Behind Closed Doors: Measurement and Analysis of CryptoLocker Ransoms in Bitcoin

Kevin Liao, Ziming Zhao, Adam Doupé, and Gail-Joon Ahn
Arizona State University
{kevinliao, zmzhao, doupe, gahn}@asu.edu

***Abstract*—Bitcoin, a decentralized cryptographic currency that has experienced proliferating popularity over the past few years, is the common denominator in a wide variety of cybercrime. We perform a measurement analysis of CryptoLocker, a family of ransomware that encrypts a victim's files until a ransom is paid, within the Bitcoin ecosystem from September 5, 2013 through January 31, 2014. Using information collected from online fora, such as reddit and BitcoinTalk, as an initial starting point, we generate a cluster of 968 Bitcoin addresses belonging to CryptoLocker. We provide a lower bound for CryptoLocker's economy in Bitcoin and identify 795 ransom payments totalling 1,128.40 BTC ($310,472.38), but show that the proceeds could have been worth upwards of $1.1 million at peak valuation. By analyzing ransom payment timestamps both longitudinally across CryptoLocker's operating period and transversely across times of day, we detect changes in distributions and form conjectures on CryptoLocker that corroborate information from previous efforts. Additionally, we construct a network topology to detail CryptoLocker's financial infrastructure and obtain auxiliary information on the CryptoLocker operation. Most notably, we find evidence that suggests connections to popular Bitcoin services, such as Bitcoin Fog and BTC-e, and subtle links to other cybercrimes surrounding Bitcoin, such as the Sheep Marketplace scam of 2013. We use our study to underscore the value of measurement analyses and threat intelligence in understanding the erratic cybercrime landscape.**

***Index Terms*—Bitcoin, CryptoLocker, cybercrime, ransomware, security.**

I. Introduction

Increasingly, Bitcoin [1] is becoming a staple utility among cybercriminals [2]—two of the digital currency's main attractions are its provisions for pseudoanonymity and its irreversible transaction protocol. Unfortunately, these provisions engender the dichotomous incentives between legitimate users, who wish to transfer money efficiently and securely, and cybercriminals, who leverage these properties to commit irrevocable and supposedly untraceable financial fraud.

Although the notion of digital currencies has existed long before Bitcoin's debut, Bitcoin was proposed by an individual under the pseudonym Satoshi Nakamoto in 2008. Nakamoto introduced a distributed *public* ledger that serializes a record of all confirmed transactions known as the blockchain. A fundamental breakthrough in technology, the blockchain enables the Bitcoin system to operate under a decentralized peer-to-peer network where users are identifiable by public keys, or more commonly referred to as Bitcoin addresses, intended to provide pseudonymity. Users can transfer digital currency, called bitcoins[1], to other addresses by issuing transactions, which are then broadcast to the public blockchain.

Since all confirmed transactions are visible to the public, the blockchain's inherent transparency has proven to be ineffective in preserving the anonymity of its users (legitimate users and cybercriminals alike). While Bitcoin addresses alone are not explicitly tied to any real-world entities, a number of recent research efforts have shown that monetary movements and address links can be traced throughout the blockchain data structure [3]–[8]. Even though there have been many attempts to enhance user privacy with varying degrees of success (i.e. generating multiple addresses, using bitcoin mixers such as Bitcoin Fog [8], or using privacy-enhancing overlays such as Coinjoin [9]), user privacy is further undermined when real-world information and quasi-identifiers found on the Internet can be imputed to users' Bitcoin addresses. Given Bitcoin's meteoric rise in popularity and scale, such a condition was inevitable and the overlap between publicly available data and blockchain data has improved identification and attribution throughout a vast, connected network of users—there are addresses tied to forum usernames, anonymous online marketplaces, Bitcoin exchanges, and popular Bitcoin services.

Privacy-preserving online services, such as the Tor hidden network [10] and the Bitcoin system, while undoubtedly useful in many aspects, play nontrivial roles in the burgeoning cybercrime landscape. The fact remains that an elegant solution for distinguishing legitimate and illicit use of these services is far from reach since the goals of Tor, Bitcoin, and the like are to protect privacy en masse. While Bitcoin does enable criminal enterprises to better obfuscate money laundering schemes compared to traditional financial systems, we have seen that digital footprints embedded in the Bitcoin blockchain can reveal salient information about its users. Given the recent prevalence of CryptoLocker [11] — a family of ransomware that encrypts files on a victim's system and demands a ransom to be paid (through MoneyPak or Bitcoin) for the decryption key — from September 2013 through early 2014, we use this as an opportunity to better understand the mechanics of a digital money laundering economy and to generate threat intelligence on brazen cybercrimes. More generally, we aim to answer the *who*, *why*, and *how* behind CryptoLocker in hopes that our findings may be extendable to future cybercrime forensics and

[1]The Bitcoin system and peer-to-peer network are referred to as "Bitcoin" while the unit of currency is referred to as "bitcoin" or abbreviated as "BTC".

analytics. Our contributions are the following:

- We design and implement a framework that collects data from the blockchain and automatically identifies ransom payments to Bitcoin addresses belonging to CryptoLocker. From this, we measure CryptoLocker's economy in Bitcoin and provide a lower-bound estimate of financial damages from September 5, 2013 through January 31, 2014.
- We present a novel approach to analyzing Bitcoin transactions by examining ransom payment timestamps both longitudinally across CryptoLocker's operating period, as well as tranversely across times of day, to distinguish trends and changes in timestamp distributions.
- We construct a non-trivial, topological network of CryptoLocker addresses and systematically examine CryptoLocker's financial infrastructure and money laundering strategies. By leveraging external, real-world data, we find connections to popular services such as Bitcoin Fog and BTC-e, and speculate connections to other Bitcoin cybercrime, such as the Sheep Marketplace heist.

II. Background

To understand our analysis of the CryptoLocker economy, we first discuss the Bitcoin protocol in Section II-A, and we next discuss the CryptoLocker ransomware in Section II-B.

A. Anatomy of a Bitcoin Transaction

Bitcoin is a decentralized cryptographic currency that was proposed by Satoshi Nakamoto in 2008 [1]. A bitcoin can be abstracted as a *chain of transactions* among owners who are identifiable by public keys generated from an asymmetric encryption scheme[2]. We will refer to these public keys as Bitcoin addresses, or simply addresses, throughout the rest of this paper.

To transfer bitcoins, a user issues a transaction, which consists of a set of inputs, a set of outputs, and a change address. The inputs are Bitcoin addresses that belong to the payer. The outputs are Bitcoin addresses that identify the payee accounts, and the change address (which is optional) is where the leftover bitcoins from the transactions are sent (the change address belongs to the payer). Bitcoin's transaction protocol stipulates that inputs to a new transaction must reference the exact value of outputs from previous transactions. In other words, users must specify from whom they received the bitcoins, thus forming a chain of transactions, and inputs to a new transaction may reference as many previous transactions needed to sufficiently fund the new output. These are known as multi-input transactions. Finally, the payer digitally signs a hash[3] of the transaction from which he or she received the bitcoins and the public key address of the payee.

We illustrate Bitcoin's transaction protocol in Figure 1. We see that there are three transactions, A, B, and C, and we will refer to their respective owners as Alice, Bob, and Charlie. In transaction A, a transaction of 2.0 BTC sent to Alice is used as an input to a transaction of 1.5 BTC to Charlie. Since the input value exceeds the output value, the remaining 0.5 BTC is sent to a change address belonging to Alice. In transaction B, we see a similar scenario in which Bob transfers 1.0 BTC to Charlie, but a change address is not necessary because the input and output values were equivalent. In transaction C, we see that Charlie transfers 2.0 BTC to an unnamed user. To issue this transaction, transaction C references the two previous transactions, from Alice and Bob, that Charlie received as inputs. Again, the total input value exceeds the required output value, so 0.5 BTC are sent to a change address belonging to Charlie.

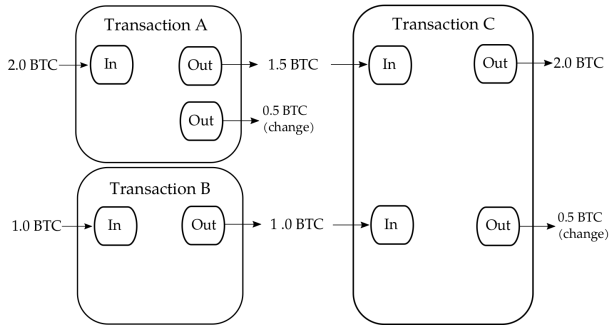

Fig. 1. The diagram illustrates the anatomy of bitcoin transactions. We have transactions A, B, and C owned by Alice, Bob, and Charlie, respectively. We can see that previous transactions to Alice and Bob are referenced in their respective transactions to Charlie, forming a chain of transactions.

