Identifying High Quality Carding Services in Underground Economy using Nonparametric Supervised Topic Model

Research-in-Progress

Weifeng Li
Eller College of Management
The University of Arizona
1130 E. Helen St. McClelland Hall 430
Tucson, AZ 85721
weifengli@email.arizona.edu

Junming Yin
Eller College of Management
The University of Arizona
1130 E. Helen St. McClelland Hall 430
Tucson, AZ 85721
junmingy@email.arizona.edu

Hsinchun Chen
Eller College of Management
The University of Arizona
1130 E. Helen St. McClelland Hall 430
Tucson, AZ 85721
hchen@eller.arizona.edu

Abstract

Over the years, cybercriminals increasingly joined the underground economy to exchange malicious services for conducting data breaches crimes. As many service providers are rippers, most cybercriminals rely on a few high quality services. To this end, cybercriminals post customer reviews evaluating the purchase experience and the service quality. To identify high quality services, researchers face two major challenges – the cybercriminal-specific language and the scale of the underground economy. This study presents a text-mining-based system for identifying high quality services by analyzing customer reviews. A novel supervised topic model is designed to accommodate the heterogeneous and uncertain nature of customer reviews. We further designed a variational algorithm for model inference. Moreover, we collected real data from two underground economy forums for English-speaking and Russian-speaking cybercriminals as our research testbed. Our research contributes to the practice of understanding and mitigating underground economy by providing cybersecurity researchers and practitioners with actionable intelligence.

Keywords: Underground economy, text mining, topic modeling, Bayesian nonparametrics, customer reviews, service quality

Introduction

Data breaches have been an emerging problem for our economy. Over the past decade, a growing body of cybercriminals joined the underground economy forums to exchange services for conducting data breaches (Graaf et al. 2013). These cybercriminals are responsible for many of the data breach incidents targeting organizations in all sectors, such as Target, JP Morgan Chase, Anthem BlueCross, etc. (Krebs 2014). A variety of carding services are being provided by cybercriminals with specialties, including writing malware writing, stealing data, counterfeiting credit cards, cashing out, etc. (Graaf et al. 2013).
However, many service providers are rippers seeking to steal from other cyber thieves. Therefore, cybercriminals tend to employ a few high quality services that are trustworthy (Lusthaus 2012). These services are popular among cybercriminals, and are therefore the backbone of cybercrimes (Clayton et al. 2015; Thomas et al. 2015). Identifying high quality carding services is a task of great interest to both cybersecurity researchers and practitioners (Peretti 2008). Researchers can study the dynamics of these services to better understand the underground supply chain for cybercrime. Law enforcement can investigate the key services and the providers behind them to deter data breach incidents.

The primary indicator of service quality in the underground economy is the textual customer reviews posted by cybercriminal customers (Radianti 2010). However, identifying high quality services from textual customer reviews is challenging. Existing methods for evaluating customer reviews are unable to provide adequate service quality estimation because of the complexity within the cybercriminal customer reviews. Particularly, the hacker customer reviews are domain-specific. Not only do these customer reviews employ a unique set of cybercriminal lexicons, but the reviews for different service categories also have distinct sets of service-specific topics. Moreover the total number of such topics within hacker customer reviews is difficult to determine. Due to our insufficient knowledge about the underground economy, there is a need for methods that can deal with uncertainties especially in terms of the number of topics. Further, as the underground economy evolves, new services will emerge introducing new topics to customer reviews.

The contribution of our paper is three-fold. First, we propose a text-mining-based system to identify high quality carding services. In our proposed system, counter anti-crawling measures are developed to collect the listings and customer reviews from the underground economy forums. Second, a novel supervised topic model is designed to address the complexity within the cybercriminal customer reviews. The model can learn from previous customer reviews for a variety of services and predict the service quality for newly obtained customer reviews. Moreover, the topics learned by our model provide important sights for exploratory studies. Third, we further design a variational inference algorithm for model inference that can scale to large datasets. To evaluate the efficacy of our system, we build our research testbed on real data from two underground economy forums for English-speaking and Russian-speaking cybercriminals as our research testbed.

Related Works

Online Underground Economy

The underground economy is the marketplace for cybercriminals, where malicious services are exchanged. These services being sold in underground economies simplify the cybercrime in such a way that any cybercriminal can conduct sophisticated crimes by purchasing services from expert cybercriminals. Our literature review focuses on prior underground economy studies from two perspectives: 1) major carding services, and 2) cybercriminal customer reviews.

