

Increasing the Efficiency of Cryptoasset Investigations by Connecting the Cases

Bernhard Haslhofer
Complexity Science Hub
Vienna, Austria
email: haslhofer@csh.ac.at

Christiane Hanslbauer
Magistrate Court of Bamberg
Bamberg, Germany
email: christiane.hanslbauer@ag-ba.bayern.de

Michael Fröwis
Iknaio Cryptoasset Analytics GmbH
Vienna, Austria
email: michael@ikna.io

Thomas Goger
Bavarian Central Office for the Prosecution of Cybercrime (ZCB)
Bamberg, Germany
email: thomas.goger@gensta-ba.bayern.de

Law enforcement agencies are confronted with a rapidly growing number of cryptoasset-related cases, often redundantly investigating the same cases without mutual knowledge or shared insights. In this paper, we explore the hypothesis that recognizing and acting upon connections between these cases can significantly streamline investigative processes. Through an analysis of a dataset comprising 34 cyberfraud and 1793 sextortion spam cases, we discovered that 41% of the cyberfraud and 96.9% of the sextortion spam incidents can be interconnected. We introduce a straightforward yet effective tool, which is integrated into a broader cryptoasset forensics workflow and allows investigators to highlight and share case connections. Our research unequivocally demonstrates that recognizing case connections can lead to remarkable efficiencies, especially when extended across crime areas, international borders, and jurisdictions.

Index Terms—cryptoassets, cybercrime

I. INTRODUCTION

The increasing prevalence of cryptoasset-related crimes has resulted in a surge of cases that criminal investigators are struggling to manage. At the Bavarian Central Office for the Prosecution of Cybercrime (ZCB), for example, five full-time prosecutors are currently investigating more than 1600 fraudulent cybertrading platforms, with thousands of individual victim cases awaiting attention. These investigations have resulted in victim losses amounting to 250 million Euros in 2022¹ and are estimated to be in the billions in Germany alone.

In addition to cybertrading fraud, investigators are faced with the challenge of addressing cryptoasset abuse, which has become a pervasive issue cutting across various forms of criminal activity. For instance, in ransomware attacks, victims are forced to pay ransom in the hope of receiving the keys necessary to unlock their data, which has been encrypted by the perpetrator [1]. Similarly, victims of sextortion spam [2] often pay using Bitcoins to prevent the disclosure of potentially

damaging adult material, which has supposedly been collected by the perpetrator.

The technique of “following the money” is a widely recognized method in forensic investigations, and it has now become a standard approach in cases involving cryptoassets as well. In response to this, a growing industry with a market value of billions of USD has emerged to offer blockchain forensics tools to investigators worldwide. Despite this progress, the current state of affairs remains less than satisfactory: these tools are often prohibitively expensive, and their availability is limited, given the high number of cases that investigators are dealing with. Consequently, numerous cases continue to pile up on the desks of a few investigators, leading to long waiting periods for results. Unfortunately, these waiting periods are frequently too extended, and in many cases, the funds have already been laundered through sophisticated networks that are challenging or impossible to trace manually.

In light of the current situation, it is becoming increasingly evident that merely augmenting manpower and expanding tool licenses is insufficient in addressing a rapidly growing problem. Instead, a more sustainable approach necessitates a period of reflection, a step back, to reevaluate and reimagine how investigations could be conducted more efficiently. One promising avenue stems from the observation that investigators frequently work on identical cases, often without the awareness of each other’s efforts or the opportunity to share findings.

In this study, we leverage this observation, hypothesizing that cryptoasset investigations can be organized more efficiently if connections between cases are known. To validate this hypothesis, we undertake an empirical examination using real case data amassed by the ZCB, focusing on cybertrading fraud and sextortion spam. The key contributions of our research are as follows:

- 1) We furnish a dataset consisting of 34 cybertrading fraud and 1793 sextortion spam cases with associated cryptoasset addresses.
- 2) We propose a straightforward method for identifying connections between cases, premised on the reuse of

¹<https://www.justiz.bayern.de/presse-und-medien/pressemitteilungen/archiv/2022/72.php>

cryptoasset addresses and the utilization of common collector wallets.

- 3) We offer empirical evidence highlighting the prevalent use of cryptoasset in cybercrime. Our findings demonstrate that we automatically find links between 41% of the cybertrading fraud and 96.9% of the sextortion spam cases.
- 4) We present a case management solution allowing investigators to collaboratively identify and share connections between cases as part of their forensics investigations.

Our findings corroborate our hypothesis and demonstrate enormous potential for efficiency improvements. Extrapolating from this modest sample of cases originating from a single institution, we assert that substantial efficiency enhancements are attainable if institutions foster collaboration and share fundamental case-related data with each other.

We advocate for a paradigm shift: given the interconnected, global nature of cybercrime, our approach should evolve from focusing on isolated instances (cases) to understanding and addressing the broader networks (case clusters). Furthermore, we illustrate that this transition can be achieved with relative ease using currently available open-source tools.

II. BACKGROUND

We begin our discussion by elaborating the prevalent role of cryptoassets in cybercrime, placing a special emphasis on the relatively underexplored area of cybertrading fraud. This area, though significant in daily operations, has seen minimal coverage in existing scientific literature. Following this, we will briefly summarize known cryptoasset tracing techniques.

A. Cryptoassets in cybercrime

Cryptoassets nowadays play a significant role in all areas of cybercrime. Their misuse and the ensuing damage, is well-documented in academic literature. Foley et al. [3] discovered an illicit annual turnover of 76 billion USD. Anderson et al. [4] estimated the direct costs of cybercrime involving cryptoassets to be around 2 billion USD, accounting for exchange hacks and direct crimes against individuals. According to Zhou et al. [5], between April 2018 and April 2022, users, liquidity providers, speculators, and operators of Decentralized Finance (DeFi) protocols collectively incurred losses totaling at least 3.24 billion USD. Oosthoek et al. [6] reported that the revenue generated by Dark Web shops amounted to a minimum of 113 million USD, with sexual abuse (94 million USD) constituting the leading illicit category. Lastly, a recent industry report by Chainalysis [7] estimated that illegal cryptoasset addresses received approximately 23 billion USD.

