Cybercriminal Markets

Assignment 1 – DRAFT

Note: This is a draft form. The final version will follow a traditional report structure with the following outline:

i. Introduction to Cybercriminal Markets Analysis
    a. Cybercriminal markets field
    b. Data
    c. Existing NLP processing
ii. Security Metrics - intro
    a. Intro to metric types
    b. Existing use in cybercriminal markets field
iii. Security Metrics – ideal
    a. Ideal metrics and how they are defined for our data set
    b. Evaluation of metrics

We are continuing to add information and will send more content by Tuesday at noon. If you have a chance to review that content as well before our session on Thursday morning, we would be grateful, however we understand that the new content will not be ready at the deadline stated of today at noon.

1. **What security issue does the data speak to?**
   The data focuses on forums known for underground cyber activities for posts related to criminal activity. This data was collected either through scraping or from complete database dumps. Forums include Blackhat World, Darkode, Hack Forums, Nulled, Antichat, Carders, and L33tCrew. The data was collected over various time periods of forum activity, and focused on commerce-related threads (or the buying and selling of goods or services).

   After the initial collection process, a natural language processing (NLP) algorithm was applied to extract data that focuses on the buying and selling of products on these forums. Our analysis will focus on this annotated and processed data regarding that topic.

   The data collected is a part of the Automated Analysis of Cybercriminal Markets project, a collaborative study between multiple US universities and institutes. This data, including background on how it was collected and processed, can be found here: https://evidencebasedsecurity.org/forums/.

2. **What would be the ideal metrics for security decision makers?**
   In order to determine what ideal metrics should be, it is important to determine who are the relevant security decision makers. We are focusing on the interest a national government-based security organization has in buying and selling activity on these forums.

There are three major categories of products to consider: those that are strictly illegal to purchase in the relevant country, those that are legal but often used in nefarious activities, and those that have little to no legal question. These three product categories will be helpful to delineate between when considering specific metrics.

Ideal metrics include tracking of specific account and overall activity over the time period of collection, with comparison of normalized and unnormalized activity over time periods that overlap between forum collections.

3. **What are the metrics that exist in practice?**
   According to the Tools for Automated Analysis of Cybercriminal Markets Report by Portnoff et. al. three metrics are defined to conduct the analysis:

   Post Type. Determines the nature of the post, specifically, whether it is an offer to buy, offer to sell, offer to exchange currency, or a post not related to trade.
   Product. Determines the product being offered or requested (buy/sell posts) or the currencies being exchanged (currency exchange posts).
   Price. Extracts the price offered in commerce (buy/sell) posts or the exchange rate (currency exchange posts).

4. **A definition of the metrics you can design from the dataset**

   To be done later (how it is calculated in our dataset)

5. **An evaluation of the metrics you have defined. This should include graphical representations of the metrics (e.g., histograms, scatter plots, time series, bar charts).**

   To Be Done Later (Python or R)