Estimation of Ransomware Payments in Bitcoin Ecosystem

Ali Raheem, Rand Raheem, Thomas M. Chen, and Ahmed Alkhayyat

***Abstract*—Ransomware is one of the malicious software that is designed to prevent access to computer system until a sum of money is paid by the victim to the attacker. During the infection, the computer will either be locked, or the data will be encrypted. Ransoms are often demanded in Bitcoin, a largely anonymous Cryptocurrency. All transactions are recorded in the blockchain and verified by peer-to-peer networks. This paper investigation collects ten recent ransomware families, which use bitcoin as a payment for their ransom. In conjunction, we identified, collected and analysed Bitcoin addresses of users combining information from a clustering model and the blockchain. We used a heuristic clustering algorithm to reveal the hidden node's payment of ransomware. Finally, we demonstrated the characteristics of ransomware encryption mechanisms that include a view of the infected process and its execution, and the distinctive demands of ransom.**

***Index Terms*—Bitcoin, Blockchain, Cybercrime, Heuistic Clustering, Ransomware, Distrbibuted Ledeger, Payment, Transaction**

I. Introduction

The expression of ransomware itself is a combination of two main words; ransom and malware. Basically, any malware that violates the functionality of a user device and requires him to pay a ransom (a sum of money) for restituting the service falls under the ransomware domain. Crypto-ransomware prevents user access to the files by encrypting them. Whereas, Locker ransomware prevents the user from having access either by locking the user desktop or the boot sector. At this time, the number of ransomware victims is dramatically increasing, for example, Cryptolocker and WannaCry infected up to 3 million victims in 150 countries [1]. Like other ransomware families, these families emphasise the extortion of potential victims who see the attack as an outright threat [2]. Ransomware presents a unique opportunity to estimate the direct financial impact of the threat: the main and the most common payments of the ransomware are done in bitcoin.

A ransomware payment can be identified correctly while the ransomware cash flows can be assessed easily. That is because the bitcoin blockchain provides a reliable basis and a good structure of a single and shared history for all users. It also provides integrity to the system mining that the data cannot be modified by an unauthorised user. New bitcoins must undergo a mining process, which adds a verified record of transactions to the blockchain. Peer-to-peer users spend the power of their computer in verifying and recording payments, thus, they earn bitcoins. Bitcoin allows pseudo anonymity; ransomware attackers do not provide their real names. Instead, they use pseudo-names that show that some entities are transacting between each other, however, the real identities remain hidden as in stock exchange operations. All transactions are publicly shown in the blockchain, meaning that all the activities of the attacks can be seen by law enforcement. In principle, the ransomware attacker identity can be revealed by linking the transaction to an off network establishment. Ransomware cannot stay completely anonymous because in the blockchain the sender and the receiver addresses are both explicitly visible. Additionally, the ransomware results are visible and accessible to other users in the blockchain via payment history records. Hence, the main contribution of this paper is listed as the following:

- *Firstly*, we provide a comprehensive investigation to ten recent ransomware families that use bitcoin as a payment method. Inclusively, the investigation collects, identifies and analyses bitcoin addresses of the same user or group of users to classify a payment as ransom using both clustering model and blockchain information.
- *Secondly*, an analysis of ransoms economic impact will be provided extorted in bitcoin based on our clustering heuristic results and blockchain information.
- *Thirdly*, we demonstrate the characteristics of ransomware encryption mechanisms that include a view of the process of infection and execution of the demands of ransom.

The rest sections of the paper are structured as the following: Section II presents the related work in literature around identifying and assessing cybercrimes in the bitcoin ecosystem. We present the clustering model used to identify the ransom payments in section III. Then, we present our clustering results in section
IV. In the V section the limitation of the blockchain will be discussed together with our proposed clustering model. Last but not least, this paper will be concluded in section VI.

II. Related Work

The literature review of this work can be divided into two groups: the cybersecurity and cybercrime analysis group; and the deanonymise the bitcoin blockchain by data analysis group.

A. Cybersecurity and Cybcrime Analysis

There are several attemps have been done by the law enforcement authorities and the research community in order

Ali Raheem, Thomas M. Chen are with City University of London (e-mail: Alihraheem@outlook.com). (e-mail: Tom.chen.1@city.ac.uk).

Rand Raheem is with Middlesex University, School of Science & Technology, , London, NW4 4BT, United Kingdom (e-mail: R.h.raheem@mdx.ac.uk).

Ahmed Alkhayyat is with College of technical engineering, the Islamic University, Najaf, Iraq (e-mail: ahmedalkhayyat85@gmail.com).

to identify and measure cybercrimes in the bitcoin blockchain. The author in [3] analysed several bitcoin addresses up to 1,872 addresses that are closely linked to the CryptoLocker ransomware. The bitcoin address and the related CryptoLocker have the same records of transactions. That is characterised by a small number of transactions and a short period of activities. Overall 83% of the analysed addresses had few transactions of around 6 transactions and only 69% of them were active for less than 10 days. Whereas another study has performed a measurement analysis of the Crypotolocker ransomware [4]. The authors investigated the addresses of two bitcoin transactions and then a cluster of 968 addresses was generated. An analysis of the bitcoin transactions of the Cryptolocker ransomware family has been done together with presenting the ransom amount and time of the ransom period payment. On the other hand, the author in [5]analysed CryptoWall and CryptoLocker using an open sourced data from bitcoinTalk and Reddit, clustering 968 Bitcoin addresses and identifying 795 ransom payments that worth up to 1,128.40 BTC. In contrast, the author in [6]analysed the mixing services (Bitcoin Fog and Bit-Laundry) and sent shared features from Blockchain.info, which obfuscated the source of the bitcoin transactions for their customers via a transaction graph analysis.