However, little is known on law enforcement actions in the field. Recently, Abramova and Böhme [8] estimated the market reaction on cryptocurrency exchanges to news about successful law enforcement actions. Also, the US Securities and Exchange Commission (SEC) lists more than 100 enforcement actions and 16 trading suspensions against cryptoasset businesses².

²<https://www.sec.gov/spotlight/cybersecurity-enforcement-actions>

B. Cybertrading fraud

Cybertrading fraud is a global issue involving misleading offers on trading platforms. Advertisements lure customers into investing significant sums in various financial instruments, including cryptocurrencies. However, the funds deposited are never genuinely invested; the visible trading platform and customer accounts are fabricated, almost always leading to total loss of the invested capital. The platforms operate with fraudulent intent, exploiting digitalization to hide their identities and locations. Investigations at the ZCB reveal that perpetrators use a global money laundering network to obscure the money flow³. Victim-transferred funds are moved across at least two further levels to accounts of pseudo-companies at various banks or financial service providers worldwide.

Over the years, cybertrading fraud has become the most common offense linked to illicit cryptocurrency use in cases filed at the ZCB. Historically, victims of cybertrading fraud would transfer their deposit amounts to domestic or foreign bank accounts. The conversion of these funds from fiat currency to cryptocurrency, aimed at obscuring the payment trail, was typically executed by the perpetrators at a later stage. Consequently, transaction investigations were feasible at least at the initial levels through bank information. However, recent investigations reveal an increasing trend wherein victims' deposits are made directly using cryptocurrencies.

This shift can be attributed to the growing popularity and acceptance of cryptocurrencies among the general public. This trend is further encouraged by commission structures within call centers, as gleaned from broker interrogations. Brokers who convince victims to deposit using cryptocurrencies, due to the inherent difficulty in tracing the recipient or sender of such transactions, receive the highest commissions. Consequently, acquiring bank information is no longer a viable starting point for investigations. Instead, the analysis of cryptocurrency transactions has become crucial for initiating investigative leads.

C. Connections between cases

The intensive investigations of the ZCB in the past have shown that the globally active criminal groups usually operate a large number of platforms with different domains, often from various call centres in different countries. As soon as negative reviews from victims about a domain accumulate on the internet, the perpetrators offer their services via a new domain. Here it is important to recognise the connection between different platforms. This is often difficult because possible connections cannot be recognised at first sight.

Currently, however, almost all criminal police units in Bavaria are entrusted with the police investigations of the trading platform cases pending at the ZCB. After all, several criminal investigation departments are working on their respective individual proceedings in isolation due to their jurisdiction of residence for the reporting injured party. This

³<https://www.justiz.bayern.de/gerichte-und-behoerden/generalstaatsanwaltschaft/bamberg/presse/2022/15.php>

dispersion at the police level, however, makes it extremely difficult to recognise platform connections and money laundering networks, especially at an early stage of the investigation.

Only an overall view of various findings from different proceedings currently allows factors such as matching recipient wallets or a jointly used money laundering network to be identified that speak for a connection between platforms. Therefore, identifying connections between cybertrading platforms as early as possible is of central importance in order to avoid duplicate investigations and to use the available resources efficiently.

D. Following the money

In recent years, the practice of tracing payment flows has gained significant traction as a forensic technique for investigating cryptoasset-related crimes. Numerous companies have emerged, offering tools capable of reconstructing payment flows using publicly available blockchain data and attribution tags that link cryptoasset addresses to real-world individuals or entities. This field of cryptoasset tracking and tracing has evolved into a thriving industry, attracting substantial investments from venture capital firms. Major tool providers, such as Chainalysis, TRM Labs, Elliptic, CipherTrace, and Merkle Science, have collectively received hundreds of millions of dollars in venture capital funding⁴.

While certain tools offer the capability to associate Bitcoin addresses with case information, this feature is typically limited to a select number of license holders within an organization. This limitation goes against the objective of identifying case connections at the earliest stage possible, as it necessitates broader availability of tools for criminal investigators within an institution. Furthermore, it is worth noting that law enforcement agencies are actively exploring alternatives to address this issue. However, these databases are typically separate from forensic tools and require extensive manual reconciliation work, which can decrease motivation and overall adoption among investigators.

III. DATA AND METHODS

This study aims to explore the connections between cases related to cybertrading fraud and sextortion. However, before proceeding with the systematic collection of the necessary data points to test our hypothesis, it is crucial to gain a deeper understanding of the patterns and terminology involved in money collection and laundering within these contexts.

A. Conceptualizing money flows

In the realm of cybertrading fraud, victims are commonly instructed to open accounts at well-established custodial cryptoasset exchanges. They transfer fiat currencies such as EUR or USD into these accounts and convert them into cryptoassets, typically Bitcoin. As illustrated in Figure 1, they

⁴According to data from Crunchbase, Chainalysis leads the pack with \$536.6 million, followed by TRM Labs with \$149.9 million, Elliptic with \$100 million, CipherTrace with \$45.1 million, and Merkle Science with \$25.6 million.

subsequently transfer the cryptoassets from their accounts to one or more addresses controlled by the perpetrator. These on-chain transactions reveal a transfer from a “victim address” to a “perpetrator” address. The former is controlled by the victim, while the latter is under the control of the perpetrator. It is worth noting that the perpetrator often directs multiple victims to send payments to the same address, potentially transforming it into a “collector address”.

