2014 IEEE Joint Intelligence and Security Informatics Conference

Identifying Top Sellers In Underground Economy Using Deep Learning-based Sentiment Analysis

Weifeng Li, Hsinchun Chen
Department of Management Information Systems
The University of Arizona
Tucson, AZ 85721, USA
weifengli@email.arizona.edu, hchen@eller.arizona.edu

Abstract— **The underground economy is a key component in cyber carding crime ecosystems because it provides a black marketplace for cyber criminals to exchange malicious tools and services that facilitate all stages of cyber carding crime. Consequently, black market sellers are of particular interest to cybersecurity researchers and practitioners. Malware/carding sellers are critical to cyber carding crime since using malwares to skim credit/debit card information and selling stolen information are two major steps of conducting such crime. In the underground economy, the malicious product/service quality is reflected by customers' feedback. In this paper, we present a deep learning-based framework for identifying top malware/carding sellers. The framework uses snowball sampling, thread classification, and deep learning-based sentiment analysis to evaluate sellers' product/service quality based on customer feedback. The framework was evaluated on a Russian carding forum and top malware/carding sellers from it were identified. Our framework contributes to underground economy research as it provides a scalable and generalizable framework for identifying key cybercrime facilitators.**

Keywords*— *cybersecurity; underground economy; carding crime; deep learning; sentiment analysis; top sellers

I. INTRODUCTION

In recent years, an increasing amount of financial information has been digitalized, such as credit/debit card verification, bank accounts, associated personal information, etc. Leaks of such information cause catastrophic collateral damages to multiple victim organizations in addition to card owners. Behind such data breaches are carders who involve in skimming credit/debit card information and cashing it out. Studies have shown that the online underground economy is closely related to cyber carding crime to the extent that attackers not only purchased the data scraper through the underground economy but also disseminate stolen data in it [1]. Product/service providers in underground economies are critical to the operation of cyber carding crime in the sense that they are the ones enabling different stages of cyber carding crimes with their special tools and services. In carding forums, they sell their specialties by advertisement threads. Customers leave their feedback in replies, leaving salient information for researchers to understand the phenomenon.

However, limited work has been done to identify and analyze these cybercrime facilitator in the online underground economy using such information. The identification and analysis of top sellers is of particular interest to cybersecurity researchers and practitioners considering their vulnerability exploiting capability, card data source, cashing channel, influence on others, etc. Developing a method for automatic identification of top sellers benefits cybersecurity researchers and practitioners; it would alleviate their required work load as well as enable further analysis to identify threats and take timely countermeasures. This paper presents a scalable and generalizable automated framework based on deep learning to identify top sellers in the online underground economy. Deep learning is an emerging, promising machine learning framework that is suitable for processing online textual content [2]. With respect to the scope of this paper, we primarily focus on two categories of sellers: malware sellers and carding sellers because they play critical roles in cyber carding crime as the former provides the programs that initiate carding crimes and the latter controls the channels to distribute the stolen data. However, the framework can be generalized to identify other top sellers.

II. LITERATURE REVIEW

We review prior literature from two streams of research: (1) underground economy and (2) sentiment analysis and deep learning.

A. Underground Economy

Underground economies refer to black marketplaces in cyberspace where fraudulent and stolen products and services are traded [3]. A variety of products and services are being sold in underground economies [1]. They are promoted through advertisement threads [4]. However, the transactions usually take place outside of the forum [1]. In order to attract the attention of potential customers, these advertisements are meticulously designed by customizing their message with capitalization, multi-colored text, ASCII flares, and repeated sales pitches across multiple lines [4]. Black market sellers cherish word-of-mouth and trust [3]. The most common way to build up trust is through customer feedback [5]. Online customer reviews as such have been shown to have a positive effect on buyer trust [6]. The customer feedback reflects the quality of the products and services [1]. Subsequent customers leverage such information in their purchase decision making

978-1-4799-6364-5/14 $31.00 © 2014 IEEE
DOI 10.1109/JISIC.2014.19

IEEE computer society

process. In forums, malicious product and service transactions are initiated within an advertisement thread followed by replies, which serve as customer reviews and accrue to become word-of-mouth [4]. Sentiment analysis is an effective technique to evaluate opinions in texts, and hence we review relevant works in the next subsection.

