

# Raising Fraud Awareness through Web Forums

Vrizlynn L. L. Thing

Cybercrime & Security Intelligence Department  
Institute for Infocomm Research, Singapore

vriz@i2r.a-star.edu.sg

**Abstract**—Topic specific forums, which contain a vast amount of information can aid in providing very useful and informative advice to their readers. Forums that are specific to scam complaints (and reporting) can also aid in raising public awareness to new scams and fraudulent activities, and provide the support for a more pro-active approach to the early detection and prevention of fraud. Accurate and efficient provision of contents from forum sites are therefore very important, to provide information to aid in preventive measures. In this paper, we acquire data from 6 popular and active scam reporting forums with varying ages (from 1.07 to 9.45 years old). We then carry out an analysis to investigate the ability to extract posts detailing victims' encounter with scams and fraud, based on the coverage via simple searches on specific keywords and keyword combinations. We also carry out an evaluation of the merchant coverage in each forum and investigate the association of keywords to support future reliable informative data provision from both topic-specific, and generic forums and online sources.

**Index Terms**—Fraud detection, fraudulent merchant, fraudulent activity analysis, scam, complaint, forum.

## I. INTRODUCTION

The widespread use and contributions of knowledge in the form of data uploaded to the Internet has made it a wealthy source of information for any conceivable topics. One of the most important platforms on the Web is the online forums. Online web forums' dynamically increasing contents, which is contributed by millions of Internet users on a daily basis, has led to its increasing richness of information. Its widespread popularity is its facilitation of global, convenient, fast and freely open discussions. Therefore, web forum data is an accumulation of a vast collection of updated human knowledge and viewpoints. Forums can thus be a highly valuable source of online information for knowledge acquisition to build up domain expertise [1], improve business intelligence [2], [3], [4], and early detection of the presence (and study) of extremist activities [5], [6], [7], [8], [9].

In [5], the authors proposed a framework for Web forum data integration to support the analysis of interactions among discussion participants. The targetted forums were Jihadist forums. The authors introduced features such as forum browsing and searching, multi-lingual translation and social network visualization in their work to support the early detection of extremism activities.

In [7], the authors carried out an analysis of U.S. and Middle Eastern extremist group forums. An affect lexicon based on probabilistic disambiguation technique was proposed to measure the presence of hate and violence related words

in the forums' contents. The authors concluded that a strong linear relationship exists between the usages of hate and violence related words in the Middle Eastern extremist group forums.

In [8], the authors evaluated the usage of stylistic and syntactic features for the sentiment classification of English and Arabic contents in Web forums. The authors concluded that the stylistic features and their proposed entropy weighted genetic algorithm (incorporating information-gain heuristic for feature selection) could significantly enhance the sentiment classification.

In [3], the authors conducted an experimental study by asking consumers to gather online information on a specific product topic by accessing Web forums. The authors concluded that consumers who acquired information from online forum discussions reported a greater interest in the selected product topic than those who acquired information from marketer-generated sources.

In [4], the authors proposed a scoring technique to evaluate specific product reviews and to summarize the opinions of the product to the user. The methodology enables the user to save time on reading all the reviews and at the same time, arrive at a generic opinion of a product based on the reviews posted on Web forums.

In [9], the authors proposed incorporating message content similarity and response immediacy to measure the degree of influence between any two users on Web forums. To ensure an accurate approach of measurement, the authors proposed the design of weight application and integration to the typical user link analysis technique. The evaluation of the proposed algorithms was carried out using the ACM Intelligence and Security Informatics KDD challenge to show the potential in identifying influential users.

However, there is no existing work which looks at the analysis of fraudulent and scam related activity reporting forums. Due to the important information they can provide, we think it is necessary to be equipped with an understanding of fraudulent and scam related activity reporting forums. Therefore, in this work, we cover the analysis of fraudulent and scam related activity reporting forums by collecting and analysing a set of popular forums that provide a platform for consumers to report their encounters as victims of scams and frauds. Our work will enable a better understanding of such forums to support the raising of public awareness to new scams and fraudulent activities in the wild, and the early detection of

such activities and potential merchants/companies involvement or association.

There are existing works in the area of forum crawling [10], [11], [12], [13], [14] and its content extraction [15], [16], [17], [18], [19]. In this work, we focus on forum content analysis, specifically in scam reporting forums. To the best of our knowledge, this is the first work which carries out an analysis of forum data on fraudulent and scam related activities.