Major Carding Services

We summarize the major carding services in the underground economy in Table 1. Three categories of services corresponding to the three stages in data breaches. First, cybercriminals solicit tool services to prepare the attacking toolkit (Thomas et al. 2015). Malware writing provides malicious programs for acquiring sensitive victim information. Phishing collects sensitive victim information by masquerading a trustworthy party. Botnets provide the control of a collection of computers and has been used to facilitate all kinds of cybercrimes. Second, cybercriminals acquire sensitive victim information from stolen data services (Holt 2012). Payment card data includes the card number, card holder, dumps, CVV, etc. Identity data includes SSN, driver's license, medical insurances, etc. Credential data includes bank account logins, paypal/ebay logins, etc. Third, cybercriminals seek supportive services to monetize stolen data (Holt and Lampke 2010). Plastic service involves counterfeiting stolen cards for card-present transactions. Drop service provides an address where the fraudulently bought goods are delivered to. COB service provides change of billing address as some high value goods are only sent to the billing address.
Table 1. Summary of Major Carding Services

<table>
<thead>
<tr>
<th>Service</th>
<th>Category</th>
<th>Description</th>
<th>Examples</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Malware Writing</td>
<td>Tool</td>
<td>Programs to collect victim financial/personal information</td>
<td>POS malware; CC skimmer</td>
<td>$300-5000</td>
</tr>
<tr>
<td>Phishing</td>
<td>Tool</td>
<td>Scam websites to acquire sensitive information</td>
<td>phishing mails; scan sites</td>
<td>$2.5-100/wk</td>
</tr>
<tr>
<td>Botnets</td>
<td>Tool</td>
<td>Compromised computers on the internet for malicious purposes</td>
<td>hosting relays for stolen cards</td>
<td>$2-60/hr</td>
</tr>
<tr>
<td>Payment Cards</td>
<td>Data</td>
<td>Information copied from credit/debit cards</td>
<td>dumps; CVV</td>
<td>$0.1-25</td>
</tr>
<tr>
<td>Identities</td>
<td>Data</td>
<td>Sensitive personally identifiable information</td>
<td>SSN; driver’s license; insurance cards</td>
<td>$1-260</td>
</tr>
<tr>
<td>Credentials</td>
<td>Data</td>
<td>Victim login credentials</td>
<td>bank accounts; Paypal accounts</td>
<td>$1-300</td>
</tr>
<tr>
<td>Plastics</td>
<td>Cashing</td>
<td>Plain card for cybercriminals to replicate stolen cards</td>
<td>blank credit cards</td>
<td>$40-110</td>
</tr>
<tr>
<td>Drop</td>
<td>Cashing</td>
<td>Location cybercriminals can have illicitly purchased goods sent to</td>
<td>mail drop</td>
<td>Royalty</td>
</tr>
<tr>
<td>COB</td>
<td>Cashing</td>
<td>Change of Billing address for cybercriminals to make purchases</td>
<td>online change; phone change</td>
<td>$35-140</td>
</tr>
</tbody>
</table>

Cybercriminal Customer Reviews

Anonymity and information asymmetry create a large deficit of trust in the underground economy (Lusthaus 2012). While customer reviews have been widely used to infer service quality in the economy “above the ground” (Gao et al. 2015), the underground economy use the same customer review mechanisms to establish “honors among thieves.” Buyers comment on their experience with the seller or the service (Holt 2012). These comments help to improve prospective buyers’ perception of the seller quality (Jensen and Nunamaker Jr. 2013). Customer review helps reduce information asymmetry in cybercriminals’ purchase decisions (Holt 2013). Prior customer reviews are a major for prospective buyers to assess the quality of the service (Holt 2013; Ma et al. 2013). On the other hand, sellers maintain positive reviews among buyers to gain reputation, trust, and credibility (Radianti 2010). Leveraging customer reviews allows us to identify high quality carding services, which has strong implications for law enforcement and cybersecurity researchers. However, cybercriminal customer reviews are heterogeneous because reviews not only contain service-specific topics but also include irrelevant discussions, which refrains existing sentiment analysis-based approaches from effectively predicting the service quality from customer reviews. Since supervised topic models can learn the topics in customer reviews to inform the prediction of service quality, we therefore review supervised topic models.