B. Sentiment Analysis and Deep Learning

Sentiment analysis is a natural language processing technique widely used to analyze online customer reviews [7]. There are generally two approaches: learning-based approach and dictionary-based approach. The former applies machine learning classifiers to determine the sentiment orientation; the latter determines the sentiment orientation based on the sentiment scores of each token given by a dictionary [8]. Sentiment analysis has been applied to a variety of user-generated review contexts [9][10]. An emerging technique for sentiment analysis is deep learning [11]. Deep learning algorithms are based on two building blocks: word embedding and the recursive neural network [12]. Word embedding builds a word vector language model by converting each word to low-dimensional continuous-valued vectors. A recursive neural network is a tree-shaped deep learning architecture that determines the semantics of a sentence by recursively merging lower level vectors. Recursive neural network applications in sentiment analysis use a sentiment distribution vector to represent each word/phrase and predict the sentiment distribution of a sentence by recursively merging the sentiment distribution vectors of lower level parse trees all the way down to each word. In [11], the Sentiment Treebank was trained on an online review corpus using a recursive neural tensor network. It outperforms state-of-the-art sentiment classifiers by 5% reaching 85% accuracy.

III. Research question

Although the underground economy provides a wealth of interesting customer feedback, limited effort has been devoted to analyzing this information. Deep learning-based sentiment analysis is capable of identifying the sentiment of online reviews with great accuracy. Motivated by this gap, this study seeks to answer the following questions:

- Who has the best rated malicious products and services in the underground economy?
- How effective is deep learning to identify these top sellers?

IV. Research testbed and design

A. Research Testbed

We identified a famous Russian hacker forum with black market postings, Zloy. We chose Russian hacker forums as Russian hackers are best known for cyber carding crimes [13]. We used web crawlers to collect the discussion content from these forums. Automatic authentication form filling technique and anonymous network access technique were used in the collection procedure since hacker forums are protective about their content. Our forum collection contains 69,385 threads containing 485,019 posts from 14,308 forum members in the time frame from 10/1/2004 to 7/6/2013.

B. Research Design

Our deep learning-based top seller identification framework comprises three components: snowball sampling, thread classification, and deep learning-based sentiment analysis (Figure 1).

The snowball sampling component retrieves threads relevant to the black market topics in a iterative fashion [1]. Hacker forum users with similar interest tend to join the discussion in the same threads [14]. Starting with a set of seeding keywords, an initial set of threads are retrieved, and then the users involved in these threads are extracted as new keywords to retrieve other threads. The threads identified in the snowball sampling component contain users with similar interest; however, the thread content may vary. The thread classification component further cleanses the threads by categorizing each thread into dedicated groups. The feature set is critical to the classification outcome [15][16]. Based on prior literature [4][17], three categories of cues have been identified to help devise the feature set for specific types of thread: topical, highlight, and hyperlink (Table I). Since we aim to extract multiple classes of threads, we use the Maximum Entropy (MaxEnt) classifier, which outperforms other classifiers in multiclass categorization [18]. The deep learning-based sentiment analysis component predicts the quality of the products/services by evaluating the sentiment score of customer feedback. It consists of four steps: translation, word vectorization, recursive neural tensor network, and sentiment scores aggregation. The translation step translates each replying post from its original language to English. It is fulfilled by Google Translate. The word vectorization step transforms each word into its vector representation using the Sentiment Treebank from [11]. It is trained on an online review dataset, which is similar to our problem. The recursive neural

