Follow the money: The relationship between currency exchange and illicit behaviour in an underground forum

Gilberto Atondo Siu
Computer Laboratory
University of Cambridge
Cambridge, United Kingdom
jga33@cam.ac.uk

Ben Collier
Science, Technology, and Innovation Studies
University of Edinburgh
Edinburgh, United Kingdom
Ben.Collier@ed.ac.uk

Alice Hutchings
Computer Laboratory
University of Cambridge
Cambridge, United Kingdom
alice.hutchings@cl.cam.ac.uk

***Abstract*—Underground forums are used to discuss and organise cybercrime (as well as more conventional social activities). These forums are also commonly used for exchanging various digital currencies, either gained through the profits of crime or through less controversial means. Understanding the link between discussions of illicit behaviour and currency exchange can provide insights to identify money laundering and other parts of the cybercrime supply chain. In this paper we use natural language processing to classify posts from HackForums by crime type over a period of more than 10 years. To the best of our knowledge, this is the first time that this type of classification has been used for this large forum dataset. Although the majority of conversations in the forum were identified as relating to non-criminal discussions, we concentrate on the types of crimes being discussed by those exchanging currencies. We find the most popular topics are related to trading credentials and bots and malware. PayPal was one of the most widely advertised digital currencies and we observe significant displacement from Liberty Reserve to Bitcoin after the former was taken down in 2013. Rather than an explicit 'cashing out' mechanism, in which cryptocurrencies gained through crime flow into state-backed fiat currencies, we instead see a circulation of capital between different forms, as cash is held and then cashed back and forward according to movements in the wider currency market. We continue our examination of discussions of cryptocurrencies and explore how the underground market has reacted to new opportunities, with a qualitative case study about Facebook's putative 'Diem' coin. We find that while most discussions are related to the technical details and potential investment opportunities, some potential cybercrime use-cases are raised.**

***Index Terms*—cybercrime, cryptocurrency, currency exchange, Facebook Diem, machine learning, natural language processing**

1. Introduction

Understanding cybercrime has become increasingly important. Criminal organisations can build complex infrastructure inaccessible to regulators and law enforcement. This convoluted supply chain supports a range of cybercrimes which cause serious harm to victims. Characterising this cybercrime supply chain is a crucial part of the academic evidence base which informs the disruption of these illicit activities.

Underground forums provide significant insights into cybercrime. They play an important role as places where individuals can learn practices and skills for committing crimes, and as marketplaces for buying and selling products and services. They also facilitate socialisation and development of trusted connections between mutually distrustful individuals [1]. Therefore, we can use forums to analyse the evolution and pathways taken by individuals involved in illegal online activities [2]. Recognising these trajectories and the types of criminal activities discussed in these platforms can help law enforcement with prevention and mitigation strategies [3].

A fundamental part of the cybercrime supply chain involves exchanging funds and the laundering of illicit proceeds. While it is difficult to quantify the underground economy, it has been estimated that at least $80 billion is laundered each year [4]. Underground forums play an important role in exchanging currencies, presumably related to online crime [5]. These include cryptocurrencies and other 'alternative' currencies, such as transferring values from PayPal accounts and Amazon gift cards. Forum members take advantage of the anonymity and lack of regulation of these digital currencies to launder profits from illicit activities. Underground forums provide insights to help understand this 'cashing out' process.

Analysing underground forums can be complex as they are often large, heterogeneous, and require substantial domain expertise. For the same reasons, manual investigation of information in these forums can be time consuming [6]. Tools such as natural language processing (NLP) and machine learning (ML) enable the automation of particularly labour-intensive aspects of this process. In this paper we use these tools to provide a quantitative analysis of the type of crimes discussed in underground forums and their links to currency exchange. Although digital currencies are often linked in media and academic discussions with cybercrime and other illicit online activities, evidence of the scale and nature of links between digital currencies and the cybercrime economy is scarce. Thus, we investigate these links through an exploration of cybercrime forum discussions.

The data for our analysis are obtained from HackForums, for many years the largest and most popular underground forum and still a major hub for cybercrime communities, which has a section dedicated to currency

exchange. To make sense of the enormous volume of discussions on HackForums, we developed a 'crime type' classifier, which categorises the types of crime being discussed by forum participants, incorporating information from post content, bulletin board title and thread title. These crime types were compiled using domain knowledge and findings from prior research [2], [7] into underground forums. To the best of our knowledge, this is the first time that underground posts classification has been done using this specific crime type criteria. We acknowledge these crime type categories are not exhaustive and are not necessarily mutually exclusive. However they provide valuable insights into the various types of illegal activities being discussed on the forum by those who engage in exchanging currencies. Therefore, we consider this approach to be one of our major contributions.

We deploy this crime type classifier to explore the links between digital currencies and crime on these forums. We also use an existing model developed by Portnoff et al. [6] to categorise currencies offered and sought for exchange. We analyse the relationship between the exchange of particular currencies and the crime types which these users discuss. We find the most advertised currencies are PayPal, Bitcoin and Liberty Reserve (until it was shut down in 2013 [8]), confirming findings from previous research [6]. We build on our objective to search for relationships between cybercrime and currency exchange and find the proposed Facebook's Diem cryptocurrency (previously known as Libra) has been a topic of particular interest on the forum. We qualitatively analyse posts related to this cryptocurrency, finding discussions are mainly centered around technical details and investment opportunities, as well as some speculation about cybercrime opportunities.

2. Related work

Underground forums have been an active area of research for some time. Motoyama et al. [1] classified forums as a subgroup of online social networks, evaluating the social interactions within them, the commercial goods and services traded, and differing levels of trust. Yip et al. [9] used anonymised private messages from underground forums, with the objective of understanding the social dynamics between those involved in cybercrime. Leukfeldt et al. [10] note that lawbreakers are able to improve their skills quickly when they can access and take advantage of these types of networks.

In relation to payment methods, Mikhaylov and Frank [11] analysed hacking and carding (which refers to using stolen credit card information to buy goods and services) forums. At that time, the cybercrime economy was mainly reliant on Webmoney and Western Union for making money transfers and cashing out digitally stolen funds. Portnoff et al. [6] categorised posts within underground forums using NLP and ML models. They classified these posts based on their commercial category, the product or currency being traded, and the price or exchange rate offered. They identified that forum participants favoured Bitcoin and PayPal, against other currencies, for transferring money and liquidating proceeds.

Pastrana et al. [2] also categorised underground forum posts using ML approaches. They identified important actors and their common interests and used social network analysis and clustering to aggregate them based on to their activities within the forum. They focused on understanding the factors that could predict future cybercrime involvement and the main trajectories into cybercrime.