Supervised Topic Modeling

In general, supervised Topic Model (STM) concerns predicting the response variables by leveraging topics in the documents. Given $D$ documents $\{w_d\}$ and associated response variables $\{y_d\}$, STM’s want to learn the underlying topics $\beta = \{\beta_k\}$ within $\{w_d\}$ in order to predict the response variable $y_d$ for a new document $w_{d^*}$, where $w_d$ is the “bag-of-words” vector of document $d$ and $\beta_k$’s are distributions over all words. We summarize major supervised topic models in Table 2, using a taxonomy of four dimensions: model, topic structure, inference, and testbeds & performance. Particularly, the topic structure dimension concerns whether the topic structure allows for new topics. Fixed topic structure implies that the model is not capable of accommodating new topics.
Table 2. Major Supervised Topic Models

<table>
<thead>
<tr>
<th>Study</th>
<th>Model</th>
<th>Topic Structure</th>
<th>Inference</th>
<th>Testbeds &amp; Performances</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blei and McAuliffe (2008)</td>
<td>Supervised LDA (sLDA)</td>
<td>Fixed</td>
<td>Variational</td>
<td>5,000 movie reviews 0.5pR2; 4,000 webpages 0.095pR2</td>
</tr>
<tr>
<td>Lacoste-Julien et al. (2008)</td>
<td>DiscLDA</td>
<td>Fixed</td>
<td>Sampling</td>
<td>20 newsgroups 17% error rates</td>
</tr>
<tr>
<td>Ramage et al. (2009)</td>
<td>Labeled LDA (L-LDA)</td>
<td>Fixed</td>
<td>Sampling</td>
<td>4K Web pages 52.12% MicroF1</td>
</tr>
<tr>
<td>Rubin et al. (2012)</td>
<td>Dependency LDA (D-LDA)</td>
<td>Fixed</td>
<td>Sampling</td>
<td>News 54.1% MicroF1; Legal docs 46.7% MicroF1</td>
</tr>
<tr>
<td>Zhu et al. (2013)</td>
<td>Logistic Supervised LDA (lsLDA)</td>
<td>Fixed</td>
<td>Sampling</td>
<td>20 newsgroups ~81% accuracy</td>
</tr>
<tr>
<td>Zhu et al. (2014)</td>
<td>MedLDA</td>
<td>Fixed</td>
<td>Sampling</td>
<td>20 newsgroups ~83% accuracy</td>
</tr>
<tr>
<td>Rabinovich and Blei (2014)</td>
<td>Inverse Regression Topic Model (IRTM)</td>
<td>Fixed</td>
<td>Variational</td>
<td>Press release 0.826L1; Yelp 0.704L1</td>
</tr>
<tr>
<td>Dai and Storkey (2014)</td>
<td>Supervised HDP (sHDP)</td>
<td>Unfixed</td>
<td>Sampling</td>
<td>Movie reviews ~0.3pR2; Webpages ~0.08pR2</td>
</tr>
</tbody>
</table>

Most of the models assume fixed number of topics. The major underlying reason is that they rely on parametric topic models such as latent Dirichlet allocation (LDA). The parametric topic models assume the number of topics to be known a priori. In most cases, it is difficult to identify the correct number of topics from the theoretical foundation of the problem space (Arun et al. 2010; Griffiths and Steyvers 2004). Model selection is therefore involved to find the optimal number of topics under a certain evaluation metrics, such as perplexity and KL-Divergence. As such, models often suffer overfitting or underfitting. Moreover, fixing the number of topics makes it impossible to identify new topics in new documents. Lately, nonparametric topic models have drawn great attention for their capability to accommodate unlimited number of topics. Hierarchical Dirichlet Process (HDP) is a leading nonparametric topic model that use the stacking of Dirichlet Process to organize topics (Teh et al. 2006). Compared with the Dirichlet distribution in LDA, which assigns proportions to the fixed number of topics, the Dirichlet Process imposes a distribution over all possible topics, so that unlimited number of topics can be modeled at the same time. Probabilistic models such as STM’s necessitate approximation inference for posteriors as they are intractable. Most of the models uses sampling-based algorithm. Sampling-based algorithms (e.g., Gibbs sampling) approximate the posteriors by empirically sampling from the distribution. Since it usually requires thousands of iterations to “burn-in” for each variable, it is not scalable to large volumes of data. Variational algorithms seek to find a tractable surrogate distribution that is closest to the posterior. Variational algorithms transform a sampling problem into an optimization problem and therefore require significantly fewer iterations than sampling-based algorithms. In terms of testbeds, prior studies mainly focused on a few application domains, such as news, movie reviews, and web pages. Given the difference in the content, language usage, and communication structure, it is unclear whether the supervised topic model extend the application to cybercriminal communities.