B. De-Anoymise the Bitcoin Blockchain

There are number of studies that concerned with challenging the bitcoin assumed pseudo-anonymity. In [7] the bitcoin anonymity has been unravelled by applying several network analysis techniques on addresses crossed with open source information. That has revealed the possibility of connecting the bitcoin user addresses with one another. The author in [8]has provided a study of direct interaction with the network by sending transactions and by clustering public keys following co-spend heuristics. that helped in identifying 1.9 million bitcoin addresses that are connected to real services or pseudo-identities. Another study has used an open-source framework in order to analyse the blockchain in the bitcoin, cluster public keys, label the clusters and visualise the network [9]. Also, the model was tested where the obtained results helped in identifying the address that contains 111,115 BTC. The previous address belongs to ransoms paid to Cryptocker with only address posted by victim on a forum as a lead. Whereas the author in [10] has used a statistical analysis to identify the patterns of sending, receiving or storing coins for bitcoin users. Results showed that most coins remain stored in addresses that have never been involved in outgoing transactions. In contrast, with high volume of transactions moving small volume of coins and the particular subject of analysis, there are hundreds of transactions that sent more than 60,000 BTC.

III. Ransom Identification

The ransoms' extortion by ransomware has been investigated together with presenting a heuristic clustering algorithm is an open source code[1] attempts to de-anonymise the bitcoin addresses of attackers through revealing all generated addresses by a single attacker. That can be done via using the derived information from the blockchain. The second source of information in which it aids in the de-anonymisation of bitcoin users is the peer-to-peer (P2P) network. The clustering algorithm, along with blockchain is used to identify ransom payments through 3 phases: (i) identifying/disclosing the addresses of the bitcoin. (ii) collecting the history and database generation details of transactions from the blockchain. (iii) setting a heuristic clustering algorithm to reveal the hidden node payment of ransomware.

A. Stage 1: Disclosing the Ransomware Addresses

Different online resources have been searched in order to identify the required addresses that belong to the ransomware: ransomware knowledge base, e.g. Kaspersky Lab, Symantec, Malware bytes and ransomware removal guides MalwareTips.com, 2-spyware.com, Bleeping computer.com, reports from the Security Operation Centers (SOC) such as online forums (e.g. Reddit.com), Phishme.com, and Dell SecureWorks. Other reports have been collected from the Incident Responses (IR) and the counter threat units where both victims and researchers post the addresses of bitcoin related to the ransomware. That is without neglecting the ransomware screenshots that are available in different image search engines (e.g. Yahoo, Google). Also, a list of ransom addresses have been obtained from ID Ransomware[2], which keeps a record of the ransomware victims with thesociated ransom addresses.

B. Stage 2: Blockchain Database generation

The Bitcoin blockchain data is publicly available. However, the height of the block of the blockchain is over 5,000,000 blocks [11] and that comes with its own high expenses in downloading/querying the entire blockchain. That is in terms of bandwidth, storage, and computations. Therefore, to solve these issues, we used blockchain wallet API[3] in MySQL to analyse transactions associated only with the determined addresses. It is to be mentioned that, in such a database, each transaction is associated with an address set, that is required to collect the hash of transaction (HASH), remitted bitcoin (BTC_to_Address), input addresses (Transaction_In_Addresses), output addresses (Transaction_out_Addresess), GMT-based data (GMT_Data) and GMT-based time (GMT_Time). The HASH field provides a primary key that can discards implicitly any duplicate transactions for multiple participating and constituting addresses. Payment that is extracted from the blockchain database. The selected addresses are palced by the algorithm to discrete values e.g. the distance between them is at least a threshold value ε. The small variable l guarantees the same accuracy of the original network, to reduce the number of cluster and at the same time expose the hidden node of ransom payment. Figure 1 shows a flowchart of the heuristic clustering algorithm:

[1]Open Source Code https://github.com/archienorman11/thesis-bitcoin-clustering

[2]ID Ransomware https://id-ransomware.malwarehunterteam.com/

[3]Its open source code used to develop and build Payment Processing, Block Chain Wallet, and Bitcoin transaction and Blocks data. https://www.blockchain.com/api/blockchain_wallet_api

Therefore, clustering algorithm has been set in three important steps of procedures as the following:

Step 1:We set the input (T_{inital}) of the collected ransomware addresses. Then ε *(0,1)*. N is the active values in hidden node (ransomware). Whereas, $V_1$is the activation value for the first pattern. So, the first cluster $C(1) = V_1$, *count= 1*, and *sum(1)*, set *N=1*.