We can see that the validity of a bitcoin is dependent on the correctness of each signature in the transaction chain, it is simple to verify a transaction's history but difficult to tamper with confirmed transactions that are deeply embedded in the blockchain (usually six blocks of confirmation is considered secure against double-spending attacks [12]). Therefore, Bitcoin transactions are essentially irreversible. This feature, coupled with Bitcoin's pseudonymity, enables cybercriminals to commit financial fraud that is virtually impossible to reverse and difficult to trace.

B. CryptoLocker Ransomware

On September 5, 2013, CryptoLocker emerged as a new family of ransomware that encrypted files on a victim's system until a ransom is paid [11]. The decryption keys were withheld by the threat actors who demanded ransoms to be paid either through MoneyPak or Bitcoin within 72 hours, otherwise the decryption keys would (allegedly) be destroyed and recovery of the encrypted files would be virtually impossible.

CryptoLocker's infection vector took two forms. In its initial release, CryptoLocker threat actors primarily targeted business professionals via spam emails taking the form of "customer complaints" against recipients' organizations. Malicious executable files were attached in ZIP archives, which would aggressively encrypt all the files on a system if opened. Later versions of CryptoLocker, beginning on October 7, 2013, were distributed through Gameover ZeuS [13], a peer-to-peer botnet that used Cutwail spam botnet to send massive amounts of spam email impersonating well-established online retailers and

[2]Elliptic Curve Digital Signature Algorithm (ECDSA)
[3]SHA-256

financial institutions. These emails typically spoofed invoices, order confirmations, or urgent unpaid balances to lure victims into following malicious links which redirected to CryptoLocker exploit kits.

From September 2013 through early 2014, CryptoLocker infections were most prevalent in the United States. A study from the Dell SecureWorks Counter Threat Unit (CTU) research team [11] shows that from October 22, 2013 to November 1, 2013, 22,360 systems were infected in the United States, constituting 70.2% of global CryptoLocker infections. During this period, CryptoLocker was also prevalent in Canada, Great Britain, India, and several countries in the Middle East and South Central Asia. In a later sample, gathered from December 9, 2013 to December 16, 2013, during a sinkhole when CryptoLocker activity was limited, CryptoLocker infections became more dispersed. The concentration of infected systems in the United States dramatically declined to 23.8% (1,540 infected systems) and CryptoLocker activity became more prevalent in Great Britain (1,228 infected systems constituting 19.0% of global infections), Australia (836 infections constituting 12.9% of global infections), several other European countries, China, India, and Brazil.

Although the United States is disproportionately represented in total global CryptoLocker infections, most ransom payments from the United States were issued through MoneyPak, an invariably more economical option than Bitcoin. This was due to bitcoin's price volatility at the time. Conversion rates soared from $120/BTC in September 2013 to well over $1,300/BTC in late November 2013. We show the exchange rate of US dollars and bitcoin throughout CryptoLocker's operational period in Figure 2. As a result, the CryptoLocker threat actors adjusted the ransom demand on several occasions to ensure that ransom demands were not exorbitant. Since, in almost all cases in the United States, there were no advantages to paying through the Bitcoin system, individuals who elected to pay via Bitcoin presumably resided in countries outside of the United States where MoneyPak was unavailable [11]. For this reason, bitcoin ransom payments represent only a small portion of the total financial damages caused by CryptoLocker.

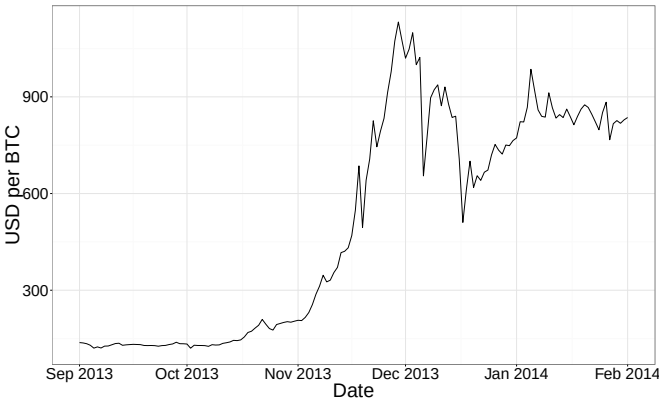

Fig. 2. The plot shows exchange rates between US dollars and bitcoins during CryptoLocker's operational period from September 2013 through January 2014.

III. Measurement Methodology

We next explain our approach to collecting information from the blockchain and various online fora in an effort to measure CryptoLocker's economy and generate threat intelligence on the CryptoLocker operation. We begin by highlighting how we generated an address cluster belonging to CryptoLocker, which we call S_{CL}, from two seed addresses in Section III-A. Based on information from previous efforts and a preliminary examination of S_{CL}, we designed and implemented a framework for identifying ransom payments from the set of all transactions sent to our cluster in Section III-B.

A. Collecting and Generating Addresses

To collect Bitcoin addresses belonging to CryptoLocker, we used an approach similar to M. Spagnuolo's study on CryptoLocker using BitIodine [7], an open source framework for forensic analysis of the blockchain. Initially, we found two known CryptoLocker addresses by manually investigating a reddit thread[4] in which victims and researchers posted Bitcoin addresses belonging to the ransomware. We will refer to one of these seed addresses, which collected 27 ransom payments, as A_{seed1}, and the other seed address, which collected 23 ransom payments, as A_{seed2}, throughout the remainder of this paper. An interested reader can find the hashes of all Bitcoin addresses mentioned in the Appendix. To expand our dataset of addresses, we use the following two clustering heuristics (based on the Bitcoin transaction protocol detailed in Section II-A) to generate a set of Bitcoin addresses controlled by the same user(s), known as *clusters* [4]:

1) *Multi-input Transactions*: A multi-input transaction occurs when a user u makes a transaction in which the payment amount p exceeds the available bitcoins (references of prior payments to u) in u's wallet[5]. In such a case, the Bitcoin protocol inputs a set of bitcoins B from u's wallet to sufficiently fund p. Therefore, we can conclude that if the bitcoins in B are owned by a set of distinct input addresses S_i, the input addresses in S_i are controlled by the same user u.
2) *Change Addresses*: The Bitcoin protocol generates a new change address in u's wallet to collect change when the sum of inputs in B exceeds p. When the set of output addresses S_o contains two addresses such that one is a newly generated address a_n and the other corresponds to a payment's destination address a_d, we can conclude that a_n is a change address and is controlled by u.

From A_{seed1}, we generate a set of 968 Bitcoin addresses belonging to the CryptoLocker cluster, S_{CL}, which happens to include A_{seed2}. Although our study does not claim to be representative of the entire CryptoLocker population within the Bitcoin ecosystem, which is difficult to quantify due to a lack

[4]https://www.reddit.com/r/Bitcoin/comments/1o53hl/disturbing_bitcoin_virus_encrypts_instead_of/

[5]A "wallet" is a collection of private keys. It is the Bitcoin equivalent of a bank account.

of *confirmed* Bitcoin addresses belonging to CryptoLocker, we can systematically measure a subset of CryptoLocker's economy, S_{CL}, and make inferences about CryptoLocker and its constituents.

B. Ransom Identification Framework

The goal of our ransom identification framework is to distinguish ransom payments from the set of all transactions to S_{CL}. Foremost, we designed and implemented our identification framework to be precise — we wanted to identify ransom payments with a high degree of confidence and minimize the number of extraneous transactions included in our dataset. Second, we built our framework to be lightweight — rather than querying the entire Bitcoin blockchain, we chose a semi-automatic approach to crawl and parse transactions to S_{CL} using the Blockchain API [14]. For each address in S_{CL}, points of interest included the total number of transactions, total sent and received bitcoins, and the number of ransom payments received. For each transaction, we were interested in input and output addresses, bitcoins transferred, and timestamps (in UNIX epoch time).