Pastrana et al. [5] used the CrimeBB dataset (which we use in this research) to study how currency exchange has evolved since 2005, tracking actors' activities in the forum before they engaged in currency exchange. They found that overall PayPal has been the most advertised currency offered for exchange. Liberty Reserve had been popular before it was closed by the United States government in 2013 due to money laundering [8]. Soon after Liberty Reserve was closed, Bitcoin became the second-most offered currency for exchange. They also found advertisements for the trade in Amazon gift cards have increased since 2015.

Some illicit activities are known to be related to particular types of digital currency. For example, eWhoring, a type of scam that is a popularly discussed topic on HackForums, commonly accepts PayPal and Amazon gift cards from victims [12]. The currencies used can also change over time. For example, Karami et al. [13] evaluated the effects of PayPal shutting down the accounts linked to booter services (which provide denial of service attacks). They found that while the intervention had a negative effect on profits, some operators displaced to Bitcoin.

Due to the size of underground forums, NLP approaches for classifying posts are becoming increasingly relevant. Established tools–often trained on media articles–struggle with users' unique lexicon, short posts, and inconsistent capitalisation and punctuation. Caines et al. [7] developed NLP tools to use in underground forums to identify post type, author's intent and addressee. Like this work, they used a mix of statistical and logical classification models to predict data tags automatically.

Bhalerao et al. [14] used NLP techniques and graph traversal algorithms to uncover and analyse cybercrime supply chains through examination of underground forums. They discovered relationships between products purchased and subsequent posts selling those same products. Their analysis shows that currency exchange was a fundamental part of the supply chains, showing up in more than 70% of validated chains.

3. Research Questions

We believe that having a more detailed overview of the topics discussed in underground forums can aid us in examining cybercrime. We contend that separating these conversations by crime type will facilitate our comprehension of illicit activities within these online platforms. We argue that a subsequent analysis of the link between these conversations and the currencies exchanged can elucidate more details of the cybercrime supply chain within the forum.

Therefore, our first research question regards the relationship between crime types being discussed by users offering to exchange currency, and the types of currency they are seeking to exchange. To answer this question, we first built a crime type classifier to categorise the activities of those involved in currency exchange elsewhere in the forum. We then graphed the currencies being offered for exchange by those discussing each crime type over time.

A deeper understanding of the use and talk of cryptocurrencies within these forums can aid in discerning fraudulent activity and money laundering. Therefore, our second research question relates to potential future use cases of Facebook's Diem cryptocurrency (previously known as Libra). This cryptocurrency was officially announced by Facebook on 18 June 2019 and was expected to be launched in 2021. We expected that this cryptocurrency would have been a topic for discussion on the forum, however, we sought to explore in particular whether these discussions related simply to general discussions of Diem/Libra, or whether forum users were actively scoping out its potential for abusive or illegal activities.

4. Method

4.1. Data

The data used in this work are a subset of the CrimeBB dataset. At the time of writing, CrimeBB contained more than 4.7 million accounts with more than 89 million posts, extending over 27 forums that have been active since 2007. This dataset was created in 2018 and is kept up-to-date by CrimeBot [5], a crawler designed specifically for scraping data from underground forums. CrimeBB is available for academic research use under a data sharing agreement with the Cambridge Cybercrime Centre.

We selected HackForums for analysis as it is the most popular, largest and longest-running underground forum. Our dataset includes more than 42 million posts in four million threads made by more than 637,000 members on 197 sub-forums (bulletin boards). Our work includes a section specifically focused on currency exchange analysis. Our dataset for this specific section includes data prior to June 2018 since the administration of marketplace contracts within HackForums changed after this date [15].

4.2. Ethical considerations

The department's ethics committee approved this research, which uses data extracted from a publicly accessible forum with more than 637,000 members. Obtaining informed consent from all forum participants is infeasible and would be considered as spamming. This work focuses on understanding aggregate information and collective behaviour; we do not analyse specific individuals, or attempt to identify users. Therefore, this work falls outside the requirement of informed consent, under the British Society of Criminology's Statement of Ethics [16].

4.3. Crime type classifier

To identify the relationship between currency exchange and crime type, we built a classifier to categorise HackForums posts by crime type. We also used heuristics to design two baselines which provided a point of comparison for performance measures.

4.3.1. Data sampling. We selected several samples from the dataset to be annotated and subsequently used to train and test our crime type classifier. By following a similar approach to Pastrana et al. [2] we selected a sample containing 2,000 forum posts. This first sample of posts included (i) 500 posts selected completely at random, and (ii) 1,500 posts randomly selected after applying a filter to the dataset. This filter focused on leveraging previous classification by Caines et al. [7], in which posts were categorised by post type. We decided to include in this filter only those posts that we contend were potentially crime related (classified as 'offer', 'request', 'exchange' and 'tutorial' in [7]).

These initial 2,000 posts were used for the first iteration of training and testing. After analysing the first iteration of results (section 5) obtained with the classifier, we decided to add a second set of 2,000 posts to improve the performance measures. For comparison purposes, we decided to randomly extract posts from forum participants actively engaged in the currency exchange section. In total, a combined sample of 4,000 posts was used for the second iteration of training and testing of the classifier (section 5.1) using a split of 70/30 per cent correspondingly.

4.3.2. Annotation and 'ground truth' definition. Each post was annotated into one category using the classification criteria for crime type shown in Table 1. These crime types were compiled using domain knowledge and findings from prior research [2], [7] into underground forums. While these crime type categories are not exhaustive and are not necessarily mutually exclusive, they allow us to dissect the types of crimes being talked about, and help us identify the types of activities discussed by those involved in currency exchange.

TABLE 1. CRIME TYPE CLASSIFICATION CRITERIA

Crime Type	Definition
Not criminal	Unrelated to crime
Access to systems	Access to systems (excluding use of malware) and SQL injection attacks
Bots & malware	Bots or malware and related services
eWhoring	eWhoring (simulation of fraudulent cybersexual encounters for financial gain)
Currency exchange	Exchanging digital currencies
DDoS & booting	DDoS attacks, booting, stressing, and stress testing
Identity theft	Online identity theft, internet fraud, online scams or credit card fraud
Spam	Sending spam, sharing email addresses or containing marketing services
Trading credentials	Trading accounts including gaming, social networks and Netflix accounts
VPN and hosting services	VPN, hosting and proxy services

The initial 2,000 posts were annotated by three observers. The second set of 2,000 was annotated by two reviewers and was incorporated into the first sample. We used Cohen's κ [17], [18] and Fleiss's κ [19] to determined the level of agreement between annotators. The Fleiss's κ of 0.889 for the first sample and the Cohen's κ of 0.948 for the combined sample represented 'almost perfect' agreement between the annotators, according to the criteria by Landis and Koch [20].

