Abstract—In this paper, the notion of anomaly detection is introduced for the first time in the area of darknet markets (DNMs). Our hypothesis is that like popular social media platforms DNMs also exhibit anomalous behaviour. However, we propose that the meaning of anomalies in DNMs differs from social media anomalies. The social media anomalies are a cause of threat to the real world, while DNM anomalies are caused by threats from the real world. We present an unsupervised learning method developed to detect anomalies. The model is based on a weighted sum of a feature set trained through an evolutionary algorithm. Our approach successfully identifies anomalies in 35 DNMs – both at the community level and at the level of its user types. Our analysis shows that most of the anomalies found align with well-known adverse events—either as a direct consequence or as a cascading effect of the root event. Moreover, the model identified additional anomalies, which we were able to link to other events through post hoc analysis. Furthermore, we show that the adverse event of market shutdown generates a two-pronged impact on the ecosystem, i.e., it not only triggers startups of new markets but it also inflict anomalies to current markets which may become fatal in some cases. We conclude that this two-pronged impact can be exploited by law enforcement agencies to produce maximum disruption in DNMs.

Index Terms—Darknet Markets, Anomaly Detection, Adverse Events, Unsupervised Learning, Evolutionary Algorithm.

1. Introduction

Understanding the dark web ecosystem is critically important to cyber security and cyber safety aims, as it typically houses a variety of illegal services such as the distribution of child abuse media, hacking, drug trafficking and weapons proliferation [6, 21]. How one may intervene appropriately and effectively in this ecosystem remains a research area in its infancy. While law enforcement organisations have developed and employed interventions [13], the effectiveness of these interventions remains unknown, with mixed consequences [21]. Positive impacts include decreasing the number of illegal sites and identifying key offenders, which subsequently reduces the benefits of offenders to continue with this type of crime. However, there is a large issue regarding the displacement of users and vendors after site take-downs or seizures — known as the “whack-a-mole” problem — for example, after the seizure of Silk Road (SR), Silk Road 2.0 (SR2.0) was launched within a month [14]. It is not unusual for marketplaces, websites, or even botnets that have been seized or taken down to reappear promptly [9, 21]. For example, the authors of [9] described the Armenian police force arresting the hacker that controlled the Bredolab botnet. However, the effect of this was minimal, with servers reactivated and running two days after the initial seizure. Similarly, work by [8] found that the number of active vendors across various Darknet Markets (DNMs) dropped substantially after Operation Onymous. However, within a month the number of active vendors had almost returned to pre-Onymous levels. This demonstrates the necessity for research into mechanisms that can help understand and model the impact of such actions on DNMs in order to effectively measure such impacts, leading to more impactful interventions. The work presented here is motivated by this necessity.

DNMs exhibit abnormal behaviour when they are disturbed by adverse events. Some examples of adverse events are: the actions taken by law enforcement agencies against DNMs, such as the Silk Road Shutdown [29], including actions taken against its management, such as the arrest of Silk Road founder Ross Ulbricht [31] and its users (e.g., Silk Road Drug dealer Cornelis Jan “Maikel” Slomp) [24], internal fights among DNMs through rumouring, DDoS attacks or hacking (e.g., the compromise
of Silk Road 2 Escrow Accounts [20]) or any negative news about DNMs in the media, such as the Gawker Blog publication about Silk Road [5]. The adverse event directed against one forum can affect the behaviour of other fora too. It is important to understand this effect as this could help law enforcement agencies to design their actions in a way to bring about maximum disruption to darknet markets. Hence, for the purpose of this study, an anomaly represents the effect of an adverse event on the behaviour of DNM users, and its detection is the estimation or measurement of this effect.

In this paper, we introduce our anomaly detection approach not only to identify the adverse events which may or may not be known a priori but also to measure their impact on the activity of the DNM users. The approach is applied on a substantial number of datasets i.e., 35 DNM communities containing over 150,000 users [18]. More specifically, our key contributions are as follows:

- We present an unsupervised learning based anomaly detection approach, which trains a weighted sum model of the selected feature set by minimisation of the standard error of estimate against the data points. Hence, the resulting model is able to identify an anomaly not only for the whole community, but also for its user types.
- Our approach enables an automated analysis of the cascading impact of disruptive events, such as a site shutdown, on the darknet ecosystem. Our results show a two-pronged impact generating not only startups of new markets but also inducing anomalies in the existing markets, which even lead to market shutdown in some cases. Such cascading impacts can be modelled and exploited by law enforcement to bring about maximum disruption to DNMs.

To our knowledge, this is the first study that investigates the feasibility of applying anomaly detection to identify the disruptive or adverse events and model and analyse their impact on DNMs. Such an anomaly detection approach provides a bottom-up focus to detect not only the immediate impact of an adverse event, but also the cascading effects, enabling a better understanding of the consequences of such events both in space, i.e., spawning or shutdown of DNMs and time, i.e., longer term impacts on the survival of markets and engagement of their users in ongoing activities.