Our main contributions in this paper are:

- 1) the collection of complete fraudulent and scam related activity reporting posts from active forums ranging from the age of 1.07 to 9.45 years
- 2) the generation of keywords relevant to fraudulent and scam related activities from the preliminary analysis of online sources of incidents reporting
- 3) the preparation of the list of companies reported in the scambook forum
- 4) the analysis of the forums and evaluation of the ability to detect posts detailing fraudulent and scam related activities and events, based on i) our single keyword based analysis, and ii) keyword combination based analysis
- 5) the evaluation of the merchants (or companies) coverage in each forum, and the investigation of keyword association with each merchant

This work will be valuable in i) providing an in-depth understanding of current popular forum sites related to fraudulent and scam related activity reporting, ii) enabling us to make recommendations based on the findings from this research, and iii) generating top relevant keywords as supporting features to detect merchants and activities related to fraud and scams in both topic-specific and generic online sources.

The rest of the paper is organised as follow. In Section II, we describe our target forums and carry out a preliminary analysis to obtain useful statistics. In Section III, we propose the analysis of the forum post data/content based on our generated keyword list, and present and discuss our results. In Section IV, we extract companies' names from the scambook forum, propose the procedural steps to clean the list to prevent high false positives and false negatives during detection, and analyse the coverage of these companies in each forum. We also investigate the association of keywords with each of these companies based on the post contents in the forums. In Section V, we provide the recommendations to improve the applicability and usefulness of the forums in raising public awareness to new scams, and to support the early detection of fraudulent merchants and activities, so as to enable a more pro-active approach in the handling of fraud and scams. We summarise the important findings in Section VI.

## II. COLLECTION OF DATA FROM SCAM REPORTING FORUMS

For our forum analysis research, we collect the contents from the following 6 scam reporting forums, namely exposeascam [20], realscam [21], scambaits [22], scambook [23], scamfound [24], and scamvictimsunited [25]. These forums allow users to post reports and complaints of their encounter

with scam related incidents. We analyse the dates of the posts in the collected contents to obtain the first date of the post (i.e., a forum's start date) and the last date of the post (till the end date of our retrieval of all the posts from each forum) per forum, and compute their ages. However, an older age does not imply that a forum is more active. We extract the total threads and posts we find in each forum, and present the information together with the forum's age in Table I. We notice that the age of the forums differs very widely. We also notice that the activeness (i.e., the average threads/posts per day/week/month, and the gaps between no posting activity) of the forums differs too. Therefore, when conducting analysis in the subsequent sections, we will carry out normalization for a fair analysis when necessary.

Forum	Total Threads	Total Posts	Age (years)
exposeascam	2910	3439	1.07
realscam	1980	27264	2.28
scambaits	1677	9848	6.67
scambook	116430	116430	1.35
scamfound	242846	244248	3.08
scamvictimsunited	3354	16418	9.45

TABLE I: Forum Statistics

## III. KEYWORD BASED ANALYSIS

To analyse the forum contents, we first generate a list of keywords to identify the applicability of scam related keywords in the detection of posts that provide details on the relevant incidents. From our observation of scam reports and consumer complaints online, we notice that the 28 keywords in Table II are often used. Therefore, we generate the list of keywords based on Table II for the keyword based analysis of the collected forums' data.

fraud	cheat	transaction	unfair
charge	liar	unauthorize	bill
rip-off	illegal	invalid	scam
fee	unethical	drug	compensate
hidden	attack	ripoff	refund
steal	unjust	unauthorise	porn
compensation	defect	damage	rip

TABLE II: Keyword List

Based on the keyword list, we analyse the forums' data and identify the posts that contain any of the keyword/s. Each word (separated by at least a word delimiter such as a space, tab, comma, full stop) is extracted from the post contents and a strict matching (i.e., not substring) with the keywords is applied. The keyword matching against the contents of the posts also provides us with the posts that are closely related to scam activities, for further analysis.

Next, we analyse the frequency of keywords found in the above identified posts. We analyse the posts to evaluate the keyword frequency by returning the post count for each keyword. In addition, we consider the large deviation in the forum sizes (i.e., number of threads/posts) and activeness, and thus

carry out a normalization of the keyword frequency against the total posts per forum, to present the percentage of the identified posts containing each keyword per forum, in Table III. Note that each post may contain more than one keyword. Therefore, the total percentage per forum may be over 100%. We also compute and show the average normalized frequency of each keyword across all the forums. We observe that the top 10 keywords, in decreasing order according to their respective average normalized keyword frequency percentage, are scam, fee, fraud, damage, charge, rip, bill, refund, transaction, and liar. We can also see that 94.99% to 100% of the detected posts contain the keyword “scam” across all the 6 forums.