The time and ransom parameters in our identification framework reflected findings from previous studies on CryptoLocker ransomware [7], [11] and our own cursory analysis of S_{CL}. The heuristics we use are as follows:

- Payments of approximately 2 BTC ($\pm$0.1 BTC) between September 5, 2013[6] (CryptoLocker release date) and November 11, 2013 to allow for a three-day ransom period after CryptoLocker authors decreased the ransom amount to 1 BTC around November 8
- Payments of approximately 1 BTC ($\pm$0.1 BTC) between November 8, 2013 and November 13, 2013 to allow for a three-day ransom period after CryptoLocker authors decreased the ransom amount to 0.5 BTC around November 10
- Payments of approximately 0.5 BTC ($\pm$0.05 BTC) between November 10, 2013 and November 27, 2013 to allow for a three-day ransom period after CryptoLocker authors decreased the ransom amount to 0.3 BTC around November 24
- Payments of approximately 0.3 BTC ($\pm$0.05 BTC) between November 24, 2013 and December 31, 2013
- Late payments of approximately 10 BTC ($\pm$0.1 BTC) between November 1, 2013 and November 11, 2013 when CryptoLocker introduced their "CryptoLocker Decryption Service" for victims who failed to pay ransoms within the given time frame
- Late payments of approximately 2 BTC ($\pm$0.1 BTC) between November 11, 2013 and January 31, 2014 when CryptoLocker decreased the cost of their "CryptoLocker Decryption Service"
- Payments of approximately 0.6 BTC ($\pm$0.1 BTC) between December 20, 2013 and January 31, 2014

[6]Time intervals are in Universal Time Coordinates (UTC) from the start of the initial day to the end of the final day.

IV. Data Overview

We begin by validating the accuracy of our data and showing how conservative our estimates are in Section IV-A. We then perform valuations of the CryptoLocker economy and measure ransom payments at different stages in its operational period in Section IV-B.

A. Data Validation

We realize that it is difficult to both comprehensively measure S_{CL}'s economy and precisely identify *all* ransom payments to S_{CL} beyond a reasonable doubt. In turn, the stringent transaction value and timestamp parameters used in our ransom identification framework provide a lower bound estimate of ransom payments and S_{CL}'s economy. To gauge how conservative our previous estimates are using the aforementioned framework, we take three different measurements with varying transaction value and timestamp parameters. The three measurement methods are as follows: Method 1) transactions to S_{CL} without any transaction value or timestamp filters, Method 2) transactions to S_{CL} filtered only by transaction values, and Method 3) transactions to S_{CL} filtered by both transaction values and timestamps (ransom identification framework).

TABLE I
Measurements by Method

Method	Transactions	BTC	USD
Method 1	1,071	1,541.39	539,080.69
Method 2	933	1,257.27	373,934.76
Method 3	795	1,128.40	310,472.38

Table I and Figure 3 show the daily volumes of bitcoins paid to S_{CL}, and their values in US dollars commensurate with daily exchange rates[7], using the three different measurements. We see that there are clear disparities between our estimates from Method 1, which accounted for all transactions to S_{CL}, compared to our estimates from Method 2 or 3, particularly in the month of November. The causes of the discrepancies on November 13 and 14 are four 7 BTC transactions, which are filtered by Methods 2 and 3, that we cannot find any conclusive evidence on. The causes of the discrepancies from November 25 through November 27 are eight 4 BTC transactions and a single transaction[8] of 15.6 BTC (November 25). We discover that these transactions eventually end up in the secondary money laundering address used in the Sheep Marketplace scam of 2013 (see Section VI-B). Comparing estimates between Method 2, which filtered transactions based on known ransom amounts demanded by CryptoLocker (i.e. 2 BTC, 0.5 BTC, 0.3 BTC), and Method 3, which filtered transactions by ransom amounts *and* their respective periods of activity, we see smaller differences between our estimates, which is a good indication that the time intervals chosen in our ransom identification

[7]There are many BTC/USD exchange rates available. We use the "24h average" BTC/USD exchange rate from http://www.quandl.com.

[8]https://blockchain.info/tx/43355544/

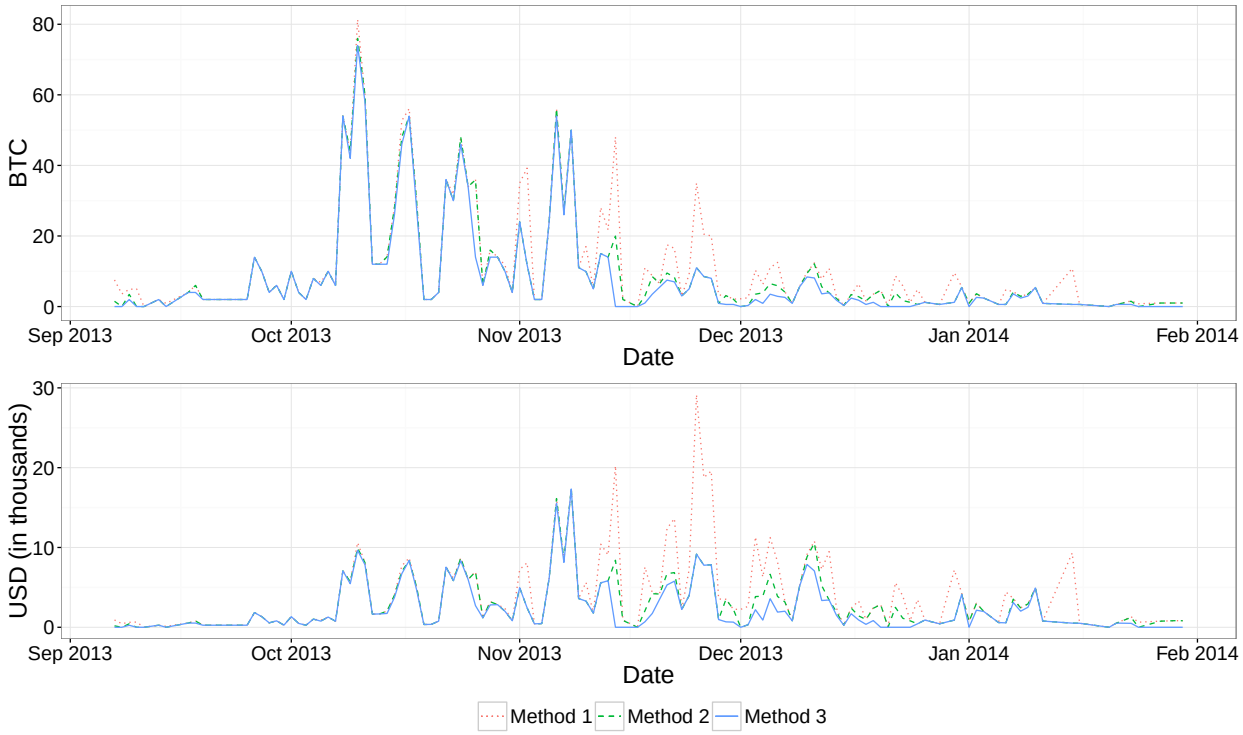

Fig. 3. The plots show longitudinal trends in the value of ransom payments to S_{CL} from September 2013 through January 2014. We compare data yielded by our identification framework to two other measurements detailed in Section III-B. Method 1 measures all transactions to S_{CL}, Method 2 measures ransom payments filtered by known bitcoin ransom demands, and Method 3 measures ransom payments filtered by both ransom demands and timestamps (ransom identification framework).

framework (including the 72-hour buffer periods) produce reliable estimates. This is important because filtering by ransom amounts is a relatively strong and straightforward heuristic, whereas filtering by time intervals is an invariably weaker heuristic. This is because the date that a ransom is paid is dependent on the date that an individual's system is infected and the 72-hour payment window, so it is entirely possible that a system could have been infected by an older version of CryptoLocker demanding an outdated ransom. We consider these anomalous ransom payments and omit these transactions by choosing Method 3 for the purpose of maintaining reliable and precise data for analyzing trends in ransom payments in Section V.

B. CryptoLocker Economy in Bitcoin

Using Method 3, we identify 795 ransom payments to S_{CL}, which contribute a total of 1,128.40 extorted bitcoins. Using daily bitcoin to USD exchange rates, we estimate that these ransom payments valued $310,472.38. However, as we mentioned before these figures are conservative and the payouts to S_{CL} may have been as much as 1,541.39 BTC and $539,080.69 based on Method 1, though we cannot be certain that the unaccounted transactions are ransom payments.