Despite the high level of agreement, the annotation process was not straightforward due the ambiguity of the posts' content. A few examples of such ambiguity, which demonstrate the difficulty of the task at hand, include:

- Some posts discuss crime types but not the actual commission of the crime being discussed. For example, one of the posts on the bulletin board called 'Suggestions and Ideas' asks for 'Death Removal of eWhoring'. The discussion focuses on whether the topic of 'eWhoring' should or could be removed from the forum. These types of posts were classified as 'not criminal'.
- Some posts discussed products or services, but did not indicate they were selling or using them. For example, some posts discussed software that can automatically increase forum participants' social network channels' subscribers by the hundreds. Where there were no indications the products were being used or sold, these posts were classified as 'not criminal'.
- Posts about 'Botting' and 'Hosting' were particularly difficult to classify as they are not always related to criminal activity. For example, the use of bots for enhanced game playing is a common topic. While this may be against a game's terms of service, no crime is being committed. Therefore, analysing the post context is crucial for categorising the post correctly. As such, these types of posts were classified as 'not criminal'.
- Some discussions move quickly onto private messaging. This pattern was observed a significant number of times during the annotation process. As the CrimeBot scraper only collects public messages, these conversations are not available to us.

The 'ground truth' was defined as the crime type chosen most frequently between three annotators or as the final agreement between two annotators where there was a disagreement encountered during the annotation. Figure 1 shows the distribution of labels after the annotation process. We observed that the distribution was highly skewed towards 'not criminal'. This was in line with our expectations, as the majority of conversations in underground forums are not of a criminal nature. However, this imbalance in the post numbers available for training can negatively impact the reliability of classification results obtained, specifically for those crime types that have a very low frequency. Table 2 shows the number of posts used in the training set for the crime type classifier. It can be seen that the categories 'eWhoring', 'Identity theft' and 'Spam' have less than 50 training observations (equivalent to 1.8% of the training set). To compensate this limitation, we used the Synthetic Minority Over-sampling Technique (SMOTE) [21] which partially offsets this limitation by increasing the representation of the outnumbered categories.

4.3.3. Baseline design. To evaluate the performance of the crime type classifier, two baselines were constructed based on rules-based methods. These rules-based systems have some limitations and disadvantages that make large-scale deployment difficult. First, they demand a significant knowledge of cybercrime taxonomies. Second, they are labour intensive and time-consuming since creating rules for a complex system demands substantial manual inspection, particularly as the language used on underground forums can include a high preponderance of colloquialisms and slang. These systems are also difficult to maintain and are not easily scalable.

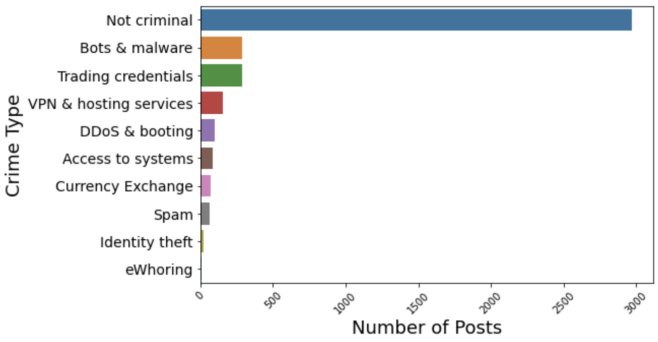

Figure 1. Distribution of Crime Type labels

TABLE 2. NUMBER OF POSTS USED FOR TRAINING

Crime type	Number of posts
Access to systems	59
Bots & malware	202
DDoS & booting	69
eWhoring	6
Identity theft	14
Not criminal	2,096
Spam	43
Trading credentials	200
VPN and hosting services	111
Total	2,800

Approximately 73% of annotated posts were 'not criminal', around 7% related to 'bots & malware' and another 7% related to 'trading credentials'. We take a simplistic approach of rounding up these percentages to obtain a total of 100%. Therefore, the proposed first baseline method is to label every post randomly with a probability of 80% as 'not criminal', 10% as 'bots & malware' and 10% as 'trading credentials'.

This approach has some disadvantages. One of the objectives of this work is to classify posts by crime type. Based on this approach, the majority of posts will be classified as 'not criminal' and only two types of crime will be considered for the classification of the remaining posts. Nevertheless, this baseline serves as a benchmark to compare the performance of the models evaluated in section 4.3.4.

The second baseline used heuristics based on insights obtained during the annotation process. A relationship was identified between the bulletin board name and the crime type assigned by the annotators. For example, we found that the majority of posts belonging to the bulletin board 'Free Ebook Hacking Tutorials' were categorised as 'Access to systems' by the annotators. Following a naive approach, the second baseline proposed assigned the crime type based on the bulletin board name as shown in Table 3.

4.3.4. Statistical Models. We pre-processed the data by getting rid of any blank inputs, changing all text to lower case, tokenising all input text, removing stop-words and performing word lemmatisation using the NLTK library.[1] To develop the classifier, we compared the performance of four statistical models, namely Support Vector Machines (SVM), Multinomial Logistic Regression, Random

1. http://www.nltk.org

TABLE 3. ANNOTATED CRIME TYPES DISCUSSED ON VARIOUS BULLETIN BOARDS

Crime type assigned	Bulletin board name
Access to systems	Free Ebook Hacking Tutorials
	SQL Injection Attacks
	PHP Development
Bots & malware	Botnets, IRC Bots, and Zombies
	Cryptography and Encryption Market
	Cryptography, Encryption, and Decryption
	Remote Administration Tools
DDoS & booting	Server Stress Testing
eWhoring	eWhoring
Identity theft	Monetizing Techniques
Spam	Free Money Making Ebooks
	Referrals
Trading credentials	Appraisals and Pricing
	Non-Free Accounts
	Online Accounts
VPN and hosting services	Hosting Services
	VPN Hosting and Services

Forests, and XGBoost. We performed hyperparameter tuning and ten-fold crossvalidation. The input data was split for training and testing using a ratio of 70/30 correspondingly. The training set was oversampled using SMOTE to deal with the highly skewed data distribution. We extracted a vector of lexical features by using the Term Frequency-Inverse Document Frequency (TF-IDF) words weighting [22] and used it in all our models.