This paper is structured as follows. We provide an overview of the related work in Section 2. The anomaly detection approach is discussed in Section 3. We describe our experiments and results in Section 4. Finally, we conclude this study in Section 5.

2. Background and Related Work

We discuss related work in two domains: anomaly detection (section 2.1) and impact of adverse events in DNMs (section 2.2).

2.1. Anomaly Detection

The term anomaly detection has been used in network intrusion detection and also in social media. Below we describe each of these separately. Later, we will emphasise why anomalies in DNMs are conceptually different from these areas and why they require a different approach.

- The anomalies considered in the area of network and cyber security are Virus, Worm, Trojan, Denial of Service (DOS), Network Attack, Physical Attack, Password Attack, Information Gathering Attack, User to Root (U2R) Attack, Remote to Local Attack (R2L), and Prob [2]. To detect these anomalies three types of systems can be deployed i.e., Misuse based, Anomaly based and Hybrid. However, Anomaly based is most popular nowadays. Network anomalies are not only security based but also performance based. Focusing on security anomalies only, these can be classified as point anomalies, collective anomalies and contexual anomalies. Various methods have been developed to identify these anomalies. These methods can be classified into three broad categories: genetic algorithm GA-based [43], Artificial Neural Network [1] and Artificial Immune System (AIS-based) [7].
- The anomalies considered in the area of Social Media are bullying, terrorist attack planning, dissemination of mis- and disinformation, hoax and rumour spreading, etc. If a single individual is involved in these acts, it is categorised as point anomaly. If several individuals are involved, then it is categorised as group anomaly [50]. The techniques developed to detect these anomalies are broadly categorized as behavior based, structure based and spectral based [26]. The example of behavior based anomaly detection is content based filtering [47]. The structure based anomaly detection consists of link mining [30]. The spectral based techniques explore the spectral graph space by different measures such as eigenvalues [49].

In this study, we extend the notion of anomaly detection to DNMs. There are very few examples where anomaly detection methods are used on darknet data. Those detection methods are aimed at identifying the threats the darknet is posing to legal communities, such as detection of distributed scan attacks [15], identification of hacker threats [40] and exploring hacker assets [39]. However, we take the perspective that the DNMs posing a threat to legal communities is their “normal” behaviour, rather than anomalous behaviour. To our best knowledge, this is the first exploration of anomaly detection with regard to anomalous behaviour of DNM users in response to events considered a threat to DNMs themselves and not the opposite, which is in contrast with previous social media and network intrusion anomaly detection approaches. Table 1 contrasts the anomaly detection in DNMs and other platforms.

Anomalies in DNMs are reactive, i.e., a change in the behaviour of DNM users in reaction to events in the real world that are considered harmful to DNMs by the users. On the contrary, anomalies in other platforms are proactive, i.e., a change in the behaviour of the users aimed at generating events that are harmful to the real world. DNM anomalies are mostly activity based, while anomalies in other platforms are of several kinds, i.e.,
behavioural, structural, spectral and performance based. Only behavioural anomalies in social media can be partly (quantitative) activity based as in the case of DNMs. However, again those activities are based on some features that are exclusive to each side, i.e., the use of coding and hacking terms exclusive to DNMs, and login/logoff records, HTTP access records and file access records exclusive to social media. Due to these contrasts, anomaly detection models used in other areas cannot simply be transferred to DNMs.

2.2. Impact of Adverse Events in Darknet Markets

The context of adverse events in cryptomarkets has been studied mostly through the lens of evaluating the effectiveness of law enforcement actions. For example, the authors of [8] studied the long-term impact of one of the largest law enforcement actions taken against cryptomarkets, Operation Onymous, and found that the overall impact of the operation was limited in scope, with the underground marketplaces adapting and recovering from the shutdowns within one to two months. This is a developing area, with some contrasting findings. Work by [4], using a difference-in-difference analysis across three forums, suggested that arrests do have a dampening effect on trade which is not fully accounted for by migration to other markets.

Recent work compared law enforcement action to other forms of adverse event affecting underground markets. In the context of high-risk opioids such as fentanyl, work by [3] studied the effect of law enforcement actions alongside voluntary market closures and exit scams, stating that law enforcement actions can have a greater impact on opioid availability than other closures, primarily by encouraging self-regulation among surviving markets.

In most cases, prior work has focused on this problem by working from known adverse events to examine their impact on an outcome measure of choice. This study, however, presents a bottom-up approach by exploring the feasibility of using anomaly detection techniques to identify when adverse events have occurred – both known and unknown – and are causing upheaval that may be worth further study or a response from law enforcement/researchers.

3. Approach

The approach for anomaly detection used in network intrusion detection and social media is that a reference model to represent normal data is developed and then new observations are tested against that model. The new observations are considered “anomalous” if they deviate from the reference model beyond the threshold line. In the case of DNMs, no reference model is available. The models developed for the social media platforms [22] cannot be used here, for the reasons discussed in section 2.1. In this section, we describe our approach to developing a model on a set of DNM data.