Next, we investigate the applicability and frequency of the combinations of keywords in the detection of scam related activities. We identify posts with contents that match any combination of the 28 keywords, and compute the number of posts matching a strict keyword combination (i.e., if the post content contains 3 different keywords, the post count will be incremented by 1 for this 3-keyword combination only. This computation is different from the 1-keyword based analysis where a post having 2 different keyword matches will have each post count incremented by 1 for each specific keyword. The keyword combination based analysis also ignores the order of the keyword appearance in the post contents.). We then extract the top 10 keyword combinations (based on the post counts) for each forum. To give a better view of the coverage of the detected posts on scam related activities based on each top keyword combination, we compute the normalized coverage in terms of percentage. The normalization is carried out over the total number of detected posts with any keyword occurrence.

The normalized coverage provided by the top 10 keyword combinations is shown in Table IV, with the total normalized coverage percentage for each forum shown in the last row of each sub-table (in bold). We observe from Table IV that with our chosen list of keywords, the top 10 keyword combination can identify 65.60% to 93.55% of the posts related to fraudulent and scam activities.

#### IV. COMPANY BASED ANALYSIS

In this section, we analyse each forum based on their ability to identify popular companies mentioned in scam reporting forums. We retrieve the popular company list from the scambook forum. The scambook forum provides a list of the most popular companies based on their site’s post data.

We retrieve the list of 1986 company names but notice that the list contains several names that may potentially generate high false positives (and false negatives) in our analysis results. Therefore, before we proceed, we clean up the company list according to the following steps (with real examples given from the original company list from the scambook forum).

- 1) Remove names with only numeric characters (e.g. 2012)
- 2) Remove all 1-character names
- 3) Remove all 2-character names if they contain only alphabetic characters (e.g. UK, SG, OK)

- 4) Remove trailing words if they are location name following a company name (e.g. “, London”, “, Oxford”) so that posts reporting a company in another location can also be detected
- 5) Remove trailing words if they are in short form and depict the company’s liability or taxation type (e.g. Int’l, Ltd, LLP, Co, Inc, LLC)
- 6) Remove top level domain name if the company name is distinct enough without it (that is, do not remove the top level domain name if the company name is for example, cars.com or lends.net)

Other than cleaning up the company list by removing the less effective detection terms, we also generate new company names based on sub-string extraction of long company names if the original company names are too specific and may potentially result in high false negatives. An example is “Gameest Int’l Network Sales” where we additionally create another company name in the list for “Gameest”. The final list contains 2019 company names.

As with the keyword based analysis, we also carry out a strict form of matching for the company based analysis. In addition, since we are carrying out the company based analysis using the company names generated from the scambook forum, we exclude this forum from some experiments in this section to remove the unfair bias in the analysis and evaluation.

First, we carry out an investigation on the number of company names that are reported in the post contents in each forum, and present the results in Table V.

We observe that there is no major overlap between the other forums and the scambook’s existing reported companies, with the exception of the scamfound forum. 47.60% of the companies in the scambook forum can be seen as reported in the scamfound forum. To better analyse and conclude on the quality of the forums, it is necessary to understand if the other forums do detect other additional companies not included in the scambook forum’s list (i.e., not reported by scambook members and users). However, the other forums do not provide a company name list compilation. A fair evaluation and comparison of all the forums should be carried out if such lists are provided in future.

Next, we identify the posts in each forum that report the companies in our list and compute the number of posts for each company per forum. We then identify the top 20 detected company names in each forum based on the number of posts reporting them. We observe from the results that some detected company names are not exactly distinctive as a company name. Some obvious examples are “not sure”, “personal”, “unknown” and “individual”. It is important to note that if forums are to provide company name lists to aid in the detection of fraudulent and scam related activities, they need to be better maintained and cleaned up. Provision of such lists will be very useful in raising the awareness on the fraudulent merchants to look out for.

