Mapping the Underground: Supervised Discovery of Cybercrime Supply Chains

Rasika Bhalerao*, Maxwell Aliapoulios*, Ilia Shumailov†, Sadia Afroz‡, Damon McCoy*
* *New York University*, † *University of Cambridge*, ‡ *International Computer Science Institute*
rasikabh@nyu.edu, maliapoulios@nyu.edu, Ilia.Shumailov@cl.cam.ac.uk, sadia@icsi.berkeley.edu, mccoy@nyu.edu,

***Abstract*—Understanding the sequences of processes needed to perform a cybercrime is crucial for effective interventions. However, generating these supply chains currently requires time-consuming manual effort. We propose a method that leverages machine learning and graph-based analysis to efficiently extract supply chains from cybercrime forums. Our supply chain detection algorithm can identify 33% and 42% relevant chains within major English and Russian forums, respectively, showing improvements over the baselines of 11% and 5%, respectively. Our analysis of the supply chains demonstrates underlying connections between products and services that are potentially useful understanding and undermining the illicit activity of these forums. For example, our extracted supply chains illuminate cash out and money laundering techniques and their importance to the functioning of these forums.**

***Index Terms*—Security, Cybercrime, Natural Language Processing**

I. Introduction

Cybercrime-as-a-Service lowers the barrier to entry for new cybercriminals by commoditizing different parts of attacks [24]. For example, without commoditization, a spammer needs to find a way to send e-mails, acquire mailing lists, create storefront websites, contract with web hosting, register domains, manage product fulfillment, accept online payments, and provide customer service. With commoditization, the spammer can outsource different responsibilities to different criminals specialized in one specific task. Cybercriminals often rely on underground cybercrime forums to establish these trade relationships that can facilitate the exchange of illicit goods and services in a supply chain fashion.

The supply chain of a cybercrime can illuminate the sequence of processes involved in the criminal activities. Prior work has shown that analyzing these supply chains can result in identifying weak points that could enable effective interventions [13]. There have been several largely manual studies exploring specific instances of these commoditized cybercrime offerings [15], [23] and how some attacks can be more effectively undermined once their dependencies to other services are understood [12], [14]. Unfortunately, extracting knowledge from cybercrime forums is currently a largely manual task. Analysts and researchers use ad hoc keyword based searching to investigate cybercrime forums to understand the product types bought and sold and then manually investigate each user on the forum to identify their expertise and connections. Machine learning has been used to automate some analysis of cybercrime forums, such as identifying products that are bought and sold [26], however, using it to discover the trade relationship between products has not been explored yet.

In this paper, we propose an approach to systematically identify relevant supply chains from cybercrime forums. Our approach classifies the product category from a forum post, identifies the replies indicating that a user bought or sold the product, then builds an interaction graph and uses a graph traversal algorithm to discover links between related product buying and subsequent selling posts. Our approach builds upon prior work on product detection [26], adding specific product classification and the supply chain algorithm.

We used our end-to-end supply chain identification pipeline to analyze two publicly available cybercrime forums. Our pipeline is able to identify 33% and 42% relevant links in our English language and Russian language forums, respectively. This is an increase from our baselines of 11% and 5%, respectively. We show how our derived supply chains can give macro information about a cybercrimal forum which is useful for research studies and can facilitate targeted analyst investigation. Please contact the authors for all of our annotations, code, data, and models, which we publicly release to enable full reproducibility.

Our analysis of the supply chains showed that currency exchange was a central activity that appeared as part of 73% and 81% of validated chains discovered on the English and Russian forums, respectively. These supply chains might enable us to better understand cash out and money laundering techniques utilized on these forums. We also discovered supply chain links where users are buying products that are likely used to facilitate subsequent product offerings (i.e., a user buying OSN reputation boosting services to groom accounts that are then sold to scammers) or users reselling products after they are no longer useful to their original owner.

The main contributions of our paper are the following:

* We develop a supervised approach for discovering cybercrime supply chains (Section IV). Our method uses natural language processing methods and graph traversal to systematically discover supply chains.
* We perform an analysis of our discovered supply chains to provide an understanding of how some commodity cybercrime products depend on other offerings within these forums (Section VI).

978-1-7281-6383-3/19/$31.00 ©2019 IEEE

* We distill our findings from the detected supply chains into several qualitative case studies (Section VII). These case studies highlight how we were efficiently able to discover supply chains exposing the connection between the purchasing of hack-for-hire services and the selling of valuable online accounts. Despite this connection being present in the forums for years, it has only recently been discovered based on manual analysis [8]. We also were able to gain new insights into cash out and money laundering methods and the central role they play on these forums.

The rest of this paper is structured in the following way. Section II discusses background and related work in this area, including past work which uses the same data. Section III outlines the data used to validate our approach. Section IV outlines our approach and contributions in classification and supply chain discovery. We evaluate our work empirically to demonstrate how our approach performs better than the baseline in Section V, and analyze the forums using our results in Section VI. Section VII outlines several real world scenarios where the generated supply chains add value to an investigation. Section VIII identifies limitations and discusses the implication of our results. We conclude in Section IX.

II. Background and Related Work

Cybercrime forums provide a unique opportunity to understand how criminal markets operate. A typical forum structure follows the `subforum > thread > post > reply` hierarchy. A subforum typically pertains to a particular subject: for example, `marketplace` or `introductions`. Users can create threads within subforums where a thread is a collection of messages. Each thread will always contain a first post, which in this paper, we will use interchangeably with `product post`. We will use the term `reply post` to refer to the posts which come after the product post in each thread.

Several prior works studied the organization of the cybercrime forums [20]–[22], [25], profiling key actors [32], products traded [5], [7], [10], evolution over time [4], [19], and ways to disrupt their business [3], [32]. However, these works either rely on the structural information on a forum or use handcrafted regular expression.

Portnoff et al. [26] demonstrated that supervised machine learning techniques can be used to automatically identify the type of a post (buy or sell), products being traded and the price of the products. Furthermore, they demonstrated that machine learning methods perform better in terms of both recall and precision over previously used keyword searching (grep). For reference, the F_1 score for their grep baseline is 0.61. For comparison, F_1 scores for our product classifiers are 0.77 for Antichat and 0.71 for Hack Forums; these metrics are further discussed in Section V. Our approach builds on their work by categorizing posts into meaningful categories, instead of identifying the exact the word representing the product. This better suits our method by providing semantic meaning to our supply chains. Caines et al. [29] recently explored classifying replies by intent, which is similar to how we classify replies as indicating buying or selling activity. Unlike their approach to classifying replies, we focus on identifying replies that strongly indicate that the forum member has actually bought or sold that product, which is a key building block for discovering likely supply chains.

Other work has explored the overall progression of illicit activities by forum members [31]. Wegberg et al. [33] analyzed longitudinal data from eight structured online anonymous marketplaces over six years to understand the value change and commoditization of the criminal markets as "cybercrime-as-a-service". Our work complements this line of research on structured forums and extends it by providing a method for detecting business-to-business transactions within unstructured forums.

Our approach goes beyond understanding the trust establishment, organization, aggregate activity, and classification performed in prior work. We use the results of our classifiers to identify semantically meaningful forum interactions and automatically discover supply chains of the products that can improve our understanding of how these markets function in practice. We analyzed the connections between products in unstructured cybercrime forums and noticed mostly business-to-business transactions[1]. This allows us to study the criminal-to-criminal supply chains that enable attacks. Some of these were previously studied from the direct attackers and victims' perspectives, such as romance scams [16]. Our new understanding of the underlying supply chains can illuminate different and potentially more effective methods of undermining these threats [11].

	Total threads	Total messages	Date range
Antichat	51,119	328,216	05/2003 - 06/2010
Hack Forums	17,298	263,832	04/2009 - 04/2015

Table I: Forum overviews including only commerce related parts of the forums.

III. Forums

To evaluate our approach, we chose two popular forums: Antichat (Russian) and Hack Forums (English) (Table I). We chose these forums because they are large, publicly available, and have been used to evaluate cybercrime analysis methodologies in prior studies [26], [28]. As was done in these prior studies, we limit our analysis to the commerce related parts of the forums. In this paper we chose to limit our evaluation to public datasets so as to enable reproducibility of our findings.

a) Hack Forums: Hack Forums is a major English-language forum covering many cybercrime-related topics. The forum has been active since 2007. We use a partial 6 year scrape between April 2009 and April 2015. The scrape is partial because it only includes the commerce related posts from the Hack Forums Marketplace.

[1]We scope this to mean "sale to the trade" where the products being bought and sold often have no value except as a building block to enable an attack

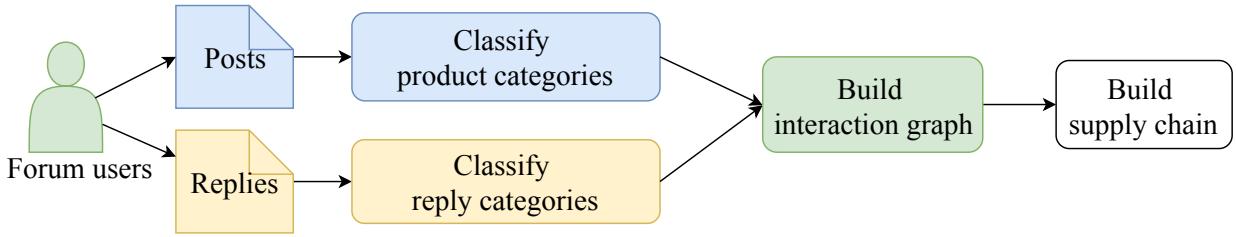


Figure 1: Our supply chain detection approach.

