Understanding money flow in ransomware

JUSTIN RUITER, University of Twente, The Netherlands

Ransomware is and has been a growing problem for years now and as Ransomware-as-a-Service (RaaS) is becoming increasingly popular, the barrier to entry into deploying ransomware attacks is lowered for anyone, and it has become feasible for users without technical knowledge to deploy a ransomware attack. This study attempts to contribute to understanding how ransom profits are split between stakeholders and focuses on crypto locker ransomware, henceforth 'ransomware'. Understanding profit split might contribute to improving the detection of Bitcoin addresses linked to ransomware attackers or to derive information on the amount of work contributed by different parties in relation to the share they receive for it. This study will focus specifically on differences in money flow between RaaS and non-RaaS attacks. The questions posed will be assessed by combining publicly available data sources, the Bitcoin blockchain, and insights from related research.

CCS Concepts: • **Security and privacy** → *Malware and its mitigation.*

Additional Key Words and Phrases: ransomware, RaaS, Bitcoin, stakeholders

1 INTRODUCTION

Although ransomware is not a new development, according to reports by cyber security companies, the amount of commodity attacks and damage done has grown in recent years [7]. The SonicWall cyber threat report 2023 finds that ransomware attacks have become increasingly common, with 2021 as an outlier [24]. Because ransomware is a global phenomenon, it is often difficult for national police to track attacks and catch the attackers.

Ransomware attacks require knowledge to set up. Filling this 'gap in the market' is a relatively new development called Ransomware-as-a-Service (RaaS) [19, p.2]. One of the first such services appeared in 2016 [13]. Cybercrime groups developing RaaS ransomware usually provide a web portal for victims to pay and their customers to negotiate with the victims. The growth of RaaS provides an incentive for this study to focus on classifying and determining differences in money flow between RaaS and non-RaaS (commodity) attacks.

Currently, it is difficult for law enforcement to track down the criminals behind ransomware attacks. Although Bitcoin is not fully anonymous, connecting Bitcoin addresses to natural persons is challenging. Many criminals are thus not punished, leaving victims with injustice. A better understanding of the ransomware process provides additional information to help punish cyber criminals.

Understanding the general order of operations and actions taken in a ransomware attack is essential for understanding and interpreting the data presented in this research. *Figure 1* provides a general overview of a successful ransom attack. Victims receive a ransom note, usually displayed by the ransomware software. If the victim decides to pay, they purchase Bitcoin. This research focuses on what happens after that, as there is no data on the ransom note in the dataset. After purchasing the Bitcoins, the victim transfers the Bitcoin to the attacker (green), where the ransom is split between multiple stakeholders or laundered first (pink). In a RaaS attack, the customer is one of the stakeholders and receives part of the split.

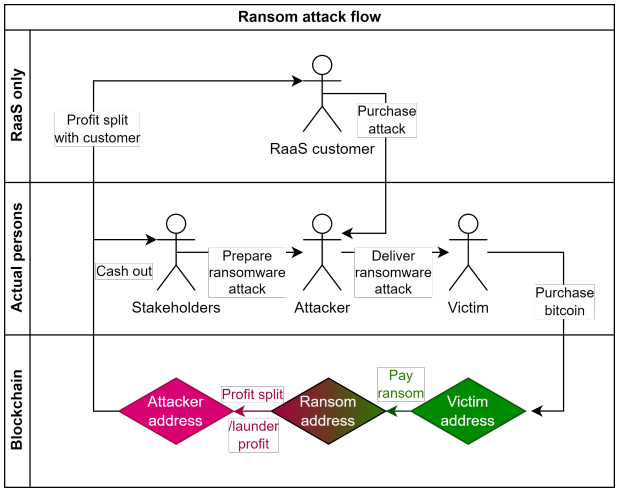

Fig. 1. This is how an average ransom attack is carried out and how the money generally flows in such an attack.

Minimal research has been done to classify different stakeholders in ransomware attacks and the profit split between stakeholders or methods to determine this statistic. Understanding money flow in ransomware is essential to prevent attacks and increase the likelihood of attackers being caught. Additionally, the money flow to different stakeholders provides insight into the inner workings of a ransomware family. Understanding what happens in cybercrime groups and on the blockchain in each step is crucial to improve detection and traceability. Answering the research question: "What conclusions about how a ransom attack generally works can be drawn from blockchain transactions that are known ransomware payments?" is likely to improve understanding of ransomware attacks, which might lead to catching more cybercriminals.

The problem is split into smaller questions to help answer the main question:

- How can a victim account be classified, and building on that, how long before payment to the attackers does a victim generally purchase the Bitcoin for payment?
  Understanding victim behavior is essential for mapping the target demographic for ransomware, which could improve education and prevent future victims.
- How is the profit distributed in a ransomware attack after a ransom payment has been made?
  Uncovering the inner workings of a cybercrime group is helped by understanding where in the group money flows. Additionally, it might indicate different stakeholders or methods of laundering the ransom.

TScIT 39, July 7, 2023, Enschede, The Netherlands

- Grouping the data collected in earlier sub-questions into ransomware families, what conclusions can be drawn about the differences in ransom payment behavior and how it is distributed after payment?
  Some ransom families might target specific demographics or targets. Understanding relevant threats can increase the effectiveness of detection and prevention measures and link multiple families to the same cybercrime group.

To answer these questions, extensive use of the ransomwhe.re [8] dataset will be used as a basis. It contains data from ransom notes uploaded by victims of ransomware attacks, with information like the ransom address and payment transactions. Additional data, like transactions associated with the victim and attacker addresses, will be added to this dataset.

This research paper will describe the technologies used for collecting, storing, and processing the data, then outline the data flow. Complications encountered during the tool's development process will be discussed, after which results will be presented and conclusions will be derived.