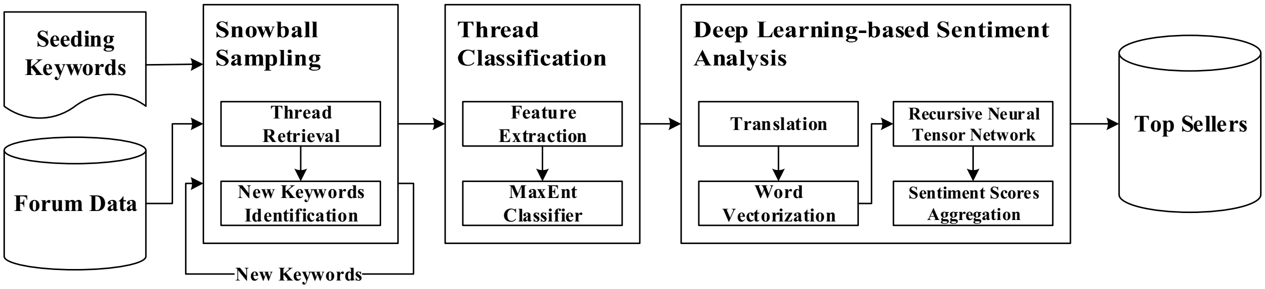

Fig. 1. Deep Learning-based Sentiment Analysis for Top Seller Identification

TABLE I. CATEGORIES OF CUES IN THREADS HELPING TO DEVELOP CLASSIFICATION FEATURES

Category	Cue	Example	Type
Topical	Monetary Lexicons	"$", "Ruble", "wmz"	Malware, Carding, Other products or services
	Domain-Specific Lexicons	"cc", "program", "v1.0", "shop"	Malware, Carding, Other products or services
	Lexical Measures	Length of the thread	Malware, Carding, Other products or services, Irrelevant
Highlight	Layout	"<center>"	Malware, Carding
	Color	"#FF0000"	Malware, Carding
	Font Style	"<strong>"	Malware, Carding
Hyperlink	External URLs	"shop address: http://octavian.su"	Carding

tensor network step predicts the sentiment orientation score of each sentence based on the vector representation. It parses the sentence first, resulting in a binary tree of phrases. In a bottom-up fashion, child phrase vectors are recursively merged into their parent phrase vectors with the same compositionality function until it reaches the vector representation of the entire sentence. The sentiment scores aggregation step collects the sentiment score predictions for each replying post and averages them into product/service feedback score and then into seller feedback score. Product/service feedback score evaluates the customer feedback for the product/service promoted in each thread. Subsequently, the seller feedback score evaluates the overall customer feedback for the products or services provided by each seller. We identify sellers with higher seller feedback scores as the top sellers.

V. RESEARCH HYPOTHESES

To test the effectiveness of our framework, we propose research hypotheses on established evaluation metrics. In [18], the Maximum Entropy classifier outperformed other classifiers in multiple established benchmark multiclass categorization datasets. We believe the Maximum Entropy classifier will also outperform others in classifying black market threads.

- *H1*: Maximum Entropy classifier will outperform Naïve Bayes, Support Vector Machine, and k-Nearest Neighbor in categorizing *malware selling* threads in terms of precision, recall, and f-measure.

- *H2*: Maximum Entropy classifier will outperform Naïve Bayes, Support Vector Machine, and k-Nearest Neighbor in categorizing *carding promotion* threads in terms of precision, recall, and f-measure.

Deep learning applications achieved better results than shallow classifiers in sentiment analysis on customer reviews [11]. We believe deep learning-based sentiment analysis will have better performance on hacker customer reviews than other state-of-the-art sentiment classification techniques.

- *H3*: Deep learning-based sentiment classifier will outperform Naïve Bayes, Support Vector Machine, and SentiWordNet in categorizing hacker product/service feedback sentiment in terms of precision, recall, and f-measure.

VI. RESULTS AND DISCUSSION

A. Evaluation

Two experiments were conducted to evaluate the effectiveness and the validity of the major components.

In the first experiment, we compared our system against widely used text classifiers: SVM, Naïve Bayes, and kNN. We trained comparison classifiers using the same feature set. Table II shows the experiment result for thread classification. The MaxEnt classifier dominated both classes on both recall and f-measure, indicating that it is able to find more actual malware selling threads and carding promotion threads and that its overall performance is the best. In the second experiment, we compared the deep learning-based sentiment classifier, recursive neural tensor network, against SVM, Naïve Bayes and SentiWordNet [19]. Learning-based sentiment classifiers were trained based on a feature set comprising bag-of-words and part-of-speech tags. Table II shows the experiment results for sentiment analysis. The deep learning-based sentiment classifier dominated positive class on precision, recall, and f-measure and negative class on precision and f-measure. The sentiment analysis performance in our experiment is around 90%, which is higher than many of the previous sentiment classification studies on other datasets.