Step 2:For every pattern *P* i represents an input address, *i= 1,2,3,4,... K*. Subsequent activation values can be clustered into one of the existing clusters where the distance between an activation value and its nearest cluster, $V\text{-}C(j)$ is computed. For distance less than ε, the activation value is put in cluster *j*. Otherwise, this activation value will form a new cluster. Assume is its active value. If there exists an index *j* such that $V - C(j)\ min_{j \in \{1,2,3,..\}} V - C(j)$ *and* $V - C(j) \leq \varepsilon$, *then set* $cont(j) = count(j) + 1$

sum (j) = sum (j) + V, else, N = N + 1,

$C(N) = 1$, $sum(N) = V$

Step 3: Replace C by the average of all activation values that have been clustered into this cluster: $C(j) = Sum(j)/count(j)$, $j = 1, 2, 3, 4,...N$.

Step 4:Finally, once the activation values of all hidden nodes (ransomware) have been obtained, the accuracy of the network is checked with the activation values at the hidden nodes replaced by their discredited values. An activation value *V* is replaced by $C(j)$, where index *j* is chosen such that $j = arg\ min_j\ V\ C(j)$. If the accuracy of the network falls below the required accuracy, then ε must be decreased and theclustering algorithm is run again, otherwise stop.

As for the verification, the sufficiently small ε, it is always possible to maintain the accuracy of the network with continuous activation values, although the resulting number of different discrete activation can be impractically large. The best ε value is one that gives a high accuracy rate. A simple way of obtaining an optimal value for ε is by searching in interval (0,1). The number of clusters and the accuracy of the network can be checked for all values of $\varepsilon = i\ \varsigma$, $i = 1, 2, 3...$ where ζ is a small positive scalar, e.g. 0.10. It is important to mention that it is not necessary to fix the value of ε equal for all hidden nodes.

IV. Clustering Results

Here the ten ransomwares are discussed. An acumen for the economic impact of this ransomware will be provided from the bitcoin payment perspective.

A. CryptoLocker

Brief Introduction: CryptoLocker appeared for the first time in September 2013; and it basically targets computers running Windows operating system using Microsoft Enhanced Rivest Shamir Adleman (RSA), proposed hybrid RSA algorithm for cloud computing, and Advanced Encryption Standard (AES) cryptography provider 'MS_ENH_RSA_AES_PROV' to create encryption keys and to encrypt user files with the strong RSA 'CALG_RSA_KEYX' and AES 'CALG_AES_256' algorithms. During the encryption process, it establishes a Connection Command and Control (C&C) server. With a unique AES key files are encrypted which in turn encrypts with RSA public key.

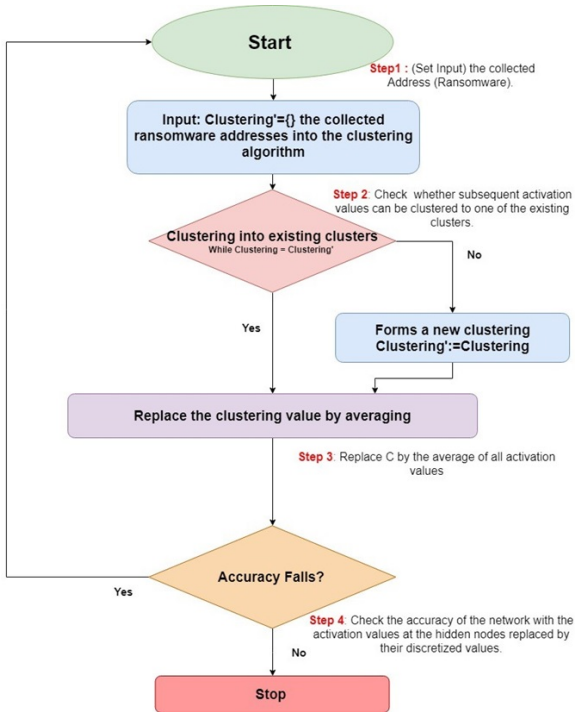

Fig. 1. Identifying address (bitcoin) using heuristics clustering algorithm

Infection: The CryptoLocker infection is spreaded via two modes. Starting 5 September 2013, the cybercriminals targeted business professionals through spam emails which appeared to be customer complaints against the recipients' firm. The emails carried a ZIP attachment containing a malicious executable Windows (exe) file. The two names of both the ZIP file and the malicious executable file were identical with 13 to 18 random alphabetical characters. In the second mode, CryptoLocker was distributed by the Gameover Zeus malware starting from 7*th* October 2013. This malware uses online cut wail botnet to send spam emails to well known online retailers and banking institutions to entice victims to be attacked by CryptoLocker exploit kits.

Ransom Demand: The ransom message asks the victim to pay within 72 hours and at the same time threatens the victim of destroying the decryption keys. The payment methods include MoneyPak, bitcoin, and UKash, but this has been changed recently where the ransoms are collected either by Moneypak or bitcoin. These payment methods are either anonymous or pseudo-anonymous to make it difficult to track both; the payer and the payee. The dates and ransom parameters in our identification clustering are based on previous studies on CryptoLocker ransomware [12]. Table I shows the amount of demanded ransom and their corresponding dates for CryptoLocker.

TABLE I
AMOUNT OF DEMANDED RANSOM (CRYPTOLOCKER)

Ransom Demanded	Time Periods
2 BTC	Between 5 September 2013 and 11 November 2013 allowing a three day ransom period.
10 BTC	Between 1 November 2013 and 11 November 2013, the payment was the fee using CryptoLocker decryption service that allowed victims who failed to pay ransoms within the given time frame to recover their files.
1 BTC	Between 8 November 2013 and 13 November 2013 allowing a three day ransom period.
0.5 BTC	Between 10 November 2013 and 27 November 2013 allowing a three day ransom period.
2 BTC	Between 11 November 2013 and 31 January 2014.
0.3 BTC	Between 24 November 2013 and 31 December 2013.
0.6 BTC	Between 20 December and 31 January 2014.