2 RELATED WORK

Ransomware and malware attacks are generally well-understood topics. Research outlines different types and tactics [6]. The history [13], attack and spread factors [25], detection and prevention [16][20] of ransomware are all well-understood topics as well. Additionally, extensive studies that focus more on defensive strategies against different families and types of ransomware exist [17]. Specific research into the cybercrime group behind the Conti ransomware has led to a better understanding of communication and cash flow between group members [14].

Ransomware-as-a-Service (RaaS) poses a unique challenge for victims defending against ransomware and an opportunity for attackers who want to set up a ransomware attack even when they do not have the technical knowledge or skills to do so. Mainly on the dark web, criminals offer tools to easily set up such an attack [18]. Understanding this market also leads to a better understanding of money flow in RaaS.

Detecting suspicious Bitcoin transactions is essential for the early detection of new ransomware attacks, and research has been done into classifying transactions as nominal or potentially malicious. Although a large dataset was used and the model reportedly has good accuracy (>98%), the researchers did not test the models in actually detecting ransomware-related transactions on the blockchain in real-time [5]. More profound research into ransomware payments and connecting related transactions and addresses has been done with a different approach and a smaller dataset [22]. This study will build upon the methods and data used in these studies.

Understanding the profit split between different stakeholders in a ransomware attack is made more difficult by the frequent use of mixers [10, p.627]. A mixer is a service that takes Bitcoin from many customers and 'mixes' these Bitcoins through multiple transactions before taking a fee and returning the coins to the customers [21, p.120]. Tracking Bitcoins through mixers is complex, making them a popular tool for laundering money. Attempts have been made to classify transactions obfuscated by different mixers [27]. However, concerns are raised that such services will change the mixing process, invalidating the current methods of detecting them.

An interesting study into methods to track the entire ransomware process from attack to payment and profit split has been done [10]. However, this research lacks focus on comparing RaaS with commodity attacks. It is challenging to map the full extent of damage done by ransomware. However, studies have been done into differences between ransomware families, and total ransom paid [19]. The ransomwhe.re tracker was created for this research, making it an important way of validating the results found in this research.

3 SOFTWARE STACK OVERVIEW

For requesting information from the Bitcoin blockchain, Bitcoin Explorer is used. Although it provides a publicly hosted instance, the number of requests required would put a significant load on this server. Therefore, the self-hosted approach was chosen. Self-hosting also provides the benefit of eliminating most network latency.
Bitcoin Explorer: Bitcoin Explorer uses Bitcoin Core and an electrum server to allow for lookup of addresses and transactions through an easy-to-use API[1] [15].
Bitcoin Core: Bitcoin Core is an open-source utility for interacting with the Bitcoin blockchain and provides the other tools with transactions and block information [4].
Electrs: electrs is a utility that uses Bitcoin Core to provide address information, such as inbound and outbound transactions and the balance [28].

4 METHODOLOGY AND DATASET OVERVIEW

The source code for building the dataset used in this research has been published in a public GitHub repository [23] and is open for contributions. An important focus of this research is to make the script export a reliable dataset, so it would also be usable for future research. The script interacts with the following data sources:
Ransomwhe.re: the ransomwhe.re dataset, used in prior research, is used as a basis for the generated dataset for this study [8].
Bitcoin Explorer: the self-hosted software stack provides an accessible interface to the Bitcoin blockchain.
Blockchain.info: blockchain.info is a tool similar to Bitcoin Explorer and is used as a backup. Bitcoin Explorer does not parse every transaction properly, and the missing data is requested from bitcoin.info [1].
PostgreSQL database: the data is stored in a high-speed, local database.

4.1 Data flow diagram

A simplified data flow diagram, providing an overview of the essential parts of the data flow, abstracting away insignificant details, can be found in *Figure 2*.

[1] An Application Programming Interface (API) provides a simple and well-documented way to interact with a software package and facilitate interconnection between software packages.

