Datahacks: Bitcoin Heist

Cedric Zhao

A New Cybersecurity Threat

Bitcoin, like other cryptocurrencies, allows cybercriminals to receive funds with a high degree of anonymity, making transactions difficult to track. Bitcoin gained notoriety as the common currency of the Dark Web, where it remains popular. It is seen as the essential cryptocurrency — easy to acquire and use, making threat actors believe victims will be more likely to pay.

https://www.marsh.com/us/insights/research/ransomware-paying-cyber-extortion-demands-in-cryptocurrency.html

BitcoinHeistData.csv

We have downloaded and parsed the entire Bitcoin transaction graph from 2009 January to 2018 December. Using a time interval of 24 hours, we extracted daily transactions on the network and formed the Bitcoin graph. We filtered out the network edges that transfer less than B0.3, since ransom amounts are rarely below this threshold.

Ransomware addresses are taken from three widely adopted studies: Montreal, Princeton and Padua. Please see the BitcoinHeist article for references.

	address	year	day	length	weight	count	looped	neighbors	income	label
0	111K8kZAEnJg245r2cM6y9zgJGHZtJPy6	2017	11	18	8.333333e-03	1	0	2	100050000.0	princetonCerber
1	1123pJv8jzeFQaCV4w644pzQJzVWay2zcA	2016	132	44	2.441406e-04	1	0	1	100000000.0	princetonLocky
2	112536im7hy6wtKbpH1qYDWtTyMRAcA2p7	2016	246	0	1.000000e+00	1	0	2	200000000.0	princetonCerber
3	1126eDRw2wqSkWosjTCre8cjjQW8sSeWH7	2016	322	72	3.906250e-03	1	0	2	71200000.0	princetonCerber
4	1129TSjKtx65E35GiUo4AYVeyo48twbrGX	2016	238	144	7.284841e-02	456	0	1	200000000.0	princetonLocky
5	112AmFATxzhuSpvtz1hfpa3Zrw3BG276pc	2016	96	144	8.461400e-02	2821	0	1	50000000.0	princetonLocky
6	112E91jxS2qrQY1z78LPWUWrLVFGqbYPQ1	2016	225	142	2.088519e-03	881	0	2	100000000.0	princetonCerber
7	112eFykaD53KEkKeYW9KW8eWebZYSbt2f5	2016	324	78	3.906250e-03	1	0	2	100990000.0	princetonCerber
8	112FTiRdJjMrNgEtd4fvdoq3TC33Ah5Dep	2016	298	144	2.302828e+00	4220	0	2	80000000.0	princetonCerber
9	112GocBgFSnaote6krx828qaockFraD8mp	2016	62	112	3.725290e-09	1	0	1	50000000.0	princetonLocky

Features

1. address: String. Bitcoin address.
2. year: Integer. Year.
3. day: Integer. Day of the year. 1 is the first day, 365 is the last day.
4. length: Integer. Quantifies mixing rounds on Bitcoin, where transactions receive and distribute similar amounts of coins in multiple rounds with newly created addresses to hide the coin origin
5. weight: Float. Quantifies the merge behavior (i.e., the transaction has more input addresses than output addresses), where coins in multiple addresses are each passed through a succession of merging transactions and accumulated in a final address.
6. count: Integer.
7. looped: Integer. Intended to count how many transactions
    a. split their coins
    b. move these coins in the network by using different paths, and finally
    c. merge them in a single address.
8. neighbors: Integer.
9. income: Integer. Satoshi amount (1 bitcoin = 100 million satoshis)
10. label: Category String. Name of the ransomware family (e.g., CryptXXX, CryptoLocker etc) or white (i.e., not known to be ransomware).

Task:

1. Data cleaning
2. Data Visualization
3. Hypothesis/Experimental Testing
4. Classification of ransomware