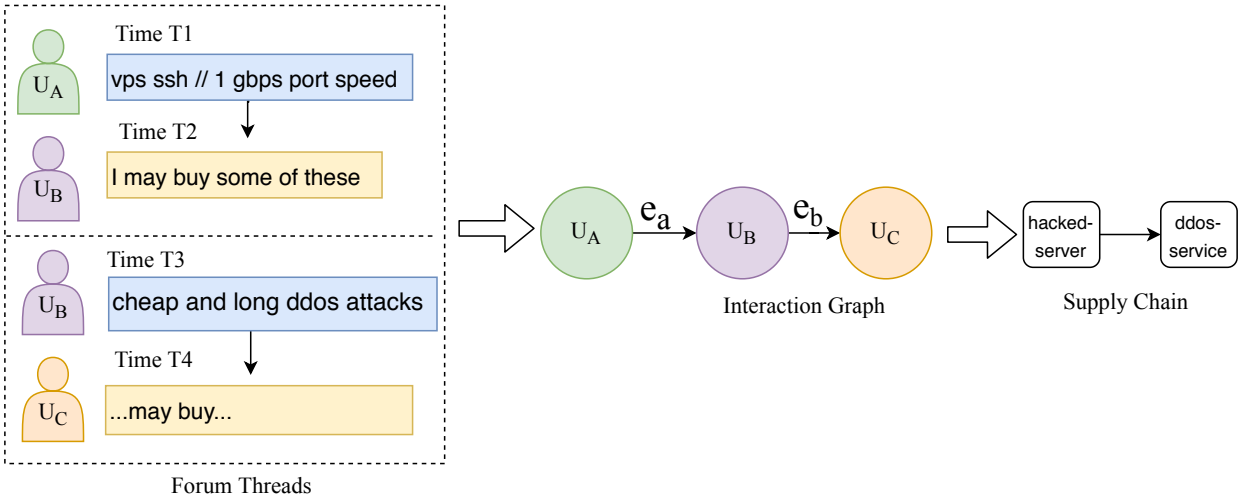


Figure 2: Example of constructing supply chains from two Hack Forums threads.

b) Antichat: Antichat is a major Russian-language forum covering cybercrime-related topics. Examples include password cracking, stolen accounts, and physical weaponry. We obtain a full database leak containing posts between January 2005 and June 2010.

A. Availability and Ethics

Both of the forums are publicly available. We did not attempt to analyze any Personally Identifiable Information (PII) in these datasets. All case studies mentioned in section VII are anonymized except for those that have been previously publicly reported, and so are any analysts mentioned. This study was exempted by our Institutional Review Board (IRB) since the datasets were publicly available and the focus of our study was not on analyzing PII in the dataset. Our methods conform to recommended best practices for ethical research pertaining to datasets of illicit origin [30]. Please contact the authors for all of the data, annotations, models and code for our study, which are publicly released in order to facilitate reproducibility.

IV. Approach

Our goal is to automatically build a supply chain for a criminal forum by analyzing the posts and replies. To build a supply chain, we need a chronological record of which products the users of a forum bought and sold. We use classifiers to categorize the posts and replies, then build an interaction graph and use the graph to build the supply chain (Figure 1).

A. Classify Products

To find a supply chain, we need to first identify the categories of the product bought or sold in forum posts. We use supervised classification to classify products into n categories and an "other" category for products that should be filtered since they are not within the scope of analysis. The specific categories and what is not within scope might depend on the forum and the analyst. Product classification performs two important functions: 1) filter products outside the scope of interest to an analyst and 2) provide semantic meaning (i.e., what products are being bought and sold) to the chains we will identify. The details of our classifier will be presented in section V.

B. Classify Replies

Once we have the product categories, we need to identify who bought or sold the products to build a supply chain. The reply classifier is similar to the product classifier. It is a supervised classifier that uses TF-IDF of character n-grams to classify replies into three categories: `buy`, `sell` and `other`. To classify only relevant replies, we run a quote removal algorithm to remove instances of a reply quoting a previous reply before feeding it to the reply classifier.

C. Build an Interaction Graph

To determine who bought what and when, we build an interaction graph for a forum. This is a directed graph, $G = (U, E)$ where each node $u \in U$ is a user who posts on the forum, and each edge $(u_a, u_b) \in E$ indicates that user u_a sold a product to user u_b.

To build this graph, we use the product and reply classifiers. For a post by u_a, the product classifier determines the category of the product sold in the post and the reply classifier determines whether user u_b's reply to u_a's post implies buying the product. We also consider the time of the buy reply as the time at which the user u_b purchased the item from user u_a.

D. Build a Supply Chain

The purpose of a supply chain graph is to illuminate the sequence of processes involved in various criminal activities. The supply chain graph of a forum is a directed graph, $S = (C, I)$, where each node, $c_i \in C$, is a product category and each edge, $(c_a, c_b) \in I$, indicates that at least one user in the forum bought a product of category c_a and sold another product of category c_b. We use breadth-first search on the interaction graph to create the supply chain. Figure 2 shows an example of creating a supply chain from two Hack Forums threads.

To create the supply chain graph, we define a supply chain link in an interaction graph as a tuple of two interactions e_a and e_b, where e_a is an edge from user u_a to user u_b, and e_b is an edge from u_b to u_c. This means that user u_a sold a product to user u_b, who then sold a product to user u_c. Our breadth-first search (Algorithm 1) follows the supply chain links in the interaction graph in chronological order to build the full supply chain.

When adding links to the supply chain graph, we do not want users who are outliers disproportionately buying or selling certain items to unfairly add to their links. For example, somebody might buy 100 items and then sell once, adding 100 links (an exaggeration of the problem). We mitigate this issue by dividing the weight that each user contributes to each link by the total number of links to which that user contributes. Our method of attenuation is: If a user appears in n edges, then that user adds $1/n$ to each edge in which they appear; so, each user adds a total of 1 to the entire graph, and each link still appears, but contributes less than if that user only created one or a few links with a single purchase. These weights after attenuation are used later in section VI. For example, if user A sold one product to user B and then user B sold 50 products to other buyers this would be attenuated to only a single supply chain.

Supply Chain Algorithm 1: Modified Breadth-First Search for Supply Chain Generation

Input: Interaction graph, $G = (U, E)$ where $u \in U \leftarrow$ user and $(u_a, u_b) \in E \leftarrow$ user u_a sold to user u_b. All u are undiscovered.
Output: Supply chain graph, $S = (C, I)$, where $c_i \in C \leftarrow$ product category and $(c_a, c_b) \in I \leftarrow$ users bought a c_a product and sold a c_b product

```
while not every user u ∈ U has been discovered do
    L1 ← undiscovered user u ∈ U
    while L1 is not empty do
        L2 ← empty list
        for each user u_i ∈ L1 do
            for each undiscovered user u_j who sold to u_i do
                W ⇐ number of items u_j bought and then sold
                for each undiscovered user u_k who sold to u_j do
                    if W > 0 then
                        (c_a, c_b) ⇐ supply chain link between (u_i, u_j) and (u_j, u_k), divided by W
                        I ⇐ (c_a, c_b) ⋃ I
                        L2 ⇐ u_j ⋃ L2
        L1 ⇐ L2
```

V. Evaluation

To evaluate our approach, we first had domain experts manually label posts from our Antichat and Hack Forums datasets into product categories and replies into reply categories. Using our labeled data as ground truth, we evaluate the performance of the classifiers and end-to-end supply chain link algorithm.

A. Labeling Ground Truth

We perform two types of manual annotation: Product labeling and Reply labeling. For product labeling, we assign each post to one of the predetermined categories. For reply labeling, we determine whether each reply indicated buying, selling, or neither.

The posts and replies were annotated by domain experts who have native fluency of the forum's primary language. For Hack Forums, disagreements in annotation were settled through discussion one by one for the final annotations. Agreement between annotators was very strong for product posts and moderate for replies, as measured by Cohen's kappa. For Antichat, we only had access to one native Russian speaker domain expert to perform all of the annotation, so we were not able to compute an agreement score; when our Antichat annotator was unsure, we discussed the translated version of the post to agree on the correct label.

For product labeling, we identified 14 product categories for our datasets (Table II). These categories were determined by domain experts based on reading posts in both forums and choosing products of interest in line with their analysis goals. To adapt our classifiers to other forums, an analyst can modify the categories to fit the forum. An alternative way of choosing product categories might be to explore unsupervised clustering methods [17], [18].

We annotated all 17,298 Hack Forums posts and a random sample of 21,996 Antichat posts. We annotated posts into a

single product category or other when it did not fit any category. The distributions of product annotations for both forums are shown in Table II.

We also annotated each reply into one of three categories: buy, sell or other. The distribution of reply types is highly dependant on the structure and rules of the forum as shown in Table III. For example, there are not many sell replies on Hack Forums, since Hack Forums is used as a marketplace and consists mainly of original posters selling products and repliers purchasing. Replies labelled as other tend to be questions about products and informational.

We selected a random sample of 6,150 Hack Forums replies and 9,993 Antichat replies which we annotated into buy, sell, or other. The distributions of reply annotations for each forum are shown in Table III.