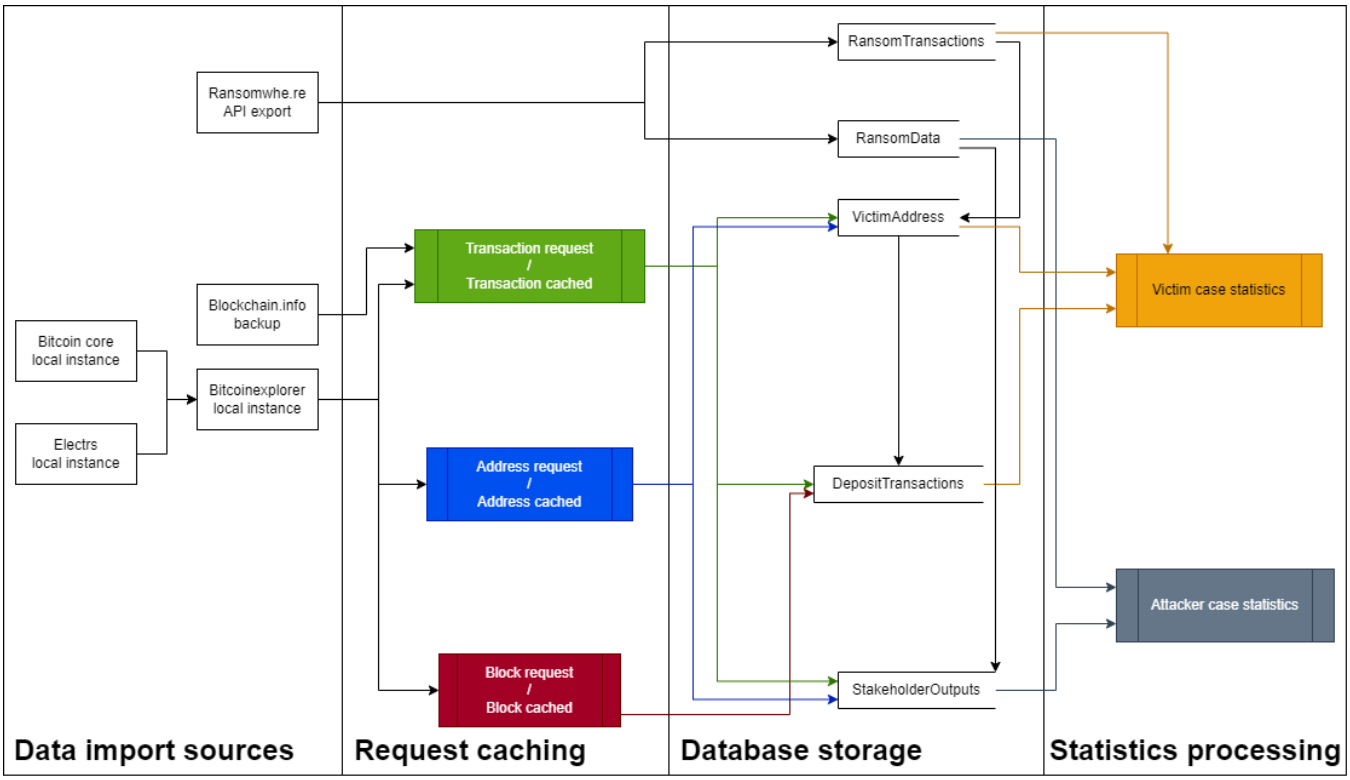

Fig. 2. A simplified data flow diagram containing all database tables used for statistics. Different colors denote the source of data.

4.2 Data import

The initial data collection step can take a day, up to a few days, limited mainly by Bitcoin Explorer, which is limited by the speed of electrs and Bitcoin Core. In addition, if a transaction cannot be deserialized properly by Bitcoin Explorer, a request is made to blockchain.info, which is rate-limited to 1 request per 10 seconds. To improve speed in subsequent imports, a cache is maintained for every requested entity. The data collection step takes only a few hours to complete using this cache and is comprised of the following steps:

- The script starts with downloading the latest data from the ransomwhe.re API.
- After this import succeeds, all addresses associated with the transactions supplied by ransomwhe.re are requested and stored in the database. Because of the scope of this research and the feasibility of building the dataset in a reasonable time, electrs is limited to only returning the transactions of addresses with 100 transactions or less.
- For all addresses within the limits, all transactions into and out of the addresses are stored in the database.
- Attacker addresses are provided in the ransomwhe.re dataset. Only transactions out of the address are stored, as this provides information about the ransom split to stakeholders or the start of money laundering. This data is also added to the database.

4.3 Data processing

During data importing, part of the data is already processed and imported into the correct tables. Additionally, most tables have fields that are processed later, as they require the full dataset. The following steps are taken to process the data:

- Profit split: the ransom division is calculated, as well as the amount left in the ransom address (which is where victims pay the ransom to).
- Deposit time delta: the time between the Bitcoin deposit into the victim account and the ransom payment.
- Is the victim a prior Bitcoin holder? This step uses the number of previous transactions and the time delta between the ransom payment and the first transaction in the account to determine whether the victim address has been used to exchange Bitcoins before the ransomware payment.

4.4 Unreliable data

There are various ways in which the data outputted by the script can become unreliable. It is important to understand which data might be affected by this and how to filter out these cases, to prevent drawing invalid conclusions.

- For performance reasons, only addresses with 100 transactions or less are returned by electrs, which causes some addresses to have no transactions according to the database. This is mitigated by adding a 'Success' field.

Table 1. Actor data

Category	Data count	Percentage
Failed imports	0	0.00%
Ransom still in address	19	0.18%
Ransom not paid	2,938	28.13%
Single case	6,639	63.57%
Extended case	848	8.12%
RaaS cases	7,389	70.75%
Non-RaaS cases	3,055	29.25%
Total cases	10,444	100%

*Note: a ransom is considered not paid if a ransomwhe.re entry does not include any transactions.

Table 2. Victim data

Category	Data count	Percentage
Failed imports	9,314	44.56%
Simple case	5,569	26.64%
Extended case	6,021	28.80%
RaaS cases	12,196	58.34%
Non-RaaS cases	8,708	41.66%
Total cases	20,904	100%

- Various checks are implemented to ensure data reliability, such as calculating that the stakeholder output percentages add up to 100% and verifying results with actual blockchain data.

All data in the database is built upon the ransomwhe.re dataset, stored in the 'RansomData' and 'RansomTransactions' tables. The former contains the 'Failed' and 'FailedVictims' columns, indicating if this data point is reliable for calculating actor and victim statistics, respectively.

4.5 Data composition

The failed imports referred to in *Table 1* and *Table 2* are cases that include addresses with more transactions than the limit. Because the limit was only applied to victim addresses, there are no failed imports for ransom addresses. Although more than 9,000 cases were marked as failed, because of how analysis is done, a ransom address is marked as failed if only one of the victim payments contains an address with more transactions than the set limit.

For this research, only the single case is considered. Failed imports and ransom still in the address cases[2] are edge cases and thus not considered either. The single case consists of ransom addresses with one input and output. *Table 1* provides the case count for these cases.

Similar to the actor data, only the 'Simple case' is considered for victim data. In this case, only transactions with one source address (one address pays the entire ransom) are considered.

[2]'Ransom still in address' cases are only cases where nothing, or part of the ransom payment is withdrawn from the address.

