THE BITCOINHEIST: CLASSIFICATIONS OF RANSOMWARE CRIME FAMILIES

Micheline Al Harrack

Marymount University, VA, USA

ABSTRACT

Ransomware attacks are on the rise and attackers are hijacking valuable information from different critical infrastructures and businesses requiring ransom payments to release the encrypted files. Payments in cryptocurrencies are designed to evade tracing the transactions and the recipients. With anonymity being paramount, tracing cryptocurrencies payments due to malicious activity and criminal transactions is a complicated process. Therefore, the need to identify these transactions and label them is crucial to categorize them as legitimate digital currency trade and exchange or malicious activity operations. Machine learning techniques are utilized to train the machine to recognize specific transactions and trace them back to malicious transactions or benign ones. I propose to work on the Bitcoin Heist data set to classify the different malicious transactions. The different transactions features are analyzed to predict a classifier label among the classifiers that have been identified as ransomware or associated with malicious activity. I use decision tree classifiers and ensemble learning to implement a random forest classifier. Results are assessed to evaluate accuracy, precision, and recall. I limit the study design to known ransomware identified previously and made available under the Bitcoin transaction graph from January 2009 to December 2018.

KEYWORDS

Ransomware, Classification, Decision Tree, Random Forest, Ensemble Learning, Bitcoin, Blockchain, BitcoinHeist, Machine Learning.

1. INTRODUCTION

With the exponential increase in ransomware attacks globally and in the U.S., cybercriminals are holding files hostages for ransom payments. Just as the criminals want to stay anonymous, payments need to be anonymous and somehow difficult to trace. The Blockchain technology and the associated cryptocurrency Bitcoin upended the monetary and banking system since its inception in 2008 by the mysterious Satoshi Nakamoto [1]. An encrypted public ledger became the perfect exchange network for trusting parties as well as malicious actors collecting on their illegal activities and transactions. The elimination of the trusted middle party like a financial institution increased the complexity of identifying recipients of cryptocurrency resulting from illegal activities [2].

In a recent study, Elliptic [3] estimated that roughly 829 million bitcoins have been spent thus far in the dark web. It is a very conservative estimate given the recent news that $1 billion of Silk Road Bitcoins are on the move after its demise [4]. No doubt the amount has significantly increased with recent ransomware attacks in addition to criminals' sophistication in circumventing detection on the blockchain. The Cybersecurity and Infrastructure Agency (CISA) in a joint statement with the Federal Bureau of Investigations (FBI) and the Department of Health and Human Services warned about Tactics, Techniques and Procedures (TTPs) targeting the health sector with ransomware for financial exploitation [5].

2. RELATED WORK

In 2016, a massive heist led to the theft of around 120,000 bitcoins being stollen amounting to around 72 million dollars leading to a Bitcoin market crash [6]. On the 28th of November 2018, The Office of Foreign Assets Control (OFAC) attributed two digital currency addresses to two individuals and placed these two bitcoins addresses on its sanction list for the first time [7]. Therefore, the need to identify these transactions and label them is crucial to categorizing them as legitimate digital currency trades and exchange or malicious activities operations. Machine learning techniques are utilized to train the machine to recognize specific transactions and trace them back to malicious transactions or benign ones. Major ransomware families identified previously have been made available under the Bitcoin transaction graph from January 2009 to December 2018. This dataset has been made public by the original researchers who worked on the Bitcoin Heist for ransomware detection on the Bitcoin Blockchain [8]. An analysis study of the hack subnetworks in the Bitcoin graph has been conducted by Chainalysis based on time series, nodes traveled, and Unspent Transaction Outputs (UTO) to cite a few of the features [9]. However, due to the sensitivity of the underlying data, these researchers have not made their specific dataset public. A Topological Data Analysis (TDA) methodology was proposed on time-series data for classification and compared with traditional classifiers such as Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) and deemed equally valid and reliable [10]. Therefore, combining the time-series dataset features studies and machine learning techniques and extending them to deep learning, I analyze the BitcoinHeist dataset for potential classifications into known and identified ransomware families tracing back transactions to specific ransomware labels, to eventually extend this classification into unidentified types of ransomwares. My proposed technique differs in that it combines the transactional features with the path activities along the length to identify the label, thereby classifying a transaction into a specific ransomware family.

