Assignment 1

Joeri Kock	s1440195	Joost Jansen	s1370030
David Stritzl	s1360752	Koen de Jong	s1367285
Ramon Houtsma	s1245228		

What security issue does the data speak to?

The existence of a large, anonymous, online marketplace creates incentives and facilitates for cyber criminals to illegally buy and sell items on cryptomarkets, thus resulting in the causing and maintaining a community of cyber criminals. Cryptomarkets decrease the effort for criminals to buy and sell illegal items and anonymity decreases the risk of getting caught. Both are commonly used factors to define opportunity reducing/increasing techniques. This imposes a security issue for governments, because the cryptomarkets encourage criminals in their country to buy and sell illegal items anonymously, leading to a potential increase of illegal activity as well as an increase of effort to track crime and consequently a decrease of public safety. However, in certain cases, tolerating cryptomarkets may have a positive impact on public safety as they may keep criminals of the streets [1]. Furthermore, since these marketplaces are not government controlled and most of the products being sold are illegal, the transactions made via these marketplaces cannot be taxed by government instances.

What would be the ideal metrics for security decision makers?

Depending on the case, a lot of different ideal metrics can be specified. Below is an list of metrics that would be ideal for the Silkroad 1 case.

- Identity of buyer: useful for knowing what kind of people buy goods on this illegal marketplace. Also useful for law enforcement in order to prosecute a suspect.
- Incentive of buyer: what will the customer do with the goods? Is this a problem for public safety?
- Identity of seller: this might be of interest when law enforcement wants to prosecute the sellers of illegal goods.
- Identity of owner of marketplace: the person/persons hosting the marketplace are accountable for the consequences of running and maintaining an illegal marketplace.
- Transaction techniques (money and goods): which techniques are being used and why? Do they provide enough anonymity and/or do these techniques differ from a normal legal marketplace?
- Shipment address: to which places are the goods being transported? Are there addresses that occur very frequently? Maybe patterns can be derived from the addresses.
- Location of servers of marketplace: it is interesting to know in which country the marketplace is hosted for prosecution purposes.
- Flow of money and goods: very useful in order to create an overview or map of the marketplace. Follow the money; find out who are the large volume sellers and to whom the money is flowing.

- Monetary loss for government and negative impact on 3rd parties per product category.

What are the metrics that exist in practice?

In practice, there is an almost infinite amount of metrics that can be used to gather intelligence about cryptomarkets or cyber crime in general. A few commonly used metrics are listed below.

- Usernames, IP addresses and other user data: data of users on such can be used to identify individuals, track them down on different websites and find related persons.
- Cryptocoin wallet IDs: having the ID of a wallet can give insights into the wallet's current balance and what and how many monetary streams the target was involved in.
- Shipment data: the source and destination of shipments can be used to identify problematic countries and gather insights into criminal organizations.
- Product categories and pricing: these give insights into what the market looks like and what the incentives of individuals using these crypto markets are.
- Forum listings, user profiles and customer feedback: communication logs can give insights into the cyber criminals incentives and connections. Possibly these logs can also contain information that can used to identify or locate said cyber criminals.
- Amount of goods in stock and timing of transactions: this information can give insights into the market supply and demand.

A definition of the metrics you can design from the dataset

The dataset contains, among others, dumps of the Silkroad 1 main webstore, the forum and the wiki, as well as data mined from these dumps.

One of the resources mined from the main site is a database containing three tables; item, price and feedback. This database provides us with several interesting columns about the products featured on the Silkroad 1 marketplace and the transaction that were made. The tables and corresponding columns are shown below.

tablename	**columns**
item	item_id | seller | ships_to | ships_from | category | first_seen | last_seen
price	item_id | price | time
feedback	item_id | feedback_time | feedback_rating | feedback_hash

A lot of metrics can be derived from this database, which we will now describe.

- Look at the progression of the price of a certain product. (TABLE: price)
- An approximation of the quality of the products by looking at the average feedback rating. (TABLE: feedback). Also, in combination with a seller (TABLE: item).
- Top 10 shipping countries. (TABLE: ships_to and ships_from)

- By looking at the last_seen field in the table 'item', we can determine if the product can be trusted (e.g. it it was last seen by the crawler 2 years ago, the product is probably not being sold anymore).

Another resource contains a part of the Silkroad 1 forum listings. It includes, among others, the username, post title, post content and post data. Using this data, connections between users can be sketched and the development of community over time can be assessed.

However, as the forum and wiki mostly contain very unstructured data, eg. forum posts, mining useful information on a large scale is difficult.

Lastly, other than this data set, the government has access to a great amount of data from, for instance, law enforcement such as information about criminals. Together with the above-mentioned data set, the demand and supply can be sketched and the impact of each sold product category on public safety can be assessed.

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

- For determining the progression of the price, a time series graph could be made showing the price of the item over time. The progression of the price is an important metric since it reveals which product go up in price or which products have become cheaper overtime. These products are of interest since obviously, a change in supply and demand has occurred.
- The average feedback rating of a product is interesting because it shows which products are rated very well and very poor. This is particularly interesting in combination with the seller, since it reveals which sellers have a good reputation and which don't.
- The Top 10 shipping countries will be displayed in a simple table. Because this is a ranking, it's not very useful to put this into a graphical representation. This ranking is important, since it lists the countries who accommodate the illegal trade of goods. Law enforcement and other instances can then focus on those countries
- The last_seen column can be presented in a bar chart in a way that a difference can be made between goods that are still being sold and goods who are not being sold anymore.

With this information, the security decision makers can keep an eye on the metrics and decide which goods, sellers or ratings are of particular interest. Together with information from law enforcement, a risk analysis on public safety can be done from which a course of action can be defined.

References

1. Nasseri, Nasseri. An investigation of cryptomarkets: assessing the online drugs trade from the perspectives of Australian health and law enforcement agencies. 2015. Available at: http://hdl.handle.net/1959.14/1050630