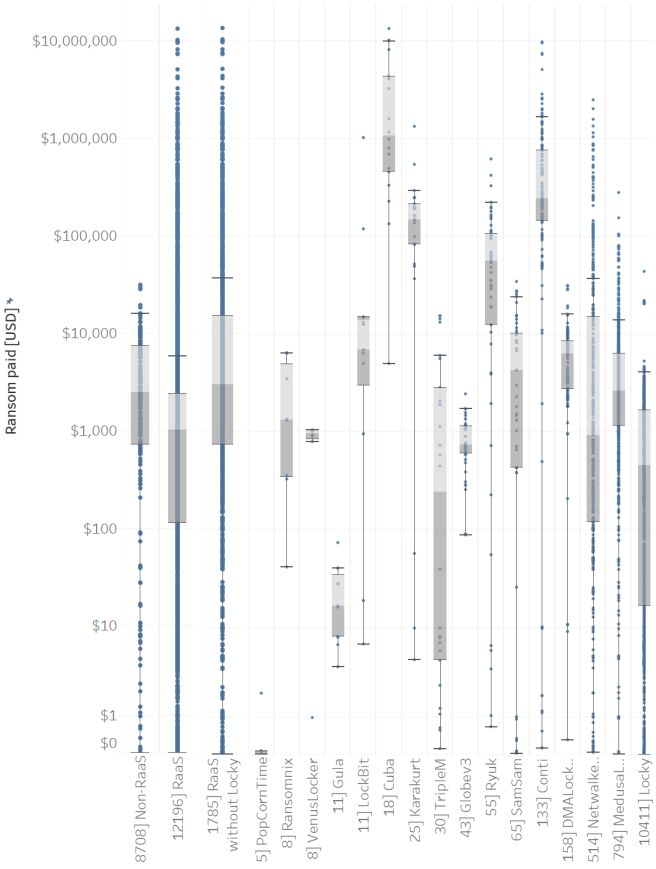

Fig. 3. A comparison between the amount of ransom paid in RaaS and non-RaaS attacks and several ransomware families and the number of cases. A logarithmic scale is used for the ransom amount.

In *Figure 3*[3], the distribution of ransom paid per family is shown. Although some ransomware families show high or low outliers, most focus on a specific ransom range. Interestingly, commodity (Non-RaaS) ransomware families generally seem to focus on a much narrower range, while RaaS families show a broader range of ransom demands, with lower average requested. Although the dataset cannot substantiate it, because RaaS has a range of different customers with ranging requirements, which likely results in a broader range of ransom amounts.

5 RESULTS

In this section, the data collected will be presented. Any ransomware flow starts with ransom payment, so first, the victim results will be analyzed, after which the ransomware actor data will be assessed.

5.1 Findings in victim behavior

Any ransom payment starts with the deposit of Bitcoin into the address from which the ransom will be paid. Only the 'simple cases', referred to as the 'dataset', will be discussed, consisting of victims

[3]For this figure, the actual ransom amount paid is used. No filter is applied to this data, except for showing only families used later in the results.

Table 3. Victim payment count across ransomware families in the simple case

Family	Simple cases	All cases
Locky	4,587	10,411
Netwalker	421	514
Conti	57	133
SamSam	31	65
TripleM	18	30
Ryuk	19	55
RaaS cases	5,186	12,196
RaaS cases without Locky	599	8,708
Non-RaaS cases	383	8,708

Table 4. Ransom to attacker address transactions count across ransomware families in the single case

Family	Single case	All cases
Locky	12,376	15,435
Netwalker	51	301
Conti	155	247
SamSam	10	104
TripleM	0	44
Ryuk	19	108
Karakurt	22	39
Cuba	34	37
RaaS cases	12,719	17,307
RaaS cases without Locky	343	1,872
Non-RaaS cases	65	9,526

whose addresses had all transactions successfully imported. Additionally, the ransom payment transaction may only have one source address. *Table 3* provides the amount of payments per family. Locky introduces a significant bias in the RaaS cases. Hence RaaS case count excluding Locky is also provided.

The dataset contains many ransomware families consisting of only a few reported payments. Therefore, except for the five largest, all families will be grouped in the 'Other' category. In *Figure 4*, a difference is apparent between the most prominent ransomware families. An outlier in the right column is the Locky family, which rarely transfers the ransom soon after the victim's payment. The ransomware actors section will elaborate on this finding.

Interesting outliers are Netwalker, which sees few instant payments, and Conti, which sees relatively many. Although some families with lower case counts show differing results, the low number of cases might cause this data to not be representative of the family.

The amount of ransom paid might affect the handling of the transaction. *Figure 5* compares the ransom paid with the time between Bitcoin purchase and ransom payment. Distinct ransom amount 'levels' can be observed, primarily with Locky, showing in the figure as 'bands' of payments that share the same ransom amount. There does not seem to be a clear correlation between the amount of ransom and the time from Bitcoin purchase to payment, although a difference in ransom paid is apparent between families. The amount of ransom requested by different ransomware families can be found in *Figure 3*.

5.2 Findings in ransomware actors

Definition: In this section, transactions refer to individual transactions from the ransom to the attacker's address.

Moving along the ransom process, in this section, the process of transferring the ransom to an attacker's address will be discussed. The single case, called the 'dataset', contains transactions from ransom addresses that have received only one ransom payment and transfer the ransom in one transaction to the attacker's address. *Table 4* shows a significant reduction of cases among the top 5 most common families, except in RaaS families.