In both forums, we acknowledge the possible ambiguity in category annotations, and we mitigate it in two ways. 1) We use a larger number of annotations: the assumption is that incorrect annotations will be incorrect in different ways while correct annotations will be correct in the same way, so an increased number of annotated posts will make a classifier more correct for each category, on average. 2) We favor precision: to have a high precision, the classifier will need multiple examples of a post of that type being annotated into that category in order to put the post in that category, which means that the same annotation mistake would have to be made several times in order for a post to be incorrectly classified.

B. Validating Product Category and Reply Classifier

Our first validation task is to determine which supervised classifier algorithms are the most precise for the product and reply classification tasks. We tested four classifiers: 1) FastText, 2) Logistic Regression, 3) Support Vector Machine (SVM), and 4) Gradient Boosted Trees (XGBoost). FastText is a sentence classification method by Facebook AI Research that uses word embeddings and a hierarchical classifier [1]. XGBoost is a scalable tree boosting system implemented in Python [2]. For Logistic Regression, SVM, and XGBoost, we extract features from posts using TF-IDF to produce a vector where each element corresponds to that term's TF-IDF score. We chose character n-grams over word n-grams based on performance on our datasets. We tuned each of these methods to reduce over-fitting of the labeled data. FastText is used as implemented by Facebook AI Research, with no tuning. Logistic Regression is used with multinomial loss over all categories (as opposed to one-versus-all). SVM is used with hinge loss and L2 penalty. XGBoost is used with a maximum tree depth of 2. Logistic Regression, SVM, and XGBoost are all used with class weights inversely proportional to the frequency of each class. We selected these four classifiers as a diverse set of classifier types but we did not test an exhaustive set of classification methods.

Both annotated datasets are highly unbalanced, making the classification task particularly hard. Unlike Hack Forums, however, our initial random Antichat data-sample had more account posts than other. The ramification of this is that when the classifier struggles to decide how to classify data points it tends to put it into the account class. Since we decided to prioritize precision for all of the classes except other, we decided to undersample account to be smaller than other in our training sets. The effect of this is an increase in precision of the account class but a decrease in recall. The numbers we report for Antichat in Table II are after undersampling the account class so that it is smaller than the other class. This undersampled data was only used for training our models. The natural distribution was used for constructing testing sets and analysis.

For our product category classifier evaluation, we used the labeled data described in Table II and classified each post into one of 14 categories or other. We performed stratified 5-fold cross validation of each classification algorithm since our classes are highly imbalanced. For highly imbalanced datasets, regular k-fold cross validation often produces a biased evaluation because the limited number of folds generated can have a class distribution that does not match the one in the actual data. We use stratified k-fold validation since it ensures an "apples-to-apples" evaluation where the same distribution of the target values that exist in the main data set are maintained for each fold [27].

In order to select the classifier used in our analysis, we use a weighted average of the precision scores across all the categories except other. We ignore the other class in this metric, since posts classified as other will be filtered out by our supply chain identification algorithms, and we prioritize having posts with the correct products of interest (i.e., precision). Prioritizing precision over all categories except other provides more certainty that the posts used in our analysis truly belong in the category output for them. We choose Logistic Regression for product classification tasks because on average, it performs better in our weighted non-other precision scoring metric, providing a weighted precision of 0.837 on Antichat and 0.714 on Hack Forums, yielding 0.778 on average. We see the performance when considering weighted precision in Table IV.

The second classification task used for identifying supply chains is to categorize each reply to the first post in each thread into one of three categories: buy, sell, or other. We chose the categories based on the requirements of the supply chain algorithm. For our reply category classifier evaluation, we used the labeled data described in Table III. We again perform stratified 5-fold cross validation of each classification algorithm since our classes are highly imbalanced.

As seen in Table IV, by our weighted non-other precision metric, FastText performed the best on average, providing 0.871 weighted precision on Antichat, 0.834 precision on Hack Forums, and 0.855 on average. Although we care about overall model performance, we want to ensure the precision of our supply chain links. So, we decided to trade slightly lower recall for improved precision of the classifications in the models used in our analysis. Please contact the authors for our annotations, code, testing and training sets, and models, which are publicly released so that others can fully reproduce our results.

Product	Description	Antichat	Hack Forums
Account	Selling or requesting an account, multiple accounts, or access codes. This also includes account creation automation software.	31%	16%
Botnet	Selling or renting access to computers infected with malicious software.	1%	1%
Crypter	A piece of software which obfuscates malware.	2%	4%
Currency exchange	Exchanging one form of currency for another.	7%	16%
DDoS service	Selling or requesting a DDoS attack.	1%	4%
Hacked server	Selling or requesting a single hacked server.	16%	1%
Hack-for-hire	Offering targeted hacking, malware coding or requesting a specific service.	4%	6%
Hosting	Hosting a website, game server, or otherwise maintaining it.This includes DDoS mitigation.	3%	3%
Malware	A piece of malicious software that is executed on a victim's machine. Examples of this include cryptocurrency miners and ransomware.	6%	7%
Proxy	Selling or requesting a proxy/VPN.	3%	1%
Social booster	Supports gaining social media attention. Examples of this are, "buying likes/views", "selling twitter followers".	2%	2%
Spam tool	Selling or requesting an email/chat service spam tool or spamming service.	7%	1%
Traffic	Selling real or fake visitors to a site. Does not include social media related "traffic".	6%	1%
Video game service	Selling or requesting any service related to video games. Includes things like mods, points, and power-leveling. Does not include selling video game accounts.	1%	8%
Other	Anything that doesn't fall into the previous categories.	10%	29%
TOTAL		21,996	17,298

Table II: Product annotation labels and distribution per source

Reply Type	Description	Antichat	Hackforums
Buy	Someone wants to buy or bought a product.	17%	19%
Sell	Someone making a sale offer to the original poster of a thread.	8%	2%
Other	Anything that didn't fall into the previous categories.	75%	79%
TOTAL		9,993	6,150

Table III: Reply annotation labels and distribution per source

	Antichat								Hack Forums							
	Product				*Reply*				*Product*				*Reply*			
Model	Prec	Prec*	Recall	F_1	Prec	Prec*	Recall	F_1	Prec	Prec*	Recall	F_1	Prec	Prec*	Recall	F_1
FastText	0.78	0.80	0.77	0.77	**0.87**	**0.87**	**0.87**	**0.86**	0.72	**0.75**	**0.72**	0.71	0.85	0.86	0.85	0.85
Logistic Regression	**0.82**	**0.84**	**0.80**	**0.81**	0.86	0.86	0.85	0.85	**0.74**	0.71	0.71	**0.72**	0.85	0.83	0.84	0.85
SVM	0.79	0.78	0.73	0.74	0.85	0.85	0.85	0.82	0.71	0.63	0.64	0.65	**0.86**	**0.87**	**0.86**	**0.86**
XGBoost	0.77	0.80	0.75	0.75	0.85	0.85	0.81	0.83	0.72	0.70	0.70	0.70	0.83	0.78	0.80	0.83

Table IV: Precision, recall and F_1 scores of classifiers across datasets and tasks, with stratified k-fold cross-validation. Prec* is the precision over all categories excluding `other`. We use Prec* to choose the classifier, since we prioritize correctness of products used in our supply chains, and we exclude `other` in our supply chains.

C. Validating Supply Chain Links

We validate whether the chain link tuples output by our algorithm describe "true" links. A "true" link is what we would consider a link in a supply chain, not where a user purchased something and then sold something else unrelated to the item they purchased, or an error in classification resulting in the lack of a purchase or even a product. For example, a user purchasing a program that adds followers to any Instagram account, and then subsequently selling an Instagram account with many followers is a "true" link. To evaluate this, we manually check the links produced by the algorithm.

a) Attenuation of supply chain links: When adding links to our supply chain alluvial graphs and to the counts in Table V, we use attenuation to prevent a few outlier users from disproportionately affecting the distribution by buying or selling multiple times. We mitigate this issue with our method of attenuation: the amount that each link contributes to its respective edge (in the supply chain alluvial graph) and relevance class (in the counts in Table V) is divided by the total number of links that the linking user creates.

Table V shows the attenuated counts and percentages of `related`, `resell`, and `unrelated` links. Table V also displays attenuated counts and percentages of links with products that should have been classified as other (`Lack of product`), or links where a reply was not a `buy` (`Lack of purchase`). Out of these, links classified as `related` and `resell` are considered true supply chain links. Agreement among Hack Forums link validators was moderate for both the algorithm output and sample baseline, as measured by Fleiss' kappa. Antichat links were annotated by a single annotator.

Link Type	**Hack Forums**		**Antichat**	
Link Type	Algorithm Output	Sample Baseline	Sample Algorithm Output	Sample Baseline
Related	48 (30%)	9 (10%)	52 (24%)	2 (2%)
Resell	4 (2%)	1 (1%)	38 (18%)	3 (3%)
Unrelated	104 (64%)	16 (20%)	31 (14%)	9 (10%)
Lack of product	2 (1%)	1 (1%)	53 (25%)	27 (31%)
Lack of purchase	5 (3%)	58 (68%)	40 (19%)	46 (53%)
TOTAL	163	84	213	86

Table V: Attenuated Link Truth Level by Forum. "Algorithm Output" provides attenuated counts for links detected by our complete method; the validation is performed for all 352 links (163 after attenuation) for Hack Forums and a random sample of 300 links (213 after attenuation) for Antichat. "Sample Baseline" are based on a random sample of 100 (84 from Hack Forumsand 86 from Antichat after attenuation) detected by only the Supply Chain Algorithm 1 without using the reply classifier to filter links (i.e., without limiting links to "buy" replies). `Related` links have a user who purchased a product and then sold another product likely using the previous one in a supply chain fashion. `Resell` links have a user who purchased and resold the same product, possibly with a different description. `Unrelated` links have a user who purchased a product and sold another product that does not logically result from the source product. `Lack of product` happens when the link does not involve a product. `Lack of purchase` happens when the link reply was not a "buy" reply. Our method of attenuation counters outliers. Antichat links were annotated by a single annotator. Agreement among Hack Forums link annotators was moderate on both the algorithm output and sample baseline.