3.1. Data

The data consists of darknet datasets and adverse events.

3.1.1. The DNM Dataset. For this analysis, we make use of over 2.5 million posts drawn from over 150,000 users from 35 cybercriminal communities, drawn from the DNM Corpus: a large dataset collected between 2013 and 2015 [18]. All the DNMs have English language as their main medium of communication. In particular, we targeted discussion fora within this collection, which acted as support areas for underground marketplaces dealing in a number of different illicit goods. Table 2 gives a breakdown of the data available for each community. Communities ranged from successfully established markets with thousands of users (though not all were always active posters) to small sites that never moved beyond a handful of initial users.

The raw data provided in [18] captures fora as scraped at several semi-regular intervals by the dataset curators. This leads to heavy redundancy within the data, as threads may be captured at multiple times. However, this redundancy is also useful, as it helps to guard against intermittent faults in the crawling process. Our approach to parsing the data takes a latest-version-first view – of all pages captured within the crawling process, we treat as canonical the most recent version, only parsing older pages where they were not captured in later scrapes. We note that capturing pages from older scrapes is an important step in handling this data, as many thousands of threads and user profile pages are not present at all in the most recent scrapes of each forum. Differences could be attributed to crawling failures in later scrapes, incomplete coverage as part of the crawling processes, or to administrator action in taking down or hiding discussion threads over time.

Parsing of the data proceeded in two stages within the scrape history of each community. First, user profile pages were processed to build a dataset of users and associated information from their profile pages (e.g., PGP public keys, membership status). Next, discussion thread pages were parsed in order to associate posts (including textual content and metadata such as posting time, subforum, etc.) with the user that authored them. Where quotations of other users could be identified within the text of a user’s post, these quotations were separated from the authored text, to avoid contamination of profiling analysis. It sometimes occurred that user profile pages were not
We observed two kinds of user types in the darknet forums: (i) group user types, i.e., one title assigned to several members and (ii) individual user types, i.e., a title exclusive for one member only. A user type can be an individual user type in one market and a group user type in another market. Every member is assigned one title, therefore there is no overlap of members in user types. It should be noted that, in some cases, we found quite atypical attributes of individual user types. We consider these anomalies as significant because they are outcome of individual action rather than the outcome of any adverse events. However, anomalies in group user types are mostly group anomalies and are given serious consideration in our analysis.

### 3.1. Adverse Events

As a baseline for evaluating our approach, we calibrated outcome measures against the impact of the adverse events (E2-E5) in the form of known law enforcement interventions shown in Table 4. Moreover, Table 5 provides a list of additional minor adverse events (E1, E6-E9) identified through a manual Internet search around the dates where unaccounted group anomalies were found. Unaccounted group anomalies refer to those anomalies that were flagged by our approach, but which could not be immediately linked to baseline adverse events in Table 4.

#### 3.1.2. Adverse Events

The approach analyses the anomalies based on unit time of one calendar month. Since calendar months are not equal, and in order to maintain equal sample sizes, the feature values are normalised for a 30-day month. We chose a sample period of a calendar month because smaller samples are not consistent in their amount of activity and therefore are bound to generate false positive anomalies. As we explained in Section 3.1, we collected data for each DNM in two files: one file representing activity of users and another file representing their meta-data. This helped us to group the data for each user type and at the community level. All features discussed below are countable, so their extraction did not warrant any additional technical difficulties. For each sample time, we extracted the following feature types:

- **Coding Terms**: This is a count of coding terms used in messages. The coding terms considered here are listed in [19]. The use of coding terms reflect the technical ability of users which comprises a very important part of user activity.
- **Hacking Terms**: This is a count of hacking terms found in messages. The hacking terms considered here are listed in [45]. Hacking is a very important service provided by darknets. Use of hacking terms reflects major activity in this area.
- **Attachments**: This is a count of attachments to the messages. This feature is used in DNMs to explain technical things which cannot be described via short messages (see [39]).
- **Quotations**: This is a count of quotations included in users’ messages. Quotations are typically used to let other DNM users know the context of the messages. This feature is also used in network analysis [37].
- **Number of posts**: This is an account of the number of messages the users post to interact with other users of the community.
- **Number of threads**: A thread represents a conversation covering a group of messages posted to discuss a particular question or statement. This feature is also used for network analysis [51].
- **Number of Active Days**: A forum is considered to have an active day if at least one message is posted to it on that day. The maximum number of

