Analyzing Bitcoin & Detecting Addresses Related to Ransomware Transactions

1.introdction

With the introduction of cryptocurrency hacking has become one of the most harming crimes in recent years. Tracing the responsible actors of these crimes is hard due to the pseudo-anonymity of such currency

1.1 Background

Creating a bitcoin address to receive and transfer funds is easy and free and there is no limit to how many wallets one can get, nor it needs any identity verification. This ease of transaction has made it very appealing for criminals and hackers who take advantage of technical vulnerabilities and wants to move their funds anonymously through the network. This increased the number of ransomwares, scamming and phishing.

1.2 Motivation

To analyze transactions and trace stolen funds via the suspicious addresses have been related to previous ransomware transactions.

2.Problem Statement

Taking speed action (Ban, Blacklisting) after identifying addresses that have been used for malicious acts to prevent future scams from these addresses.

3.Objective

- To identify addresses that was involved in ransomware transactions.
- Preventing and decreasing future fraud and scam acts.

4.Dataset Description

Data Set Information

A parsed dataset of the entire Bitcoin transaction graph from 2009 January to 2018 December.
Using a time interval of 24 hours, and daily transactions.
Ransomware addresses are taken from three widely adopted studies: Montreal, Princeton and Padua[1].

Attribute Information

Features

Address: String. Bitcoin address.
Year: Integer. Year.
Day: Integer. Day of the year. 1 is the first day, 365 is the last day.
Length: Integer.
Weight: Float.
Count: Integer.
Looped: Integer.

Neighbors: Integer.
Income: Integer. Satoshi amount (1 bitcoin = 100 million satoshis).
Label: Category String. Name of the ransomware family (e.g., Cryptxxx, cryptolocker etc) or white (i.e., not known to be ransomware).

5.Requierments

Tools

- Anaconda3
- Jupyter Notebook
- Python

Libraries

- Pandas for data analysis and manipulation.
- Numpy for mathematical functions.
- Matplotlib for data visualization.
- Seaborn for data visualization.
- Datetime for manipulating date and time.

6.MVP

A clean dataset after performing some EDA processes that is ready to use for the next step.

7.Referneces

[1] Archive.ics.uci.edu. 2021. UCI Machine Learning Repository: BitcoinHeistRansomwareAddressDataset Data Set. [online] Available at: <https://archive.ics.uci.edu/ml/datasets/BitcoinHeistRansomwareAddressDataset#> [Accessed 2 October 2021].