3. DATASET AND METHODOLOGY

In this section, I go over the dataset selection, features, and the methods used to train and test the sets to predict the corresponding ransomware label.

3.1. Datasets Description and Features

A ransomware is a type of malicious software denying access to a computer system by locking data until a ransom is paid [5]. For this study, the ransomware classification will be limited to the families of ransomware defined in the BitcoinHeist dataset. The original dataset lists 2,916,697 daily transactions spanning from January 2009 to December 2018 [8]. Networks edges featuring amounts less 0.3 Bitcoins have been removed as ransomware payments are generally substantially higher. This resulted in 1,048,576 remaining transactions. Each transaction included an address which is similar to a physical address or an email address and needed for a Bitcoin payment, typically for one transaction [11]. Loops count how many transactions split the coins and travel using different paths in the network to finally converge at one specific address. Ultimately, Bitcoins at their final address can be sold and converted to different currencies. Weigh aims to quantify the merging behavior whereby transactions would have more input nodes than output nodes. Count quantifies the merging behavior as to the number of the merging transactions compared to the amounts in weighs. Length refers to mixing rounds in bitcoins where nodes receive and distribute equal amounts of coins over several rounds using newly created addresses to hide the transactions origins. Neighbors of a transaction t are the number of transactions that converge into the t transaction address. Income is in satoshi amounts where 1

bitcoin equates to 100 million satoshis. Ransomware labels are adopted from the famous studies Montreal, Padua, and Princeton [12], [13].[14].

3.2. Data Preprocessing

The dataset included bitcoin addresses identified as white meaning that have not been confidently labeled as ransomware. I removed all white labels, then limited the timeframe between 2014 and 2017 which resulted in 28265 transactions of labeled ransomware as shown in Table 1. Certain ransomware such as WannaCry have been excluded as they fall into unique and specific occurrences. I included all Princeton identified ransomware families (2), all Padua ones (3), and selected XTPlocker and XTPlockerv5.0 from Montreal to normalize the distribution and test if variations of the same XTPlocker family can successfully be classified. data[label]= ['princetonCerber', 'princetonLocky', 'paduaCryptoWall', 'paduaKeRanger', 'paduaJigsaw', 'montrealXTPlocker', 'montrealXTPlockerv5.0']

Table 1. Summary of ransomware labels, counts, and sum of income between 2014 and 2017

Column Lal	2014		2015		2016		2017		Total Coun	Total Sum of
Label	Count of addr	Sum of income	Count	Sum of income	Count o	Sum of income	Count of add	Sum of income		
montrealXLockerv5.0							7	1299975690	7	1299975690
montrealXTPLocker					8	2109946870			8	2109946870
paduaCryptoWall	9157	7.86128E+12	3233	8.32035E+11					12390	8.6933E+12
paduaJigsaw					2	96788152			2	96788152
paduaKeRanger					10	999900000			10	999900000
princetonCerber					6043	7.00482E+11	3180	2.5156E+11	9223	9.5204E+11
princetonLocky					6585	1.60829E+12	40	1.099E+10	6625	1.6193E+12
Grand Total	**9157**	**7.86128E+12**	**3233**	**8.32035E+11**	**12648**	**2.31198E+12**	**3227**	**2.6385E+11**	**28265**	**1.1269E+13**

3.3. Methodology

The processed dataset is used for training and testing for ransomware classes predictions. The training and testing is conducted with a 0.9 to 0.1 proportion. The selected features in the processed dataset are address, year, day, length, weight, count, looped, neighbors, income, and label. An illustration of a sample the data features address, income and label is shown in Table 2.

Table 2. A sample of the data set to be used for training and prediction

address	income	label
111K8kZAEnJg245r2cM6y9zgJGHZtJPy6	100050000.0	princetonCerber
1123pJv8jzeFQaCV4w644pzQJzVWay2zcA	100000000.0	princetonLocky
112536im7hy6wtKbpH1qYDWtTyMRAcA2p7	200000000.0	princetonCerber
1126eDRw2wqSkWosjTCre8cjjQW8sSeWH7	71200000.0	princetonCerber

Table 3 shows a description of the remaining features used to train and predict the labels.