Since the exchange rates for bitcoins were quite volatile throughout the months of CryptoLocker's operation, the value of these extorted bitcoins would have seen a meteoric increase in price should they have been exchanged at the height of bitcoin's exchange rate. In Figure 4, we show a valuation in US dollars of the CryptoLocker economy throughout its operational period. We can see our estimate of $310,472.38 with the assumption that the CryptoLocker threat actors cashed out ransom proceeds at the end of each day ("daily cash in"). In the "cumulative cash in" curve, we assume that the CryptoLocker threat actors aggregated the bitcoins collected from ransom payments, and we perform a valuation commensurate with the USD/BTC exchange rate on each day. Based on the latter assumption, we estimate that the peak valuation of the CryptoLocker economy occurred on November 29, 2013 when they had collected a total of 1,044.13 BTC worth upwards of $1.18 million. The USD/BTC exchange rate also reached its peak on that day at $1,332.26/BTC. This estimate is corroborated by the valuation of CryptoLocker provided by Spagnuolo *et al.* at $1.1 million taken on December 15, 2013 [7].

It is difficult to accurately measure or visualize the rate of CryptoLocker infections from Figure 3 due to the changing ransom demands and exchange rates for bitcoins, so we construct a time series plot showing the frequency of ransom payments to S_{CL} throughout CryptoLocker's operational period. From Figure 5, we begin to see low levels of activity starting on September 9, 2013 when the first ransom[9] of 1.99 BTC is

[9]https://blockchain.info/tx/33208314/

TABLE II
SUMMARY OF CRYPTOLOCKER RANSOM TYPES

Type	Time period	No. ransoms	BTC	USD
2 BTC	Sep. 5 - Nov. 11	422	843.77	146,623.33
10 BTC (Late)	Nov. 1 - Nov. 11	9	89.80	23,780.19
1 BTC	Nov. 8 - Nov. 13	43	43.04	15,904.49
0.5 BTC	Nov. 10 - Nov. 27	116	58.01	46,415.23
2 BTC (Late)	Nov. 11 - Jan. 31	10	20.00	14,492.46
0.3 BTC	Nov. 24 - Dec. 31	144	43.19	37,759.44
0.6 BTC	Dec. 20 - Jan. 31	51	30.59	25,497.24
Total	Sep. 5 - Jan. 23	795	1,128.40	310,472.38

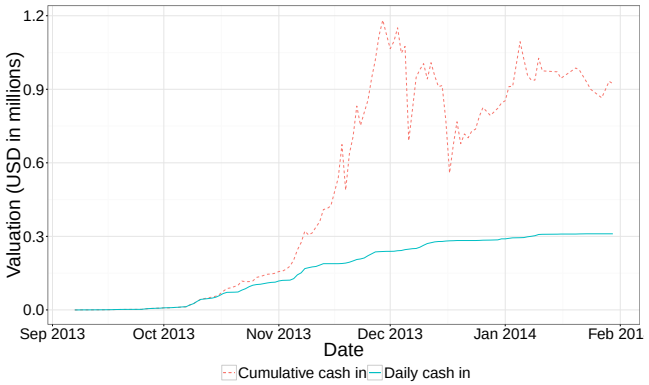

Fig. 4. The plot shows valuations of S_{CL} (using Method 3) in USD. The "cumulative cash in" curve performs a valuation commensurate with the total bitcoins extorted in USD while the "daily cash in" curve performs a valuation commensurate with daily bitcoins extorted to USD.

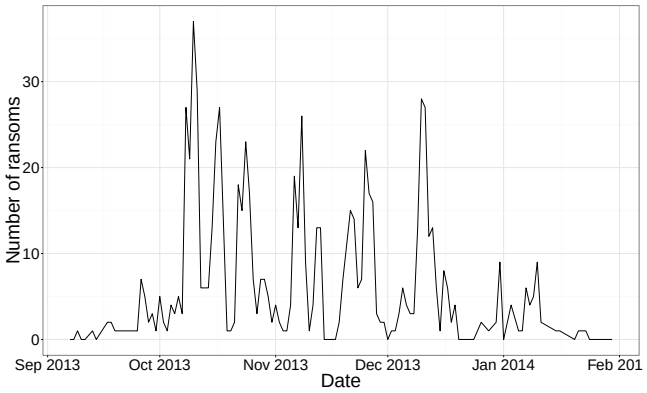

Fig. 5. The plot shows number of ransoms paid to S_{CL} on each day from September 2013 to January 2014.

paid to S_{CL}. On October 8, 2013 through October 11, 2013, we see a sharp increase in ransom payments (27, 21, 37, and 29, respectively), which is consistent with our knowledge on CryptoLocker's use of the spam botnet Gameover ZeuS starting on October 7, 2013 [13]. As a result, we see that the original ransom of 2 BTC constituted 422 of the 795 identified ransoms and nearly half of the total transaction volume in US dollars. This was undoubtedly CryptoLocker's most prolific period in terms of successful infections. Over the next two months, we see undulating periods of activity which leads us to believe that CryptoLocker may have been distributed in several batches throughout its operation. S_{CL} experiences a significant decline in ransom payments starting in mid-December of 2013 and eventually comes to a close by the end of January 2014; we are not aware of any further CryptoLocker addresses after this period.

V. DATA ANALYSIS

Our goal is to gain insight on CryptoLocker's targets and changes in targets throughout its operation by statistically determining distinct distributions in the times of day that different ransom types (i.e. 2 BTC, 1 BTC) were paid to S_{CL}. We achieve this by performing Kolmogorov-Smirnov (goodness of fit) tests on ransom payment timestamps to determine whether different ransom types come from different populations in Section V-A. We analyze these trends and explain our findings in Section V-B.

A. Kolmogorov-Smirnov Tests

The Kolmogorov-Smirnov test for goodness of fit is based on the maximum difference between either an empirical and a hypothetical cumulative distribution (one-sample) or two empirical cumulative distributions (two-sample) [15], [16]. For our study, we use two-sample Kolmogorov-Smirnov tests to determine, using ransom payment timestamps, whether or not samples from different ransom types come from different parent populations at the 99.5% confidence level. Our sample populations include 2 BTC, 1 BTC, 0.5 BTC, 0.3 BTC, and 0.6 BTC ransoms, but exclude 10 BTC (Late) and 2 BTC (Late) ransoms, because we do not have sufficient sample sizes. We provide metadata on each type of ransom in Table II.

We begin by stating the null hypotheses of our Kolmogorov-Smirnov tests: $f_{n_1}(x)$ and $f_{n_2}(x)$ are samples of two empirical cumulative distribution functions $f_1(x)$ and $f_2(x)$, and that

$$H_0 : f_1(x) = f_2(x) \qquad -\infty \leq x \leq +\infty \tag{1}$$

The alternative hypotheses are that

$$H_1 : f_1(x) \neq f_2(x) \qquad -\infty \leq x \leq +\infty \tag{2}$$

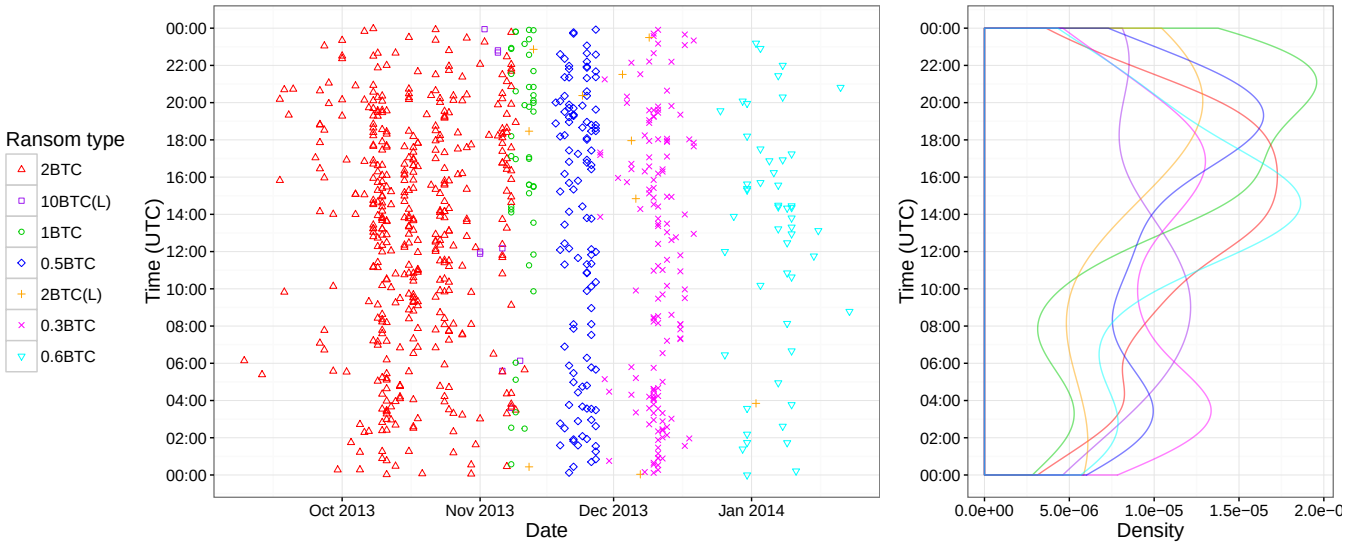

Fig. 6. The plot shows trends in the times of day that ransoms were paid to S_{CL}.