### Table 2. Breakdown of the communities targeted for this study.

<table>
<thead>
<tr>
<th>Community</th>
<th>Posts</th>
<th>Users</th>
</tr>
</thead>
<tbody>
<tr>
<td>Silk Road 2</td>
<td>882,418</td>
<td>26,163</td>
</tr>
<tr>
<td>Silk Road</td>
<td>846,077</td>
<td>52,383</td>
</tr>
<tr>
<td>Evolution</td>
<td>509,225</td>
<td>33,743</td>
</tr>
<tr>
<td>Abraxas</td>
<td>276,300</td>
<td>1,607</td>
</tr>
<tr>
<td>Agora</td>
<td>84,914</td>
<td>6,153</td>
</tr>
<tr>
<td>Black Market Reloaded</td>
<td>80,467</td>
<td>7,006</td>
</tr>
<tr>
<td>Nucleus</td>
<td>65,175</td>
<td>9,478</td>
</tr>
<tr>
<td>The Hub</td>
<td>58,642</td>
<td>7,357</td>
</tr>
<tr>
<td>Panda</td>
<td>49,023</td>
<td>8,729</td>
</tr>
<tr>
<td>Black Bank</td>
<td>32,817</td>
<td>2,381</td>
</tr>
<tr>
<td>The Majestic Garden</td>
<td>26,121</td>
<td>1,510</td>
</tr>
<tr>
<td>Utopia</td>
<td>14,548</td>
<td>4,392</td>
</tr>
<tr>
<td>Diabolus</td>
<td>11,456</td>
<td>2,151</td>
</tr>
<tr>
<td>Kingdom</td>
<td>10,285</td>
<td>856</td>
</tr>
<tr>
<td>Project Black Flag</td>
<td>6,131</td>
<td>330</td>
</tr>
<tr>
<td>Cannabis Road2</td>
<td>5,842</td>
<td>2,139</td>
</tr>
<tr>
<td>Cannabis Road3</td>
<td>4,905</td>
<td>1,903</td>
</tr>
<tr>
<td>Bungee54</td>
<td>3,325</td>
<td>1,510</td>
</tr>
<tr>
<td>Panacea</td>
<td>2,241</td>
<td>520</td>
</tr>
<tr>
<td>Tor Bazaar</td>
<td>2,205</td>
<td>902</td>
</tr>
<tr>
<td>The Real Deal</td>
<td>1,049</td>
<td>115</td>
</tr>
<tr>
<td>Hydra</td>
<td>937</td>
<td>276</td>
</tr>
<tr>
<td>Kiss</td>
<td>933</td>
<td>145</td>
</tr>
<tr>
<td>Andromeda</td>
<td>894</td>
<td>1,601</td>
</tr>
<tr>
<td>Outlaw Market</td>
<td>689</td>
<td>2,007</td>
</tr>
<tr>
<td>Revolver</td>
<td>660</td>
<td>85</td>
</tr>
<tr>
<td>Tor Escrow</td>
<td>490</td>
<td>294</td>
</tr>
<tr>
<td>Dark Bay</td>
<td>332</td>
<td>484</td>
</tr>
<tr>
<td>Doge Road</td>
<td>300</td>
<td>118</td>
</tr>
<tr>
<td>Darknet Heroes</td>
<td>190</td>
<td>793</td>
</tr>
<tr>
<td>Havana</td>
<td>181</td>
<td>77</td>
</tr>
<tr>
<td>Tom</td>
<td>144</td>
<td>4,120</td>
</tr>
<tr>
<td>Grey Road</td>
<td>43</td>
<td>24</td>
</tr>
<tr>
<td>Tortuga</td>
<td>37</td>
<td>7</td>
</tr>
<tr>
<td>Mr Nice Guy</td>
<td>25</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 2: Breakdown of the communities targeted for this study.
### TABLE 3. USER TYPES IN DNM FORUMS