R*ansom payments in bitcoin:* To evaluate the economic impact of CryptoLocker, we used a clustering algorithm and the information derived from the blockchain, 964 addresses were found belonging to CryptoLocker ransomware. The analysis of transactions to CryptoLocker clustering shows that the total amount received is over 63,000 payments, which accounts for over 138,000 BTC. With further analysis of the CryptoLocker clustering, we found that approximately 86.26% of bitcoin addresses received a maximum of two payments, whereas 12.37% of bitcoin ad- dresses received no more than one bitcoin. Table II shows the three bitcoin addresses discovered by the clustering algorithm where the maximum number of payments and bitcoins were collected.

TABLE II
NUMBER OF RANSOMS | AND BITCOIN RECEIVED PER ADDRESS IN CRYPTOLOCKER CLUSTERING

#	CryptoLocker Bitcoin Address	Payments	BTC
1	16R14EH4v8A9GPXkAAP8geMFBA8oxA8nbY93		120.83
2	1H6doatrp959BBx3R3x6QaR73q4xXt8j13	81	112.94
3	1JfTGgVz1vmrgyGjE29mgCPFNdmR4LrBvj	72	101.70
	Total	**246**	**335.47**

The clustering algorithm discovered 823 ransom payments to CryptoLocker, in total 1455.7467 BTC. We cannot be assured that the unaccounted transaction is not a ransom payment because attackers keep changing the Bitcoin addresses or redirect the payments to a different wallet, which makes the tracking more difficult. Although, the results are adjusted with the previous research [13] [14] , the authors have used the Bitcoin price on the day of their evaluation. On the other hand, we can trust the methodology of our research for evaluating other ransomware, where a baseline for comparison is not available as resources on this topic are limited. However, we used the blockchain technology to verify the clustering results.

TABLE III
TOTAL RANSOMS PAID TO CRYPTOLOCKER

Ransom Demanded	Time Periods		Payments	BTC
	From	To		
2 BTC	05-09-2013	11-11-2013	448	895.7432
10 BTC	01-11-2013	11-11-2013	19	190.000
1 BTC	08-11-2013	13-11-2013	41	40.0000
0.5 BTC	10-11-2013	27-11-2013	120	63.0000
2 BTC	11-11-2013	31-01-2014	109	223.0000
0.3 BTC	24-11-2013	31-12-2013	32	9.2473
0.6 BTC	20-12-2013	31-01-2014	53	34.7562
Total	**05-09-2013**	**31-01-2014**	**823**	**1455.7467**

B. CryptoDefense

Brief Introduction: CryptoDefense first appeared at the end of February 2014, which revealed since then, the significant number of ransomware and the capability of its system. For instance, it used bitcoin as a payment method, Tor networks for anonymity, RSA-2048 based public key cryptography for strong encryption, and the typical pressure tactics, such as a short deadline for payment, with threats of increasing the ransom after the deadline. CryptoDefense encrypts mainly Windows system files. It encrypts them using the AES-256 algorithm. The Windows CryptoAPI library is used to generate the encryption key on the victim's computer. When the encryption process is completed, the AES is itself encrypted using the RSA-2048 public key.

Infection: CryptoDefense spreads mainly through spam emails that contain a malicious PDF file. It contacts its C&C stealthily and send information about the infected system. The encryption process starts immediately after receiving an acknowledgment from the C&C server.

Ransom Demand: CryptoDefense asks for a ransom of the equivalent of 500 USD or EUR in bitcoin with four days as the period of time to decrypt the files. The cost of decrypt files increases to USD/EUR 1,000 after four days. The victims can see a screenshot of their compromised system and a decrypted file as a proof of the conditional imminent decryption.

Overall *on ransom payments in bitcoin:* The analysis of the CryptoDefense Clustering transaction are verified using the blockchain information. We have collected 126 payments.The total value of these payments is above 142.5183 BTC. We have verified each payment to CryptoDefense Clustering through the blockchain as shown in Table IV

TABLE IV
NUMBER OF RANSOMS AND BITCOIN RECEIVED PER ADDRESS IN CRYPTODEFENSE CLUSTERING

#	CryptoDefense Bitcoin Address	Payments	BTC
1	1EmLLj8peW292zR2VvumYPPa9wLcK4CPK1	82	97.1722
2	19DyWHtgLgDKgEeoKjfpCJJ9WU8SQ3gr27	34	41.1501
3	1Mz7153HMuxXTuR2R1t78mGSdzaAtNbBWX	10	4.196
	Total	**126**	**142.5183**

Overall, we found 105 ransom payments to CryptoDefense between 26 February 2014 and 12 April 2014. The total was 122.1601 BTC. It should be noticed that the CryptoDefense in nature has a built-in flaw because of the poor design and implementation of Microsoft's cryptographic infrastructure. When generating the asymmetric key pair on the victim's machines that leaves a local copy of the key that helps the ransomware to encrypt the victim system.