TABLE III
FIVE-NUMBER SUMMARY AND MEAN (H:M:S UTC)

Sample	Min.	Q_1	Median	Q_3	Max.	Mean
2 BTC	00:01:20	08:55:36	13:55:06	18:01:00	23:58:33	13:12:42
1 BTC	00:34:31	14:11:09	17:07:49	21:11:29	23:54:33	16:28:47
0.5 BTC	00:07:27	05:51:07	15:16:48	19:23:24	23:55:19	13:14:47
0.3 BTC	00:06:59	04:15:56	11:33:30	17:29:14	23:54:45	11:16:19
0.6 BTC	00:00:01	08:27:37	13:53:14	16:54:00	23:11:33	12:36:46

	2 BTC	1 BTC	0.5 BTC	0.3 BTC	0.6 BTC
2 BTC		0.3028 | 0.2770	0.1449 | 0.1815	0.1802 | 0.1671	0.1068 | 0.2566
1 BTC			0.2476 | 0.3089	0.3532 | 0.3006	0.3461 | 0.3582
0.5 BTC				0.1865 | 0.2158	0.2067 | 0.2907
0.3 BTC					0.1908 | 0.2819
0.6 BTC					

Fail to reject H_0 $D \mid D_{crit}$ Reject H_0 $D \mid D_{crit}$

Fig. 7. The table compares the test statistic D and the critical value D_{crit} for all permutations of ransom types. If $D > D_{crit}$, we may reject the null hypothesis and assume that the samples come from two different parent populations at the 99.5% confidence level. These are shown in red.

We then compute empirical cumulative distribution functions, $f_1(x)$ and $f_2(x)$, from the two samples. To compute the test statistic D from $f_1(x)$ and $f_2(x)$, we find the maximum absolute difference over all values of x given by

$$D = \max_x |F_{n_1}(x) - F_{n_2}(x)| \tag{3}$$

After computing D for all permutations of our sample populations, we compute critical values, D_{crit}, at the 99.5% confidence level for each permutation given by

$$D_{crit} = 1.73\sqrt{(n_1 + n_2)/n_1 n_2} \tag{4}$$

In Figure 7, we compare D and D_{crit} for each permutation. For permutations where $D < D_{crit}$, we fail to reject the null hypothesis, which tells us that the two samples did not come from different populations at the 99.5% confidence level. For permutations where $D > D_{crit}$, we may reject the null hypothesis, which tells us that the two samples come from two different populations at the 99.5% confidence level.

B. Analysis of Timestamp Data

Before we go into detail on the time series and density plots in Figure 6 and a statistical summary of time distributions in Table III, we put into perspective a hypothetical time distribution that CryptoLocker victims from the United States paid primarily in bitcoin (which we know to be false from the CryptoLocker study by CTU researchers [11]). What we do know is that CryptoLocker targeted business professionals, so we would expect ransom payments to be transacted during typical working hours, or at the very least, during the daytime. Let us assume a typical 9:00 a.m. to 5:00 p.m. working day for business professionals in the United States. In Pacific Time (PT, UTC-08:00), 9:00 a.m. to 5:00 p.m. would correspond to 17:00 UTC to 1:00 UTC on the following day. In Eastern Time (ET, UTC-05:00), 9:00 a.m. to 5:00 p.m. would correspond to 14:00 UTC to 22:00 UTC. Based on Figure 6 and Table III,

the only sample that could follow such a distribution is the 1 BTC sample, however, we cannot conclude that 1 BTC ransom payments were primarily paid by victims from the United States without further evidence. Because we have no reason to believe that the marginal number of CryptoLocker victims in the United States who opted to pay using bitcoin would choose to pay in the nighttime, we assume that ransoms paid to S_{CL} came from outside of the United States.

From our Kolmogorov-Smirnov tests, we find that the timestamp sample of 2 BTC ransoms differs from samples of 1 BTC and 0.3 BTC ransoms. Additionally, we find that the sample of 1 BTC ransoms differs from the sample of 0.3 BTC ransoms. Thus, we know that all three samples come from different populations at the 99.5% confidence level. In the 2 BTC sample, the median time of payment is 13:55:06 UTC with an interquartile range ($Q_3 - Q_1$) of 09:05:24 hours. From the density curve in Figure 6, we see that it has a unimodal distribution, which suggests that a large portion of ransom payments during this time came from the same region. Referring back to the CTU researchers' study, we see that from October 22, 2013 to November 1, 2013, they measure 1,767 infections from Great Britain, which has the second highest percentage of total global infections (5.5%) after the United States (70.2%). If we use our rough 9:00 a.m.–5 p.m. "working day conjecture", we would expect ransom payments from Great Britain (GMT, UTC±00:00) to be concentrated between 9:00 UTC and 17:00 UTC. This is consistent with the first and third quartiles from the 2 BTC sample (08:55:20 UTC and 18:01:00 UTC), but again, we make no solid claims without further evidence.

Turning to the 1 BTC sample, the median time of payment is 17:07:49 UTC with an interquartile range of 07:00:20 hours. We see that it also has a unimodal distribution and restate that the distribution of timestamps from this sample roughly models what we would expect if a majority of these ransoms were paid from the United States. Next, we take a look at the 0.3 BTC sample, which has a median time of payment of 11:33:30 and an interquartile range of 13:13:18. We see that the large spread can be attributed to the 0.3 BTC sample's bimodal probability distribution, which suggests that the ransom sample came from two distinct sources. One of these sources closely resembles the time distribution from our 2 BTC sample, which we imputed to ransom payments from Great Britain. The second source of ransoms in early December falls between a 6-hour time interval from 23:00 UTC to 5:00 UTC on the following day. Referring back to the CTU researchers' study, from December 9, 2013 to December 13, 2013, their measurements show that CryptoLocker infections become more dispersed with 23.8% from the United States, 19.0% from Great Britain, and 12.9% from Australia. We continue to see CryptoLocker infections in Great Britain, which is consistent with one of the peaks in the bimodal 0.3 BTC sample. If we convert the time interval from 23:00 UTC to 5:00 UTC using our "working day conjecture" and set 23:00 UTC as 9:00 a.m., we would expect these ransoms to come from time zones with an offset of UTC+10:00. This corresponds to the Australian Eastern Time Zone, which is, again, consistent with measurements by the CTU researchers' CryptoLocker study, however we make no solid claims that a large portion of the 0.3 BTC ransoms undoubtedly came from Australia.

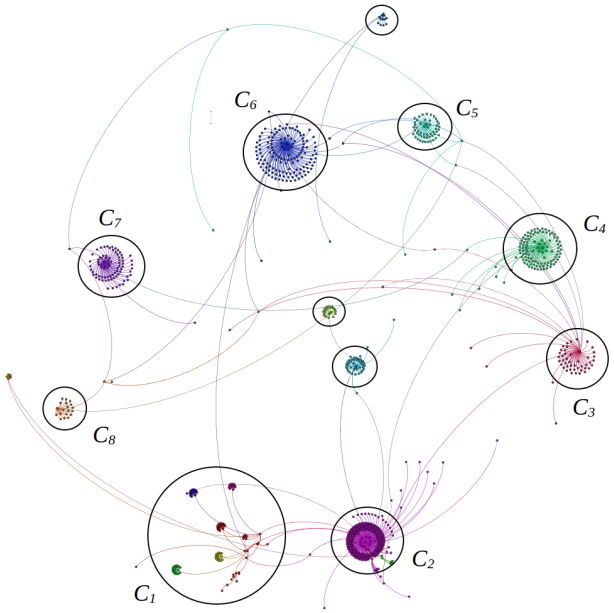

Fig. 8. A visualization of outgoing bitcoin transactions from the CryptoLocker cluster, S_{CL}, network. Based on Louvain Modularity for community detection, we distinguish 17 distinct sub-communities in the S_{CL} network. We see that the ransom balances from all addresses within a community are transferred to a single *aggregate address* at the center. We group several sub-communities together based on shared addresses and the times they were active (i.e. Community 1, or c_1) and identify eight communities of interest (c_1 through c_8).