Another interesting observation from the results is the detection of legitimate companies such as McDonalds, Walmart

	exposeascam	realscam	scambaits	scambook	scamfound	scamvictimsunited	Average Normalized Percentage
attack	2.09	3.34	0.70	0.21	0.03	0.59	1.16
bill	13.99	5.19	4.49	15.25	2.20	2.99	7.35
charge	17.53	3.65	1.95	42.09	6.04	4.64	12.65
cheat	2.33	0.61	0.28	0.78	0.69	0.63	5.32
compensate	0.26	0.22	0.28	0.14	0.01	0.11	0.17
compensation	0.44	1.12	0.48	0.33	0.06	0.31	0.46
damage	2.27	1.51	0.41	100.00	0.47	0.54	17.61
defect	0.90	0.09	0.03	0.93	0.47	0.06	0.41
drug	0.87	0.78	0.47	0.28	0.15	0.35	0.48
fee	18.67	22.35	10.18	16.26	97.64	11.52	29.44
fraud	23.18	12.33	10.33	11.96	5.76	13.82	12.90
hidden	1.63	0.58	0.12	0.30	0.15	0.20	0.50
illegal	1.63	4.14	0.55	1.83	0.53	0.85	1.59
invalid	0.23	0.14	0.40	0.28	0.05	0.08	0.20
liar	2.65	2.85	0.84	1.08	0.41	1.21	1.51
porn	0.23	2.65	0.15	0.19	0.03	0.03	0.55
refund	14.36	1.53	0.20	13.89	2.40	1.73	5.69
rip	18.26	18.41	4.54	10.85	3.18	3.67	9.82
rip-off	0.55	0.21	0.02	0.19	1.31	0.03	0.39
ripoff	3.05	0.45	0.00	0.75	0.34	0.18	0.80
scam	100.00	99.99	99.92	100.00	100.00	94.99	99.15
steal	5.55	1.40	0.92	1.67	0.39	1.13	1.84
transaction	2.30	0.98	5.15	6.91	0.27	2.85	3.08
unauthorise	0.26	0.01	0.01	0.91	0.31	0.01	0.25
unauthorize	1.22	0.13	0.16	10.18	2.22	0.13	2.34
unethical	1.42	0.46	2.73	0.18	0.36	0.06	0.87
unfair	0.81	0.45	0.11	0.23	0.30	0.14	0.34
unjust	0.06	0.12	0.03	0.04	0.03	0.07	0.06

TABLE III: Normalized Keyword Frequency - Post Count Per Keyword (in Percentage)

exposeascam	realscam	scambaits
scam: 30.13	scam: 45.94	scam: 69.24
fraud,scam: 10.32	rip,scam: 10.63	fraud,scam: 5.59
fee,scam: 5.21	fee,scam: 10.20	fee,scam: 4.81
rip,scam: 5.21	fraud,scam: 5.47	bill,scam: 3.01
refund,scam: 4.01	bill,fee,scam: 1.48	fraud,scam,unethical: 2.52
charge,scam: 3.95	fee,rip,scam: 1.40	rip,scam: 2.46
bill, fraud, scam: 3.75	porn, scam: 1.20	fee, scam, transaction: 1.88
bill, scam: 1.48	illegal, scam: 1.26	scam, transaction: 1.36
attack, bill, fraud, scam: 1.42	fee, fraud, scam: 1.18	charge, scam: 0.57
charge, fee, scam: 1.31	bill, scam: 0.95	scam, steal: 0.43
<b>66.79</b>	<b>79.71</b>	<b>91.85</b>
scambook	scamfound	scamvictimsunited
damage, scam: 28.23	fee, scam: 75.96	scam: 64.80
charge, damage, scam: 12.63	fraud, scam: 4.61	fraud, scam: 8.22
damage, fee, scam: 3.76	charge, fee, scam: 3.52	fee, scam: 5.81
damage, refund, scam: 3.37	fee, refund, scam: 1.91	rip, scam: 1.56
bill, charge, damage, scam: 3.21	scam: 1.91	charge, scam: 1.49
damage, fraud, scam: 3.12	bill, fee, scam: 1.54	bill, scam: 1.24
bill, damage, scam: 3.04	charge, fee, scam, unauthorize: 1.23	fee, fraud, scam: 1.13
charge, damage, scam, unauthorize: 3.03	fee, rip, rip-off, scam: 1.20	scam, transaction: 1.09
damage, rip, scam: 2.95	fee, rip, scam: 1.05	fee: 0.90
charge, damage, fee, scam: 2.28	fee, scam, unauthorize: 0.63	fraud: 0.73
<b>65.60</b>	<b>93.55</b>	<b>86.96</b>