C. CryptoWall

Brief Introduction: CryptoWall can be recognised by its strong encryption algorithm. The CHM wich helps compiling and saving documentation in a compressed HTML format. It may include text, images, and hyperlinks; viewable in web browser. File infection mechanism and C&C activity over the anonymous Tor network makes it much harder to analyse its communications and to take down malware servers. According to Semantic security report [19] and Dell SecureWorks Counter Threat Unit (CTU) research team [20]. The CryptoWall infection starting from mid November 2013, impersonated the appearance and the behavior of the CryptoLocker.

Infection: There are many infection vectors through which CryptoWall has spread. These include email attachments, browser exploit and drive downloads. It is to be mentioned here, that the email used a standard mimicked message from a government institutions or financial agencies are used by the above mentioned links. These institutions and agencies are used because they have links payload hosted over cloud services.

Evolution: The CryptoWall evolution is as the following version:

- **CryptoWall 1.0:** This initial variant of CryptoWall first appeared in the early 2014. That is why it does not have a special name.
- **CryptoWall 2.0:** Malware developers released this version in October 2014. This version solved some problems in the original version like developer run.
- **CryptoWall 3.0:** This CrptoWall variant is Web-to-Tor gateways. It has a unique bitcoin address for each victim and the ability to delete original unencrypted files.
- **CryptoWall 4.0:** This version include minor changes in the above mentioned versions. These changes summarised in filenames, new Tor gateways and increase in the initial ransom deadline.
- **CryptoWall 5.1:** This version appeared in November 2015. Here file names are changed concerning the way of encrypting and deleting them. Further more, new HTML ransom note file name and new payment gateways are redesigned. This appears to be the latest version of the CryptoWall virus, which is the most dangerous ransomware in the world. This new variant uses AES- 256, one of the strongest encryption algorithms available. The fact that the ransom note is written in Italian demonstrates that Italian users are the main target. The victims should pay the ransom within 48 hours.

Besides, these kinds of attacks can accept the payment either in Litecoin or bitcoin. However, according to SecureWorks [15]observation, Litecoin address LTv4m4y7NKHCXdw31dSEpTJmP6kXTinWDy never received any payments.

Ransom Demand: The amount of ransom fluctuates frequently and Table V shows this fluctuation.

TABLE V
AMOUNT OF DEMANDED RANSOM (CRYPTOWALL)

Ransom Demanded	Time Periods
$200	Between 1 March 2014 and 5 November 2015.
$500	Between 2 March 2014 and 23 December 2015
$600	Between 4 March 2014 and 6 November 2015. This kind of payment was three times the original ransom amount.
$1000	Between 5 March 2014 and 4 November 2015. This kind of payment was two times the original ransom amount.
$700	Between 10 March 2014 and 11 December 2015.
$1,400	Between 11 March 2014 and 21 December 2015. This kind of payment was two times the original ransom amount.

Overall on ransom payments in bitcoin: The publicly known bitcoin address of CryptoWall is used, which has generated 3,127 addresses belonging to CryptoWall clustering. The ransom payments reached 3,873, which contributed to 5,509.3203 BTC extorted as Table VI shows.

TABLE VI
TOTAL RANSOMS PAID TO CRYPTOWALL

Ransom Demanded	Time Periods		Payments	BTC
	From	To		
$200	01-03-2014	05-11-2015	598	212.5736
$500	02-03-2014	23-11-2015	1,824	2430.8276
$600	04-03-2014	06-11-2015	375	432.6321
$1000	05-03-2014	04-11-2015	420	834.4302
$700	10-03-2014	11-12-2015	458	956.5321
$1,400	11-03-2014	21-12-2015	198	642.3247
Total	**01-03-2014**	**21-12-2015**	**3,873**	**5,509.3203**

D. DMA Locker

Brief Introduction: DMA Locker first appeared in December 2015. The components of the DMA locker continuously change by Cyber Crooks. However, for file encryption the Symmetric key cryptography is used. This version proofs to be the strongest for combining both the AES-256 and the RSA-2048 encryption algorithms.

Infection: DMA Locker affects the systems running Windows operating system of the victims and the links are distributed through email spamming.

Evolution: DMA Locker evolution is as the following versions:

- ***DMA Locker 1.0:*** This version appeared in late 2015. It manipulated both English and Polish languages. To encrypt victim files, AES-256 algorithm in ECB mode is used to encrypt and delete them.
- ***MA Locker 2.0:*** The version appeared on 3 February 2016, It was updated by ransomware attackers to use a separate key for each file. It used AES for the encryption.
- ***DMA Locker 3.0:*** This new version has been developed in early 2016 to fix a weakness in the random number generator, AES key. Nerveless, the same RSA key pair is used for decrypting other infected systems.
- ***DMA Locker 4.0:*** This version appeared in 19 May 2016. It has the advantage of working offline because it can download asymmetric public key from the server.

Ransom Demand: DMA Locker uses bitcoin to pay the ransom. The function of the DMA Locker 4.0 is to give payment instructions on the hosted website, using the same IP address and updating the ransom amount like other components. Instead of the four days to pay the ransom, this version makes it seven on condition the ransom ill be increased. The time and ransom parameters in our identification clustering were derived and reflected from previous studies and reports on DMA Locker ransomware [16]. The following table shows the amount of ransom and the deadline of the time payment.