**Research Design**

We propose the design of a text-mining-based system for identifying key services in the underground supply chain. Our proposed system comprises four major steps: 1) data collection, 2) pre-processing, 3) review analysis, and 4) evaluation. The major methodological contribution and innovation of our
Identifying High Quality Carding Services in Underground Economy

The proposed system is the review analysis step, in which we design a nonparametric supervised topic model (NSTM) to learn the customer reviews for predicting service quality. NSTM is capable of capturing the uncertainty and the heterogeneity of topics within customer reviews.

Data Collection

The data collection step focuses on collecting service listings along with customer reviews. To avoid the law enforcement, cybercriminals usually employ several anti-crawling measures that entangles the collection of underground economy forums. Common anti-crawling measures includes IP checking, request monitoring, and authentication. To this end, we have designed a set of counter measures to collect the forums, including relayed connection, random request interval, browser behavior emulation, and cookie deployment (see Table 3). Based on these counter measures, we develop targeted crawlers to collect the listings from underground economy forums. For each listing, we further extract corresponding customer reviews. In underground forums, listings are the threads and the customer reviews are the posts replying to the listing threads.

Table 3. Counter Anti-crawling Measures

<table>
<thead>
<tr>
<th>Anti-crawling Measure</th>
<th>Description</th>
<th>Counter-measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>User-agent Check</td>
<td>Shops verify the HTTP request comes from a legitimate user-agent (browser.)</td>
<td>We use packages that mimics the behavior of mainstream browsers.</td>
</tr>
<tr>
<td>User/password Authentication</td>
<td>Shops require users to register and login before accessing the data. CAPTCHA is widely used to verify the user inputting the credential is a human-being.</td>
<td>We manually login the shop first and extract the corresponding cookies. With these cookies carried with HTTP request, we can bypass the login process.</td>
</tr>
<tr>
<td>Session Timeout</td>
<td>Shops automatically logout users that have been in the shop for too long.</td>
<td>Need human involvement to acquire and deploy renewed cookies.</td>
</tr>
<tr>
<td>IP Check</td>
<td>CloudFlare verifies the HTTP request comes from a legitimate IP address rather than a public known proxy, such as Tor.</td>
<td>We setup a private, dedicated proxy server to reroute our connections. The proxy server can be deployed in Digital Ocean as it is easily to deploy a new one once the current ip got banned.</td>
</tr>
<tr>
<td>DDoS Prevention</td>
<td>CloudFlare detects possible DDoS signs and bans the suspicious IP address.</td>
<td>We set intervals between two successive requests; allow the private proxy server change IP address easily.</td>
</tr>
</tbody>
</table>

Preprocessing

In this step, the proposed system prepares the customer reviews for subsequent analysis by building a language model. The preprocessing pipeline includes text cleaning, text tokenization, and language model building.
building. In text cleaning, we use regular expression-based approach to filter out web links, duplicated punctuations, and personal contacts, which are not relevant to the service quality. Further, we tokenize customer reviews for each service. Here, we aggregates all customer reviews for the same service to avoid the biases from single reviews. Following prior text-mining studies (Blei 2012; Blei and Mcauliffe 2008), we build a unigram language model from tokenized customer reviews. The unigram language model captures the exchangeability for the words in a document, which is a critical assumption in topic modeling (Blei 2012).

**Review Analysis**

The review analysis step trains a novel nonparametric supervised topic model for predicting service quality based on customer reviews. As we mentioned, the topic-model-based approach accommodates for the heterogeneity in the customer reviews for a variety of services. Moreover, the nonparametric approach not only relieves us from the model selection complications but also allows us to capture unknown topics in customer reviews for emerging services.

Following prior topic modeling literature (Blei and Mcauliffe 2008), we design NSTM under a hierarchical Bayesian modeling framework (Gelman et al. 2014). The structure of NSTM can be represented by a graphical model as shown in Figure 2. Consistent with prior graphical model representations, the grey nodes represent observed variables such as customer reviews $w_d$ and service quality $y_d$ for service $d$; the white nodes represent the hidden variables used to model the customer review generating process; the rectangle plate indicates an independent instance of the variables within. Overall, NSTM has three conceptual components.