4.3.5. Performance Measures. We used measures of precision, recall and F-measure to evaluate and compare each of the baselines and models performance. Precision identifies the percentage of posts the classification model predicted correctly out of the total number of posts that it predicted for a given crime type label. Recall refers to the percentage of posts the classification model predicted for a given crime type label out of the total number of posts it should have predicted for that given crime type label. F-measure is a weighted average of precision and recall. All of these scores range from 0 to 1, with 1 being the best possible score and 0 the worst, and are calculated as follows [7]:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{1}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{2}$$

$$F = 2 \cdot \left(\frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \right) \tag{3}$$

Where:

- TP = True positive, if the predicted crime type label for each post is "Not Criminal" and is equal to the ground truth category which is "Not Criminal".
- TN = True negative, if the predicted crime type label for each post is not "Not Criminal" and is equal to the ground truth crime type.
- FP = False positive, if the predicted crime type label for each post is not equal to the ground truth crime type category.
- FN = False negative, if the predicted crime type label for each post is not "Not Criminal" and is not equal to the ground truth category which is "Not Criminal".

To complement these performance measures we also used accuracy. This is defined as the percentage of posts classified correctly with the right crime type label as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

Accuracy can also range from 0 to 1, with 1 being the best possible score and 0 the worst. This score by itself is not a very good measure for classification problems with skewed distributions, therefore we used all four performance measures.

4.4. Currency exchange analysis

One of our objectives is to examine the use and discussion of cryptocurrencies within underground forums. This can help us identify fraudulent activity and money laundering, which is an important element of the supply chain of cybercrime. For this reason, we analysed advertisements for currency exchanges in HackForums per crime type over the course of more than eight years. Our dataset included posts by 11,397 actors who were active in the 'Currency Exchange' sub-forum. We extracted all posts from these actors prior to June 2018 across all 197 sub-forums in HackForums. More than 11.7 million posts were obtained and categorised using our crime type classifier.

We used the tools developed by Portnoff et al. [6] to parse and tokenise posts from the 'Currency Exchange' sub-forum. Their classification model was also used to extract the currencies being offered and searched for by currency exchange actors from December 2009 to June 2018. We linked each of these currencies being advertised for exchange to the other forum posts by the same actors, classified by crime type.

Additionally, we obtained the net currency flow over time, calculated as the number of posts aiming to buy a particular currency minus the number of posts aiming to sell that currency, establishing a rough measure of the net flow towards or away from a given currency at a given time. We focus our modelling in particular on Bitcoin's net currency flow (defined as number of posts wanting Bitcoin minus number of posts trying to sell Bitcoin) and compare this to the inverse base-10 logarithm of the Bitcoin price (we use this transformation to minimise the volatility observed).

4.5. Qualitative analysis

We continue our examination of potential relationships between cybercrime and currency exchange by analysing conversations in HackForums about the proposed Facebook Diem/Libra cryptocurrency. A Python script was used to extract all posts related with Facebook's proposed cryptocurrency. The terms used for this purpose included: 'Facebook Diem', 'Facebook Coin', 'Facebook Libra', 'Globalcoin' and 'Calibra'.

A total of 25 posts from 22 actors were found between May 2019 and April 2021, which were analysed qualitatively by hand. The posts were grouped according to the main topic of the discussion. All quotations are provided verbatim.

5. Results

5.1. Crime type classifier

Table 4 shows the performance results for both baselines described in section 4.3.3. Our aim is for the classifier to outperform these baselines.

TABLE 4. BASELINE PERFORMANCE

Model	Precision	Recall	F-Measure	Accuracy
Baseline 1	0.678	0.722	0.700	0.545
Baseline 2	0.833	0.948	0.886	0.822

The first baseline provides an F-measure of 0.700, which is weighted relatively equally across precision and recall. While labelling the majority of posts as 'not criminal' at random with a probability of around 80% provides a fairly good F-measure, the accuracy is poor at 0.545. This naive implementation is unsuitable for this research as one of the objectives is to predict the crime type being discussed. This also reflects the nature of conversations that participants have in HackForums, with the majority not being overtly related to illegal activities, at least before moving into private messaging or other means of communication. The inclusion of the bulletin board title as a heuristic in the second baseline improves all the measures. In particular, the F-measure increases by 27% to 0.886 and the accuracy increases by 50% to 0.822. This also reflects the nature of HackForums, with threads sub-divided by topic-specific bulletin boards.

5.1.1. Using 'Post' content for feature extraction. Our first attempt, using post content for feature extraction, performed better than baseline 1 but worse than baseline 2. Table 5 shows the performance results for all the statistical models. The results show that the XGBoost and the Random Forest models have the best performance.

TABLE 5. PERFORMANCE OF STATISTICAL MODELS - FEATURE EXTRACTION INCLUDES 'POST' CONTENT

Model	Precision	Recall	F-Measure	Accuracy
SVM	0.733	0.719	0.723	0.719
Logistic Regression	0.770	0.728	0.744	0.728
Random Forest	0.741	0.771	0.747	0.771
XGBoost	0.750	0.769	0.754	0.769

5.1.2. Using 'Bulletin Board Title' for feature extraction. As the baseline 2 results show a significant improvement compared to those of baseline 1 when bulletin board titles are included, we decided to incorporate this information into the statistical models. Our second approach includes only information from the Bulletin Board Title for feature extraction. Table 6 shows the performance results for all the statistical models.

Compared to Table 5, these results show an improvement in all performance indicators for the Random Forest model but an improvement in precision only for the rest of the models.

TABLE 6. PERFORMANCE OF STATISTICAL MODELS - FEATURE EXTRACTION INCLUDES 'BULLETIN BOARD TITLE'

Model	Precision	Recall	F-Measure	Accuracy
SVM	0.842	0.608	0.677	0.608
Logistic Regression	0.843	0.616	0.685	0.616
Random Forest	0.846	0.863	0.852	0.863
XGBoost	0.838	0.702	0.749	0.702

5.1.3. Using 'Post' content and 'Bulletin Board Title' for feature extraction. Table 7 shows the performance results for all models for our third attempt, which incorporated the information of both post content and the bulletin board title into the feature extraction.

TABLE 7. PERFORMANCE OF STATISTICAL MODELS - FEATURE EXTRACTION INCLUDES 'POST' CONTENT AND 'BULLETIN BOARD TITLE'

Model	Precision	Recall	F-Measure	Accuracy
SVM	0.837	0.846	0.840	0.846
Logistic Regression	0.862	0.848	0.853	0.848
Random Forest	0.858	0.862	0.850	0.862
XGBoost	0.854	0.861	0.854	0.861

The results show an improvement between 5% and 16% compared to those obtained including only post information into the feature extraction. The results for all the models continue to outperform those of baseline 1. Additionally, accuracy for the models SVM, Random Forest and XGBoost, is higher than that for baseline 2 of 0.822. Nevertheless, baseline 2 F-measure continues to outperform all the statistical models.

5.1.4. Using 'Post' content, 'Bulletin Board Title' and 'Thread Title' for feature extraction. One final evaluation incorporated the thread title into the feature extraction containing the post content and the bulletin board title. Table 8 shows the performance results for all models.