Although Locky is the most prominent family of the dataset, it only contains 298 unique transactions to an attacker's address. This behavior is, with a few exceptions, not observed in other families and is also apparent in *Figure 6*. This figure shows the number of victim payments clustered in transactions to an attacker's address, compared to the time from ransom payment to the laundering transaction. A family called 'QLocker', a non-RaaS family applies a similar technique[4] to Locky. By contrast, most ransomware families start laundering only a day to a few days after the ransom has been paid, but generally not much later.

Comparing only RaaS and non-RaaS clusters yielded no interesting results besides the discussed outliers. The range of ransom amount is chosen because the average time to laundering is, with only a few exceptions, not longer than 31 days, and the total ransom amount in a cluster is not filtered.

The dataset contains four laundering transactions that seem to combine attacker addresses from two ransomware families. Three such transactions have sources classified as APT and TripleM family attacker addresses, while the others are classified as Cryptowall and JigSaw family addresses. There does not appear to be a clear link between these families, so this is a suspected misclassification in the ransomwhe.re dataset. These transactions were not in the single case.

Interestingly, although the Conti (and its predecessor Ryuk) generally do not cluster ransom payments, laundering starts within the first day after the ransom payment. Its clusters, consisting of only one victim payment, are worth more than most Locky clusters. The Cuba ransomware family shows similar behavior and contains even more valuable clusters. From *Figure 3*, it becomes apparent that these families generally target a higher ransom amount from their victims, whereas Locky primarily targets lower ransom amounts, explaining the observed difference in the total cluster amount.

To determine whether there is a pattern in the split between attacker addresses, enough cases are required to discern a pattern. Because of this, only Locky, Conti, Netwalker, Cuba, Karakurt, and Ryuk are considered. *Figure 7* shows that Conti's ransom split forms a distinct pattern that changes over time. The split percentages used by Ryuk before the Conti family emerged are similar to what Conti used in the first attacks. Additionally, all of Ryuks cases precede every Conti case. This is not surprising, as Conti and Ryuk originate

[4] QLocker clusters are the two larger circles in the 'Other' category, to the right of the bulk of the Locky data points

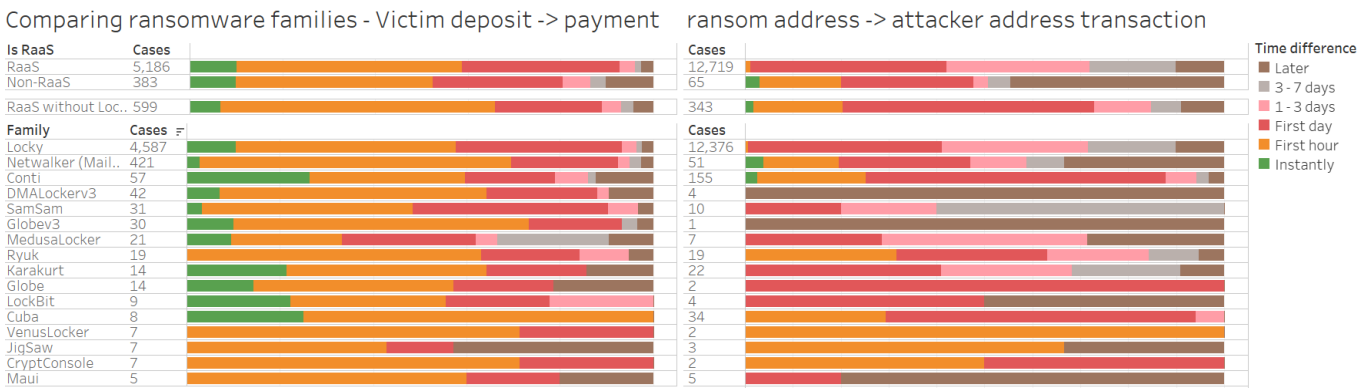

Fig. 4. Timing of different stages of the ransom flow. Only families with data for both victim payment and stakeholder split, with a minimum of 5 payments are shown. *Figure 8* provides a version of this figure with all families included.

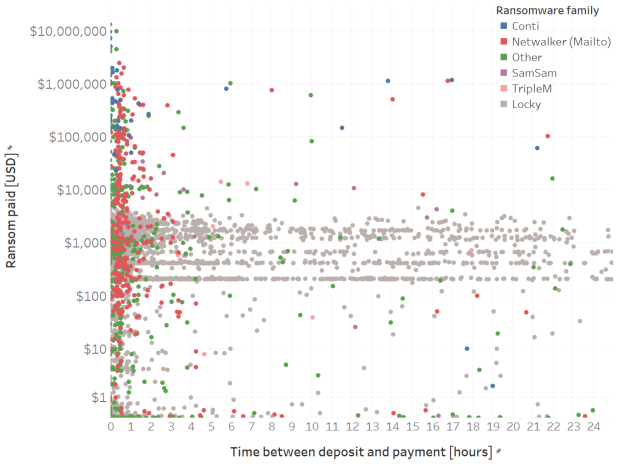

Fig. 5. A comparison of the ransom amount with the time between Bitcoin deposit and ransom payment.

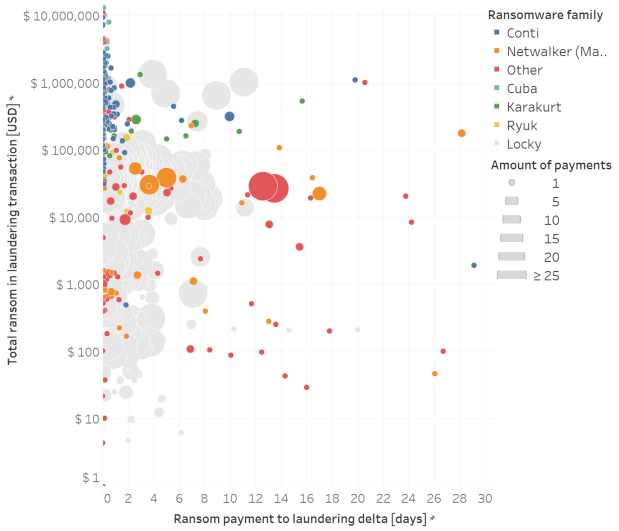

Fig. 6. Ransom payments grouped by the laundering or stakeholder split transaction, comparing the sum of ransom payments laundered in a transaction against the average time between ransom payment and laundering for that cluster. The size of the circles denotes the number of ransom payments present in a laundering transaction.

from the same cybercrime group [14, p.6]. Trickbot, which is also interconnected according to this research, does not occur in the ransomwhe.re dataset as of yet. Time of day analysis did not yield any interesting results.