Our algorithm outputs 30% related and 2% resell in Hack Forums and 24% related and 18% resell in Antichat.

"Algorithm Output" shows the attenuated counts and associated percentages of links detected by our complete method that fell into each relevance level; the validation is performed for all the links for Hack Forums and a random sample of 300 links out of the 17,402 total links (before attenuation) for Antichat. "Sample Baseline" shows the same values for a random sample of 100 links detected by only our supply chain algorithm (the modified breadth-first search), without using the results of our reply classifiers to filter links (i.e., without limiting links to "buy" replies).

To determine if using our classifiers increases the density of valid links by disproportionately filtering out invalid links, we manually validated all of the links we discovered after filtering. For a baseline comparison, we used the supply chain algorithm to find links, but without filtering using the results of the reply classifier. We then manually evaluated a random sample of 100 of these unfiltered supply chains (note the total after attenuation is 84 for Hack Forums and 86 for Antichat). The results of our baseline supply chain evaluation are in Table V as "Sample Baseline." Considering `related` and `resell` links as relevant, we find that our classifiers improve the rate of relevant links from 11% to 33% for Hack Forums and from 5% to 42% for Antichat. It is infeasible to compute the recall since there is no existing efficient method for identifying all relevant links in our datasets.

VI. Analysis

We performed time series analyses of the classified and annotated product posts using the taxonomy from Figure 3. The dataset analyzed is the one described in Table I where all of the 17,298 Hack Forums posts and 21,996 of the Antichat posts were manually annotated; the rest were classified using an XGBoost classifier trained on the annotated posts.

The Hack Forums product task is unique because we were able to annotate all the posts. In order to reclassify them, we performed k-fold classification where all of the annotated product posts were split into 5 folds. We then classified all of the posts in one fold using a model trained on posts from the remaining four folds and repeated this until all of the posts in the five folds were classified. We classified the Hack Forums posts using this method so that we could provide a realistic end-to-end assessment of our supply chain detection method that included the likely misclassification error, assuming it is not possible to label all of the posts.

A. Product Analysis

We demonstrate how a product-level trend analysis gives insight into what activity is present on a forum and how forums change over time. Normally this requires manually reading through hundreds of posts to get a sense of changes.

The interesting similarity between the two forums is that as volume increases we see that the product offerings become more diverse. This indicates that these forums evolve into ecosystems where specialized and likely more efficient sellers start to organize into supply chains where one seller of a higher level service, such as DDoS attacks, depends on hacked servers supplied by other sellers. Similar to normal business ecosystems, this likely enables increasingly efficient and sophisticated attacks to emerge. These product category trend analyses only show what is being sold but they do not illuminate the connections between products.

B. Supply Chain Analysis

The data used for our supply chain analysis is as follows: we had 51,119 posts for Antichat and 17,298 for Hack Forums. Out of these, only posts with products classified into categories outside `other` were used for supply chain analysis; this was 42,993 for Antichat and 11,143 for Hack Forums. The total

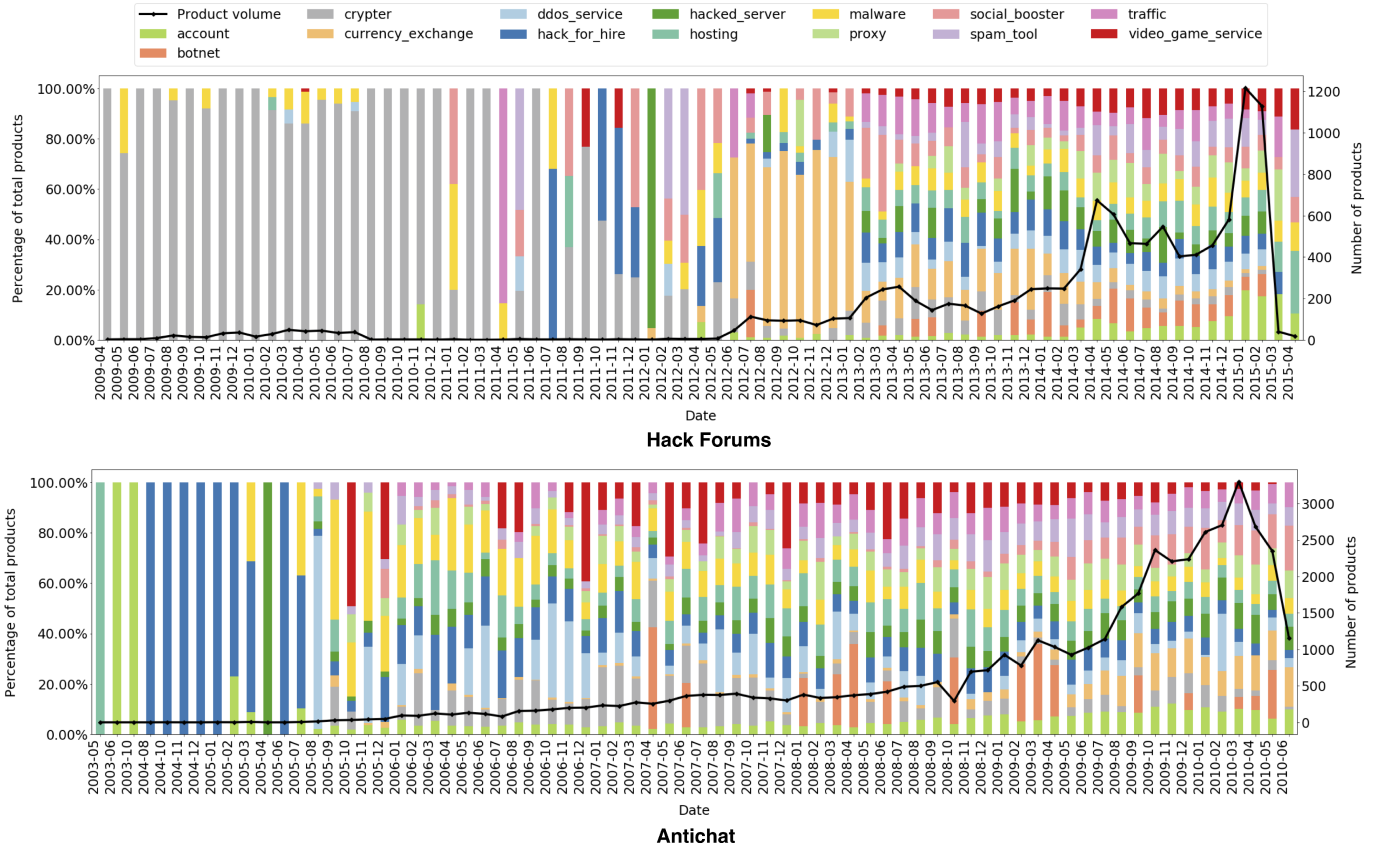


Figure 3: Products trends. The right hand y-axis is the actual volume of posts, and the left y-axis is the percentage of the total number of posts. The black line signifies the volume of post and is associated with the right y-axis.

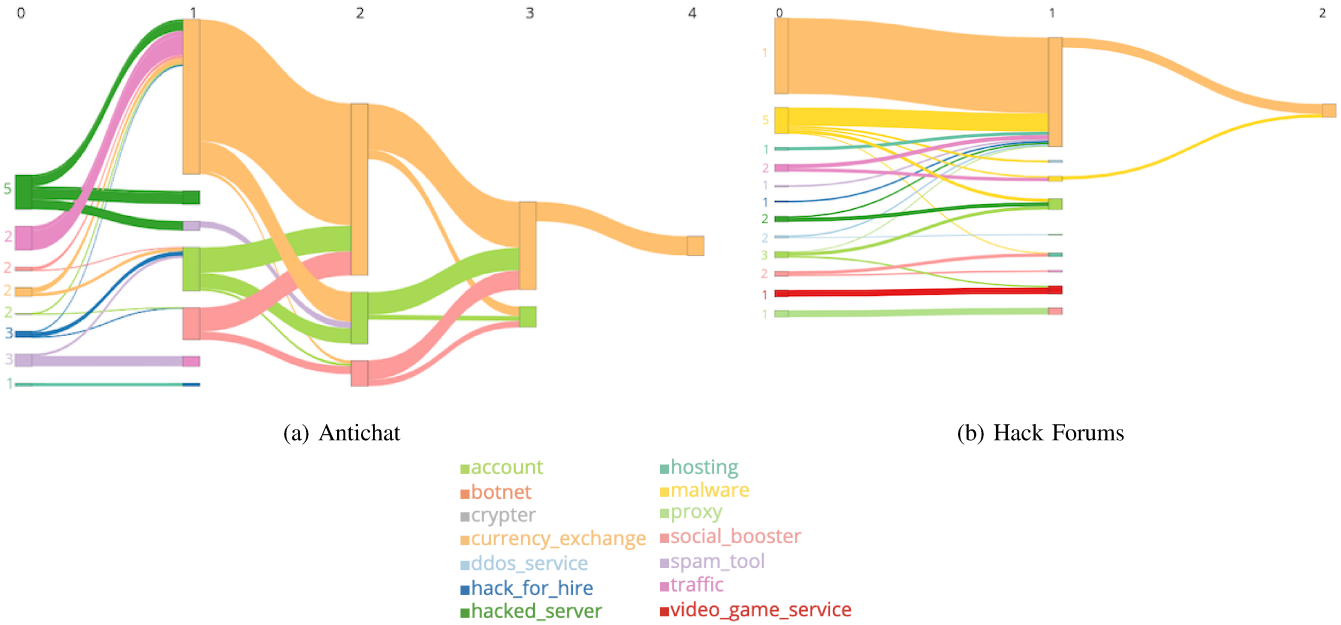
number of links, when allowing posts with `other` products (but still limiting to "buy" replies), is 19,915 links in Antichat and 429 links in Hack Forums and when filtering out `other` product posts is 17,402 Antichat and 352 in Hack Forums. The same links with filtering out `other` products, through attenuation, become 2,216 for Antichat and 163 for Hack Forums.