Once the funds are collected, the perpetrator initiates the money laundering process, which typically involves various steps such as placement, layering, and integration. Cryptoasset exchanges, both centralized (CEXs) and decentralized (DEXs), are frequently utilized for placement and integration. Layering, on the other hand, often encompasses a multitude of addresses since their creation incurs negligible costs. In recent years, there has been a shift towards decentralized exchanges (DEXs) due to their lack of regulation and absence of Know-Your-Customer (KYC) measures (cf. [9]). Additionally, perpetrators, who are increasingly technologically adept, now transfer funds across multiple chains and automate their money laundering procedures to evade traceability. Consequently, the outcome is a complex network of money laundering that is arduous, if not impossible, to trace manually.

Much like cybertrading fraud, sextortion incidents can be viewed within a similar overarching framework. Victims likewise create accounts to transfer funds to the perpetrator, who then amasses these funds and starts the money laundering process.

B. Case data collection

Our data collection was based on filed cases of cybertrading fraud and sextortion at the Bavarian Central Office for the Prosecution of Cybercrime between January, 2021 and July, 2023. Each case is assigned a unique file number, such as BY1234-010123-22/6, which consists of the county code (BY for Bavaria), a distinctive identifier for the police station where the case was filed (1234), a sequential number within that station (010123), the year of filing (2022), and an integrity check number (6). In total, we gathered 34 distinct cases of cybertrading fraud and 1793 cases of sextortion.

For cybertrading fraud cases, we were able to obtain the victim addresses and the names of the cryptoasset exchanges where the victims opened their accounts. However, in our analysis of case connections, we focused on the addresses controlled by the perpetrators and disregarded victim addresses. In sextortion cases, we only have information about the cryptoasset addresses used in the emails received by the victims. Overall, we collected 83 and 828 distinct addresses, all of which are Bitcoin addresses. To align with previous studies (see [1], [2], [6]), we refer to these addresses, which are controlled by the perpetrators, as *seed addresses*.

C. Address expansion and network abstraction

In the next phase, we expand the range of these addresses by computing address clusters, which encompass additional addresses likely controlled by the same real-world entity.

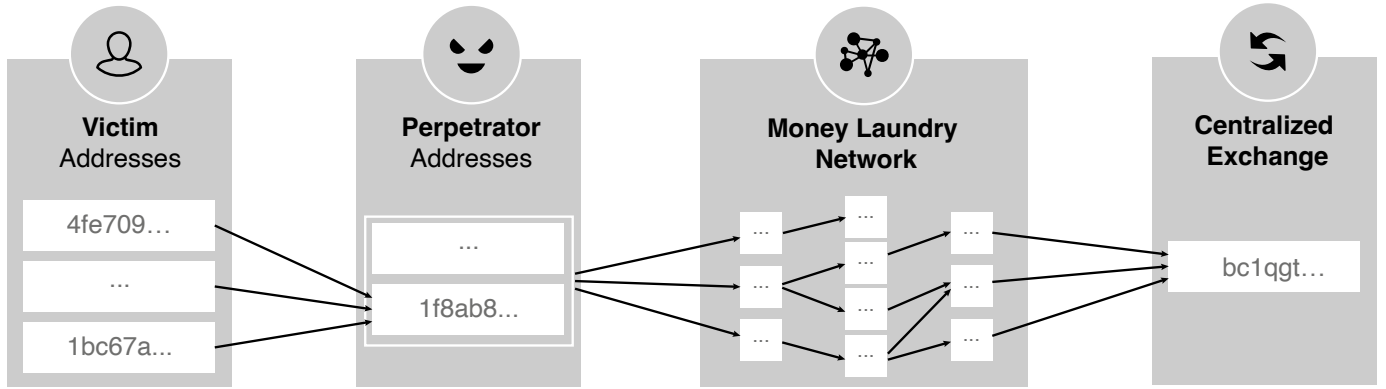


Fig. 1: **Conceptual money flow.** Funds are transferred from victim to perpetrator addresses. From there, funds are funneled through a complex money laundry network before being reintegrated and cashed out at some centralized cryptoasset exchange.

To achieve this, we leverage the multi-input heuristics, pre-filtering CoinJoin transactions, using the toolset from GraphSense [10]. We opt against other clustering heuristics due to their propensity for producing false positives [11]. Through this augmentation process, we pinpoint 69 clusters containing 1563 addresses linked to cybertrading fraud, and 147 clusters with 968 addresses tied to sextortion spam. We label the addresses within these clusters as *expanded addresses*. We can confidently infer that even though they weren't in the initial dataset, they are likely under the control of the perpetrators.

Once we have identified the pertinent clusters and addresses, we proceed to map them onto the entity graph. In this graph, each node corresponds to an address cluster, and an edge connecting two nodes signifies the combined financial transactions between the addresses encompassed within the source and target clusters. This abstraction allows us to trace the combined money inflows and outflows for a given perpetrator wallet.

Table I summarizes our seed and expanded datasets by case category. In total, it comprises 1827 different cases including 911 perpetrator addresses. For 554 of the addresses in the dataset we did not find any on-chain activity, after filtering them we end up with 1184 cases where we identified activity on the bitcoin chain, which can be used to further analyse and connect the cases.

D. Basic money flow statistics

Having collected the pertinent addresses, we next extract the transactions involving these addresses from the publicly accessible Bitcoin blockchain. For the Cyberfraud cases, we found 4909 transactions from the period between January, 2018 and August, 2022. For the Sextortion spam cases, we pinpointed 2599 transactions covering the period from January, 2019 till July, 2023. The inflow of funds for both the initial seed addresses and the expanded dataset are shown in Figures 2 and 3. We removed transactions going directly to known cryptoasset exchanges and services entities.