Pair-wise t-tests were conducted to test the research hypotheses for both experiments (Table III). In the first experiment, the MaxEnt classifier significantly outperformed comparison methods on recall and f-measure in both malware selling threads classification and carding promotion threads classification. On precision, the MaxEnt classifier significantly outperforms Naïve Bayes and kNN in classifying both categories of threads. Overall, the results mostly support *H1* and *H2*. In the second experiment, the deep learning-based sentiment classifier significantly outperformed its comparison methods on f-measure. It also significantly outperformed the Naïve Bayes and SentiWordNet on recall. Overall, the results mostly support *H3*.

TABLE II. THREAD CLASSIFICATION PERFORMANCE COMPARISON

	Experiment 1					
	Malware			**Carding**		
	Precision	**Recall**	**F-measure**	**Precision**	**Recall**	**F-measure**
ME	94.12%	**50.00%**	**65.31%**	98.51%	**74.16%**	**84.61%**
NB	18.00%	28.13%	21.95%	56.31%	65.17%	60.42%
SVM	**100%**	18.75%	31.58%	**100%**	61.80%	76.39%
kNN	28.13%	28.13%	28.13%	73.33%	74.16%	73.74%
	Experiment 2					
	Positive			**Negative**		
	Precision	**Recall**	**F-measure**	**Precision**	**Recall**	**F-measure**
RNTN	**93.87%**	**98.00%**	**95.89%**	**94.90%**	85.32%	**89.86%**
NB	92.89%	88.80%	90.80%	76.67%	84.40%	80.35%
SVM	89.34%	97.20%	93.10%	91.95%	73.39%	81.63%
SWN	93.68%	64.96%	76.72%	50.26%	**88.99%**	64.24%

[a] ME: Maximum Entropy; NB: Naïve Bayes; SVM: Support Vector Machine; kNN: k-Nearest Neighbor; RNTN: Recursive Neural Tensor Network; SWN: SentiWordNet

B. *Top Sellers*

Our deep learning-based top seller identification framework was performed on the research testbed. We were able to quantify the overall product/service quality of each seller. We normalize our prediction into a 1-to-5 Likert scale. The top/worst 5 malware sellers and carding sellers and their quality scores are listed in Table IV. A major finding is that malware sellers generally tend to have higher ratings than carding. This phenomenon is attributable to the property of the goods they are selling. The quality of malwares is easier to determine than that of carding information. We censored their screen names to protect their identities.

VII. CONCLUSION AND CONTRIBUTION

Carders providing top products/services are of particular interest to cybersecurity researchers and practitioners. However, little research has attempted to identify these product/service providers using automated methods. In this paper, we present a deep learning-based sentiment analysis framework for identifying top sellers in an underground economy. In general, our proposed techniques outperformed their comparison methods and were applied to identify top/worst malware sellers and carding sellers. Our research contributes to underground economy research as it provides a scalable and generalizable automated framework to identify key sellers in the underground economy. Future directions for this research include expanding the dataset to IRC channels and extending analytics to provide more intelligence.

TABLE III. P-VALUES FOR PAIR-WISE T-TESTS FOR PROPOSED METHOD VERSUS ALTERNATIVE TECHNIQUES

	H1		
	NB	**SVM**	**kNN**
Precision	<0.001	0.058	<0.001
Recall	0.023	0.005	0.063
F-measure	<0.001	0.005	0.001
	H2		
	NB	**SVM**	**kNN**
Precision	<0.001	0.343	0.001
Recall	0.349	0.188	0.857
F-measure	0.002	0.210	0.078
	H3		
	NB	**SVM**	**SWN**
Precision	0.908	0.109	0.993
Recall	<0.001	0.306	<0.001
F-measure	<0.001	0.093	<0.001

[b] MaxEnt: Maximum Entropy; NB: Naïve Bayes; SVM: Support Vector Machine; kNN: k-Nearest Neighbor

TABLE IV. TOP/WORST MALWARE AND CARDING SELLERS

	Malware		**Carding**	
Rank	**User**	**Score**	**User**	**Score**
	Top 5			
1	Perf****rypt	5	Dre***own	5
2	Di***Man	4	Bi***yer	4
3	Dj***lf	4	Bu***s11	4
4	D**X	4	DEDIC***LE123	4
5	Rob***ood	4	se**cc	4
	Worst 5			
1	r**t	1.5	rio***ray	1
2	w3***370	1.6	m***an	1
3	g***en	2	j**aa	2
4	ma***ll	2	GGr***rez	2
5	q***3b	2	jD***am	2

ACKNOWLEDGMENT

This work was support in part by the National Science Foundation under Grant No. SES-1314631 and DUE-1303362.