6 DISCUSSION

In this section, the data presented in the previous section will be analyzed, and observations will be discussed.

An interesting difference in instant ransom payments between Netwalker, Locky, and Conti can be observed in *Figure 4*. Instant payments could suggest the use of tools to transfer purchased Bitcoins to the ransom address in one action. In screenshots of the Netwalker online interface, no such system nor a reference to a similar, external system seems to be made however [3]. Locky's interface provides the user with more detailed instructions on where to purchase Bitcoins and even mentions how simple it is [2]. Conti does provide a chat service, however, similar to Netwalker; no mention is made of a simple method to automate Bitcoin purchasing and transferring [11].

Comparing the families, the hypothesis would suggest Conti provides such a system, whereas Netwalker would not. However, neither seems to offer or mention a simple method to purchase and transfer Bitcoins in one transaction. The years the family was active might also influence the number of instant payments. *Figure 7* shows that Locky precedes Netwalker and Conti by a few years. However, there appears to be no clear correlation. Additionally, uncovering conclusions from this data does not seem to be a focal point of related research, so no additional insights were drawn from related studies.

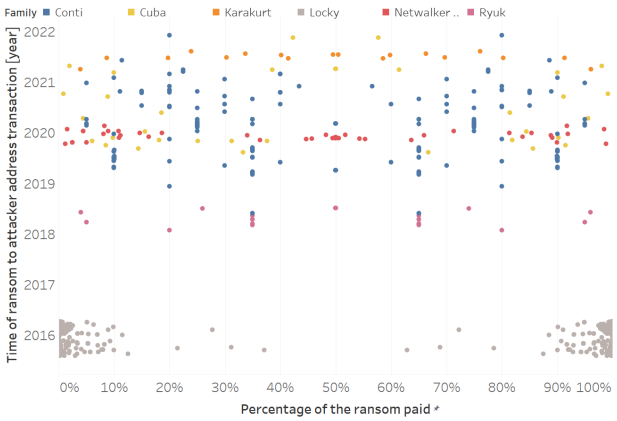

Fig. 7. A comparison between the time of profit split transactions and the percentage of the split.

In *Table 4*, a notable difference in case reduction between RaaS and non-RaaS cases was observed. This could mean that commodity ransomware reuses the ransom payment addresses. Oosthoek et al. concluded that RaaS families typically generate a unique address per victim, whereas commodity ransomware indeed typically uses a single address [19, p.5]. In this research, it is suggested that this might be to improve operational security, and their data seems to indicate a focus on operational security in RaaS attacks. However, our dataset cannot substantiate that claim.

Moving on to the second part of a ransomware attack: laundering the victim's payment. Locky differentiated its method by grouping many ransom addresses into a single laundering transaction. Research focussing on the financial impacts of ransomware payments found similar behavior for Locky [22, p.6]. Their dataset included additional transactions after the transaction from the ransom to the actor's address. They found that the transactions were sent to exchanges and mixers in their respective clusters. However, no apparent reason was presented for the different methods used for Locky attacks. Tracing one of these transaction chains uncovers a long chain of so-called 'scrape off' transactions. This refers to a popular technique to make tracing payments to an actual person more challenging. This technique results in a long chain of transactions where in every transaction, a small amount of the input is sent to a different address than the more significant part, so a part of the input is scraped off. Later, through many transactions, the money is regrouped in one or multiple addresses to cash out or buy goods or services with [19, p.7].

The effects of the scrape-off technique are also seen in *Figure 7*. As the input amount represents 100% of the amount laundered, the transactions with only two outputs will show as being symmetrical. In the dataset, only Locky (with a few exceptions in families with fewer cases) seems to be using the scrape-off technique. The symmetry it causes is also visible for Conti and Ryuk laundering, with a few exceptions. Because the split percentages for these families are larger and more consistent, it doesn't seem to indicate the use of the scrape-off technique.

Figure 1 shows ransom payments usually flow into one or multiple attacker addresses. Multiple stakeholders contribute to a ransomware attack, such as programmers, general managers, and spammers [14, 7-9]. Although a clear split in the Ryuk and Conti ransomware is evident, the data does not prove this is a split among multiple stakeholders. However, prior research has found similar split percentages, such as a study into the inner workings of the cybercrime group behind Conti [14, p.4] and elliptic, using a smaller dataset [12]. The focus of the study into Conti enabled the researchers to conclude the ransom is split between the Conti collective and affiliates. From the blockchain data alone however, they could similarly not conclude the possible roles of the affiliates.

7 PROBLEMS ENCOUNTERED

During the research process, some problems were encountered. In this section, these issues and their solutions will be described.

Setting up the self-hosted software stack took some time as I was unfamiliar with the tools used for this research. However, after a few days of troubleshooting the electrs server and Bitcoin Explorer and re-indexing the blockchain with Bitcoin Core with different settings, the software stack was up and running.