By examining the payment values transferred to the perpetrator, we can glean insights into the pricing models of two distinct criminal schemes. Figure 4 illustrates the distribution

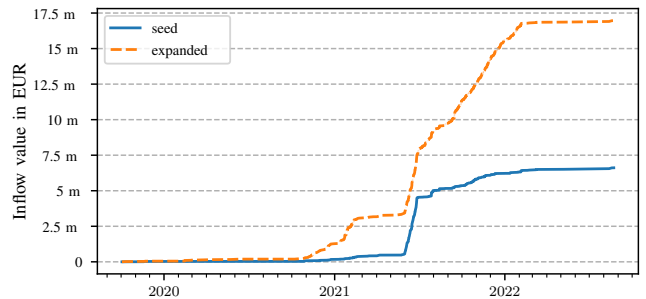


Fig. 2: **Cyberfraud inflows.** Cumulative payments received by addresses involved in cyberfraud cases.

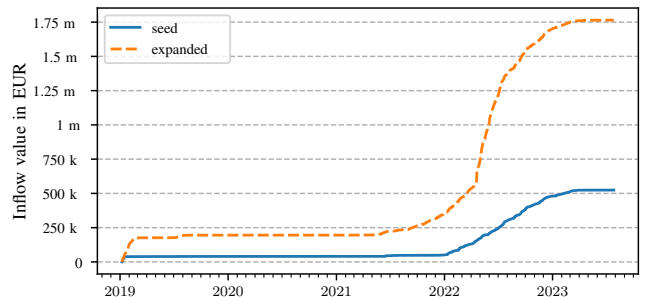


Fig. 3: **Sextortion spam inflows.** Cumulative payments received by addresses involved in sextortion spam cases.

of payments related to cyberfraud. It is evident that the majority of these payments fall within the low thousands, although there are some outliers exceeding 100k EUR. On the other hand, payments associated with sextortion spam, as depicted in Figure 5, are considerably lower, typically under 1000 EUR.

TABLE I: Summary statistics of by case category.

	Raw		Filtered		Expanded		
	# Cases	# Addr.	# Cases	# Addr.	# Cases	# Addr.	# Entities
Cyberfraud	34	83	34	83	34	1563	69
Sextortion Spam	1793	828	1150	275	1150	968	147
Σ	1827	911	1184	358	1184	2531	2531

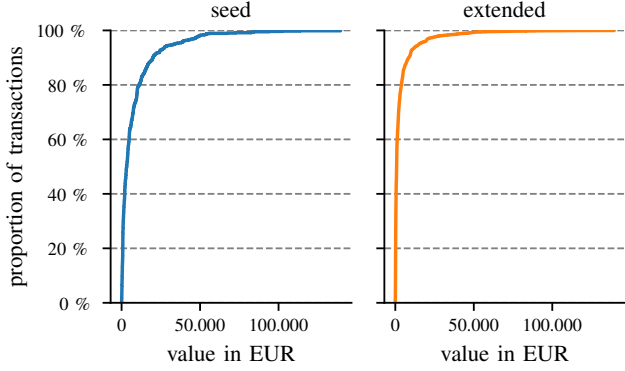


Fig. 4: **Cyberfraud payments.** Distribution of incoming cyberfraud transaction values.

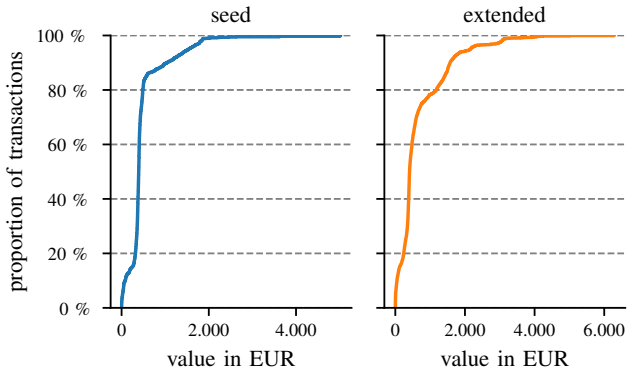


Fig. 5: **Sextortion spam payments.** Distribution of incoming sextortion spam transaction values.

IV. IDENTIFYING CASE CONNECTIONS

A. Common address heuristic

The same cryptoasset addresses appearing across multiple cases are a robust and strong indication of linked activities. In our filtered dataset, which encompasses 1184 cases, we observed that several cases already share identical addresses. By linking these cases, we identified 242 distinct *case clusters*, all of which have the same perpetrator address. Notably, the largest cluster identified using this straightforward address-matching approach connects 80 cases.

When considering case categories, it becomes evident that this method is particularly effective in linking sextortion spam cases collapsing 1150 cases into 208 case clusters. This

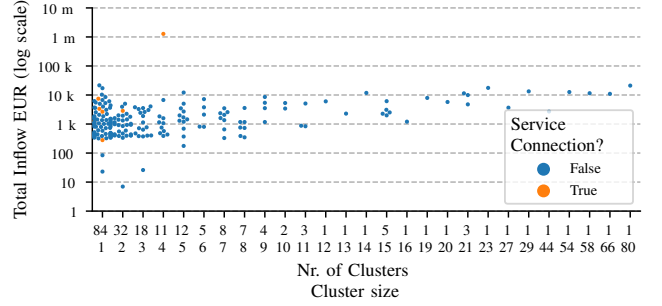


Fig. 6: **Breakdown of sextortion spam case clusters based on common addresses.** The x-axis categorizes and enumerates cases by cluster size, while the y-axis represents the inflow for each cluster. Clusters with addresses managed by an exchange are marked by orange dots. For these, inflow values may represent hot wallets shared among multiple users and should be interpreted with caution.

outcome aligns with the known fact that sextortion spam campaigns are extensively automated, target victims globally, and often reuse cryptoasset addresses within and across campaigns [2]. In contrast, for cybertrading fraud, there aren't any readily apparent links between cases. This can be explained by the intrinsic characteristics of the crime: perpetrators target victims on an individual basis and usually generate unique addresses for each victim.

Figure 6 offers a detailed breakdown of the case clusters we identified. On the x-axis, we rank and enumerate case clusters by their size, while the y-axis illustrates the inflow for each cluster to highlight its significance. For example, the first column on the x-axis represents 84 case clusters, each of size one, indicating no established connections to other cases. The subsequent column represents 32 case clusters where two cases are connected by a shared address, and so forth. It's evident that only 84 cases (represented by the first column on the x-axis) remain isolated. For all the remaining cases, there's at least one identifiable link to another case, specifically a shared address. This suggests that even such a rudimentary approach enables us to associate roughly 93% of the cases.

