>
<td></td>
<td>#tradesecrets</td>
</tr>
<tr>
<td></td>
<td>#embezzlement</td>
</tr>
<tr>
<td></td>
<td>#embezzling</td>
</tr>
<tr>
<td>Nation-State</td>
<td>#militarysecrets</td>
</tr>
<tr>
<td></td>
<td>#spy</td>
</tr>
<tr>
<td></td>
<td>#spying</td>
</tr>
<tr>
<td></td>
<td>#spies</td>
</tr>
</tbody>
</table>
Collection Method

- Use Python package twarc to retrieve tweets from Twitter Search API based on hashtag query
- Tweets collected between March 27th and April 15th 2020 (data has some gaps)
- Import Twitter JSON data into ORA – ORA handles creating derived networks and basic stats.
- Use ORA for reporting and visualization

Data Description

- 5 nodesets: Agent, Hashtag, Location, Tweet, URL

<table>
<thead>
<tr>
<th>Network</th>
<th>Twitter JSON AE Hashtags</th>
</tr>
</thead>
<tbody>
<tr>
<td>First tweet date</td>
<td>2020-04-13 07:06:07-04</td>
</tr>
<tr>
<td>Last tweet date</td>
<td>2020-04-13 08:45:03-04</td>
</tr>
<tr>
<td>Number of tweets</td>
<td>13840</td>
</tr>
<tr>
<td>Number of tweets with geo</td>
<td>15</td>
</tr>
<tr>
<td>Number of tweets with URL</td>
<td>4939</td>
</tr>
<tr>
<td>Number of mentions</td>
<td>1521</td>
</tr>
<tr>
<td>Number of retweets</td>
<td>628</td>
</tr>
<tr>
<td>Number of verified users</td>
<td>145</td>
</tr>
<tr>
<td>Number of users agency mention</td>
<td>17</td>
</tr>
<tr>
<td>Number of citations</td>
<td>4233</td>
</tr>
<tr>
<td>Number of distinct hashtags</td>
<td>6783</td>
</tr>
<tr>
<td>Number of distinct hashtags used more than once</td>
<td>3212</td>
</tr>
<tr>
<td>Number of distinct retuits</td>
<td>0</td>
</tr>
<tr>
<td>Number of distinct retuits used more than once</td>
<td>0</td>
</tr>
<tr>
<td>Number of distinct locations</td>
<td>9</td>
</tr>
</tbody>
</table>
Inside Threat Tweets

ALL CATEGORIES
Overall – Super Spreaders

Overall – Super Friends
Overall Takeaways

• Difficult to find anything of note in the whole collection
• “Spy” hashtag has a lot of out-of-scope discourse
  – Movie and TV
  – Video games (Team Fortress 2)
  – Novels, books, stories, etc.
  – ES Futures vs SPY (refuse to look deeper into this)

Inside Threat Tweets

“GENERAL” GROUPING
Insider Threat - General – Super Spreaders
Insider Threat - General – Super Friends

High degree centrality suspended user
Bot or not?

Twinybots
Artificial Intelligence made to help with social media

Inside Threat Tweets
“CORPORATE” GROUPING
Insider Threat - Corporate – Super Spreaders

Carnegie Mellon

11 June 2020  Osterritter

11 June 2020  Osterritter
"NATION-STATE" GROUPING

Inside Threat Tweets

Osterritter
Insider Threat - Nation – Super Spreaders
Findings

- Much of the conversation around insider threat are news aggregators and companies marketing services
- ...but, there is more to do!
Next Steps

- Bot analysis
- NetMapper
- Network comparison (corporate vs nation-state vs general)
- Get list of disabled users in data collected

Future Work
- Explore other hashtags (APT28, APT29, APT41, etc.)
- Possibly cross-reference with other social media (Facebook, YouTube) Maltego?

Questions for future thought

- What other insights would be useful to show?
  - Other analyses from ORA Twitter report?
  - Other network visualizations?
- What would we want to know about this conversation?
  - Possibly: Geographic or group attribution of conversation drivers – how to divine this?
  - What companies are present here?
- Best practices for analyzing a conversation?
  - Overall methods to go from large set of Twitter data to meaningful insights
ORA Walkthrough
ORA Walkthrough

* Can choose to anonymize tweeter names if needed for real data

'Derived Networks' tab - you can choose non-default networks if desired.

At this point, click 'Finish' to import your data
Select ‘Hashtag x Hashtag – Co-occurrence’ network, then choose ‘Visualize this Network’.
ORA Walkthrough

[Image: ORA Walkthrough diagram]

ORA Walkthrough

[Image: ORA Walkthrough diagram]
ORA Walkthrough

* Leave defaults for initial exploration
ORA Walkthrough

Report will save to local machine and open in default web browser

Explore Data Statistic, Super Friends report, and Super Spreaders report