Table 3. A sample description of the Bitcoin dataset analyzed for testing

	year	day	length	weight	count	looped	neighbors	income
0	2017	11	18	0.0083331	0	2		100050000.0
1	2016	132	44	0.0002441	0	1		100000000.0
2	2016	246	0	1.0000001	0	2		200000000.0
3	2016	322	72	0.0039061	0	2		71200000.0
4	2016	238	144	0.072848456	0	1		200000000.0

3.4. Data Analysis

I ran different features selection tests and assessed that length, and weight of a transaction in combination with income are the best predictors of the class label. Using the features of length, weight, and income to predict the label, I used two classification algorithms. First, I ran a decision tree classifier as decision trees are attractive due to their high speed [15]. The overall accuracy was 0.99

Table 4. A classification report for the predicted values by the Decision Tree Classifier

	precision	recall	f1-score	support
montrealXLockerv5.0	0.00	0.00	0.00	0
montrealXTPLocker	0.00	0.00	0.00	1
paduaCryptoWall	1.00	1.00	1.00	209
paduaJigsaw	0.00	0.00	0.00	1
paduaKeRanger	0.00	0.00	0.00	0
princetonCerber	0.99	0.98	0.98	924
princetonLocky	0.97	0.99	0.98	692
accuracy			0.99	827
macro avg	0.42	0.42	0.42	2827
weighted avg	0.99	0.99	0.99	2827

Table 5. A confusion matrix of the decision tree classifier

```
 6535    paduaCryptoWall
25563    paduaCryptoWall
14798    paduaCryptoWall
 12083    princetonCerber
3675     paduaCryptoWall
                ...
 25220     princetonLocky
18089    paduaCryptoWall
 15242     princetonLocky
 27624     princetonLocky
 2247     princetonCerber
Name: label, Length: 2827, dtype: object

array([[   0,    0,    0,    0,    0,    1,    0],
       [   0,    0,    0,    0,    0,    0,    0],
       [   0,    0, 1209,    0,    0,    0,    0],
       [   0,    0,    0,    0,    0,    0,    0],
       [   0,    0,    0,    0,    0,    0,    2],
       [   0,    0,    0,    1,    0,  902,    8],
       [   0,    1,    0,    0,    0,   21,  682]], dtype=int64)
```

Using the same features for training and testing, I implemented an Ensemble Learning algorithm for Random Forest Classifier. The model performed well with an accuracy of 0.94 as shown in Table 6.

Table 6. Predictions classification report

	precision	recall	f1-score	support
montrealXTPLocker	0.00	0.00	0.00	1
paduaCryptoWall	1.00	1.00	1.00	1209
paduaJigsaw	0.00	0.00	0.00	1
princetonCerber	0.85	0.98	0.91	924
princetonLocky	0.97	0.77	0.86	692
accuracy			0.94	2827
macro avg	0.56	0.55	0.55	2827
weighted avg	0.94	0.94	0.94	2827

Table 7. Confusion Report for the Random Forest Classifier

```
26535    paduaCryptoWall
25563    paduaCryptoWall
14798    paduaCryptoWall
12083    princetonCerber
3675     paduaCryptoWall
                ...
25220     princetonLocky
18089    paduaCryptoWall
15242     princetonLocky
27624     princetonLocky
2247     princetonCerber
Name: label, Length: 2827, dtype: object
array([[   0,    0,    0,    0,    0],
       [   0, 1209,    0,    2,    2],
       [   0,    0,    0,    0,    0],
       [   0,    0,    1,  904,  154],
       [   1,    0,    0,   18,  536]], dtype=int64)
```

3.5. Results Discussion

For each algorithm, a classification report was generated with precision, recall, and accuracy metrics and a confusion report to distinguish where the algorithm failed to correctly classify specific labels of ransomware. Precision is a measure of true positives overall all positive classifications including true and false positives:

$$\text{Precision} = TP/(TP+FP) \quad (1)$$

Recall is a measure of true positive over the totality of true positives and false negatives

$$\text{Recall} = TP/(TP+FN) \quad (2)$$

F1 score sees a measure between precision and recall:

$$F1 = 2 * (\text{Precision}*\text{Recall})/(\text{Precision} + \text{Recall}) \quad (3)$$

Looking at the accuracy scores, we see that the decision tree classifier performed better than the random forest model at first sight with 0.99 accuracy compared to a 0.94 but given that some transactions are supported by one label and the random forest classifier had a depth of 2, the random forest might perform as well as the decision tree if more data is available both for training and testing. Nevertheless, the results were promising to correctly classify a ransomware family based on transactional features and path activities combined with the length. This model is both simple in design and powerful in predicting ransomware families in supervised context. To seek high confidence, a transaction classified by both models as belonging to the same ransomware family would justify further investigation and tracing of the individuals and entities behind the identified addresses.