TABLE VII
AMOUNT OF DEMANDED RANSOM (DMA LOCKER)

Ransom Demanded	Time Periods
1 BTC	Between 26 December 2015 and 24 July 2016
1.5 BTC	Between 15 January 2016 and 30 May 2016
2 BTC	Between 28 January 2016 and 24 July 2016
4 BTC	Between 22 February 2014 and 04 June 2016
8 BTC	Between 21February 2014 and 06 August 2016 to allow a three day ransom period
1.7 BTC	Between 17 May 2016 and 12 July 2016
3 BTC	Between 24 May 2016 and 26 August 2016

Overall on ransom payments in bitcoin: In order to understand the economic impact of DMA Locker, our clustering analysis identified 103 ransom payments to the DMA Locker

clustering, which resulted in 323.9207 extorted BTC, as shown in Table VIII

TABLE VIII
TOTAL RANSOMS PAID TO DMA LOCKER

Ransom Demanded	Time Periods		Payments	BTC
	From	To		
1 BTC	26-12-2015	24-07-2016	14	12.98353
1.5 BTC	15-01-2016	30-05-2016	2	3.2682
2 BTC	28-01-2016	24-07-2016	12	27.9723
4 BTC	22-02-2014	04-06-2016	31	127.8593
8 BTC	21-02-2016	06-08-2016	3	31.3930
1.7 BTC	17-05-2016	12-07-2016	5	7.8231
3 BTC	24-05-2016	26-08-2016	36	112.6213
Total	**26-12-2015**	**26-08-2016**	**103**	**323.9207**

E. WannaCry

Brief Introduction: WannaCry appeared for the first time on 12 May 2017. This type of malicious software is known by different names such as WannaCryptor, WannaDecryptor 2.0, and WCry. Explicitly, this type affects Windows systems using a combination of the AES and the RSA algorithms. Eventually, each file is encrypted with a separate 128 bit AES key in a CBC mode and the RSA-2048 encryption algorithm.

Infection: Here the WannaCry aims at scanning the presence of the DoublePulsar backdoor of the target; if it is not there the WannaCry tries to compromise the system using the Eternal Blue exploit. The shadow brokers is a hacker group that attacks the WannaCry. To terminate a program execution a kill switch is often used.

Ransom Demand: The victim is asked to pay 300 US Dollar of ransom in bitcoin within 3 days. Also, the amount will be doubled within 7 days, otherwise, all the encrypted files will be deleted.

Economy of ransom payments in bitcoin: To enable victims pay the ransom Cybercriminals create a special bitcoin payment address, at the same time a race condition bug will prevent the correct extraction of one of the three hardcoded bitcoin ransom addresses. The WannaCry clustering, the analysis found based on these five addresses and WannaCry cluster received 336 payments. These transactions are worth 67.5213 BTC. We verified each payment to WannaCry Cluster through blockchain analysis. Table IX demonstrates the maximum number of both bitcoins and the collected ransom. Consequently, based on our analysis on WannaCry Clustering, we identified 361 ransom payments between 12 May 2017 to 14 February 2018, with a total of 50.9557 extorted BTC.

TABLE IX
NUMBER OF RANSOMS AND BITCOIN RECEIVED PER ADDRESS IN WANNACRY CLUSTERING.

#	WannaCry Bitcoin Address	Payments	BTC
1	13AM4VW2dhxYgXeQepoHkHSQuy6NgaEb94	102	19.9832
2	12t9YDPgwueZ9NyMgw519p7AA8isjr6SMw	94	19.2721
3	15p7UMMngoj1pMvkpHijcRdfJNXj6LrLn	83	14.680
4	15zGqZCTcys6CjDkE3DypCjXi6QWRV6V1	43	10.3439
5	1Hydrt8E3ybCWS7YBtggFFNn1AyrqWaBzz	14	3.2439
	Total	**336**	**67.5213**

F. Crypo-TorLocker2015

Brief Introduction: Crypto-Tor-Locker2015 was discovered by Symantec on 5 February 2015 as a low-level threat for Windows operating system. This attack uses only public key cryptography for file encryption. The Crypto-Tor-Locker2015 uses the RSA-2048 encryption algorithm, and the public key RSA downloads from the attacker C&C.

Infection: being Trojan, the Crypto-Tor-Locker2015 spreads through classical infection mechanism such as drive-by download.

Ransom Demand: Crypto-Tor-Locker2015 asks to pay a ransom of 0.5 BTC, which is equivalent to EUR/USD 100 within five days of infection to decrypt the files.

Overall on ransom payments in bitcoin: Our analysis clustered collected eight new addresses belonging to the Crypto-Tor-Locker2015. These addresses received around 4.7205 BTC with 136 payment.

G. TeslaCrypt

Brief Introduction: In February 2015, TeslaCrypt getting momentos. Obviously, it targets game-related user content like custom maps used for saving files together with other personal documents like pictures. On the other hand, videos, audio files and removable USB storage are completely neglected in the TeslaCrypt. To encrypt files and confuse the victims AES algorithm is used. This is performed by attaching 'exe' extension. This is done simultaneously with the ransom note message using the RSA-2048 encryption algorithm. Needless to say the TeslaCrypt C&C attack in particular Tor anonymity network that demands SSL encrypted connection. However, preventing interaction with TeslaCrypt cannot stop locally the encryption keys.

Infection: Using the Nuclear browser and Angler exploiting kits for the process of distributing this attack.