![Figure 2. The Nonparametric Supervised Topic Model](image)

The first component builds the nonparametric topic model. Following prior literature (Blei and Mcauliffe 2008), we define a topic $\beta$ as a probabilistic distribution over all words. Specifically, $\beta$ is a $|W|$-dimensional vector whose elements $\beta^{(w)}$ sum up to one, where $|W|$ is the number of all words and $\beta^{(w)}$ is the probability of word $w$ being generated from the topic. While LDA-based models assumes $K$ topics ($\beta_1, ..., \beta_K$) and each topic $\beta_k$ arising with probability $\theta_k$, we assume the topic space $\Theta$ encompassing all possible topics and impose a distribution $G$ onto this topic space. The topics space assumption provides the flexibility of any probabilistic distributions over words. Particularly, the distribution $G$ is a distribution over topics, which are themselves distributions over words. In other words, $G$ is the distribution over distributions. As suggested in statistical literature (Teh 2010), this distribution-over-distribution idea can be implemented with a Dirichlet Process (DP). DP is a stochastic process whose marginal distributions are Dirichlet distributed according to the base distribution $H$. This implies 1) some topics can be shared across different draws from DP and 2) new topics (i.e., topics that have never been drawn) can appear. For each service $d$, the distribution $G_d$ of the topics in its customer reviews under unigram language model can be considered as a DP such that topics can be shared across different words. Further, $G_d$ can be further considered as draws from a corpus-level DP such that some topics can be shared across different services and customer reviews for each service may have unique topics. Formally, we define the generative process for modeling topics as follows:
1. The corpus-level distribution of topics $G_0 \sim DP(\omega, H)$, where $\omega$ is the concentration parameter, and $H$ is a symmetric Dirichlet distribution.
2. For each service $d$,
   a. The service-level distribution of topics $G_d \sim DP(\alpha, G_0)$, where $\alpha$ is the concentration parameter.
   b. For each word $w_{dn}$ in customer reviews for $d$,
      i. Topic of the word $\beta_{dn} \sim G_d$.

The second component concerns the influence of service quality on customer reviews. This is intuitive in the sense that the customer reviews for high-rated services are more positive than for low-rated services. We use multinomial logistic model to implement this intuition (Taddy 2013). Under unigram model, each word in the reviews can be considered as chosen from the vocabulary by the service quality. High-rated quality services tend to choose words such as “good,” “fast,” and “pleasant.” Multinomial logistic model helps to examine the effect $\phi_w$ of service quality $y_d$ on the choice of word $w$: $p(w) \propto \exp(\alpha_w + \phi_w y_d)$. In this formula, $p(w)$ is the probability of word $w$ appears, $\alpha_w$ is the bias term of word $w$, and $\phi_w$ is the effect of service quality $y_d$ on word $w$. Further, since many of the words are not affected by service quality $y_d$, we want to shrink the $\phi_w$’s of these words to zero. This can be achieved by assuming $\phi_w$ to follow Laplace distribution with zero mean: $\phi_w \sim \text{Laplace}(\lambda)$ (Park and Casella 2008).

In the last component, we jointly leverage the intuitions from the previous two components. Recall that in the nonparametric topic model, topic $\beta_k$ is a distribution over all words. Equivalently, the probability of word $w$ arising under topic $\beta_k$ is $p(w|\beta_k) = \beta^{(w)}_k \cdot \beta^{(w)}_k$, represents the occurrence of word $w$ regardless of the service quality. In the second component, we model the probability of word $w$ arising as follows: $p(w) \propto \exp(\alpha_w + \phi_w y_d)$, where $\alpha_w$ is the bias term that is not affected by service quality. Replacing this term with $\beta^{(w)}_k$ allows us to account for topics and the service quality simultaneously. Therefore, we use an extended multinomial logistic model to reflect this intuition: $p(w|\beta_k) \propto \beta^{(w)}_k \cdot \exp \phi_w y_d$. Further, we identify the service category based on the interpretation of topics through a two stage process. In the first stage, we examine all the topics extracted by NSTM and interpret each topic $\beta_k$ based on the words with the top 10 highest probability. In the second stage, we study the topics associated with each listed service and use these topics to define the features of this particular service.