We produced alluvial graphs of the supply chains resulting from our algorithms only including links manually labeled as `related` or `resell`. Note that for Hack Forums, all links were manually validated, and for Antichat, a sample of 300 links were manually validated, a subset of which are the relevant links depicted in the alluvial graph. Figure 4 depicts only the supply chains we were able to validate as alluvial graphs for Hack Forums and Antichat, respectively. There are 163 links in Hack Forums and 213 links in Antichat, after attenuation.

There are several ways the alluvial plots demonstrate the value of mapping these underground supply chains. At a macro view, they give an understanding of the interaction activity of a forum. Antichat is more interconnected than Hack Forums, as shown by its 17,402 total unvalidated links as compared to Hack Forums's 352 total unvalidated links, indicating the relative sophistication of Antichat users over Hack Forums users. This results in more steps to their supply chains, indicating that interactions are not simply one-off purchases, but instead reflect several actions taken and reliance on different product type industries.

Our supply chains enrich the forum activity with supply chain related metadata to streamline an analyst's process into a micro view. Further, while product trend plots show how types of products change over time, the supply chain based alluvial adds another dimension. We are able to see in Figure 4 instances where an individual in Hack Forums purchases malware and sells DDoS services, hosting, or accounts. This could lead to targeted investigation where an analyst studying account stealing malware could hone in on specific forum posts to understand the facilitators of that activity or even obtain a binary.

These supply chains allow researchers to understand the sophistication of the users operating on a forum and the main products important to an ecosystem. When viewing this through the lens of a potential mitigation, the supply chains give way to an understanding of the potential bottlenecks on a forum. For instance, in both forums, currency exchange is a popular interaction: 49% and 69% of validated Antichat links have it as their source and destination, respectively, and 49% and 73% of validated Hack Forums links have it as



(a) Antichat

(b) Hack Forums

(c) Legend for alluvial graphs

Figure 4: Supply chains limited to links manually validated as "Related" or "Resell" using all Hack Forums links and a random sample of 300 Antichat links (213 after attenuation). Antichat is in Figure 4a, and Hack Forums is in Figure 4b. Edges are colored according to forum's product category, and have widths determined by the number of users who purchased the source product and sold the destination product, with attenuation. Numbers at the top correspond to the level in the modified breadth-first search algorithm at which the node was discovered. The number of chains originating with each product category is denoted next to the start of those chains. Color-coding follows legend in Figure 4c

.

their source and destination, respectively. In the entire forum data, only 7% of Antichat product posts and 16% of Hack Forums posts are related to currency exchange, indicating that currency exchange posts are disproportionately included in supply chain links. The interconnection of currency exchange with almost every other category can be seen in the alluvial graphs of the validated supply chains for both of the forums in Figure 4. This phenomenon is also present in the larger set of potential supply chains from Antichat shown in Appendix C. The alluvial graphs enable researchers to compare and contrast the activity on these forums based on the product relationships as well as understand the overall supply-chain profile of a particular site.

In the case of Antichat, the abundance of currency exchange operations outlined a clear need in the market to move internal currency in and out (more on this in the case studies section). However, it was unclear what allowed attackers to successfully earn and collect currency. The supply chains explained what happened there: the spam tools were used to perform phishing and grow bot networks, stealing currency on the accounts that already had it. The newly acquired accounts were used to further send spam. The chains also showed that people buying currencies were buying quite a few internal applications and we later found that that was the only way to actually pass currency between different users without consuming it. Finally, a lot of currency was earned through the SMS traffic also bought on the forum. To summarize, in Antichat, the supply chains highlighted the an important and integral part of the criminal enterprises. In practice, one should be able to also use those chains to track down the effect of the introduced features on the market and find the required components of a successful product compromise.

VII. Case Studies

In this section, we demonstrate how an analyst can use the supply chains to explore the criminal markets, understand how products are derived, and use popular supply chains to describe an underground forum marketplace. All the following case studies are based on the derived supply chains. These case studies demonstrate the value of extracting supply chains in determining the origins of a specific crime or the use cases for products in our taxonomy.

A. Hack Forums

1) Money laundering: Currency exchange in Hack Forums is present on 81% of all individual supply chains and represents 49% and 73%, respectively, of sources and destinations of supply chain links output by the algorithm. Currency exchange to currency exchange supply chains comprise 53% of all currency exchange related supply chains. An example

of a currency exchange to currency exchange chain might be a scenario where $user_a$ is offering to exchange bitcoin for PayPal and $user_b$ posts a reply indicating they exchanged currency with $user_a$. Then $user_b$ might post a message offering to exchange bitcoin for Webmoney where $user_c$ replies indicating that they complete an exchange. This could simply be a method of profiting by charging currency exchange fees similar to legitimate exchanges. However, this might be an example of $user_b$ money laundering stolen PayPal currency into irreversible cryptocurrency or of $user_a$ cashing out stolen bitcoin into PayPal which can then be converted into fiat currency while avoiding Know Your Customer (KYC) regulations. These supply chains provide a starting point for additional investigation by researchers or analysts.

In 47% of currency exchange related chains on Hack Forums the source was another product, such as 22% malware, 10% traffic, and 7% DDoS service. For these currency exchange related chains, researchers and analysts can often gain some insights into the likely origins of the currency being exchanged. For example, we see an actual example where a Hack Forums member indicated purchasing a cryptocurrency miner, and the purchaser later started selling bitcoin for Paypal. Also of note is that the destination currency exchange for 70% in these cases are users requesting PayPal for some other form of cryptocurrency. This supports the hypothesis that this type of chain might be related to cashing out ill-gotten bitcoin into PayPal which can then be converted into fiat currency while avoiding KYC regulation. These chains would be an interesting starting point for researchers and analysts to further investigate topics such as how money laundering is performed on underground forums or the scale of cryptojacking.

2) Valuable accounts: One common type of business in Hack Forums, made apparent through supply chain link analysis, is selling boosted social accounts. This situation occurs when a user buys a social-booster to boost the "follower" or "like" count of a social media account, and then later sells the account for a premium because of the higher social status. Of the 589 supply chain links extracted from the classified Hack Forums data, 3% were instances of social boosting and selling an account. In an example that appeared in the Hack Forums dataset, a user purchased a service which promised Twitter followers, and later sold a "pre-made" Twitter with 2k followers. Some of these groomed accounts were "eWhore" accounts (they are intended to appear to be owned by either an attractive man or woman), which we discovered are sold to romance scammers. This illuminated a connection between social-booster services and romance scammers which was not mentioned in prior work studying these scams [16].

The value of a social media account is also dependent upon the "rarity" of the handle. Similar to a domain name, a handle is more valuable if it is shorter or if it includes a popular word or phrase. On Hack Forums, these accounts are referred to as "OG," which stands for "original gangster", and 25% of our account links mention this term. Our supply chains depict these "OG" accounts exchanging hands, and 14% of the links where "OG" is mentioned in the destination category come from an account source category. In order to obtain an "OG" account if an actor is not purchasing it directly, they must discover who the owner of the account is so that they can attempt to take over the account using methods such as phishing or SIM swapping attacks. Furthermore, going from an "OG" username to personally identifiable information (PII) can happen through doxing [34]. In our taxonomy, we categorized doxing under hack-for-hire. There is an example link where the source category is an actor advertising a doxing service (hack-for-hire) and the purchaser of the service then sells a stolen "OG" account.

3) DDoS, botnets, and their crypter roots: It is possible to understand the botnet supply chains through supply chain links in Hack Forums. We can see from our product category classifier results that many DDoS and botnet related criminal activities originate on Hack Forums. Suppose there is a cybersecurity analyst interested in what type of crypter a botnet master is using. The analyst could determine this based on the supply chain links, on Hack Forums which indicate what cryptor(s) that operator has purchased. The categories which flow into botnet are malware, proxy and account and the categories which flow into ddos-service are account, hosting, ddos-service, hack-for-hire, traffic, proxy, malware, video-game-service, and crypter. Assuming further that the analyst is interested in the technical aspects of the ddos-service, they may be inclined to discover which crypters were purchased before a service is offered. These crypter to ddos-service chains are rare, and in fact this one type makes up less than 1% of the found chains. Thus, an analyst would have difficulty discovering this specific supply chain manually. With the help of our derived chains, we can see that the specific crypter works via Java drive-by download, which could lead an analyst to further investigate which systems are susceptible to this kind of exploit, infected with botnet malware obfuscated by the crypter, and carrying out DDoS attacks.