TABLE 8. PERFORMANCE OF STATISTICAL MODELS - FEATURE EXTRACTION INCLUDES 'POST' CONTENT, 'BULLETIN BOARD TITLE' AND 'THREAD TITLE'

Model	Precision	Recall	F-Measure	Accuracy
SVM	0.881	0.886	0.882	0.886
Logistic Regression	0.895	0.890	0.891	0.890
Random Forest	0.867	0.872	0.860	0.872
XGBoost	0.871	0.880	0.873	0.880

The results for all the models continue to outperform those of baseline 1. They also show an improvement between 1% and 3% compared to those obtained including only post information and bulletin board title into the feature extraction. Accuracy for all the models outperforms the baseline 2 score of 0.822. The F-measure for the logistic regression model also outperforms both baselines. When analysing the prediction results, we found the SVM model and the Logistic Regression model were overfitting the training data. The XGBoost model showed the best prediction performance without over-fitting the data, and therefore was subsequently selected for the final classification.

5.2. Currency exchange analysis

We used our crime type classifier to categorise more than 11.7 million posts across all of HackForums prior to

June 2018 made by actors who were active on the currency exchange board. The results showed that 97% of the posts analysed were classified as 'not criminal' and 3% were of a criminal nature. Out of those posts discussing illicit topics, we find the majority of them to be related to trading credentials (42%) and bots and malware (32%). After the categorisation, we linked the currencies offered and sought for exchange to the types of crime being discussed by the same actors. Figures 2 to 9 show the evolution over time for the four most highly advertised currencies for each crime type. Each figure shows the number of users on a relative basis (normalised by dividing the number of users by the aggregated number of posts for all crime types over time) who had ever discussed that crime type, advertising the most popular currencies by month. The currencies offered and wanted are shown in dashed and solid lines respectively. The common features observed amongst all different crime types shown in Figures 2 to 9 are the following:

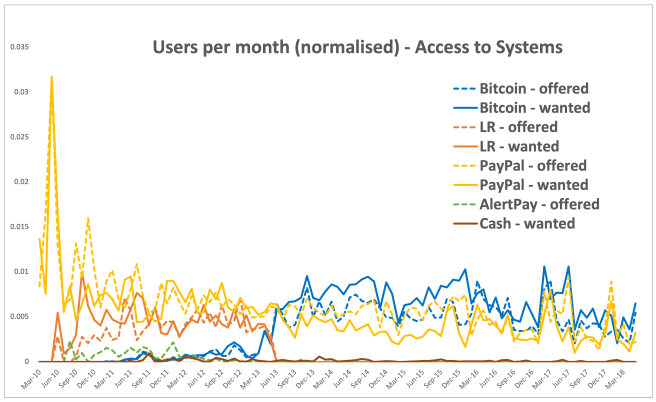

Figure 2. Currencies exchanged by actors posting about access to systems

- Overall, before June 2013, PayPal was the most highly advertised (offered and wanted) digital currency followed by Liberty Reserve. This is consistent with results found by Pastrana et al. [5].
- After Liberty Reserve was shut down in May 2013, Bitcoin became the most highly wanted currency amongst all crime types (blue solid lines). Similarly, PayPal was the most highly offered currency (yellow dashed lines) in the whole period observed. This is consistent with previous research by Portnoff et al. [6] who identified Bitcoin and PayPal to be the preferred methods for cashing out or transferring money.

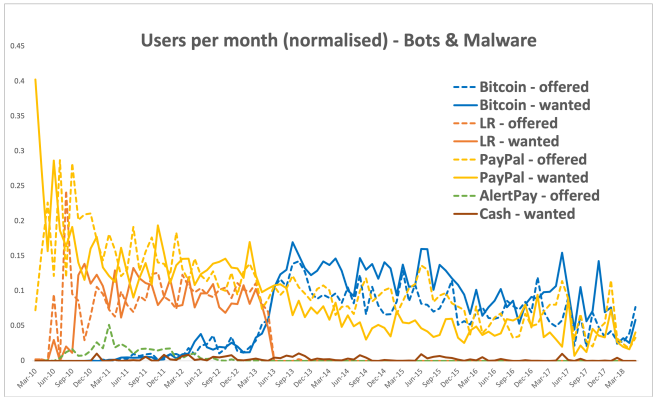

Figure 3. Currencies exchanged by actors posting about bots & malware

We can also observe that the most highly advertised currencies (after Bitcoin, PayPal and Liberty Reserve) are AlertPay, OmniCoin and Cash. Figures 2, 3, 7, 8 show that forum participants actively involved in conversations about Access to Systems, Bots & Malware, Spam and Trading Credentials are looking for Cash and offer AlertPay. Figure 4 shows that users discussing DDoS & Booting are also offering AlertPay but they want OmniCoin (OMC) instead.

Prior research by Pastrana et al. [12] found Amazon gift cards and PayPal are the most common payment platforms used for eWhoring. This is also confirmed by our results in Figure 5. Similarly, Figures 6 and 9 show that AlertPay is the most highly advertised currency for people that are actively talking about Identity Theft and VPN & hosting services.