<table>
<thead>
<tr>
<th>Forum</th>
<th>Group User Type</th>
<th>Individual User Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hack hound</td>
<td>Advanced, Advanced Member, Banned, Beginner, Expert, Intermediate Member, Member, Newbie, Titleless¹</td>
<td>[curtailed] Intelligence Service, Intermediate, RDG Soft products, Retarded, Suspended, Ulb3moob</td>
</tr>
<tr>
<td>Cracking fire</td>
<td>Active Member, Banned, Cracking Team, Cracking Crew, Ex Staff, Guest, New Member, Titleless¹ Very New Member, Well-Known Member</td>
<td>Android Codex – Retired, CF: Cracker, cracking is my life, Fucker, GEX Expert, Hacking Team Br, Legacy, Moderator, Redyooh approved!, Retired, Risk is My Business, Super Master, VIP Member</td>
</tr>
<tr>
<td>The Hub</td>
<td>Brand Spankin New, Full Contributor, Hero Contributor, Jr. Contributor, Newbie, New Contributor, Senior Contributor, Titleless¹</td>
<td>Ful Member, Hero Member</td>
</tr>
<tr>
<td>Evolution</td>
<td>Administrator, Banned, Forum Moderator, Market Moderator, Member, Moderator, Public Relations, Troll, Vendor</td>
<td>Guest, Resident Medical Expert</td>
</tr>
<tr>
<td>Nucleus</td>
<td>Administrator, Banned, Member, Moderator</td>
<td>Guest, Scammer</td>
</tr>
<tr>
<td>Black Market Reloaded</td>
<td>!!!!!!Scammer!!!!!!, Administrator, Banned, BMR Vendor, Hero Member, Jr. Member, Member, Moderator, New Member, Sr. Member</td>
<td>Unregistered¹</td>
</tr>
<tr>
<td>Bungee54</td>
<td>Administrator, Junior Member, Member, Newbie, Titleless¹</td>
<td>Bastard Administrator, Bungee54 Team, Customer Support One, Customer Support Two, Moderator, Senior Member, Your worst nightmare</td>
</tr>
<tr>
<td>Abraxas</td>
<td>Administrator, Full Member, Hero Member, Newbie, Sr. Member, Vendor</td>
<td>Jr. Member</td>
</tr>
<tr>
<td>Black Bank</td>
<td>Administrator, Banned, Member, Newbie, Vendor</td>
<td>Moderator</td>
</tr>
<tr>
<td>Diabolus</td>
<td>Freedom Fighter, Global Moderator, Newbie, Silk Road Vendor, Titleless¹ We rise from the ash</td>
<td>Administrator, Destiny will guide me</td>
</tr>
<tr>
<td>Pandora</td>
<td>Administrator, Full Member, Hero Member, Junior Member, Newbie, New Newbie, Pandora Support, Sr. Member, Titleless¹ Vendor, We rise from the ash</td>
<td>Unregistered¹</td>
</tr>
<tr>
<td>Ism</td>
<td>Newbie</td>
<td>Administrator, Global Moderator</td>
</tr>
<tr>
<td>Utopia</td>
<td>Administrator, Banned, Member, Moderator, Vendor</td>
<td>SR Moderator</td>
</tr>
<tr>
<td>The Darknet Heroes</td>
<td>Daemon, Heroes, Newbie, Root</td>
<td>Member</td>
</tr>
<tr>
<td>Havana</td>
<td>Administrator, Member, Titleless, Vendor</td>
<td>Banned</td>
</tr>
</tbody>
</table>

¹ These users do not have any title.
² This name is curtailed because it was very long.
³ Not a member.

### TABLE 4. LIST OF KNOWN INTERVENTIONS BY LAW ENFORCEMENT

<table>
<thead>
<tr>
<th>Adverse Event</th>
<th>Date of Event</th>
<th>Agency Involved</th>
<th>Event Breakdown</th>
</tr>
</thead>
<tbody>
<tr>
<td>E2: Silk Road Shutdown</td>
<td>October 2013</td>
<td>FBI</td>
<td>Seizure of US$3.6 million of funds in escrow. Arrest of the founder and chief operator of the site, Ross Ulbricht (undercover name Dread Pirate Roberts) [29].</td>
</tr>
<tr>
<td>E3: Arrests of Silk Road Ad-</td>
<td>December 20, 2013</td>
<td>FBI</td>
<td>Arrest of three admins – Two were working in Silk Road 2.0. The new Dread Pirate Roberts surrendered control of the site. Defcon took over the site and promised to bring it back to working order [31].</td>
</tr>
<tr>
<td>mins</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E4: Silk Road 2.0 escrow ac-</td>
<td>February 13, 2014</td>
<td>Silk Road Rivals</td>
<td>Bitcoin in escrow accounts worth US$2.7 m, reported stolen [31].</td>
</tr>
<tr>
<td>counts compromised</td>
<td></td>
<td>Dutch police</td>
<td>Launch and closure of Utopia. Servers located in Germany. Five people were arrested [16].</td>
</tr>
<tr>
<td>E4: Operation Commodore</td>
<td>February 2014</td>
<td>Europol’s EU3, FBI, ICE, HIS, Eurojust</td>
<td>410 hidden services including Silk Road 2.0 servers taken down. 17 vendors and administrators arrested. US$1 m Bitcoin, EUR 180K cash, drugs, narcotics, gold and silver seized [12].</td>
</tr>
<tr>
<td>E5: Operation Onymous</td>
<td>November 2014</td>
<td>Europol’s EU3, FBI, ICE, HIS, Eurojust</td>
<td></td>
</tr>
</tbody>
</table>
TABLE 5. LIST OF MINOR ADVERSE EVENTS