B. Common entity heuristic

Clustering cases based on common addresses provides an intuitive and effective approach. However, this method does not account for the possibility that a single entity, such as a perpetrator, might control multiple cryptoasset addresses. As detailed in Section III-C, we can group addresses that are

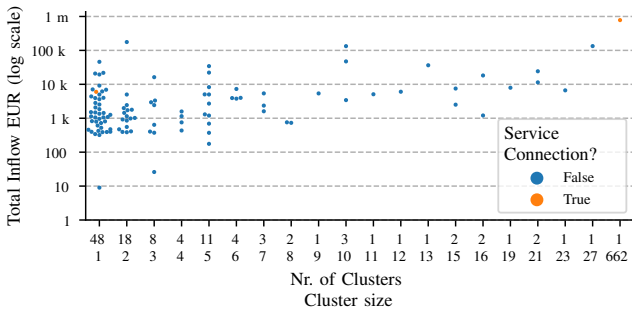


Fig. 7: **Breakdown of sextortion spam case clusters based on common entities.** The x-axis categorizes and enumerates cases by cluster size, while the y-axis represents the inflow for each cluster. Clusters with entities managed by an exchange are marked by orange dots.

likely controlled by the same real-world entity. We term such a group of addresses as an *entity*. Extending the previous logic, we can elevate our analysis from an address-centric perspective to an entity-centric one, thereby linking cases based on these shared entities.

Applying this heuristic to our refined dataset reveals that 1184 cases can be grouped into just 148 case clusters. The most extensive of these clusters connects a remarkable 662 distinct cases. Much like the address-level analysis, the majority of connections are found within sextortion cases, consolidating them from 208 down to 115 case clusters. For fraud-related cases, our method resulted in the amalgamation of only two cases, leaving us with 33 clusters at the entity level. Notably, we didn't identify any overlaps between fraud and sextortion cases, a finding that aligns with the unique modus operandi characterizing each type of crime.

When we revisit the breakdown of case clusters for sextortion as depicted in Figure 7, we observe that a mere 48 cases remain unlinked. This enhancement boosts our efficiency from the initial 92% achieved with the basic common address heuristic to approximately 95%.

Transitioning from addresses to entities through the multi-input heuristic proves generally effective [11], [12], provided specific preconditions are satisfied, such as the filtering of CoinJoin transactions. However, caution is advised when perpetrators deploy custodial addresses managed by service providers, like cryptoasset exchanges. Such actions could lead to the clustering of addresses belonging to unrelated entities. Therefore, we excluded entities tagged as exchange (11 instances) or having more than 10,000 addresses (2 instances). Consequently, we eliminated 13 clusters from our inflow data, presuming that these addresses were associated with exchanges or other service providers in the ecosystem.

C. Common collector heuristic

Perpetrators frequently generate specific addresses designated for receiving funds from one or multiple victims. After collecting the funds, they often consolidate these victim

payments into so-called *collector addresses*. These funds are subsequently funneled through intricate money laundering networks. Notably, these collector addresses are directly connected to the perpetrator addresses, being just one hop away on the outgoing side. As a result, it is logical to infer that any address or entity transmitting funds to a collector address has some association with the criminal activity. Extending this logic to individual cases suggests that cases involving addresses or entities forwarding funds to the same collector address are likely interconnected or related

Applying the common collector heuristic on our sextortion spam cases, we observe a significant consolidation: our cases are distilled down to just 81 case clusters. Only 35 cases remain isolated, lacking connections to other instances. This approach enables us to enhance the linkage of sextortion cases, elevating the connectivity rate from 95% (as achieved using the common entity heuristic) to 96.9%.

When this method is employed for our cyberfraud cases, we observe notable enhancements in linkage. By focusing on the immediate entities that receive funds from recognized perpetrator addresses, our initial 34 cases condense down to 23 case clusters. While this approach may not be as potent as it was for the sextortion spam cases, it still successfully connects 41% of our cases, leaving only 20 without any connections.

We must clearly proceed with caution when using collectors as a basis for linking cases. These collectors might not always directly be controlled by the perpetrators and sometimes be the targets of benign payments. Nevertheless, when treated with the necessary caution, collector addresses and entities can offer valuable insights into the interconnections between cases. Figure 8 illustrates to what extent sextortion spam and cyberfraud cases can be grouped into smaller heuristics using the heuristics described above.

D. Case networks

Figure 11 and 10 visualize the complete case networks. Although, the two figures give a good overview of the connectedness in general it is hard to parse exactly how clusters are formed. Therefore, Figure 9 illustrates the formation of one single cluster which is the the second largest case cluster in the sextortion cases. Its easy to see that one common collector (which is marked as being involved in a spam campaign) connects 3 smaller clusters formed via the common entity heuristic, and 5 clusters that are apparent from the address linking alone. In total the cluster links 32 different cases.

E. Overall damage estimation

In Figures 3 and 2, we presented the cumulative inflow into both the seed and expanded addresses for sextortion and cyberfraud cases, respectively. To prevent overestimation of victim flows, we excluded transactions that directly reached addresses likely associated with a service. As a result, the depicted values may be an underestimate of the actual flows.