B. Antichat

The following analysis is based on the extracted supply chains from Antichat. 89% of validated links (without attenuation) involve vk.com[2] so one of our case studies focuses these links. It appears that much of the observed supply chains are centered around the Russian internal market rather than those outside of Russia.

1) VK.com-related Supply Chains: We consider the supply chains where either the seller or the buyer provided a service on the VK.com platform. Figure 5 depicts these supply chains. By far the most bought and sold asset is currency exchange services. The second largest category are accounts. From what we observed, the currencies and accounts are a crucial part of the underground forum infrastructure, and all of the services bought and sold end up sinking into them in at least one of the supply chain stages.

We find that the larger chunk of those currency exchanges was the exchange of VK.com-internal currency called 'vote'

[2]VK is a popular social networking site in Russia that is similar to Facebook.

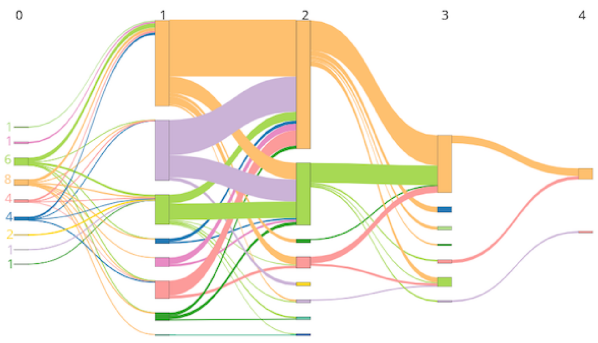


Figure 5: Supply chains limited to links manually validated as relating to VK.com using a sample of 300 Antichat links (213 after attenuation). Edges are colored according to source product category, and have widths determined by the number of users who purchased the source product and sold the destination product, with attenuation. Numbers at the top correspond to the level in the modified breadth-first search algorithm at which the node was discovered. The number of chains originating with each product category is denoted next to the start of those chains. Color-coding follows legend in Figure 4c.

(golos). The authors of the social network did not allow common users to cash out their accumulated currency and the only way to do it was by creating games/apps. When it was possible, the cash out fee was at least 50% + 13% tax of the collected currency [3]. It seems that the lack of an official way to cash out became a lucrative business for both legitimate and illegitimate entrepreneurs.

According to the derived supply chains, forum members buying exchange services also purchased hacked servers, traffic, and social boosters, likely to further acquire and sell more currency. Popularity of these categories in the supply chains suggest that the combination of the above provide an efficient way to make money on the underground market which to the best of our knowledge has not been reported on before. Finally, extracted supply chains illuminate that about a fifth of all transactions were reselling. This reselling appears to sometimes be related to a user reselling products, such as accounts, after they are done using them. The other common cause of reselling is pricing arbitrage where a lesser ranked member would sell products to a higher reputation user that resold it for a profit.

2) Hacked Server Supply Chains: The largest chunk of the remaining supply chains is centered around hacked servers, their operation and spamming. The second largest group is account; however, those accounts are mostly email accounts, and not from VK.com. People acquire dedicated servers in order to brute-force either email accounts or other dedicated servers. Similarly, hijacked email accounts being sold and later service-specific (e.g. torrent-tracker accounts are being sold, presumably found from email access). We also find that people consume tools to automate spamming and later on start selling traffic, presumably having spread their malware. Some of the dedicated servers have clear indications that they were previously used for poker or spam, suggesting that reselling might be due to blacklisting. Finally, some dedicated servers were consumed by a user, who then was selling SMS from infected smartphones to particular prefixes, suggesting that those are useful for botnet operation.

[3]For example, see discussion at https://habr.com/ru/post/112669/

VIII. Discussion

a) Limitations: Despite both forums operating for almost a decade, we were only able to identify a few hundred supply chains. Yet, the analysis presented should be considered a lower bound estimation of the supply chains on the forums for the following reasons. First, the biggest reduction in the number of links considered was the choice of categories. Without limiting the links to those describing products that we are interested in (i.e., without filtering posts classified as `other`), there were 429 links in Hack Forums and 19,915 links in Antichat, which were reduced down to 352 (82%) and 17,402 (87%), respectively. Second, as we chose to prioritize the precision of the classifiers, our models were conservative. Third, to re-emphasize the importance of precision, when choosing what counts as "evidence" of purchasing a product, our reply annotators were conservative, excluding mere indications of slight interest as "evidence" of purchasing. Our annotators were also likely conservative while annotating links' relevance, since they were limited by their imagination of how products can be related. Another way of explaining this is that there could potentially be more related interactions because a user could have acquired multiple products in order to make a sale, but because of our cap of two interactions for a valid link, we are not always able to see the whole picture. Finally, our supply chains were constructed from the public part of the forums that feature only a subset of interactions between criminals. Moreover, we only considered the replies users left under the corresponding selling post and not the reply under the member account.

b) Practical Usage: Prior work demonstrated the problem with cross-domain prediction in underground forums [28]. Therefore, we had to annotate posts and replies to generate a separate model for each forum. In order to extend our algorithm to other forums, a domain expert would need to annotate additional posts and replies from these forums. Additionally, experiments show that forum shifts over time result in classifier performance deterioration, which can be fixed with re-annotation; the datasets used for our analysis are already years old. From our learning curves (shown in Appendix Figure 6), we estimate that building a well-performing classifier would likely require labeling of around 6,000 – 8,500 posts by domain experts, which takes about 2 person-days. We envision that in production a tool could exist to facilitate analyst annotation while preforming their normal tasks. For example, annotations could be done while analysts are reading

posts as part of their normal workflow. Then, after enough browsing and annotating, the analysts and other researchers would have access to a powerful supply chain derivation tool to augment their investigation.

c) Generalizability: One remaining question is if our supply chain identification algorithm generalizes to other forums. We will provide some arguments and evidence why most of our algorithm should generalize to many cybercrime forums. The forums we used for our evaluation and analysis were two of the larger English and Russian cybercrime forums. They are structured similarly to most of the other known major cybercrime forums with a few subforums dedicated to commerce related activity. Other parts of the forum are reserved for exchange of information and informal conversations. Moderators of the forums will remove commerce related activities from these non-commerce parts of the forum. This simplified our approach since we did not have to filter out posts that were not buying or selling products.[4]

The features we have selected for our classifier work well for English and Russian and should extend well to other similarly structured languages, but it is unclear if other features would be required for less related languages such as Chinese. Another likely issue, is that many forums focus on buying and selling products that might be ban in other forums, such as stolen credit cards. This would require an analyst developing a new set of product categories. It would also require annotation of additional posts, which as we mentioned above is already likely required when extending our algorithm to any other forums.

Finally, text chat based systems such as IRC [7] and more recently Discord and Telegram are being used by cybercriminals to buy and sell products [9]. The high level ideas of our algorithm, such as using classifiers to categorize messages and our graph based chain reconstruction, would likely generalize. However, the classifiers would need to be highly modified and additional techniques such as chat text thread disentanglement algorithms [6] would need to be adapted to the cybercrime text chat domain.

IX. Conclusion

In this paper, we proposed, implemented, validated, and analyzed a set of methods that can identify underground cybercrime forum supply chains. Our approach is the first step toward leveraging machine learning in the discovery of supply chains, which can significantly reduce the manual effort required to analyze these forums.

We have shown how those supply chains can be used to understand the collaboration in cybercriminal forums and help with providing insights into major security incidents. Our analysis of these supply chains enabled us to better identify and understand several illicit activities that occur on cybercrime forums, such as cash out, money laundering, romance scams, and targeted valuable account hijacking. While our study is a first step towards leveraging machine learning for the task of supply chain detection, more research is needed to fully automate the discovery of supply chains.

[4]It is rare but there are some forums that are not well moderated or that do not separate out commerce related posts. For these forums, we would need to use a classifier that could filter out non-commerce related posts similar to the one developed in prior work [26].

Acknowledgments

We thank Vern Paxson and Kirill Levchenko for their valuable feedback and discussion that helped us shape this research, and the Cambridge Cybercrime Center for assisting with data storage and advice. This work was supported in part by the National Science Foundation under grants CNS-1717062, CNS-1237265 and CNS-1619620, DHS S&T FA8750-19-2-0009, by the Office of Naval Research under MURI grant N000140911081, by the Center for Long-Term Cybersecurity, and by gifts from Google. We thank all the people that provided us with forum data for our analysis; in particular Scraping Hub and SRI for their assistance in collecting data for this study. Finally, we would like to thank our anonymous reviewers for their invaluable feedback. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsors.

References

[1] Armand Joulin, Edouard Grave, Piotr Bojanowski, and Toma Mikolov. "Bag of Tricks for Efficient Text Classification." In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers, pages 427-431. 2017.

[2] Tianqi Chen and Carlos Guestrin. "XGBoost: A Scalable Tree Boosting System." In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 785-794. 2016.

[3] David Décary-Hétu and Dominique Laferrière. "Discrediting vendors in online criminal markets." Disrupting criminal networks: Network analysis in crime prevention, 2015.

[4] Thomas J Holt. "Examining the forces shaping cybercrime markets online." Social Science Computer Review, 2013.

[5] Kyle Soska and Nicolas Christin. "Measuring the longitudinal evolution of the online anonymous marketplace ecosystem." In 24th USENIX Security Symposium (USENIX Security 15), 2015.

[6] Shikib Mehri and Giuseppe Carenini. "Chat disentanglement: Identifying semantic reply relationships with random forests and recurrent neural networks." In Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pages 615–623, Taipei, Taiwan, November 2017. Asian Federation of Natural Language Processing.