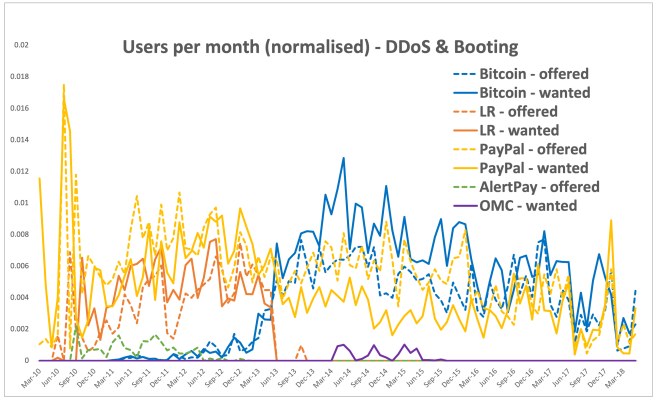

Figure 4. Currencies exchanged by actors posting about DDoS & booting

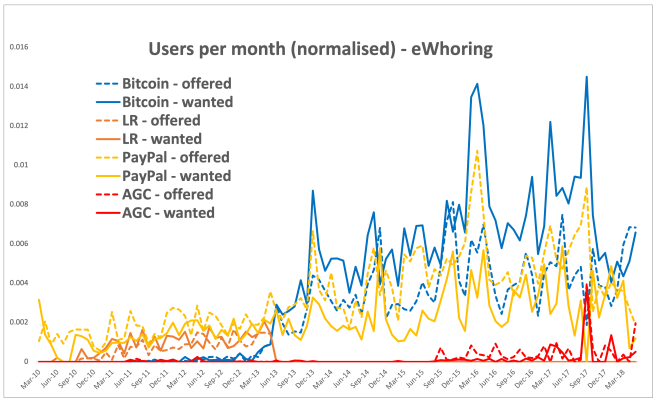

Figure 5. Currencies exchanged by actors posting about eWhoring

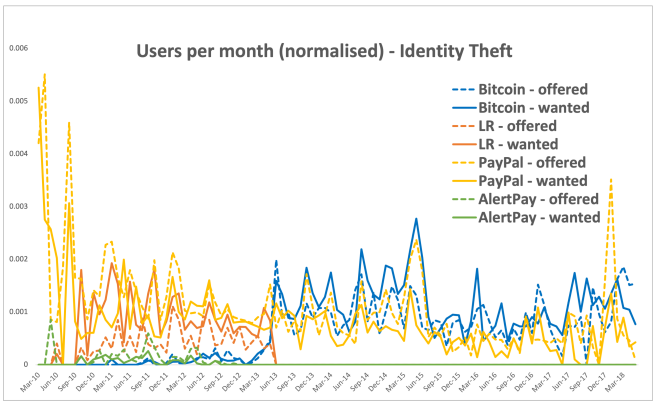

Figure 6. Currencies exchanged by actors posting about identity theft

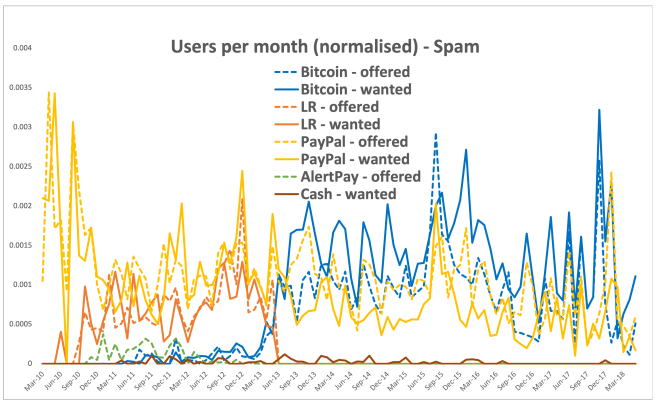

Figure 7. Currencies exchanged by actors posting about spam

Our objective with these figures is not to prove causality or to identify reasons behind the preferences found. Instead we intend to find patterns that could help us elucidate possible relationships between the crime types discussed in these forums and the alternatives that actors are using to 'cash out'.

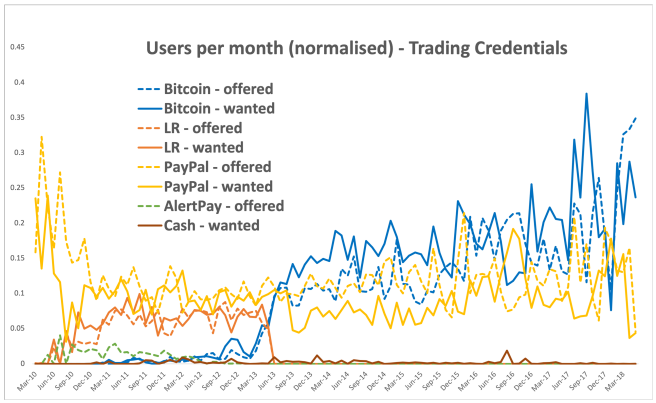

Figure 8. Currencies exchanged by actors posting about trading credentials

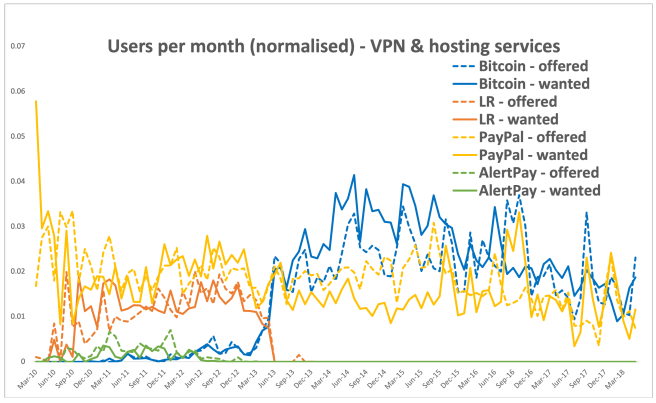

Figure 9. Currencies exchanged by actors posting about VPN and hosting services

The net currency flow over time is shown in Figure 10. Although there are generally a large number of posts seeking to buy and a large number seeking to sell a particular currency at a given time, there is also generally a weighting towards buying or selling a given currency at a particular time. Overall, we observe a clear flow from one currency to another across the whole period analysed. For instance, we see that there is a flow from PayPal to Liberty Reserve around July 2010 which then reverses several times in 2011. Our results also show that Bitcoin started being advertised amongst HackForums users around May 2011. In 2013, we see a net flow from PayPal into Bitcoin. This could be related with the exponential increase observed in Bitcoin's price when it rose from $13.30 in 1 January 2013 to $770.44 in 1 January 2014 [23].

Our findings suggest that as Bitcoin's price increases, the flow from PayPal to Bitcoin tends to decrease (or even reverse), with some steep changes around particularly strong perturbances in the Bitcoin dollar value. Our modelling finds that the net currency flow (defined as number of posts wanting Bitcoin minus number of posts trying to sell Bitcoin) varies closely with the inverse base-10 logarithm of the Bitcoin price, as shown in Figure 11.

There is a clear relationship on the forums between some minor currencies (such as Amazon gift cards) and particular crime types (in this case, eWhoring), however

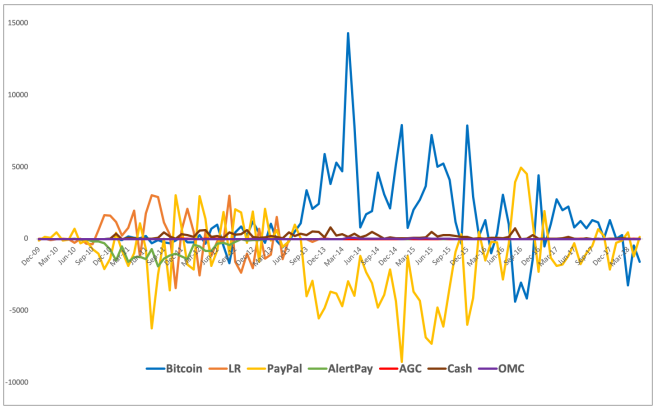

Figure 10. Net Currency Flow

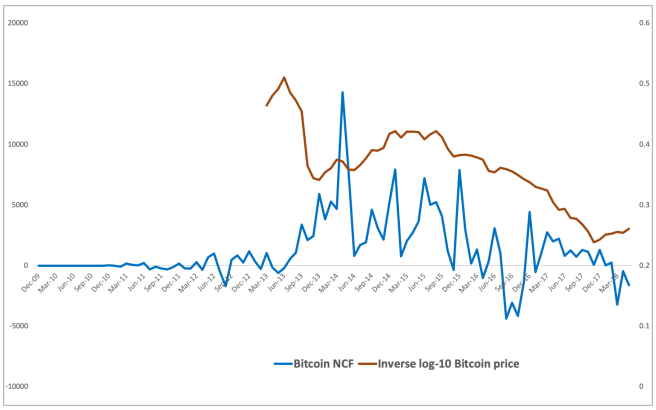

Figure 11. Bitcoin Net Currency Flow v. Inverse Base-10 logarithm of Bitcoin price