Although I have written scripts like the script used for building the dataset, the amount of data processed by this script and the variety of data did lead to a few issues further in development. There were many edge cases in the data returned by Bitcoin Explorer. One example is addresses with more than 1000 transactions, which would take a long time to produce a result because of electrs. These addresses were filtered out of the dataset before analyzing the data. Additionally, I did not anticipate that improving the reliability of the dataset would take a couple of weeks, as new edge cases kept appearing.

Although the amount of data to import posed some problems, primarily for electrs, there might be gains from expanding the dataset. After gathering the initial dataset with a maximum of 50 transactions per address, it was expanded to include up to 100 transactions per address. Because of this, the victim dataset decreased failed imports with 1,704 cases to 9,314. This expansion was made possible by a new version of Bitcoin Explorer which can also parse input addresses to transactions, drastically reducing the requests needed to fill the dataset. Because of this increase, the data became more representative of the entire ransomwhe.re dataset, and the simple victim cases increased from 4,996 to 5,569, including many non-RaaS cases.

8 LIMITATIONS AND FUTURE WORK

Many limitations originate from the amount of time that could be dedicated to this research. As described in the 'Data composition' section, although the dataset is extensive, only simple cases were considered for this research. This means a significant part of the data was not analyzed. With more time, dedicated processing cases and procedures could also be written to consider this data. Because the dataset still includes other cases, it is possible for future research to process this data further and expand upon the conclusions drawn. Only using part of the data also introduced biases in the study, as seen in previous sections. Few non-RaaS simple cases remained, and RaaS cases consisted mainly of Locky. The latter was mitigated by splitting RaaS cases in RaaS with and without Locky. In comparable research, Locky is a similarly large percentage of the total dataset [26, p.7:19][19, p.4][22, p.7].

Importing the data through Bitcoin Explorer is the biggest bottleneck when building the dataset, as it relies on and is limited by the tools it receives data from. Directly interacting with the electrs and Bitcoin Core RPC API would remove one source of latency, but it is difficult to predict the impact of this change. While this research focuses entirely on ransom payments collected through the Bitcoin blockchain, other cryptocurrencies are also used for ransom payments [9, p.10].

To fully map a ransom attack, which starts with infecting a victim device and delivering the ransom note, information about the ransom note might yield additional insights. The time delta between the delivery of the ransom note and ransom payment might indicate the difficulty a victim had paying.

A different limitation is that using only this dataset, it is impossible to differentiate between a personal address with prior activity and a shared address. This can partially be solved by adding a lookup table to the dataset with shared addresses that belong to certain Bitcoin exchanges. However, finding and combining this data would take additional time and not all exchanges publish this data. Additionally, because ransomware is a global phenomenon, many exchanges would need to be added to the dataset to draw any useful conclusions. Determining whether a victim was already active in trading Bitcoin before the ransom payment is not a primary objective of this research, so collecting this data was not pursued.

A technique mentioned in the stakeholder results section is 'scrape off'. This technique makes it difficult to follow the ransom from payment to exchange. It should be possible to detect this behavior and follow the 'money trail'. Additionally, only cases where a unique address per victim is used and a single transaction is used to split profits between stakeholders are considered. As a result of this, only a few Non-RaaS attacks were considered. In future research, better data processing methods can be developed.

Because in *Figure 6* only the simple case is considered, this graphic does not give a complete overview of all ransomware families in the dataset. The simple case contains significantly fewer non-RaaS cases (*Table 3*), not many interesting insights could be concluded comparing RaaS with commodity cases. Processing more of cases in the extended case is likely to increase the usefulness of this comparison.

9 CONCLUSION

This research aims to improve understanding of the money flow in ransomware attacks, focusing on comparing Ransomware-as-a-Service with commodity attacks. To achieve this goal, a tool was developed to extend the ransomwhe.re dataset [8] with victim and attacker address data.

After applying the filters, Locky ransomware is the most common family in the datasets for victim behavior and ransom split classification. Comparing different ransomware families, an interesting difference in the time between Bitcoin deposit and ransom payment was found, with Conti as an outlier having significantly more instant payments than different families. No correlation could be determined between the amount of ransom and the time between deposit and payment.

A different yet interesting finding is that the Locky family of ransomware is the largest of only a few families that group different ransom addresses in a single transaction to scrape off the profits. Although a sub-goal of this research is to identify stakeholders in a ransomware attack by the split of the ransom, because most ransomware families use a laundering technique before dividing the ransom between stakeholders, it is difficult to identify different stakeholders based on the dataset available. Although Locky uses unique ransom addresses, it groups multiple addresses in a single transaction to an atttacker's address, an exception in the dataset.

An interesting pattern in ransom split was observed for the Conti and Ryuk ransomware families when comparing how the ransom is split over time. The pattern was observed to not change between Conti and Ryuk operations, making it possible to identify Conti as the successor of Ryuk, a conclusion also drawn in the literature discussed.