Ransom Demand: TeslaCrypt accepted the ransom through different methods of payment. Usually the ransom amount in bitcoin was 1.5 BTC with seven days of period time, otherwise, the amount increased to 2.5 BTC. The victims infected by TeslaCrypt ransomware in North American region can select USD 1000 with PayPal, whereas the European victims might pay EUR 600 with Paysafecard or Ukash.

Overall on ransom payments in bitcoin: Our analysis identified 78 ransom payments to the TeslaCrypt clustering between 02 February 2015 to 15 July 2015 time of periods, which aggregated to 117.5 extorted BTC.

H. Jigsaw

Brief Introduction: Jigsaw was designed in April and released a week after creation in March 2016, and this kind of attack affects the running Windows operating system.

Infection: It was designed to spread through malicious attachments in spam emails. It contains different language versions, whereas each type is hard-coded that is only executable after a certain data. Jigsaw employs a unique strategy, where a deleting process for hundreds of files is done every hour for the first 24 hours. Hence, if the ransom is not paid by the victim within three days, then Jigsaw will be deleting all the rest of files. Also, if the victim tries to restart or shut down his/her machine for any reasons, Jigsaw destroys a thousand files as a punishment, in order to put the victim under-pressure to pay the ransom.

Ransom Demand: Jigsaw demands variant ransom amount from 23 to 500 US Dollar paid in the digital currency (bitcoin). On the other hand, the cyber crimes that host the payload on free cloud storage services distribute the links to the malicious payload via spammed emails. Jigsaw may work offline and victim's files can be encrypted via the AES-128 encryption algorithm.

Overall on ransom payments in bitcoin: We have found through the clustering analysis 59 payments between March 2016 to August 2016 time of periods, which equal 2.601 BTC.

I. ZCrypto

Brief Introduction: ZCryptor appeared first on 24 May 2016 targeting computers running Windows operating system. The user's files are encrypted by the RSA encryption algorithm after obtaning the victim-specific algorithm key from the C&C. One of the ransomware families is the ZCryptor that can self-propagate on other connections on different computer networks devices, without using spamming and exploit kit.

Infection: To be infected, conventional distribution techniques are used by this attack i.e.fake software, email spamming, macro malware in Microsoft office suite and Adobe flash updater.

Ransom Demand: The ransom message is displayed on the ZCryptor where it had asked for 1.2 BTC.

Overall on ransom payments in bitcoin: We have found through the clustering analysis, 15 payments between 24 May 2016 to 28 June 2016, which equals 63.926 BTC.

J. VenusLocker

Brief Introduction: VenusLocker appeared in August 2016, as a type of ransomware family. This type of attack targets Windows-based systems. The AES-256 algorithm is used in the VenusLocker in order to encrypt data files. The AES encryption key is generated on the victim's system from a cryptographically strong random number generator and is encrypted with an embedded RSA-2048 public key before sending the C&C. A unique ID C&C is created by the attacker where it is required to identify the infected system.

Infection: VenusLocker spreads through either drive or download, usually the attacker allows three days for the payment of the ransom in bitcoin.

Ransom Demand: The first launched attack where its ransom demand is a100 US Dollar. The ransom amount was settled on 1 BTC; however, that was updated in December 2016.

Overall on ransom payments in bitcoin: We found 2 addresses based on our analysis and 10 received the payment, which is worth 6.8 BTC shown in Table X.

TABLE X
NUMBER OF RANSOMS AND BITCOIN PER ADDRESS IN VENUSLOCKER CLUSTERING

#	CryptoDefense Bitcoin Address	Payment	BTC
1	1Dj9YnMiciNgaKuyzKynygu7nB21tvV6QD	3	0.001452
2	16jvWspVfvhjRgJhGCDETf29cjQAyNmx9G	7	6.8
	Total	**10**	**6.8014**

V. BLOCKCHAIN LIMITATIONS

The paper is concerned with tracking the ransom payments by using clustering algorithm and blockchain, which are presented in section III and VI. The main issue of concern is the quality of the collected data which is from public sources where the bitcoin addresses are collected from these sources. Collecting ransomware binaries is another alternative way that could be done through the many time execution of the ransomware in a virtual environment. That will help to obtain the required bitcoin addresses. Based on the nature of the problem, the used approach in [4] [17], [18] has been followed while taking an extreme precaution when the addresses are collected from public sources. There are three main flaws in the fundamental principles of the bitcoin protocol in our address identification model shown as below:

Overestimation: In a single transaction multiple users can pool their own transaction. For instance 'Mixing or laundering' happens when a third service is used to break the connection between two entities; the address the bitcoin is sent to and the source address of the bitcoin. It is to be mentioned that the blockchain of the bitcoin is a public ledger which keeps a record of every transaction. That is an issue for the ransomware because mixing coins is critical for this attack where the attacker does not want everyone to know the two parties of communication; the source and destination addresses of the bitcoin transactions together with where the BTC is stored.

Underestimation: Studies have proven that there is no conjunction between the owned address of a user in the blockchain with any other addresses of the same user. Though, in an exceedingly given state of affairs, the blockchain might report additional correct results in comparison to the prevailing approaches because of its attributes in the classification of ransom.

Access to external data: Blockchain services cannot inherently make arbitrary network requests to access data outside the network, e.g. if the blockchain service is retrieving information from an external source, this retrieval must be done iteratively and separately by each node. As the source is out of the blockchain, it cannot be guaranteed that the same answer will be received by every node. The response will be changed by the source in the time between requests from several nodes, or could become temporarily unavailable. Therefore, blockchain interactions are limited on chain data.