In both crime categories, expanding our seed dataset more than doubles the observable inflow. The inflow from sextortion

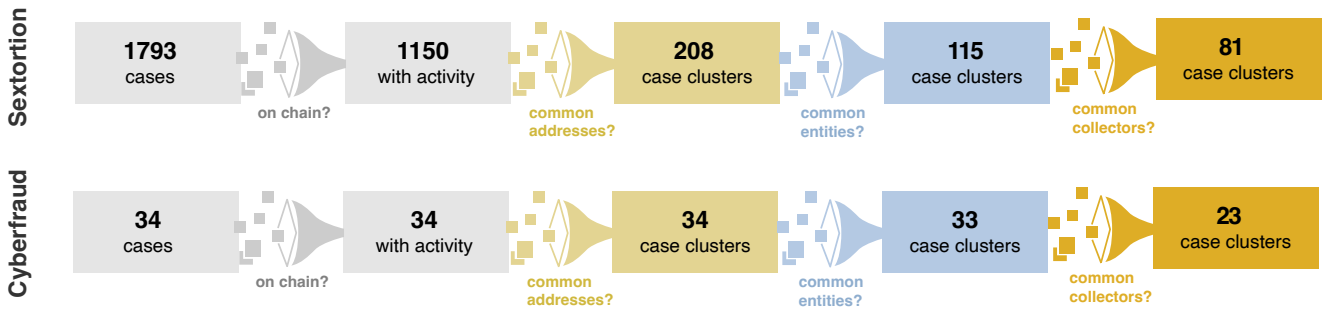


Fig. 8: **Grouping cases into case clusters.** Collected sextortion spam and cyberfraud cases are first filtered for on-chain activity and then grouped into case clusters using the common address, common entity, and common collector heuristic.

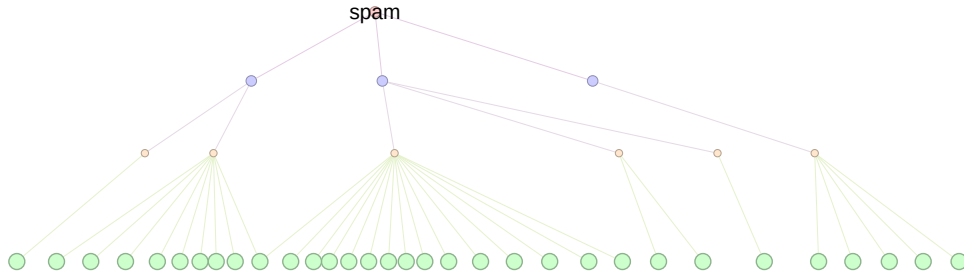


Fig. 9: **Sextortion fraud: formation of the second largest cluster using the common collector heuristic.** The cluster shown connects 32 different cases. The color coding is as follows: green nodes \odot are cases, orange nodes \circ , purple nodes are entities \odot computed by the common entity heuristic and red nodes \odot represent common collector entities.



Fig. 10: **Cybertrading fraud case network.** The color coding is as follows: green nodes \odot are cases, orange nodes \circ , purple nodes are entities \odot computed by the common entity heuristic and red nodes \odot represent common collector entities.

spam cases in the expanded dataset is estimated at approximately 1.7 million EUR, while the inflow from cyberfraud cases totals nearly 17 million EUR.

To estimate the number of potential victims, we can count the unique addresses found in incoming transactions. By focusing solely on the seed addresses, we identified at least 2359 unique addresses that sent money to a perpetrator in sextortion spam cases. Likewise, for fraud cases, we found 2431 unique sending addresses. However, this is a rudimentary estimate of the actual victim count. Some addresses might have executed multiple transactions to the perpetrator, which is often observed when funds are directly transferred from an exchange wallet.

V. CASECONNECT TOOL

Our empirical research reveals that investigators often concurrently work on identical cases, unbeknownst to each other. The pressing question then arises: how can we facilitate these investigators in collaboratively pinpointing connections among these cases and subsequently disseminating this crucial information both intra- and inter-organizationally? The ultimate objective of implementing such a mechanism is to substantially enhance the efficiency of cryptoasset investigations. This can be achieved by detecting case interconnections at an early stage, ideally when the cases are first registered by police.

In the conception of our cryptoasset case management tool, we considered the following design guidelines:

- 1) Recognizing that cases are not centrally filed but rather in a decentralized fashion by victims at individual police

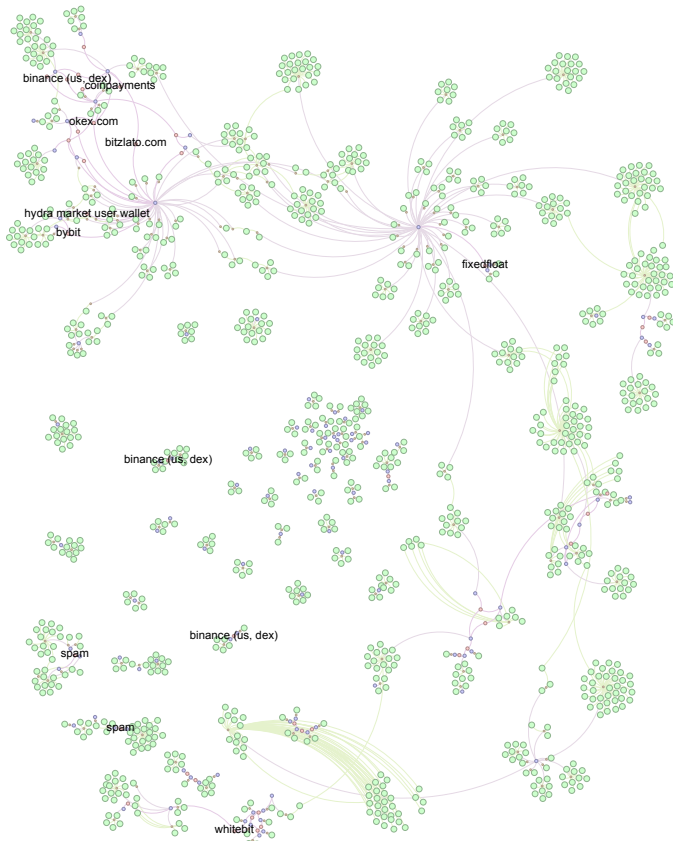


Fig. 11: **Sextortion spam case network.** The color coding is as follows: green nodes \odot are cases, orange nodes are addresses \circ , purple nodes are entities \bullet computed by the common entity heuristic and red nodes \ominus represent common collector entities.

stations, our tool must be broadly accessible instead of being restricted to a handful of centralized units.