REFERENCES

[1] [n. d.]. Blockchain Data API.
[2] 2016. Locky Ransomware Information, Help Guide, and FAQ.
[3] 2020. NetWalker Ransomware in 1 Hour.
[4] 2022. Bitcoin Core.
[5] Qasem Abu Al-Haija and Abdulaziz A. Alsulami. 2021. High Performance Classification Model to Identify Ransomware Payments for Heterogeneous Bitcoin Networks. *Electronics* 10, 17 (Aug. 2021), 2113. https://doi.org/10.3390/electronics10172113
[6] Amir Atapour-Abarghouei, Stephen Bonner, and Andrew Stephen McGough. 2019. Volenti Non Fit Injuria: Ransomware and Its Victims. In *2019 IEEE International Conference on Big Data (Big Data)*. 4701–4707. https://doi.org/10.1109/BigData47090.2019.9006298
[7] Gabriel Bassett, C. David Hylender, Philippe Langois, Alex Pinto, and Suzanne Widup. 2022. Data Breach Investigations Report.
[8] Jack Cable. 2022. Ransomwhere: A Crowdsourced Ransomware Payment Dataset (1.0.0). https://doi.org/10.5281/zenodo.6512123
[9] Lin William Cong, Campbell R. Harvey, Daniel Rabetti, and Zong-Yu Wu. 2022. An Anatomy of Crypto-Enabled Cybercrimes. https://doi.org/10.2139/ssrn.4188661
[10] Danny Yuxing Huang, Maxwell Matthaios Aliapoulios, Vector Guo Li, Luca Invernizzi, Elie Bursztein, Kylie McRoberts, Jonathan Levin, Kirill Levchenko, Alex C. Snoeren, and Damon McCoy. 2018. Tracking Ransomware End-to-end. In *2018 IEEE Symposium on Security and Privacy (SP)*. 618–631. https://doi.org/10.1109/SP.2018.00047
[11] editor. 2021. Conti Ransomware.
[12] Elliptic. 2021. Conti Ransomware Nets at Least $25.5 Million in Four Months. https://www.elliptic.co/blog/conti-ransomware-nets-at-least-25.5-million-in-four-months.
[13] Danyal Farhat and Malik Shahzad Awan. 2021. A Brief Survey on Ransomware with the Perspective of Internet Security Threat Reports. In *2021 9th International Symposium on Digital Forensics and Security (ISDFS)*. 1–6. https://doi.org/10.1109/ISDFS52919.2021.9486348
[14] Ian W. Gray, Jack Cable, Benjamin Brown, Vlad Cuiujuclu, and Damon McCoy. 2023. Money Over Morals: A Business Analysis of Conti Ransomware. arXiv:2304.11681 [cs]
[15] Dan Janosik. 2021. Bitcoinexplorer.
[16] Lorenzo Fernández Maimó, Alberto Huertas Celdrán, Angel L. Perales Gómez, Félix J. García Clemente, James Weimer, and Insup Lee. 2019. Intelligent and Dynamic Ransomware Spread Detection and Mitigation in Integrated Clinical Environments. *Sensors* 19, 5 (March 2019), 31. https://doi.org/10.3390/s19051114
[17] Timothy Mcintosh, A. S. M. Kayes, Yi-Ping Phoebe Chen, Alex Ng, and Paul Watters. 2022. Ransomware Mitigation in the Modern Era: A Comprehensive Review, Research Challenges, and Future Directions. *Comput. Surveys* 54, 9 (Dec. 2022), 1–36. https://doi.org/10.1145/3479393
[18] Per Håkon Meland, Yara Fareed Fahmy Bayoumy, and Guttorm Sindre. 2020. The Ransomware-as-a-Service Economy within the Darknet. *Computers & Security* 92 (May 2020), 101762. https://doi.org/10.1016/j.cose.2020.101762
[19] Kris Oosthoek, Jack Cable, and Georgios Smaragdakis. 2022. A Tale of Two Markets: Investigating the Ransomware Payments Economy. https://arxiv.org/abs/2205.05028v1.
[20] Harun Oz, Ahmet Aris, Albert Levi, and A. Selcuk Uluagac. 2022. A Survey on Ransomware: Evolution, Taxonomy, and Defense Solutions. *Comput. Surveys* 54, 11s (Sept. 2022), 238:1–238:37. https://doi.org/10.1145/3514229
[21] Jaswant Pakki, Yan Shoshitaishvili, Ruoyu Wang, Tiffany Bao, and Adam Doupé. 2021. Everything You Ever Wanted to Know About Bitcoin Mixers (But Were Afraid to Ask). In *Financial Cryptography and Data Security (Lecture Notes in Computer Science)*, Nikita Borisov and Claudia Diaz (Eds.). Springer, Berlin, Heidelberg, 117–146. https://doi.org/10.1007/978-3-662-64322-8_6
[22] Masarah Paquet-Clouston, Bernhard Haslhofer, and Benoît Dupont. 2019. Ransomware Payments in the Bitcoin Ecosystem. *Journal of Cybersecurity* 5, 1 (Jan. 2019), 10. https://doi.org/10.1093/cybsec/tyz003
[23] Justin Ruiter. 2023. Bitcoin Tool.
[24] SonicWall. 2023. 2023 SonicWall Cyber Threat Report. *2023 SonicWall Cyber Threat Report | Charting Cybercrime's Shifting Frontlines* 11 (2023), 1–69.
[25] Kutub Thakur, Thaier Hayajneh, and Jason Tseng. 2019. Cyber Security in Social Media: Challenges and the Way Forward. *IT Professional* 21, 2 (March 2019), 41–49. https://doi.org/10.1109/MITP.2018.2881373
[26] Kai Wang, Jun Pang, Dingjie Chen, Yu Zhao, Dapeng Huang, Chen Chen, and Weili Han. 2021. A Large-scale Empirical Analysis of Ransomware Activities in Bitcoin. *ACM Transactions on the Web* 16, 2 (Dec. 2021), 7:1–7:29. https://doi.org/10.1145/3494557
[27] Jiajing Wu, Jieli Liu, Weili Chen, Huawei Huang, Zibin Zheng, and Yan Zhang. 2022. Detecting Mixing Services via Mining Bitcoin Transaction Network With Hybrid Motifs. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 52, 4 (April 2022), 2237–2249. https://doi.org/10.1109/TSMC.2021.3049278
[28] Roman Zeyde. 2023. Electrs.

A EXTENDED DATA FOR FIGURE 4

Fig. 8. Timing of different stages of the ransom flow.