<table>
<thead>
<tr>
<th>Adverse Event</th>
<th>Date of Event</th>
<th>Agency Involved</th>
<th>Event Breakdown</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1: Silk Road Notoriety</td>
<td>June 2011</td>
<td>Gawker Blog</td>
<td>Publication of article in GAWKER blog on 1st of June 2011</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>US Senetor Charles Schumer asked federal authorities to shut down the market on 5th of June 2011.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Publication of article in The Sydney Morning Herald on 12th of June 2011.</td>
</tr>
<tr>
<td>E6: Conviction of Silk Road Founder</td>
<td>February 2015</td>
<td>Federal Court Manhattan.</td>
<td>Ross Ulbricht got 30 years in prison Sentence.</td>
</tr>
<tr>
<td>E7: Arrest of two Federal Agents</td>
<td>March 2015</td>
<td>US District Court Northern California</td>
<td>Carl Mark and Staun Bridges were arrested for working as informants of Ulbricht.</td>
</tr>
<tr>
<td>E8: Conviction of Silk Road Drug Dealer</td>
<td>May 2015</td>
<td>Federal Court Chicago</td>
<td>Cornelis Jan “Maikel” Slomp sentenced to 10 years in prison.</td>
</tr>
<tr>
<td>E9: Ulbricht appeal denied</td>
<td>May 2017</td>
<td>US Court of Appeals for the 2nd Circuit</td>
<td>Ulbricht appeal was declined.</td>
</tr>
</tbody>
</table>

- Number of Active Users: This feature is measured in different ways based on the degree of activeness of individual users for network analysis [51] and in the identification of roles [44]. However, since we are not performing any network analysis, we do not measure degree of activeness of users. We simply consider users as active users if they post a message during a sample period.

- Number of Memberships: Membership of the forum is a mandatory requirement for any user of the forum. This feature is used by the forum administration for self-regulation [48]. The number of memberships is the count of new members registered during a sample period.

- Average Message Length: This is an average number of characters in a message. This feature was extracted to measure users’ engagement with the forum. Characters rather than words are used as units of length because words are unequal and messages are not properly punctuated in darknet datasets.

3.3. Model Description

The anomaly detection approach offers a model that can be used for training purposes in order to detect anomalies in a darknet forum. The model is characterised by feature values, and consists of their weighted sum. Weighted sum is a useful model, which can be used for classification [36], clustering [42], and regression [33] – herein used for regression with regard to prediction of the effect of adverse events for anomaly detection by exploiting the differences along the time series in activity of DNMs. The weighted sum model was selected, because it is the most widely used standard linear representative model for vast variety of real-world applications, such as mechanics [28], business [10], Chemistry [23], Decision Support Systems [32], and statistics [25].

The model is trained by optimising the weights of a feature set through minimisation of a standard error of the model obtained through linear regression. This is done through the application of an evolutionary algorithm. Evolutionary algorithms are a useful optimisation procedure and have been used previously to optimise weighted sum in other applications [38]. This algorithm has proven its efficiency to optimise weighted sum models, for example, in crop planning [41], environment and economy [27], and engineering [34]. Fig. 1 depicts a flowchart of the overall procedure.

The learning process starts by taking data from the forum. The input data includes a dataset, which consists of two files, one file representing the activity of users and another file representing users’ metadata. The input also includes the configuration parameters, which consist of model type, feature types, sample size, threshold limit and output options. The input also includes other info such as a list of coding terms [19] and a list of hacking terms [45]. These lists are used to extract the terms present in the user messages. The input data is then processed, i.e., feature set samples are collected from the two files representing the dataset. The data is then used in the evolutionary algorithm (EA) which generates a random population of solutions, which means a random generation of weights of features. The weights are uniformly generated between 1.0 and 2.0. There was no upper limit on weights as far as evolutionary process is concerned. They could be evolved up to any limit. However, their lower limit was fixed to 1.0. This is because the evolutionary process was tailored to minimise the standard error. If weights were allowed
to go below the limit of 1.0, all the weights would have
evolved to 0.0 ending up in the model exactly over the x-
axis with 0.0 standard error. Hence, controlling the lower
limit of weights was essential. The EA then applies linear
regression to each individual solution to compute slope, y-
intercept and standard error. Next, the EA reproduces the
next generation by applying the reproduction procedure
on the most promising individuals – the solutions with
comparatively lesser standard error among the individuals
within the population. The reproduction procedure cons-
ists of the application of genetic operators, i.e., mutations
and crossovers, over the promising solutions selected from
the population. Linear regression is applied again on the
new generation to produce the next generation. This iter-
ative procedure continues until a termination condition is
met. The termination condition is met when there is no
improvement (no further minimisation of standard error)
in any of the solutions within a population for a certain
number of generations. This is a very efficient method to
train the weights of the feature set. This trained set of
feature weights represents a linear model of user activity
of the forum. Any sample activity that deviates from this
model, beyond a threshold limit, can be categorised as
anomalous activity.

The activity of a sample is considered anomalous if
it surpasses the threshold limit. The threshold limit is
measured in terms of standard error units. If the weighted
sum of a feature set which represents activity of the month
is above two standard units away from the model estimate,
then the sample value is beyond the threshold limit of
the model value and it is considered anomalous. The
threshold limit of two standard error units was chosen
after experimentation with different DNMs, where we
observed distance of data points from the model estimate
against well-known adverse events. We observed that at
two standard error units, the model yielded minimum false
positive outliers and maximum true positive outliers. It
should be noted that minor adverse events 5 were not part
of this analysis. For the darker forum with fewer samples
standard error is normally large. Hence, anomalies are
rarely found in such communities.