Nevertheless, in this work, we have given a scenario that gives more accuracy in terms of results in comparison to nowadays approaches because of its attributes in the classification of ransom.

VI. CONCLUSION

This paper has investigated the ransomware ecosystem which is considered as one of the cybercriminal phenomenons. It has stated that bitcoin is the most used payment method in ransomware. Hence, it was needed to understand the involved operations in such transactions and gain a key insight into the financial inner workings of these operations. In this paper, we also provide an investigation of ten recent ransomware families

using bitcoin payments. We used a clustering algorithm alongside blockchain information to collect, identify and analyse bitcoin addresses belonging to cybercriminals (ransomware). Besides, we demonstrate the characteristics of ransomware encryption mechanisms that include a view of the infection and execution process, and the distinctive demands of ransom.

ACKNOWLEDGMENT

The authors would like to thank all participants in this work. Thanks are extended to School of Mathematics, Computer Science & Engineering at the City University of London that has indeed played a significant role in supporting and backing this work. The authors also wants to thank the Faculty of Science and Technology at Middlesex University and the College of Technical engineering at the Islamic University of Al-Najaf, Iraq for their invaluable contribution to the success of this work. Last but not least, the authors also want to thank EPSRC for project EP/P011772/1, on the EconoMical, PsycHologicAI and Social Impact of Ransomware (EMPHASIS), which supported this work. Also, Thanks are extended to the science and technology department at Middlesex University that has played indeed a significant role in supporting and backing this work.

REFERENCES

[1] P. Dworak, , and A. C. and, "Modernizacja systemów ESD na tłoczniach gazu," pp. 558–563, 2016. [Online]. Available: 10.18668/ng.2016.07.10;https://dx.doi.org/10.18668/ng.2016.07.10

[2] Y. Guillot and A. Gazet, "Automatic binary deobfuscation," *Journal in Computer Virology*, vol. 6, no. 3, pp. 261–276, 2010. [Online]. Available: 10.1007/s11416-009-0126-4;https://dx.doi.org/10.1007/s11416-009-0126-4

[3] A. Kharraz, W. Robertson, . D. Balzarotti, K. E, and . ', "Cutting the Gordian Knot: A Look under the Hood of Ransomware Attacks," *International Conference on Detection of in- intrusions and Malware, and Vulnerability Assessment*, pp. 3–24, 2015.

[4] K. Liao, Z. Zhao, . A. Doupe, and Ahn, "Behind Closed Doors: Measurement and Analysis of CryptoLocker Ransoms," *Electronic Crime Research (eCrime)," IEEE, APWG Symposium on*, pp. 1–13, 2016.

[5] K. Caba, P. Gawkowski, . K. Grochowski, and Osojca, "Network activity analysis of CryptoWall ransomware," *Przeglad Elektrotechniczny*, vol. 91, pp. 201–2015, 2015.

[6] Christin, "Traveling the Silk Road: A measurement analysis of a large anonymous online marketplace," *Proceedings of the 22nd international conference on World Wide Web*, pp. 213–224, 2013.

[7] C. N, "Traveling the Silk Road: A measurement analysis of a large anonymous online marketplace," pp. 213–224, 2013.

[8] P. S. M. Meiklejohn, J. G. K. Levchenko, D. Mccoy, G. M. Voelker, and Savage, "A fistful of bitcoins: characterizing payments among men with no names," *Proceedings of the 2013 conference on Internet measurement conference*, pp. 127–140, 2013.

[9] M. M. F. Spagnuolo and Zanerobitiodine, "Extracting intelligence from the bitcoin network," in *International Conference on Financial Cryptography and Data Security*. Springer, 2014, pp. 457–468.

[10] S. D. A. Reid and Shamir, "Quantitative analysis of the full bitcoin transaction graph," *International Conference on Financial Cryp-tography and Data Security*, pp. 6–24, 2013.

[11] 2017. [Online]. Available: https://www.blockchain.com/btc/block-height/500000

[12] 2014. [Online]. Available: https://www.europol.europa.eu/

[13] K. Savage, . P. Coogan, and Lau, 2015. [Online]. Available: http://www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/the-evolution-of-ransomware.pdf

[14] M. M. F. Spagnuolo and Z. S. Bitiodine, "Extracting Intelligence from the Bitcoin Network," *Springer Financial Cryptography and Data Security (LNCS)*, pp. 457–468, 2014.

[15] "Dell SecureWorks Counter Threat Unit Threat Intelligence," *CryptoWall Ransomware Threat Analysis,webpage*, 2014.

[16] 2016. [Online]. Available: https://blog.malwarebytes.com/threat-analysis/2016/05/dma-locker-4-0-known-ransomware-preparing-for-a-massive-distribution

[17] D. Huang, A. D. M. Mccoy, L. Invernizzi, E. Bursztein, L. K. J. Mcroberts, . K. Levchenko, and Snoeren, "Tracking Ransomware End-to-end," *IEEE Symposium on Security and Privacy (SP)*, pp. 793–806, 2018.

[18] E. Bursztein, . K. Mcroberts, and Invernizzi, "Tracking Desk- top Ransomware Payments," *IEEE*, pp. 1–6.