[7] Jason Franklin, Vern Paxson, Adrian Perrig, and Stefan Savage. "An inquiry into the nature and causes of the wealth of internet miscreants." In Proceedings of the 14th ACM Conference on Computer and Communications Security, CCS '07. ACM, 2007.

[8] Brian Krebs. Busting sim swappers and sim swap myths, 2018.

[9] BALAJI N. "Hackers now switching to telegram as a secret communication medium for underground cybercrimes." https://gbhackers.com/telegram-communication/, 2018.

[10] M. Yip, N. Shadbolt, and C. Webber. "Structural analysis of online criminal social networks." In Intelligence and Security Informatics (ISI), 2012 IEEE International Conference on, 2012.

[11] Kirill Levchenko, Andreas Pitsillidis, Neha Chachra, Brandon Enright, Márk Félegyházi, Chris Grier, Tristan Halvorson, Chris Kanich, Christian Kreibich, He Liu, Damon McCoy, Nicholas Weaver, Vern Paxson, Geoffrey M. Voelker, and Stefan Savage. "Click trajectories: End-to-end analysis of the spam value chain." In Proceedings of the 2011 IEEE Symposium on Security and Privacy, SP '11, 2011.

[12] Damon McCoy, Hitesh Dharmdasani, Christian Kreibich, Geoffrey M. Voelker, and Stefan Savage. "Priceless: The role of payments in abuse-advertised goods." In Proceedings of the 2012 ACM Conference on Computer and Communications Security, CCS '12, 2012.

[13] R. Clayton, T. Moore, and N. Christin. Concentrating correctly on cybercrime concentration. In Proceedings (online) of the Fourteenth Workshop on the Economics of Information Security (WEIS), June 2015.

[14] Mohammad Karami, Youngsam Park, and Damon McCoy. "Stress testing the booters: Understanding and undermining the business of ddos services." In Proceedings of the 25th International Conference on World Wide Web, WWW '16, 2016.

[15] Juan Caballero, Chris Grier, Christian Kreibich, and Vern Paxson. "Measuring pay-per-install: The commoditization of malware distribution." In Proceedings of the 20th USENIX Conference on Security, SEC'11. USENIX Association, 2011.

[16] JingMin Huang, Gianluca Stringhini, and Peng Yong. "Quit playing games with my heart: Understanding online dating scams." In Magnus Almgren, Vincenzo Gulisano, and Federico Maggi, editors, Detection of Intrusions and Malware, and Vulnerability Assessment. Springer International Publishing, 2015.

[17] Do kyum Kim, Geoffrey Voelker, and Lawrence Saul. A variational approximation for topic modeling of hierarchical corpora. In Proceedings of the 30th International Conference on Machine Learning, Proceedings of Machine Learning Research. PMLR, 2013.

[18] Chao Zhang, Fangbo Tao, Xiusi Chen, Jiaming Shen, Meng Jiang, Brian Sadler, Michelle Vanni, and Jiawei Han. "Taxogen: Unsupervised topic taxonomy construction by adaptive term embedding and clustering." In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery; Data Mining, KDD '18. ACM, 2018.

[19] Luca Allodi, Marco Corradin, and Fabio Massacci. "Then and now: On the maturity of the cybercrime markets the lesson that black-hat marketeers learned." IEEE Transactions on Emerging Topics in Computing, 2016.

[20] Jonathan Lusthaus. "How organised is organised cybercrime?" Global Crime, 2013.

[21] E Rutger Leukfeldt, Edward R Kleemans, and Wouter P Stol. "Cybercriminal networks, social ties and online forums: social ties versus digital ties within phishing and malware networks." British Journal of Criminology, 2016.

[22] Ahmed Abbasi, Weifeng Li, Victor Benjamin, Shiyu Hu, and Hsinchun Chen. "Descriptive analytics: Examining expert hackers in web forums." In Intelligence and Security Informatics Conference (JISIC), 2014 IEEE Joint. IEEE, 2014.

[23] Haitao Xu, Daiping Liu, Haining Wang, and Angelos Stavrou. "E-commerce reputation manipulation: The emergence of reputation-escalation-as-a-service." In WWW 2015 - Proceedings of the 24th International Conference on World Wide Web. Association for Computing Machinery, Inc, 2015.

[24] Kurt Thomas, Danny Huang, David Wang, Elie Bursztein, Chris Grier, Thomas J. Holt, Christopher Kruegel, Damon McCoy, Stefan Savage, and Giovanni Vigna. "Framing dependencies introduced by underground commoditization." In Workshop on the Economics of Information Security, 2015.

[25] Marti Motoyama, Damon McCoy, Kirill Levchenko, Stefan Savage, and Geoffrey M Voelker. "An analysis of underground forums." In Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference. ACM, 2011.

[26] Rebecca S Portnoff, Sadia Afroz, Greg Durrett, Jonathan K Kummerfeld, Taylor Berg-Kirkpatrick, Damon McCoy, Kirill Levchenko, and Vern Paxson. "Tools for automated analysis of cybercriminal markets." In Proceedings of the 26th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, 2017.

[27] George Forman and Martin Scholz. "Apples-to-apples in cross-validation studies: Pitfalls in classifier performance measurement." SIGKDD Explor. Newsl., 2010.

[28] Greg Durrett, Jonathan K. Kummerfeld, Taylor BergKirkpatrick, Rebecca S. Portnoff, Sadia Afroz, Damon McCoy, Kirill Levchenko, and Vern Paxson. "Identifying products in online cybercrime marketplaces: A dataset for fine-grained domain adaptation." In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP). Association for Computational Linguistics, 2017.

[29] Andrew Caines, Sergio Pastrana, Alice Hutchings, and Paula J. Buttery. "Automatically identifying the function and intent of posts in underground forums." Crime Science, 2018.

[30] Daniel R. Thomas, Sergio Pastrana, Alice Hutchings, Richard Clayton, and Alastair R. Beresford. "Ethical issues in research using datasets of illicit origin." In Proceedings of the 2017 Internet Measurement Conference, IMC '17. ACM, 2017.

[31] Sergio Pastrana, Daniel R Thomas, Alice Hutchings, and Richard Clayton. "Crimebb: Enabling cybercrime research on underground forums at scale." In Proceedings of the 2018 World Wide Web Conference on World Wide Web. International World Wide Web Conferences Steering Committee, 2018.

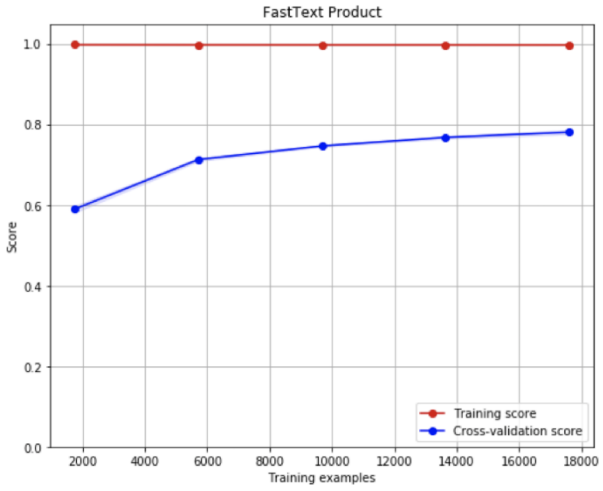
[32] Sergio Pastrana, Alice Hutchings, Andrew Caines, and Paula Buttery. "Characterizing eve: Analysing cybercrime actors in a large underground forum." In Michael Bailey, Thorsten Holz, Manolis Stamatogiannakis, and Sotiris Ioannidis, editors, Research in Attacks, Intrusions, and Defenses, pages 207–227, Cham, 2018. Springer International Publishing.

[33] R. van Wegberg, S. Tajalizadehkhoob, K. Soska, U. Akyazi, C. Hernandez Ganan, B. Klievink, N. Christin, and M. van Eeten. "Plug and prey? measuring the commoditization of cybercrime via online anonymous markets." In Proceedings of the 27th USENIX Security Symposium (USENIX Security'18), 2018.

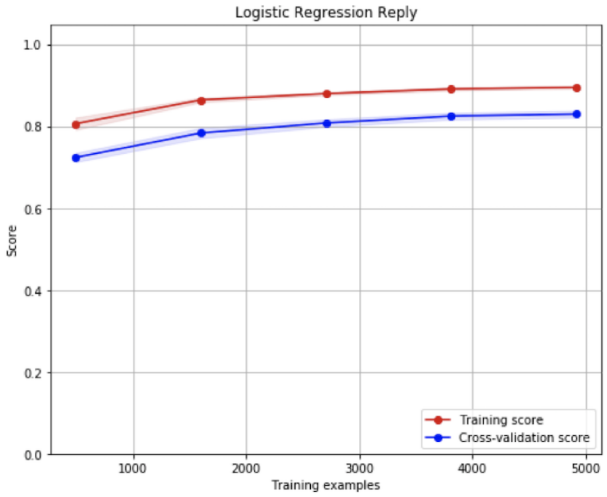
[34] Peter Snyder, Periwinkle Doerfler, Chris Kanich, and Damon McCoy. "Fifteen minutes of unwanted fame: Detecting and characterizing doxing." In Proceedings of the 2017 Internet Measurement Conference, IMC '17, 2017.