It should be noted that point anomalies are caused
by the action or inaction of only one individual (see
Section 3.1 and [50]). Therefore, they are not considered
a consequence of any adverse event. However, group
anomalies are caused by the collective action or inaction
of individuals in a group, therefore they are considered a
consequence of the adverse events. Hence, in experiments
it is expected that point anomalies may not align with the
dates of the adverse events in most of the cases. However,
group anomalies should. Some exceptions can be allowed
due to unexpected anomalies that can happen due to events
not directly relevant to DNMs. The proposed model is
designed to comply with this criterion, where fulfillment
of this test can be considered a success of the model.

4. Experiments and Results

Our approach was applied on 35 DNM communities
along with their several user types (see Section 3.1), which
resulted in the identification of the 10 adverse events
labeled (E1-E9) in Tables 4 and 5. There are nine labels
for 10 events because two events have the same label (E4)
due to their occurrence in the same sample period. The
approach is based on unsupervised learning as described
in Section 3.3.

In our first attempt, we tried different feature com-
binations and different weighted sum models, such as a
model based on relative weighted sum as compared to past
average. This model has been applied in different appli-
cation areas such as text mining [46]. All these attempts
were aimed at finding anomalies against baseline adverse
events. However, we found that no specific combination
of features and parameters could be achieved on the wide
spectrum of datasets. Therefore, we decided to use all the
features described in section 3.2 and their absolute
weighted sum as a model to represent the amount of
activity. We also decided to look for other adverse events
which might be affecting the activities of users causing
anomalies elsewhere. Furthermore, we investigated po-
tential cascading effects, i.e., shutdowns and startups of
DNMs triggering anomalies in other forums.

Figure 2 lists anomalies detected in all 35 commu-
nities, along with the 10 adverse events E1-E9. Events
E2-E5 are baseline adverse events listed in Table 4 while
the rest of the events are minor adverse events listed in
Table 5. Two types of anomaly are represented: anom-
alous behaviour of the whole community is represented
by the larger filled circles, while anomalous behaviour
of one or more defined subgroups is represented by
smaller points. The green circles represent anomalies that
are directly aligned with either known or minor adverse
events, whereas red circles represent anomalies not di-
rectly aligned with any adverse event. However, several of
those are found aligned to some adverse event indirectly
through a cascading effect. The startup and shutdown
of a market place is represented through arrows <—
> and the start and end point of data availability are
represented through a vertical bar |. The dates of startups
and shutdowns of market places were taken from [17].
The data available is not always in accordance with the
startup and shutdown of a marketplace. This is because, in
such cases, a forum starts before the start of a marketplace
and ends after the end of a marketplace.

As can be seen in Figure 2, our model has detected
anomalies that are aligned with the both major and minor
adverse events (E1-E9). It can be seen that most of the
anomalies are found in the context of major adverse
events (E2-E5), while minor adverse events have lesser
effect. This can be observed in the graph against minor
adverse events (E6-E9), where anomalies are represented
by the smaller circle. This shows minor events only cause
anomalies in user types rather than in whole communities.
It is also interesting to see that lots of startups and
shutdowns are aligned with these adverse events. There
are also several anomalies and startups and shutdowns
that are not exactly aligned with these adverse events, e.g.,
anomalies, startups and shutdowns between E2 and E3.
These can be explained by the delayed effect of adverse
event E2. Similarly, anomalies, startups and shutdowns
between events E3 and E4 can be seen as the delayed
effect of events at E3 and a cascading effect of events
at E2. This reasoning is based on the evidence from the
output of our model.

As an example, Fig. 3 depicts the level of activity
in terms of the weighted sum of the feature set in each
month for the Black Market Reloaded DNM. It shows a considerable increase in activity in October 2013. This aligns with event E2. Fig. 4 demonstrates the distance of activity points from the model in terms of standard error units. It can be seen that activity in October 2013 crosses the threshold line of two standard units. Therefore, this is anomalous activity caused by event E2. Table 6 represents the use of each feature during the anomalous month compared to its long term average. It can be seen that the largest increase is in the use of hacking terms i.e., 718%, followed by new memberships 549%, use of coding terms 515%, number of posts 500%, number of threads 464%, and number of active users 444%. This is an extraordinary increase in activity with the largest portion taken by use of hacking terms escalated by new memberships. This made administrators of Black Market Reloaded suspicious. In November 2013, they announced that the marketplace would be closed soon and advised members to close their crypto currency escrow accounts. Eventually, the market was closed in December 2013. This was a direct effect of event E2.