VI. Financial Infrastructure

We next turn to understanding CryptoLocker's financial infrastructure by analyzing communities in the CryptoLocker cluster, S_{CL}. We begin by overviewing S_{CL}'s topology in Section VI-A. Examining S_{CL} at a lower level, we impute real-world data found on various online fora and Bitcoin services to distinct communities in Sections VI-B to VI-C. We find connections to BTC-e and Bitcoin Fog and speculations to the Sheep Marketplace scam and other bitcoin cybercrime.

A. Transaction Graph

To better understand CryptoLocker's financial infrastructure, we use Gephi[10], an open source software for network visualization and exploration, to visualize outgoing transactions from S_{CL} (Figure 8). In total, the network contains 993 nodes (addresses) and 1,020 weighted (by BTC), directed edges, which delineate one or more transactions between two addresses. The 993 addresses in the network are comprised of 968 addresses from S_{CL} and 25 additional addresses that are recipients of outgoing transactions from S_{CL}, which we

[10]http://gephi.github.io/

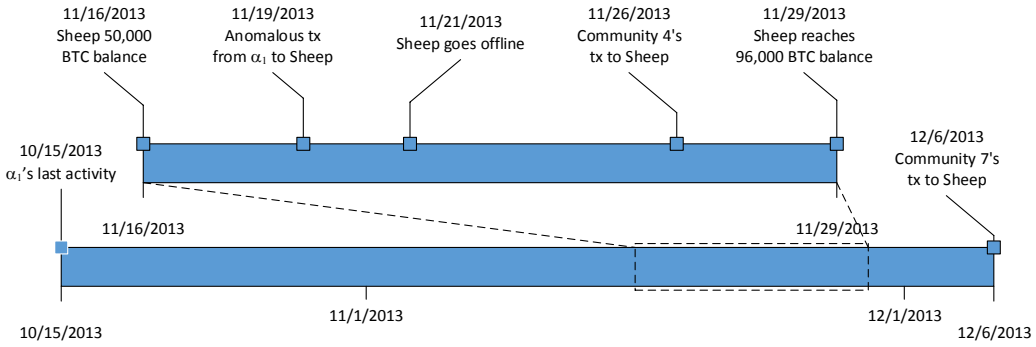

Fig. 9. Timeline of activity related to the Sheep Marketplace.

will call *aggregate addresses*. Using the Louvain Method for community detection, we distinguish 17 distinct communities in the CryptoLocker network [17]. An interested reader can see that the typical community is comprised of addresses in S_{CL} (child nodes) that report to a central aggregate address to which all collected ransom payments are sent to (parent node). We characterize eight communities of interest (arbitrarily and not chronologically named c_1 through c_8) that yield external information on addresses in S_{CL} from our analysis. Communities 2 through 8 are canonical communities in that they each correspond to only one aggregate address. Community 1, on the other hand, is not centralized around one single aggregate address, but rather is comprised of multiple small communities and aggregate addresses. We will expand on c_1's topology in Section VI-B, which we learn to be CryptoLocker's earliest community.

B. BTC-e and the Sheep Marketplace

Before we begin our analysis of CryptoLocker's communities, it helps to provide background on one pattern we found in several communities, namely an indirect connection to the Sheep Marketplace scam of 2013. The Sheep Marketplace was the successor of the Silk Road [18], [19], an anonymous online marketplace specializing in the trade of narcotics, after its infamous takedown in February 2011. Launched in March 2013, the Sheep Marketplace quickly gained traction as the leading anonymous online drug marketplace until, on November 21, 2013, the owners shut down the site and absconded with 96,000 bitcoins (over $100 million) belonging to its users [20]. We find that several aggregate addresses and child addresses make transactions to BTC-e[11], one of the largest bitcoin exchanges available where users can convert bitcoins into other cryptocurrencies, Dollars, Euros, and Rubles. From BTC-e, the bitcoins are then transferred to two money laundering addresses used in the Sheep Marketplace scam. While it is unclear what sort of connection the CryptoLocker operation and the Sheep Marketplace may share, if there is any connection at all, we continue by pointing out instances in which the two are somehow linked. A timeline of S_{CL}'s activity in relation to the Sheep Marketplace is provided in Figure 9 (with the assumption that these speculations are true).

1) Case study on A_{seed1}: We first manually analyze A_{seed1} to understand CryptoLocker's early operation. We discover an anomalous transaction sharing the aforementioned Sheep Marketplace pattern, which prompts us to look for this same pattern in other communities.

Incoming transactions: On September 7, 2013, A_{seed1} receives a payment of one BTC from a temporary address a_1 active for only that day. An outgoing payment for the same amount is processed four days later. We believe this was a test trial for CryptoLocker's payment system. On September 13, 2013, A_{seed1} receives its first ransom payment of two BTC. Throughout the rest of the month, A_{seed1} receives 26 additional ransom payments of approximately two BTC each, which totals to 53.9081 BTC (worth about $6,600 at the time of payment). On September 29, 2013, A_{seed1} receives an *anomalous payment of 0.0002 BTC* from an inconclusive single-use address a_2.

Outgoing Transactions: From September 27, 2013 to October 15, 2013, A_{seed1} makes nine outgoing payments amounting to 53.9081 BTC (the exact number of bitcoins collected from ransom payments) to aggregate addresses in c_1. A_{seed1}'s remaining balance after its distribution of ransom payments is the unaccounted 0.0002 BTC. Considering that the exact amount of bitcoins from ransom payments is transferred to aggregate addresses, as opposed to the entire balance, we hypothesize that transfers from child addresses to aggregate addresses were automated, rather than performed manually.

An Anomalous Transaction: On November 19, 2013, over a month after A_{seed1}'s last activity and right before the Sheep Marketplace shutdown, we find that the unaccounted 0.0002 BTC is transferred as part of a multi-input transaction of 15 BTC to an address $A_{exchange1}$ belonging to BTC-e. From the BTC-e address, the 15 BTC are included in a multi-input transaction of 1,000 BTC to an address that we later learn to be Sheep's primary money laundering address A_{sheep1}, which has processed over 466,000 BTC as of February 2015.

Our efforts to find external data on A_{sheep1} leads us to a reddit thread that details how user $sheeproadreloaded2$ [21] traced the $100 million of stolen bitcoins from the Sheep Marketplace, through Bitcoin Fog, to the money laundering

[11] https://btc-e.com/

address A_{sheep1}. Examining the transaction history of A_{sheep1}, we find that it has a balance of 50,000 BTC on November 16, 2013. Throughout the rest of the month, we notice a massive increase in activity as it receives a total of 96,500 BTC while sending a total of 50,000 BTC to auxiliary addresses, most notably a secondary money laundering address, A_{sheep2}.

Considering the time of A_{seed1}'s final transaction just two days prior to the Sheep Marketplace's alleged theft, and its discontinued use after transferring its balance to aggregate addresses, we speculate that A_{seed1}'s anomalous 0.0002 BTC may have been part of the Sheep scam. Posted within a week of A_{seed1}'s anomalous transaction on November 22, 2013, we find a reddit thread[12] detailing a similar occurrence in which a user reports having 3.9 BTC stolen from his or her Mt. Gox [22] account and transferred to an intermediate address before ending up in A_{sheep2}. This indicates that A_{seed1}'s anomalous transaction is not an isolated incident and we use this finding as a point interest for other CryptoLocker addresses.

2) Community 1: While canonical communities c_2 through c_8 each correspond to only one aggregate address, we group five small communities, which we will call sub-communities, and six aggregate addresses into c_1. The largest sub-community in c_1, containing 42 addresses (including addresses A_{seed1} and A_{seed2}), is the connecting component between the four surrounding sub-communities. We combine these five sub-communities into a single community, c_1, based on connectivity and the times they were active.

We find that addresses in c_1 were active during the onset of CryptoLocker's attacks between September 9, 2013 and October 15, 2013. According to c_1's decentralized topography and the reuse of addresses for ransom collection (i.e. A_{seed1} and A_{seed2}), we can assume with a high degree of confidence that CryptoLocker's threat actors did not dynamically generate new addresses for its nascent attacks. The value in doing so would be to obfuscate relationships between CryptoLocker addresses and money laundering addresses.