the predominance of a few major currencies for each crime type suggests a well-established set of cash-out practices for this illict economy. PayPal and AlertPay are generally the currencies being cashed out – those with a negative flow, which people are generally trying to get rid of, and Bitcoin, Liberty Reserve and fiat currencies in the form of cash are generally the desired currencies, towards which the cash-out process flows. However, Figure 10 only shows the net currency flow – it is clear from Figures 2 through 9 that there is generally a large interest in both buying and selling each of these currencies at any given time, despite the existence of a net flow in one direction.

Thus, our results imply that either all of these currencies are a potential 'end state' for the cash-out process (which is rather unlikely), or, more likely, this process is generally cyclical in nature, with capital circulating in a variety of different forms in the market, only being cashed out at particularly opportune times, such as when the Bitcoin price is particularly favourable, for example.

5.3. Qualitative analysis of discussions about Facebook Diem/Libra

We were interested in how forum users perceived new crime opportunities that might arise due to the forthcoming launch of the Facebook's cryptocurrency. It was anticipated Diem (formerly known as Libra) would have been a topic for discussion on the forum and we wanted to know if it was being discussed in relation to crime. The broad topics of discussion the 25 posts fell into were technical details, investment opportunities and price speculation, and the potential cybercrime use cases.

5.3.1. Technical details. More than a third of the posts (nine posts out of 25) mentioned or discussed technical features of Facebook's proposed cryptocurrency. For instance:

> Libra will be a stable coin although it is not sure what it will be capped to. Libra is not a pegged to a single currency. Libra will have its own Blockchain and will use a BFT approach to consensus as opposite to POW and POS.

We can see that some users have conversations that are technically complex and not necessarily obvious for people that are not familiar with cryptocurrencies. The following quote provides a few points summarising some of the details originally announced by Facebook:

> Designing and using the Move programming language. Using a Byzantine Fault Tolerant (BFT) consensus approach. Adopting and iterating on widely adopted blockchain data structures.

Another post debates some of the technical elements stated in the technical paper and shows his discontent with them. Other posts focus on the economic and financial specifics related with Facebook's proposal and other digital and cryptocurrencies. For example, a few posts talk about the scrutiny that Facebook has received from government regulators in the United States and the European Union to ensure regulatory compliance and consumer protection.

These excerpts show that underground forums can be used to discuss complex topics on a timely manner. They also show that some users are familiar with cryptocurrency concepts such as 'BFT', 'POW', 'POS'.

5.3.2. Investment - price speculation. A third of the posts were focused on the potential of profiting from price speculation, which has occurred with Bitcoin and other cryptocurrencies. For example, one of the posts mentions the potential 'bullish impact' that Facebook's globalcoin will have on Bitcoin's price.

> The launch of globalcoin from facebook will have bullish impacts in my opinion and with the binance sell off and the upcoming LTC halving, we are likely to see a continued rise. We need to see a continued rise in trading volume however to support an estimate of 14k in the near future, but I don't see a drop to lower than 6k unless there is some sort of triggering event.

Another post predicts the price of Bitcoin will go to zero, while Facebook's coin price will increase significantly. However, we can see that underground forums are a place for debate, with another post on the same thread contradicting these posts, pointing out that Facebook's coin is a stable coin and therefore they will not be able to profit from price movements.

5.3.3. Potential cybercrime use cases. Less than a fifth of posts had some reference to cybercrime. For instance, one post seems to make a comparison of Facebook's coin to another cryptocurrency that is heavily 'carded':

> facebook coin will be a cancer shitcoin inb4 that shit gets carded to the next dimension

On the other hand, another post mentions the challenges for carding which are related to the use of Blockchain and higher security protocols to fight fraud:

> carding nowadays has become wayyyy harder. if you aint a pro better leave this topic alone. and soon carding is coming to an end due to blockchain tech and cryptos like the Libra coin which mastercard, visa n paypal were among the first to announce as partners, in the project. their main goal is to fight fraud and you can't defraud a blockchain network. say goodbye to carding, bitches.

Another example refers to the reliability and security of Facebook's coin, noting it can be 'reversed in case of scammers':

> facebook coin, its private, fast, pegged to USD, can be reversed in case of scammers, and conTROLLed by the genius john cuckerzorg

One post mentions the potential vulnerabilities that Facebook's cryptocurrency will introduce to the security of Facebook's servers:

> there are always new features, that INEVITABLY introduce new vulnerabilities. Facebook Libra, if not stopped by Congress and other gov branches, will be one

Another post talks about scams and payment fraud:

> Facebook has even been a target itself, with fake pages on its social networks claiming to be official entities for the upcoming Libra cryptocurrency. They offered the chance for consumers to supposedly buy the stablecoin at heavily discounted prices, even though they hadn't been launched.

These posts indicate that at least some forum participants are thinking about the various opportunities for crime Facebook's cryptocurrency might enable. They also provide a snapshot of the ideas and topics debated in underground forums related to new cryptocurrencies. However, these conversations do not generate–at this early stage–a significant amount of discussion or curiosity about novel crime opportunities.

6. Conclusion

Research into underground forums has provided insight into the tools used and pathways taken by those who commit cybercrime. Underground forums serve a number of crucial functions in the cybercrime economy. However, as confirmed by this work's analysis and results, the majority of the discussions in these forums are not of an illegal nature. Our results indicate that conversations between individuals involved in cybercrime can start in underground forums, but then move into private messaging and other platforms. We argue that these actors potentially come back to underground forums to launder proceeds using digital currencies.

