Data Capture and Analysis of Darknet Markets

Цель  
The Australian National University’s (ANU) Cybercrime Observatory8 provides a method for ‘crawling’ (the automated act of exploring a site to discover and download website pages) a darknet market, and ‘scraping’ (the automated act of extracting a HTML file to export the information contained) the relevant market, product or vendor information for analysis. These methods facilitate the collection of the following: overview of product listings; prices and quantities of different products, shipping information (where available), and other parameters that vary from one market or product listing to another; and, vendor information, including: alias, number of completed sales, stated vendor location (where available) and PGP key.

While the methodology is illustrated on the Apollon9 darknet market, the method can be generalized to fit other marketplaces, and has been successfully applied to 42 omnibus and niche markets observed by the Cybercrime Observatory (see Appendix 1). Building upon the existing crawling and scraping procedures, a naive image capture method can also be implemented using the image links procured through the aforementioned crawling and scraping methods.

Метод

Data Capture Data capture from websites is made easier when utilising ‘crawler’ and ‘scraper’ technologies. Such tools are commonly used in the clearnet or open web, and web-crawlers have been widely used since the 1990s such as the web robot and W3Catalog.

The ANU’s Cybercrime Observatory has written a specific set of crawlers designed to capture product and vendor information from Tor darknet markets. The process through which these crawlers operate is typically a three-step process: perform a ‘page crawl’ of a listing of products (and scraping of the page(s) for later purposes); then a ‘product crawl’ based on the scraped listing pages; and finally a ‘vendor crawl’ based on the scraped product listing pages.

Анализ данных

**Data Analysis For data analysis, the Jupyter Notebook20 application is employed. The Jupyter Notebook reads the data from a CSV file into a ‘Data Frame’ (a two dimensional data structure; a column and a row of equal length featured within the Python Pandas module).**

**Результат**

**The current iteration of darknet data capture tools are operational and usually reliable, however, there are limitations such as prolonged downtime, DDoS and other operational security defences can amplify problems of missing and incomplete data.**

**Чтобы усилить оперативную безопасность, чтобы избежать потенциальных событий DDoS и исключить использование автоматических ботов и других систем отслеживания, многие даркнеты использовали использование автоматического тайм-аута для учетных записей. Поскольку для просмотра списков продуктов требуется учетная запись, тайм-аут может стать проблематичным в зависимости от времени, затраченного на выполнение обхода.**