Examining c_1 at a low level, we find that the addresses in c_1's largest sub-community report to two single-use aggregate addresses which collect 40 BTC and 20 BTC. From here, we see that these balances are processed through long chains of single-input transactions with fractional bitcoins chipped away in each transaction, which is characteristic of the Bitcoin Fog mixer [8]. Bitcoin Fog is a mixing service accessible via Tor and allows users to deposit their bitcoins to up to five newly generated addresses. To deter obvious indications of using Bitcoin Fog, the service takes a randomized fee between 1–3% of the transaction value and processes bitcoins over a randomized timespan between 6 and 96 hours. The four remaining sub-communities in c_1 are all similar in size and structure. Each sub-community consists of between 14 and 18 addresses which lead to aggregate addresses holding 20 to 40 BTC in preparation for tumbling.

[12]https://www.reddit.com/r/Bitcoin/comments/1r9rtp/i_just_had_39_btc_stolen_from_my_mtgox_account/

3) Community 4: We next take a look at c_4, which is comprised of 110 addresses. Throughout mid-November (when CryptoLocker decreased the ransom amount due to the rising value of bitcoins), addresses in c_4 receive numerous ransom payments of 0.5 BTC. On November 26, 2013, the balances from addresses in c_4 are transferred to a single-use aggregate address A_{c4} belonging to BTC-e, which collects a total of 96.6 BTC. On the same day, the aggregate address transfers its balance into a multi-input transaction of 1,000 BTC to the Sheep Marketplace's secondary money laundering address, A_{sheep2}.

4) Community 7: Community c_7 is a medium-sized community containing 83 addresses. Its aggregate address A_{c4} also belongs to BTC-e and collects 82.2 BTC from December 5 to December 9, 2013. On December 6, 9, and 10, we see that the aggregate address transfers its bitcoins in multi-input transactions of 400 BTC, 500 BTC, and 300 BTC, which are also sent to Sheep's secondary address, A_{sheep2}.

C. Botnets, Speculations, and Misc.

From communities c_2, c_3, c_5, c_6, and c_8, we characterize how CryptoLocker's operation evolved over time and we find subtle connections to other bitcoin cybercrime.

1) Community 2: By far the largest community in the CryptoLocker network, c_2 accounts for one-third of the addresses in S_{CL}. We find that c_2 quickly follows c_1 and is active from October through mid-November, CryptoLocker's most prolific period of operation. Consistent with CryptoLocker's use of Gameover ZeuS and the Cutwail botnet starting on October 7, 2013, the increased distribution of ransomware called for more addresses to collect ransom payments. Out of the 328 addresses in c_2, 318 are single-use addresses. Thus, we infer that the CryptoLocker threat actors opted to dynamically generate addresses, instead of reusing addresses, to obfuscate their activity.

At the center of c_2 is an aggregate address A_{c2} belonging to BTC-e, which has received a total of 5,332.8 BTC. There is a notable disparity between how the extorted bitcoins were handled between c_1 and c_2. In c_1, we see a clear indication that the CryptoLocker threat actors decide to use Bitcoin mixers, particularly Bitcoin Fog, to launder their proceeds. In c_2, however, we see that the threat actors decide to generate new addresses for each ransom payment and directly transfer their proceeds to BTC-e without any means of obfuscation. The downfalls of the former method, which include a prolonged return on proceeds, transaction fees, diminished anonymity for large transactions, and trust in a third party, may have deterred CryptoLocker's growing enterprise from continued use of Bitcoin Fog. In the latter method, the CryptoLocker authors might have assumed that newly generated addresses would be sufficient in preserving privacy.

2) Community 3: Community c_3 is a medium-sized community consisting of 52 addresses. On February 12, 2014, 14.63 BTC are transferred to c_3's single-use aggregate address A_{c3} predominantly in 0.3 BTC ransom payments. Considering the time frame of 0.3 BTC ransom payments, we presume that

c_3 is one of CryptoLocker's later communities. The balance from c_3's aggregate address is transferred on March 3, 2014 as part of a 100 BTC multi-input transaction to an address a_3 which has processed 374 BTC as of April 2014. Further analysis of c_3 and its aggregate address does not yield any useful information, however, we note that this transaction is one of S_{CL}'s last movements.

3) Community 5: c_5 is a medium-sized community containing 57 addresses. On January 14, 2014, c_5's aggregate address A_{c5} collects a sum of 50 BTC from its component addresses. After just one hour, the balance is transferred as part of a multi-input transaction of 1,134.99 BTC to an address, a_4, that has processed over 277,000 bitcoins as of early 2014. In our search for external information, we find another reddit thread[13] posted on February 3, 2014 explaining how the reddit user's coins were subject to an unauthorized withdrawal attempt to a_4 (though the transaction was unsuccessful due to insufficient funds). We also find a BitcoinTalk thread[14] alleging that the address is related to the BitPay hack while another comment in a separate blogpost[15] suggests that the address belongs to Mt. Gox.

4) Community 6: Community c_6, the second largest community in the CryptoLocker cluster, is comprised of 141 addresses. The community's aggregate address A_{c6} collects a total of 100 BTC in ransoms payments on February 12, 2014 (another one of CryptoLocker's later communities). One day later, the balance is transferred as part of a 528.74 BTC multi-input transaction to an address, a_5, which has processed over 11,000 bitcoins as of February 2014. One reddit post [16] mentions that a_5 is related to the potential Mt. Gox address, a_4. Thus, we assume that a_5 is, at the very least, tied to some kind of Bitcoin cybercrime.

5) Community 8: c_8 is comprised of just 22 addresses. The aggregate address A_{c8} receives just 6.67 BTC on February 12, 2014. One day later, we see that the balance is transferred to a_5. Again, we assume that this community may be linked to some kind of Bitcoin cybercrime.

VII. Related Work

The public's interest in Bitcoin has continually grown throughout the years, undeterred by countless hacks and scams. Therefore, it is important for users to fully understand the implications and limitations of the Bitcoin system. Numerous studies have thoroughly examined the flawed provisions for privacy inherent in the Bitcoin system. Reid and Harrigan performed one of the first analyses of anonymity in the Bitcoin system and were able to attribute external identifying information to addresses using a nascent representation of the Bitcoin transaction network [3]. Androulaki *et al.* tested Bitcoin's privacy provisions, in both the actual Bitcoin environment and a simulation of Bitcoin in a university setting, and proved that behavior-based and transaction-based clustering techniques could effectively deanonymize up to 40% of Bitcoin users in the simulation [4]. Ron and Shamir perform an analysis of the entire transaction graph and examine how users spend bitcoins, how bitcoins are distributed among users, and means by which users protect their privacy in the Bitcoin system [5]. Meiklejohn *et al.* explored Bitcoin wallets by using clustering heuristics (similar to Androulaki *et al.*) in order to classify the owners of Bitcoin wallets and discussed the growing discrepancy between potential anonymity and actual anonymity in the Bitcoin protocol [6].

In the growing literature on measuring cybercrime, our study on CryptoLocker is related to a number of works aimed at analyzing cybercrime in the Bitcoin ecosystem. Christin performed a comprehensive measurement analysis of the Silk Road, detailing the anonymous marketplace's constituents and discussing socioeconomic and policy implications of the results [18]. In a later study, Soska and Christin perform a longitudinal study of online anonymous marketplaces, which includes the Sheep Marketplace, and examines how these virtual marketplaces have grown and evolved [19]. Spagnuolo *et al.* created an expandable framework, called BitIodine [7], used for forensic analysis of the Bitcoin blockchain; they investigate addresses corresponding to the Silk Road owner, known as Dread Pirate Roberts, and CryptoLocker ransomware. Möser evaluated the effectiveness of several mixing services, commonly used to launder bitcoins, on preserving privacy [8]. Vasek *et al.* investigated the impact of distributed denial-of-service attacks on popular Bitcoin services [23]. Ron and Shamir used the blockchain to timeline events leading to Silk Road owner Dread Pirate Robert's accumulation of wealth before his arrest [24]. In a similar fashion to many of the aforementioned works, we have performed the first in-depth, systematic analysis of the CryptoLocker network in the Bitcoin ecosystem.

VIII. Discussion and Future Work

Our study corroborates findings from previous studies on CryptoLocker ransomware, but we want to emphasize that the measurements and analysis in our own study were solely drawn from blockchain analysis and crawling publicly available data. In some cases of cybercrime analysis, this can be a substantial limitation, but in others, this can be quite a valuable resource. We concede that our findings are, at best, controvertible assumptions unless further concrete evidence on the CryptoLocker operation proves otherwise; although we show that our findings are consistent with other studies on CryptoLocker, our study cannot standalone prove any of the conjectures made beyond a reasonable doubt. However, considering how rapidly the cybercrime landscape is changing, there will be circumstances in which we, as researchers, do not possess substantial evidence on nascent cybercrimes. In this regard, we may use some of the techniques discussed in our study to form a rudimentary understanding on these new schemes.