REFERENCES

[1] T. J. Holt, “Examining the Forces Shaping Cybercrime Markets Online,” *Soc. Sci. Comput. Rev.*, vol. 31, no. 2, pp. 165–177, Sep. 2012.

[2] Y. Bengio, “Learning deep architectures for AI,” *Found. Trends® Mach. Learn.*, vol. 2, no. 1, pp. 1–127, 2009.

[3] C. Herley and D. Florêncio, “Nobody sells gold for the price of silver: Dishonesty, uncertainty and the underground economy,” in *Economics of Information Security and Privacy*, T. Moore, D. Pym, and C. Ioannidis, Eds. Boston, MA: Springer US, 2010, pp. 33–53.

[4] M. Fossi, E. Johnson, and D. Turner, “Symantec report on the underground economy,” *Symantec Corp.*, vol. 3, no. 1, pp. 77–82, 2008.

[5] T. Zeller, “Black market in stolen credit card data thrives on Internet,” *N. Y. Times*, 2005.

[6] P. Pavlou and A. Dimoka, “The nature and role of feedback text comments in online marketplaces: Implications for trust building, price premiums, and seller differentiation,” *Inf. Syst. Res.*, vol. 17, no. 4, pp. 392–414, 2006.

[7] B. Pang and L. Lee, “Opinion mining and sentiment analysis,” *Found. Trends Inf. Retr.*, vol. 2, no. 2, pp. 1–135, 2008.

[8] B. Liu, “Sentiment analysis and subjectivity,” *Handb. Nat. Lang. Process.*, pp. 1–38, 2010.

[9] A. Pak and P. Paroubek, “Twitter as a Corpus for Sentiment Analysis and Opinion Mining.,” in *LREC*, 2010.

[10] M. Taboada, J. Brooke, and M. Tofiloski, “Lexicon-based methods for sentiment analysis,” *Comput. Linguist.*, vol. 37, no. 2, pp. 267–307, 2011.

[11] R. Socher, A. Perelygin, J. Y. Wu, J. Chuang, C. D. Manning, A. Y. Ng, and C. Potts, “Recursive deep models for semantic compositionality over a sentiment treebank,” in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2013, pp. 1631–1642.

[12] L. Deng, “A tutorial survey of architectures, algorithms, and applications for deep learning,” *APSIPA Trans. Signal Inf. Process.*, vol. 3, no. January, 2014.

[13] T. J. Holt, “Exploring the social organisation and structure of stolen data markets,” *Glob. Crime*, vol. 14, no. 2–3, pp. 155–174, May 2013.

[14] J. Zhuge, T. Holz, C. Song, J. Guo, X. Han, and W. Zou, *Studying malicious websites and the underground economy on the Chinese web*. Boston, MA: Springer US, 2009.

[15] V. A. Benjamin and H. Chen, “Machine learning for attack vector identification in malicious source code,” *2013 IEEE Int. Conf. Intell. Secur. Inform.*, pp. 21–23, Jun. 2013.

[16] H. Chen, “IEDs in the Dark Web: Genre classification of improvised explosive device web pages,” in *Intelligence and Security Informatics, 2008. ISI 2008. IEEE International Conference on*, 2008, pp. 94–97.

[17] T. J. Holt and E. Lampke, “Exploring stolen data markets online: products and market forces,” *Crim. Justice Stud.*, vol. 23, no. 1, pp. 33–50, Mar. 2010.

[18] C. Manning and D. Klein, “Optimization, maxent models, and conditional estimation without magic,” in *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology: Tutorials-Volume 5*, 2003, pp. 8–8.

[19] S. Baccianella, A. Esuli, and F. Sebastiani, “SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining.,” *LREC*, vol. 10, pp. 2200–2204, 2010.