We built a crime type classifier using statistical models. To the best of our knowledge, this specific type of classifier has not been used before to categorise posts in HackForums. Although the crime type categories are not exhaustive and not necessarily mutually exclusive, they provide valuable insights into the types of crimes being discussed, and allow us to segregate topics and better understand the types of activities being discussed, specifically by those actors offering to exchange currencies. The XGBoost model showed the best performance without over-fitting the data whilst outperforming the two baselines. The inclusion of bulletin board title and thread title in the feature extraction, as well as post content, provided a significant improvement to the performance of all the models trained and tested.

We explored the relationships between currencies advertised for exchange (classified using Portnoff et al.'s [6] model) and crime type by categorising the posts made by those advertising currency exchange elsewhere in the forum. We observed that the most advertised currencies have been Bitcoin, PayPal and Liberty Reserve, consistent with previous research [5], [6]. Pastrana et al. [12]'s findings that eWhoring users prefer PayPal and Amazon gift cards was also supported by this research. Crucially, the cybercrime economy appears as dependent on PayPal as it is on cryptocurrencies. Rather than the explicit 'cashing out' mechanism which is often assumed, in which cryptocurrencies gained through crime flow into state-backed fiat currencies, we instead see a circulation of capital between different forms, as cash is held and then cashed back and forward according to movements in the wider (legitimate) currency market.

Finally, the qualitative analysis focused on discussions around Facebook Diem, confirms this proposed cryptocurrency has been a topic for discussion on underground forums, even though it has not yet been officially released. The analysis of posts extracted for this work showed that themes discussed in these posts were centered around technical details, investment or potentially related to cybercrime.

7. Acknowledgments

As part of the open-review model followed on WACCO this year, all the reviews for this paper are publicly available at https://github.com/wacco-workshop/WACCO.

We thank the Cambridge Cybercrime Centre for access to the CrimeBB dataset. We also thank Sergio Pastrana, Andrew Caines, Anh Viet Vu, Jack Hughes, Richard Clayton, Ross Anderson, and our other colleagues at the Cambridge Cybercrime Centre for advice and support. Additionally, we thank Michael Green for his support during the annotation process.

This work is supported by the Engineering and Physical Sciences Research Council (EPSRC) [grant number EP/V026178/1] and the Economic and Social Research Council (ESRC) [grant number ES/T008466/1].

References

[1] M. Motoyama, D. McCoy, K. Levchenko, S. Savage, and G. Voelker, "An analysis of underground forums," in *Proceedings of the 2011 ACM Internet Measurement Conference*, ser. IMC '11. ACM, 2011, pp. 71–80.

[2] S. Pastrana, A. Hutchings, A. Caines, and P. Buttery, "Characterizing Eve: Analysing cybercrime actors in a large underground forum," in *Proceedings of the 21st International Symposium on Research in Attacks, Intrusions and Defenses (RAID)*, 2018.

[3] A. Hutchings and T. Holt, "Cybercrime trajectories: An integrated theory of initiation, maintenance, and desistance," *Crime online: Correlates, causes, and context*, pp. 117–140, 2016.

[4] Bromium, "Cybercriminals launder up to $200 billion annually research reveals," 2018. [Online]. Available: https://www.bromium.com/press-release/cybercriminals-launder-up-to-200-billion-annually-research-reveals/

[5] S. Pastrana, D. R. Thomas, A. Hutchings, and R. Clayton, "CrimeBB: Enabling cybercrime research on underground forums at scale," in *Proceedings of the 2018 World Wide Web Conference*, ser. WWW '18, 2018, p. 1845–1854.

[6] R. Portnoff, S. Afroz, G. Durrett, J. Kummerfeld, T. Berg-Kirkpatrick, D. McCoy, K. Levchenko, and V. Paxson, "Tools for automated analysis of cybercriminal markets," in *Proceedings of the 26th World Wide Web Conference*, 2017, pp. 657–666.

[7] A. Caines, S. Pastrana, A. Hutchings, and P. Buttery, "Automatically identifying the function and intent of posts in underground forums," *Crime Science*, vol. 7, no. 1, pp. 1–14, 2018.

[8] BBC, "Liberty reserve digital money service forced offline," 2013. [Online]. Available: https://www.bbc.co.uk/news/technology-22680297

[9] M. Yip, N. Shadbolt, and C. Webber, "Structural analysis of online criminal social networks," in *Proceedings of the IEEE International Conference on Intelligence and Security Informatics*, 2012, pp. 60–65.

[10] E. R. Leukfeldt, E. R. Kleemans, and W. P. Stol, "Cybercriminal networks, social ties and online forums: Social ties versus digital ties within phishing and malware networks," *British Journal of Criminology*, vol. 57, no. 3, pp. 704–722, 2017.

[11] A. Mikhaylov and R. Frank, "Cards, money and two hacking forums: An analysis of online money laundering schemes," in *2016 European Intelligence and Security Informatics Conference (EISIC)*. IEEE, 2016, pp. 80–83.

[12] S. Pastrana, A. Hutchings, D. Thomas, and J. Tapiador, "Measuring eWhoring," in *Proceedings of the ACM Internet Measurement Conference*, ser. IMC '19, 2019, p. 463–477.

[13] M. Karami, Y. Park, and D. McCoy, "Stress testing the booters: Understanding and undermining the business of DDoS services," in *Proceedings of the 25th International Conference on World Wide Web*, 2016, pp. 1033–1043.

[14] R. Bhalerao, M. Aliapoulios, I. Shumailov, S. Afroz, and D. McCoy, "Mapping the underground: Supervised discovery of cybercrime supply chains," in *2019 APWG Symposium on Electronic Crime Research (eCrime)*, 2019, pp. 1–16.

[15] A. V. Vu, J. Hughes, I. Pete, B. Collier, Y. T. Chua, I. Shumailov, and A. Hutchings, "Turning up the dial: the evolution of a cybercrime market through set-up, stable, and COVID-19 eras," in *Proceedings of the 2020 ACM Internet Measurement Conference*, ser. IMC '20. ACM, 2020.

[16] British Society of Criminology, "Statement of ethics," 2015. [Online]. Available: https://www.britsoccrim.org/ethics/

[17] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and Psychological Measurement*, vol. 20, no. 1, pp. 37–46, 1960.

[18] ——, "Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit," *Psychological Bulletin*, vol. 70, no. 4, pp. 213–220, 1968.

[19] J. L. Fleiss, "Measuring nominal scale agreement among many raters," *Psychological Bulletin*, vol. 76, no. 5, pp. 378–382, 1971.

[20] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, vol. 33, no. 1, p. 159, 1977.

[21] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, p. 321–357, Jun 2002.

[22] K. Sparck Jones, "A statistical interpretation of term specificity and its application in retrieval," *Journal of documentation*, vol. 28, no. 1, pp. 11–21, 1972.

[23] I. Webster, "Bitcoin historical prices," 2020. [Online]. Available: https://www.officialdata.org/bitcoin-price