[13]https://www.reddit.com/r/Bitcoin/comments/1wvz66/who_is_taking_my_bitcoins/

[14]https://bitcointalk.org/index.php?topic=399024.0

[15]http://btcanalytics.blogspot.com/2014/02/bitcoins-most-mysterious-wallet.html

[16]https://www.reddit.com/r/DarkNetMarkets/comments/1xw39e/ok_so_everyones_trying_to_find_out_the_giant/

We have performed an in-depth, measurement analysis on CryptoLocker's economy and financial infrastructure. Initially, from two seed CryptoLocker addresses gathered from online fora, we generate a cluster of 968 addresses belonging to the CryptoLocker enterprise. From there, we identify and quantify ransoms paid (in bitcoin) by victims and produce an estimate upwards of $310,472.38 in financial damages. Based on our analysis of ransom timestamps, we form conjectures on regions where CryptoLocker infections were prevalent from trends in the times that ransom payments were made and compare our findings with other studies on CryptoLocker. Finally, we visualize the CryptoLocker operation's underlying financial infrastructure and analyze the topology in a modular fashion. We find subtle links between CryptoLocker, the Sheep Marketplace heist, and other Bitcoin cybercrime schemes. Our findings suggest that Bitcoin cybercrime may be much more interconnected than originally considered and that further analysis could uncover and confirm any existing relationships in the underground cybercrime landscape.

Given the increased prevalance of ransomware in the cybercrime landscape as of late, we believe that our study may enable further research in longitudinally measuring the evolution, economies, and strategies of ransomware criminal enterprises. We plan to further develop our heuristics for identifying ransom transactions from the Bitcoin blockchain in order to comprehensively measure the economies of CryptoLocker, as well as newer families of ransomware, and better understand their underlying financial infrastructures and money laundering strategies.

ACKNOWLEDGEMENTS

This research was supported in part by grants from Army Research Office and Center for Cybersecurity and Digital Forensics at Arizona State University. The information reported here does not reflect the position or the policy of the funding agencies.

REFERENCES

[1] S. Nakamoto. (2012) Bitcoin: A peer-to-peer electronic cash system. [Online]. Available: http://www.bitcoin.org/bitcoin.pdf

[2] S. T. Ali, D. Clarke, and P. McCorry, "Bitcoin: Perils of an unregulated global p2p currency," in *Security Protocols XXIII*. Springer, 2015, pp. 283–293.

[3] F. Reid and M. Harrigan, *An analysis of anonymity in the bitcoin system*. Springer, 2013.

[4] E. Androulaki, G. O. Karame, M. Roeschlin, T. Scherer, and S. Capkun, "Evaluating user privacy in bitcoin," in *Financial Cryptography and Data Security*. Springer, 2013, pp. 34–51.

[5] D. Ron and A. Shamir, "Quantitative analysis of the full bitcoin transaction graph," in *Financial Cryptography and Data Security*. Springer, 2013, pp. 6–24.

[6] S. Meiklejohn, M. Pomarole, G. Jordan, K. Levchenko, D. McCoy, G. M. Voelker, and S. Savage, "A fistful of bitcoins: characterizing payments among men with no names," in *Proceedings of the 2013 conference on Internet measurement conference*. ACM, 2013, pp. 127–140.

[7] M. Spagnuolo, F. Maggi, and S. Zanero, "Bitiodine: Extracting intelligence from the bitcoin network," in *Financial Cryptography and Data Security*. Springer, 2014, pp. 457–468.

[8] M. Moser, R. Bohme, and D. Breuker, "An inquiry into money laundering tools in the bitcoin ecosystem," in *eCrime Researchers Summit (eCRS), 2013*. IEEE, 2013, pp. 1–14.

[9] S. Meiklejohn and C. Orlandi, "Privacy-enhancing overlays in bitcoin," in *Financial Cryptography and Data Security*. Springer, 2015, pp. 127–141.

[10] P. Syverson, R. Dingledine, and N. Mathewson, "Tor: the second-generation onion router," in *Proceedings of the USENIX Conference on Security Symposium*. USENIX Association, 2004.

[11] K. Jarvis. (2014) Cryptolocker ransomware, 2013. [Online]. Available: http://www.secureworks.com/cyber-threat-intelligence/threats/cryptolocker-ransomware/

[12] G. O. Karame, E. Androulaki, and S. Capkun, "Double-spending fast payments in bitcoin," in *Proceedings of the 2012 ACM conference on Computer and communications security*. ACM, 2012, pp. 906–917.

[13] B. Stone-Gross. (2012) The lifecycle of peer-to-peer (gameover) zeus. [Online]. Available: https://www.secureworks.com/research/the_lifecycle_of_peer_to_peer_gameover_zeus

[14] Blockchain api. [Online]. Available: https://blockchain.info/api/blockchain_api/

[15] F. J. Massey Jr, "The kolmogorov-smirnov test for goodness of fit," *Journal of the American statistical Association*, vol. 46, no. 253, pp. 68–78, 1951.

[16] I. T. Young, "Proof without prejudice: use of the kolmogorov-smirnov test for the analysis of histograms from flow systems and other sources." *Journal of Histochemistry & Cytochemistry*, vol. 25, no. 7, pp. 935–941, 1977.

[17] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, 2008.

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

[19] K. Soska and N. Christin, "Measuring the longitudinal evolution of the online anonymous marketplace ecosystem," in *Proceedings of the 24th USENIX Conference on Security Symposium*. USENIX Association, 2015, pp. 33–48.

[20] S. Kerner. Sheep, bitcoin and the $100 million heist. [Online]. Available: http://www.eweek.com/security/sheep-bitcoin-and-the-100-million-heist.html

[21] sheepreloaded2. I just chased him through a bitcoin tumbler, and when he came out with 96,000 btc, i was waiting for him... [Online]. Available: https://www.reddit.com/r/SheepMarketplace/comments/1rvlft/i_just_chased_him_through_a_bitcoin_tumbler_and

[22] R. McMillan. The inside story of mt. gox bitcoin's $460 million disaster. [Online]. Available: http://www.wired.com/2014/03/bitcoin-exchange/

[23] M. Vasek, M. Thornton, and T. Moore, "Empirical analysis of denial-of-service attacks in the bitcoin ecosystem," in *Financial Cryptography and Data Security*. Springer, 2014, pp. 57–71.

[24] D. Ron and A. Shamir, "How did dread pirate roberts acquire and protect his bitcoin wealth?" in *Financial Cryptography and Data Security*. Springer, 2014, pp. 3–15.

APPENDIX

Table of Bitcoin Addresses	
A_{seed1}	1KP72fBmh3XBRfuJDMn53APaqM6iMRspCh
A_{seed2}	18iEz617DoDp8CNQUyyrjCcC7XCGDf5SVb
$A_{exchange1}$	161yYpWYCx8cWGYW95QaZ9NUuR3fd5n4xt
A_{sheep1}	1CbR8da9YPZqXJJKm9ze1GYf67eKAUfXwP
A_{sheep2}	174psvzt77NgEC373xSZWm9gYXqz4sTJjn
A_{c2}	1AEoiHY23fbBn8QiJ5y6oAjrhRY1Fb85uc
A_{c3}	1D5DoY5KxGtcatoxR5M3fSEeRpxfkgbsA3
A_{c4}	15F4g9sou83dNojKBikHx9vAkgnr1AhAEH
A_{c5}	1C1ypgQSiFdkwkRN7F9VrYFJG4wjyGPsfB
A_{c6}	1KpYNzAjTXZHwY782S4JMQbAXD8Ymyyw8s
A_{c7}	1AzMZUvHuakMgEaukDuzdAMujkXxgrHCGZ
A_{c8}	1LXbUwPDaBGAwxAQzutsLoTQJhTeyGskQb
a_1	1wffa72YQFTYJGMwuxbvWHFNAFGNKX3Bm
a_2	1Jfb6hbqfzwdLdR6Bspfh4HkND2mFj8LL
a_3	14CR4HYnJpvNFmtpoq34JUdHcA2rXeehkd
a_4	1Facb8QnikfPUoo8WVFnyai3e1Hcov9y8T
a_5	15WtLXz24WitRWtfdEzWPWZJYYDEBjjhUf