Consistent with Blei (2012), we fit our NSTM model by designing a variational algorithm (Hoffman et al. 2013). Since the posterior of the hidden variables cannot be estimated in an analytical form, the variational algorithm seeks to approximate the posteriors so that the approximated posterior can be further used for predicting the service quality from the new customer reviews. The variational algorithm approximates the posteriors by finding the closest surrogate distribution to the posterior using coordinate ascent. While sampling-based method such as Gibbs sampling requires thousands of iterations over all variables before the convergence can be reached, the variational inference method significantly reduce the complexity by converting sampling into optimization, which is more suitable for “big data” problems. Further, we also develop a variational prediction algorithm that predicts new service quality from customer reviews. Based on the approximated posterior, the variational prediction algorithm seeks to optimize the posterior of the new service customer reviews using gradient ascent. Interested readers are recommended to Hoffman et al. (2013) for detailed exposition of the variational-based method.

Evaluation

Research Testbed

Our research testbed is built on two underground economy forums (Table 4). The two forums are chosen based on the following considerations. First, abundant data breach-related transaction activities have been observed in both forums. Listings in both forums encompasses Point-of-Sale (POS) skimmer, stolen credit cards, personal identifiable information (PII), counterfeit cards, and mules for monetizing. Further, both forums are from the English-speaking and Russian-speaking cybercriminal communities, who constantly involve in data breach incidents. Lastly, both forums have been consistently used by cybercriminals over time. Recent activities have been observed, including listings from the cybercriminal who is believed to be responsible for the Target data breach (Krebs 2014). Using the counter anti-crawling...
measures, we were able to collect all the members, threads, and posts from both forums. We list the metadata of our collection in Table 4. Forum names are censored to avoid possible complications.

<table>
<thead>
<tr>
<th>Forum</th>
<th>Members</th>
<th>Threads</th>
<th>Posts</th>
<th>Language</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>C*****b</td>
<td>4,972</td>
<td>4,834</td>
<td>38,335</td>
<td>English &amp; Russian</td>
<td>8/7/2013 ~ 3/11/2016</td>
</tr>
</tbody>
</table>

To build our research testbed, we select the listings that have service quality ratings voted by cybercriminal buyers from both forums. The service quality rating ranges from 1.0 to 5.0, with 1.0 being the lowest quality and 5.0 being the highest quality. Further, we consider the services with rating greater than 3.0 as high quality services and the services with rating equal to or less than 3.0 as low quality services. We use bootstrapping to ensure the balance between the two categories in our research testbed.

**Experiment Design**

We evaluate the effectiveness of our proposed system in identifying key services in the underground supply chain. Particularly, we examine how well our proposed system can predict the service quality based on the service reviews (Shmueli and Koppius 2011). For each forum, we split the service quality \( Y \) and corresponding customer reviews \( w \) into \( T \) time spells: \( \{Y_1, ..., Y_T\} \) and \( \{w_1, ..., w_T\} \). Using service quality and reviews from earlier stage: \( \{Y_1, ..., Y_t\} \) and \( \{w_1, ..., w_t\} \), we want to predict the service quality \( \{Y_{t+1}, ..., Y_T\} \) of \( \{w_{t+1}, ..., w_T\} \). This allows us to realistically assess the predictive performance of our proposed system as we want to learn from the past service customer reviews to inform understating new service customer reviews. We vary \( t \) from 1 to \( T – 1 \) for two reasons. First, we want to know how well our proposed system can capture new topics. Second, we want to know how much training data we need in order to achieve plausible performance. We quantify the predictive performance using precision, recall, and F-measure. We conduct two sets of experiments to assess the proposed system from two perspectives. In the first set of experiments, we seek to gauge the improvement from incorporating topics. We compare the proposed system with multinomial logistic models (Taddy 2013). In the second set of experiments, we seek to examine the benefit of accounting for new topics. We compare the proposed system against supervised LDA with 50 topics, 100 topics, and 150 topics (Blei and Mcauliffe 2008). Further, to demonstrate the practical impact of our proposed system, we showcase a series of identified key tasks in our research testbed.

**Conclusions**

Targeting the key cybercriminal services can help mitigate the underground supply chain. This paper proposed a system for identifying key services in the underground supply chain. The system is capable of crawling the underground economy forums with counter anti-crawling measures. Moreover, we designed a novel nonparametric supervised topic model to accommodate for the heterogeneity in customer reviews for various services. We proposed to perform a series of experiments to assess the effectiveness of NSTM.

**Acknowledgements**

This work was supported by the National Science Foundation under Grant No. SES-1314631 and also under Grant No. DUE-1303362.

**References**


Identifying High Quality Carding Services in Underground Economy