- 2) The process of linking cryptoasset addresses to specific cases needs to be seamlessly incorporated within a foundational cryptoasset forensic workflow. This integration enables investigators to rapidly evaluate a case and determine potential links at the outset. The immediacy of this function is crucial, particularly when it comes to time-sensitive tasks like freezing assets.
- 3) To preserve the integrity of the evidence in investigations, it is essential to avoid black box solutions. Additionally, adherence to the GDPR data minimization principle⁵ is imperative.

In alignment with our overarching goal and design considerations, we opted to base our case management tool on the open-source cryptoasset analytics platform, GraphSense [10]. Technically, our tool leverages the platform’s customization mechanism and is implemented as add-on on top the foundational platform. This approach ensures its seamless integration within the comprehensive cryptoasset forensic workflow.

⁵<http://data.europa.eu/eli/reg/2018/1725/oj>

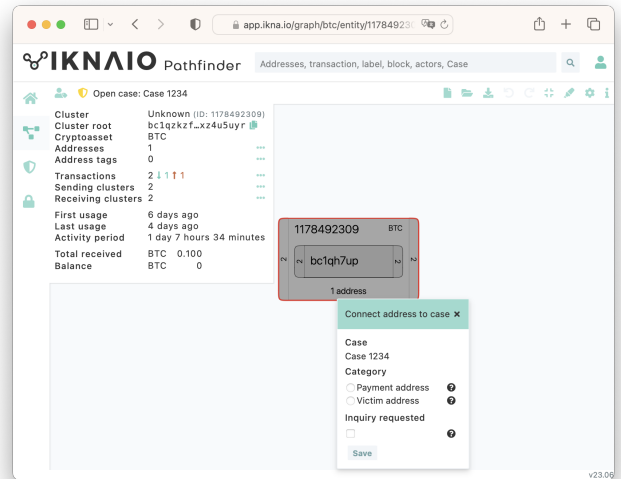


Fig. 12: **Associating address with cases.** Users can create cases and associate addresses with these cases.

Adhering to the principle of data minimization, our tool meticulously records only those data points pivotal to case linkage. Specifically, we capture case identifiers and map their associations with relevant cryptoasset addresses. Figure 12 provides an illustrative snapshot, depicting how users can associate addresses with respective cases through a straightforward annotation mechanism. To ensure clarity and eliminate ambiguities, we restrict annotations to only victim and perpetrator addresses, as these details are the primary facts presented when a case is filed. The detailed investigative trails and the resulting tracing graphs are archived within the case management system, accessible to other members within the same zone. Recognizing that investigations often necessitate information solicitations from law enforcement to cryptoasset exchanges, our system diligently notes which addresses have already been queried. This step is crucial to prevent redundant and time-consuming repeated requests.

Further bolstering data security and organization, case-related information is stored in an external relational database, which is organized into distinct *zones*. Each zone can be envisioned as an institution, where users have the privilege of viewing annotations related to cases within their specific zone. There’s also flexibility built into the system, allowing for the potential readability across zones. Figure 13 offers a visual representation of the myriad of cases and demonstrates their aggregation into what we term as *case clusters*.

VI. DISCUSSION

Our study is motivated by the pressing challenge faced by prosecutors in Bavaria, who are confronted with an escalating number of cryptoasset-related cases. In their daily work, they observed that investigators often unintentionally delve into overlapping or interconnected cases, leading us to postulate that identifying these links could substantially enhance in-

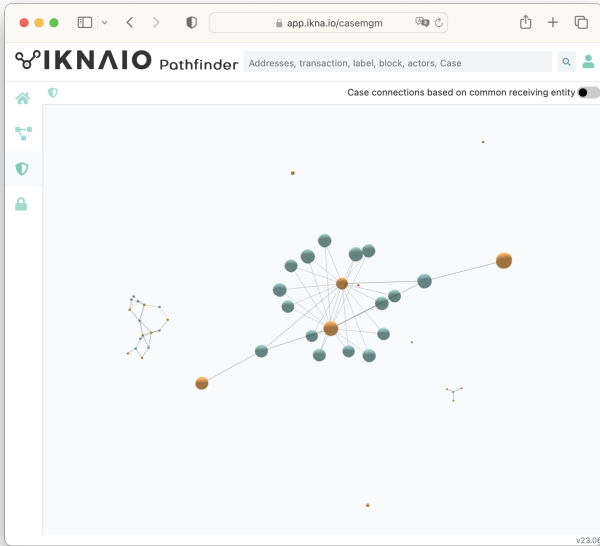


Fig. 13: **Cases and case clusters.** Users can view all cases filed within a zone (organization) and visually inspect connections to other cases.

investigation efficiency. To validate this, we analyzed sample cases from two distinct crime areas, namely cybertrading fraud and sextortion spam, aiming to establish heuristics for detecting inter-case connections. The findings from our study compellingly indicate the widespread nature of such connections, thus confirming our hypothesis. This is also in-line with previous findings showing that concentration appears to emerge frequently in cybercrime [13].

In response, we have developed a simple yet potent tool to unveil these case linkages. If adopted extensively, this tool can not only streamline investigative processes but also render economic benefits. Specifically, by focusing on clusters of interlinked cases instead of individual ones, we can achieve tangible cost savings at both the police and prosecutorial levels, allowing for a more targeted use of public funds.

While our research has provided insightful findings and a solution for a pressing challenge, it does come with certain acknowledged limitations: notably, the sample size, particularly for cybertrading fraud, is somewhat limited and geographically confined to Bavaria. Additionally, our methods using common entity and common collector heuristics operate under the assumption that perpetrators utilize non-custodial wallets to amass and collect payments from victims. To avoid false positives, we have implemented safeguards, such as filtering out recognized exchanges and excluding clusters with substantial address counts, as a practical approach to distinguish services like exchanges. Nevertheless, the possibility of a few false positives remains, which may diminish the reported numbers.