APPENDIX A
LEARNING CURVES

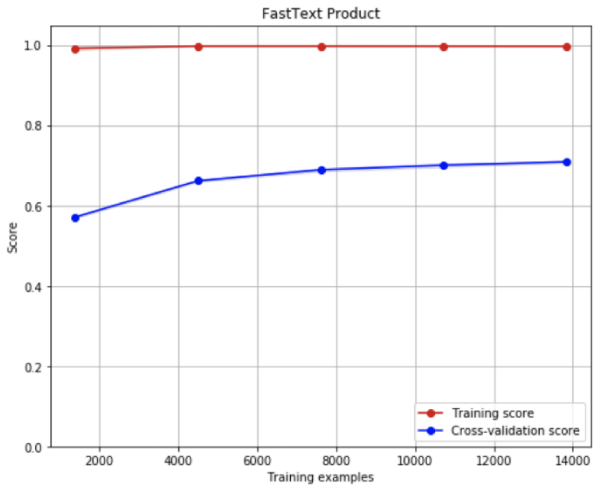
Learning curves for the product and reply classifiers.



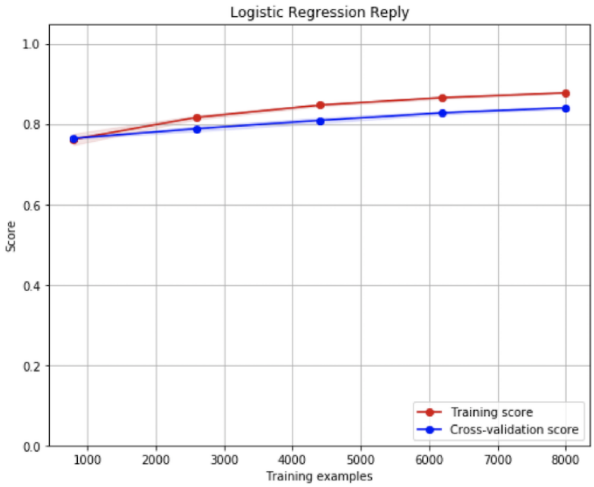
Antichat Product Learning Curve



Hack Forums Reply Learning Curve



Hack Forums Product Learning Curve

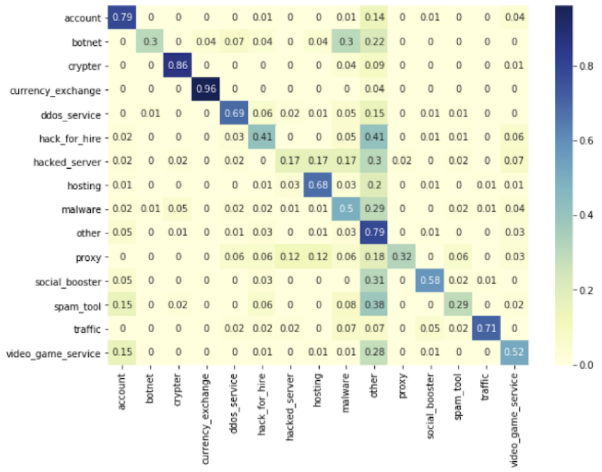


Antichat Reply Learning Curve

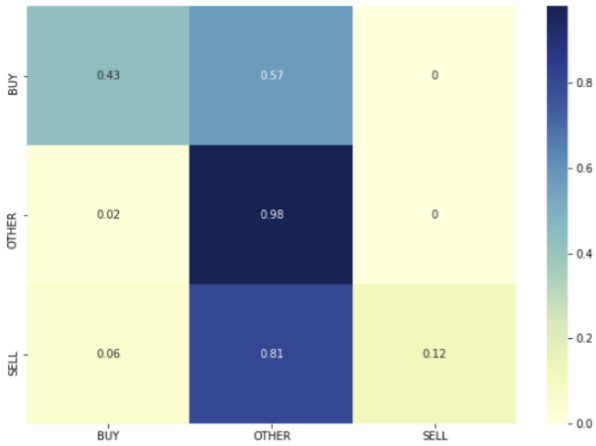
Figure 6: The learning curves for the FastText product classifiers and for the Logistic Regression reply classifiers.

APPENDIX B
CONFUSION MATRICES

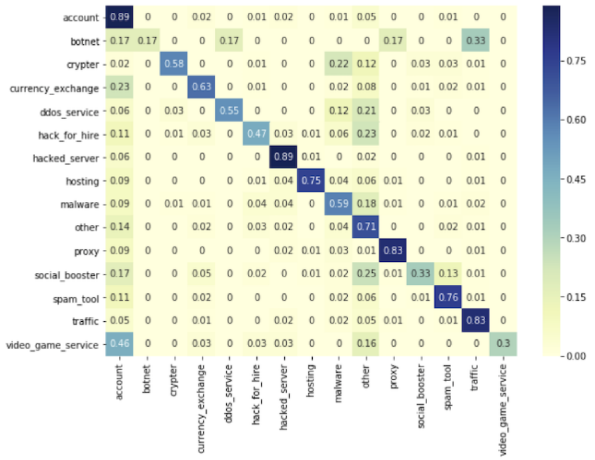
Confusion matrices for the product and reply classifiers.



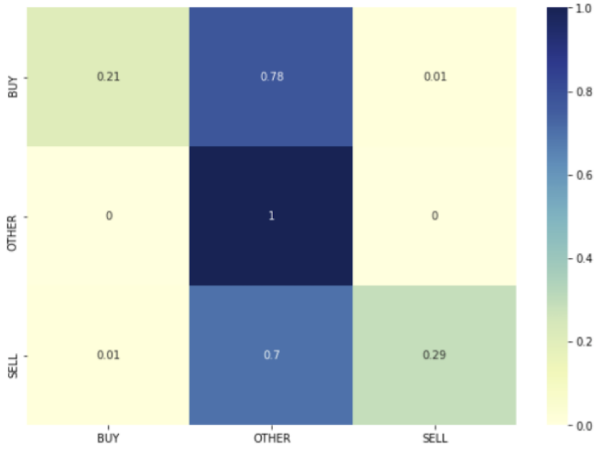
Hack Forums Product FastText Confusion Matrix



Hack Forums Reply Logistic Regression Confusion Matrix



Antichat Product FastText Confusion Matrix



Antichat Reply Logistic Regression Confusion Matrix

Figure 7: The confusion matrices for the FastText product classifiers and for the Logistic Regression reply classifiers.

APPENDIX C
ALLUVIAL GRAPH WITH ALL LINKS FOR ANTICHAT

Entire unvalidated alluvial graph for Antichat.

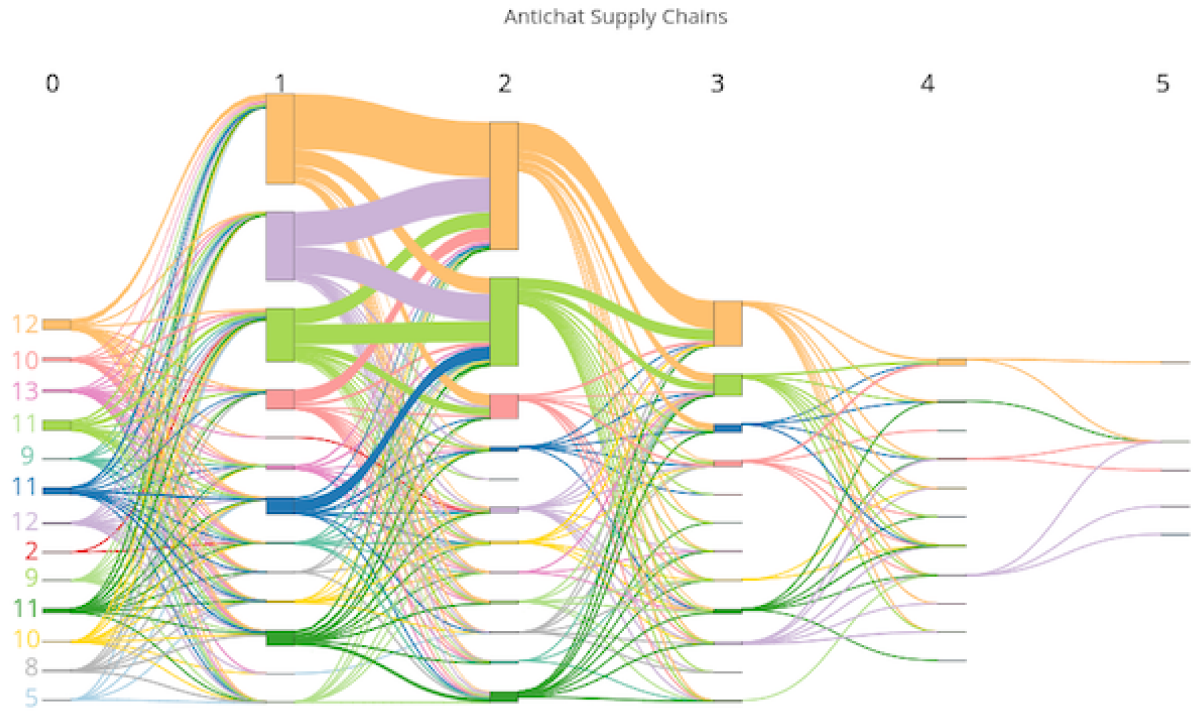


Figure 8: All supply chains found in Antichat, unvalidated. Edges are colored according to source product category, and have widths determined by the number of users who purchased the source product and sold the destination product, with attenuation. Numbers at the top correspond to the level in the modified breadth-first search algorithm at which the node was discovered. The number of chains originating with each product category is denoted next to those chains. Color-coding follows legend in Figure 4c.