4. CONCLUSIONS

This study was limited to years 2014-2017 and analyzed specific ransomware as identified by Padua, Princeton, and Montreal. All white labels were excluded. Possible future work can include covering the transactions from 2011 until 2018 with all classes of ransomware to be tested and classified. Another stage would incorporate white labels to test against an existing class of ransomware or possibly different classes of unidentified ransomware based on transactional features, length, and path activities. An approach to apply is clustering the BitcoinHeist transactions to identify any possible commonalities among families of ransomwares.

REFERENCES

[1] Nakamoto, S., (2008) "Bitcoin: A Peer-to-Peer Electronic Cash System", http:// bitcoin.org/bitcoin.pdf

[2] Soska, K. & Christin, N ., (2015) "Measuring the longitudinal evolution of the online anonymous marketplace ecosystem", in 24th{USENIX} Security Symposium ({USENIX}Security 15), 33–48.

[3] Elliptic.co, (2019) " Bitcoin Money Laundering: How Criminals Use Crypto" (elliptic.co)

[4] Fbi.gov, (2014) Dozens of Online 'Dark Markets' Seized Pursuant to Forfeiture Complaint Filed in Manhattan Federal Court in Conjunction with the Arrest of the Operator of Silk Road 2.0, https://www.fbi.gov/contact-us/field-offices/newyork/news/press-releases/dozens-of-online-dark-markets-seized-pursuant-to-forfeiture-complaint-filed-in-manhattan-federal-court-in-conjunction-with-the-arrest-of-the-operator-of-silk-road-2.0

[5] Cisa.gov, (2020) Ransomware Activity Targeting the Healthcare and Public Health Sector | CISA

[6] Smith, M., (2016) "Another huge bitcoin heist: Bitcoin worth $72 million stolen from bitfinex", Network World (Online)

[7] U.S. Department of The Treasury, (2018) Cyber Activity and for the First Time Identifies Associated Digital Currency Addresses, https://home.treasury.gov/news/press-releases/sm556

[8] Akcora, C., Li, Y., Gel, Y., & Kantarcioglu, M., (2019) "BitcoinHeist: Topological Data Analysis for Ransomware Detection on the Bitcoin Blockchain", https://arxiv.org/abs/1906.07852

[9] Goldsmith, D., Grauer, K. & Shmalo, Y., (2019) "Analyzing Hack Subnetworks in the Bitcoin Transaction Graph", arXiv:1910.13415v1 [physics.soc-ph]

[10] Rivera-Castro, R., Moustafa, S., Pilyugina, P., & Burnaev, E., (2020) "Topologically-based Variational Autoencoder for Time Series Classification" (latinxinai.org)

[11] Bitcoin.org, (n.d.) https://bitcoin.org/en/vocabulary#bitcoin

[12] Paquet-Clouston, M., Haslhofer, B., Dupont, B., (2018) "Ransomware payments in the bitcoin ecosystem", arXiv preprint arXiv:1804.04080

[13] Conti, M., Gangwal, A., Ruj, S., (2018) "On the economic significance of ransomware campaigns: A bitcoin transactions perspective", Computers & Security

[14] Huang, D., McCoy, D., Aliapoulios, M., Li, V., Invernizzi, L., Bursztein, E., McRoberts, K., Levin, J., Levchenko, K., Snoeren, A., (2018) "Tracking ransomware end-to-end", IEEE, pp. 1–12.

[15] Ho, T., (1995) "Random Decision Forests", in the Third International Conference 1995 on Document Analysis and Recognition (Volume 1) - (acm.org).

AUTHOR

Micheline Al Harrack

Micheline is a Faculty at Marymount University. Her research interests lie at the intersection of Machine Learning Applications, Cybersecurity, and Linguistics.