We also see several potential avenues for future work: cryptoassets have unmistakably emerged as universal payment methods across various domains of cybercrime. Consequently,

an immediate extension of our research should include other cybercrime arenas such as ransomware, child sexual abuse, and romance scams. Additionally, the scope of our study was limited to Bavaria, a mere fraction when considering that it is one among 16 provinces in Germany, which itself is just one European country, hinting at the vastness of the global stage. This microscopic focus underscores the vast potential we can unleash by expanding our methodology and approach across international boundaries and jurisdictions, especially given the pervasive and global implications of cybercrime.

On a broader scale, our study highlights a prevailing challenge: the super-linear growth of cyber-related incidents and the resulting investigations that arise in combating cyber and cyber-enabled crime. While augmenting human resources might seem like a solution, this approach realistically scales linearly at best and often less efficiently due to the organizational challenges that come with expanding teams. In the dynamic realm of cyber threats, the significance of automated assistance, combined with network-oriented thinking, addressing inherent complexities, and leveraging understood connections, will be of paramount importance.

VII. CONCLUSIONS

Cybertrading fraud and sextortion spam are witnessing an alarming surge in case numbers. Given the limited resources at hand, it is important for law enforcement to reevaluate and innovate their approach to these challenges. Merely augmenting human resources and purchasing additional licenses for centralized units is not a sustainable solution, as such strategies often scale linearly at best. Our research clearly showcases the interconnectivity and overlap among these cases. We are confident that there is a potential for tremendous efficiency improvements when this approach is extrapolated across different crime domains and transcends international borders. This study underscores the necessity of rethinking entrenched procedures, embracing a network-centric perspective, and deploying supportive tools on a broader scale. These systems, designed to discern connections and foster knowledge-sharing, represent the future of efficient and collaborative investigations. In essence, by harnessing these insights, investigators can optimize their workflows and benefit from the synergistic effects of shared knowledge and expertise.

ACKNOWLEDGEMENT

The authors would also like to thank Lukas Knorr and Barbara Krüll for their support in gathering the data for this study. This work is partially funded by the Austrian security research program KIRAS of the Federal Ministry of Finance (BMF) under the project DeFiTrace (grant agreement 905300) and the COMET Centre ABC — Austrian Blockchain Center — managed by the FFG (grant agreement 909237).

REFERENCES

- [1] M. Paquet-Clouston, B. Haslhofer, and B. Dupont, "Ransomware payments in the bitcoin ecosystem," *Journal of Cybersecurity*, vol. 5, no. 1, p. tyz003, 2019.
- [2] M. Paquet-Clouston, M. Romiti, B. Haslhofer, and T. Charvat, "Spams meet cryptocurrencies: Sextortion in the bitcoin ecosystem," in *Proceedings of the 1st ACM conference on advances in financial technologies*, 2019, pp. 76–88.
- [3] S. Foley, J. R. Karlsen, and T. Putnins, "Sex, drugs, and bitcoin: How much illegal activity is financed through cryptocurrencies?" *Review of Financial Studies*, vol. 32, no. 5, pp. 1798–1853, 2019. [Online]. Available: <https://EconPapers.repec.org/RePEc:oup:rfinst:v:32:y:2019:i:5:p:1798-1853>.
- [4] R. Anderson, C. Barton, R. Bölme, R. Clayton, C. Ganán, T. Grasso, M. Levi, T. Moore, and M. Vasek, "Measuring the changing cost of cybercrime," in *The 2019 Workshop on the Economics of Information Security*, 2019.
- [5] L. Zhou, X. Xiong, J. Ernstberger, S. Chaliasos, Z. Wang, Y. Wang, K. Qin, R. Wattenhofer, D. Song, and A. Gervais, "Sok: Decentralized finance (defi) attacks," Cryptology ePrint Archive, Paper 2022/1773, 2022, <https://eprint.iacr.org/2022/1773>. [Online]. Available: <https://eprint.iacr.org/2022/1773>
- [6] K. Oosthoek, M. Van Staalduinen, and G. Smaragdakis, "Quantifying dark web shops' illicit revenue," *IEEE Access*, vol. 11, pp. 4794–4808, 2023.
- [7] Chainalysis Inc., "The chainalysis 2023 crypto crime report," 2023. [Online]. Available: <https://go.chainalysis.com/2023-crypto-crime-report.html>
- [8] S. Abramova and R. Böhme, "Out of the dark: The effect of law enforcement actions on cryptocurrency market prices," in *2021 APWG Symposium on Electronic Crime Research (eCrime)*, 2021, pp. 1–11.
- [9] R. Auer, B. Haslhofer, S. Kitzler, P. Saggese, and F. Victor, "The technology of decentralized finance (defi)," Bank for International Settlements, Tech. Rep., 2023, working paper.
- [10] B. Haslhofer, R. Stütz, M. Romiti, and R. King, "Graphsense: A general-purpose cryptoasset analytics platform," *Arxiv pre-print*, 2021. [Online]. Available: <https://arxiv.org/abs/2102.13613>
- [11] M. Möser and A. Narayanan, "Resurrecting address clustering in bitcoin," in *Financial Cryptography and Data Security*, I. Eyal and J. Garay, Eds. Cham: Springer International Publishing, 2022, pp. 386–403.
- [12] M. Harrigan and C. Fretter, "The unreasonable effectiveness of address clustering," in *2016 Intl IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress.*, 2016, pp. 368–373.
- [13] R. Clayton, T. Moore, and N. Christin, "Concentrating correctly on cybercrime concentration," in *14th Workshop on the Economics of Information Security*, 2015. [Online]. Available: http://www.econinfosec.org/archive/weis2015/papers/WEIS_2015_clayton.pdf