Additionally, the BMR shutdown has produced anomalies in Hack Hound, Cracking Arena and Silk Road 2.0, as shown in Fig 2. However, the event E3, i.e., the compromise of Silk Road escrow accounts also occurred during this month. Hence, this event may also have con-

tributed to these anomalies. The data inside these anomalies may give some indication of their cause. According to this data, in Cracking Arena Market in December 2013, the number of memberships were increased by 254% compared to long term average. This effect can be attributed to both events, i.e. BMR shutdown and compromise of accounts in Silk Road 2.0. The members from BMR and Silk Road 2.0 may have joined Cracking Arena. In Hack Hound, the anomaly is only found in one user type (small circle). The membership increase here is 132%, which can also be caused by both events. Similarly, in Silk Road 2.0, there is 152% membership increase in its user type Newbie. It stands to reason that this is likely to be an impact of the BMR Shutdown.

As far as Event E3 is concerned it is unlikely that the compromise of escrow accounts could contribute to the increase in membership of the same market where the negative incident had happened. However, none of these anomalies were fatal enough to cause shutdown of these markets. The reason behind this is that Cracking Arena and Hack Hound were not Crypto Currency Markets and Silk Road 2.0 was a freshly opened market, so its administrators did not have any long term past record to judge changes in the data. Unfortunately, we do not have data of a well established crypto currency market during that time that could be tested for anomalies. However, a Europol report [16] shows that there was a well established market 'Buy It Now', at the time of the BMR shutdown, which voluntarily closed a couple of months after the BMR shutdown. Such volunteer shutdowns are a typical consequence of an unusual hype in memberships, as we already highlighted in the example of BMR at the time of Silk Road Shutdown. It is likely that the 'Buy it Now' shutdown is a cascading effect of Silk Road shutdown through anomaly infliction effect reaching 'Buy It Now' via anomalies in BMR.

The second impact of site shutdown, i.e., start up of new markets can also be witnessed from October 2013 (Silk Road Shutdown) and onwards in Fig. 2. Therefore, our analysis shows that the shutdown of any site has a two-pronged effect. It gives birth to new sites and it also produces anomalies in the current sites. The anomalies may cause shutdown of those sites and the shutdowns again, followed by the same two-pronged effect. This results in the chain reaction of startups and shutdowns which can be witnessed in the Fig. 2. Keeping in mind this observation, we have proposed some suggestions in the conclusion section for disruption of the markets by emulating effect of market shutdowns.

### TABLE 6. ANOMALIES IN BLACK MARKET RELOADED (WHOLE COMMUNITY)

<table>
<thead>
<tr>
<th>Feature Name</th>
<th>10/2013</th>
<th>Long Term Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Hacking Terms</td>
<td>425,871</td>
<td>51,8179</td>
</tr>
<tr>
<td>Number of Coding Terms</td>
<td>151,958</td>
<td>26,548,39</td>
</tr>
<tr>
<td>Average Message Length</td>
<td>347.277</td>
<td>335.547</td>
</tr>
<tr>
<td>Number of Posts</td>
<td>18982.3</td>
<td>3165.34</td>
</tr>
<tr>
<td>Number of Active Days</td>
<td>30</td>
<td>28.9457</td>
</tr>
<tr>
<td>Number of Attachments</td>
<td>0</td>
<td>0.0774194</td>
</tr>
<tr>
<td>Number of Threads</td>
<td>2060.32</td>
<td>385.368</td>
</tr>
<tr>
<td>Number of Memberships</td>
<td>180</td>
<td>277.268</td>
</tr>
<tr>
<td>Number of Active Users</td>
<td>2559.08</td>
<td>4.07.34</td>
</tr>
</tbody>
</table>
nity. Where our analysis identified anomalies that did not directly align with the timing of the events, we analysed if they represented cascading effects. We discovered a two-pronged effect of shutdowns which causes alternating startups and shutdowns of forums. This two-pronged effect of shutdown of market can be exploited by law enforcement agencies. While law enforcement agencies can do little about new startups, they can exacerbate anomalous effect of shutdown by introducing fake memberships and use other disruption techniques like rumours, spam and DDOS attacks.

Our current approach is based on user types which depends on titles assigned to users during their registration. These titles sometimes do not reflect the actual roles users play in the market. Our future work will focus on refining the the anomaly detection approach by replacing user types with actual roles of the users which could be identified by utilising social network analysis techniques [35] which are currently under study. Further to this, the approach requires detailed analysis regarding parametric optimisation with specific emphasis on smaller sample sizes to see the impact of adverse events at a micro level, in order to further our understanding of darknet ecosystem.

6. Acknowledgement

This work was supported by the UK Home Office through a National Cyber Security Programme research grant, AMoC: Advanced Modelling of Cyber Criminal Careers. The authors would also like to thank all members of the Advisory Group that have contributed to this project, our shepherd, Sergio Pastrana, and the reviewers for their thoughtful comments and efforts towards improving our manuscript.

References


1. 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


[20] Hackterms. All terms related to coding. URL: www.hackterms.com/about/all.

[21] Martin Horton-Eddison and Matteo Di Cristofaro. “Hard interventions and innovation in crypto-drug...


[39] Weiguo Sheng et al. “A weighted sum validity function for clustering with a hybrid niching ge-