TABLE IV: Top 10 Keyword Combinations for Each Forum (with Normalized Post Coverage in Percentage)

Forum	Number of Companies	Percentage of Companies
exposeascam	121	5.99
realscam	92	4.56
scambaits	54	2.67
scamfound	961	47.60
scamvictimsunited	79	3.91

TABLE V: Number and Percentage of Companies Detected in Forum

and Apple, with a high number of reported cases (i.e., in terms of the number of posts). A highly probable reason is the use of these legitimate platforms and their resources by scammers and fraudulent merchants to carry out scam related activities (e.g., advertising). In the case where these companies offer legitimate and highly popular products, the reports may also be associated with counterfeit products being advertised or sold as legitimate ones by the fraudulent merchants.

Other than that, we also observe that some companies are actually reported to be directly linked to complaints of scams and fraudulent activities. Some examples are C2 and C15 (as shown in Table VI), which have been reported in the forums to be associated with feedbacks such as delivering skin-care products that caused serious negative reactions, charging customers' credit cards without authorization, and/or recursively, or that they are uncontactable for feedback/refund thereafter.

Next, we carry out an analysis to identify the keywords associated with a selected set of the detected companies. By "associated", we do not mean that the keywords are indicative of the description of the company's activities. We mean that the keywords as well as the company name are within the contents of a same post.

For the company name and associated keyword analysis, we eliminate detected companies which do not have distinctive company names, or are well-established legitimate, high set-up cost companies, financial institutions and multi-national companies. We notice that the remaining companies are mainly online merchants or shops associated with multi-level marketing, pharmaceutical products, dating/matchmaking, advertising, etc. We extract the top keyword combination associated with each selected company name from each forum, consolidate them across all the forums, and present the selected companies from the top 20 detected companies and the associated keywords (as found in the post contents) in Table VI. In this table, company names are modified to preserve their identities as the objective of this analysis is simply to identify useful keywords associated with companies being flagged or complaint against.

While searching for the top keyword combination for the selected companies in the forums, we notice that even though some companies are not in the top 20 results of some forums, they do exist within the forums' post contents. We compile a list to indicate the presence or absence of the selected companies within each forum, and present the results in Table VI. Since the company list is generated from the scambook forum, it is excluded from this analysis.

From Table VI, we can see that there is a significant overlap in the presence of the detected top companies among the forums. However, the total overlap for most companies is minimal in some forums. Therefore, to enable a better detection of the fraudulent and scam related merchants and activities, we should rely on the detection results from multiple sources and carry out correlations, to obtain a better detection accuracy with a low false positive rate. There is also a need to eliminate false positives due to the wide presence of well-known legitimate sources. This elimination can be through a whitelist configuration and should only be implemented when it is definite that these companies do not provide resources that may be exploited by fraudulent merchants and scammers. However, a scenario that may not be avoidable is when scammers exploit the well-established reputation of such legitimate companies and use these company names to carry out malicious activities such as scamming and phishing.

## V. DISCUSSION

Based on this research and the observations, an important recommendation is the need for the provision of well-maintained fraudulent merchants or company list by scam reporting forums. The availability of this resource will enhance the value of these forums and fulfill their main purpose in providing readers with valuable information on the fraudulent merchants and scams to avoid. In addition, such lists and information could also be used by companies providing e-payment services to monitor the on-going status and reputation of their registered merchants, so as to take immediate action in the event of any violation of their terms and policies.

As our work is to investigate the possibility to raise public awareness to scams and fraud, and to enable the early detection of such malicious activities in the wild, it is necessary to ensure the quality of the detected results to prevent false triggering for investigations. The first and most important step would be to ensure the reliability and trustworthiness of the information from the online sources and forums. Scam reporting forums can incorporate moderation of the posts submitted to their forums to ensure that they are accurate through the provision of concrete supporting evidence (e.g. legal incident report, transaction statement) from the incident reporting user. This step may incur an additional overhead but is essential in ensuring the quality of the data in the forum.

## VI. CONCLUSION

In this paper, we have carried out a fraudulent and scam related activity reporting forum data analysis. We collected the posts from 6 popular and active scam reporting forums, and generated a list of relevant keywords based on our preliminary analysis and knowledge of online sources on scam incident reporting. We then carried out an investigation on the ability to detect posts relevant to fraud and scams based on different keyword-based analysis scenario. We showed through our analysis that the choice of a single keyword can have an average coverage of 99.15% of the posts. However, the single keyword based detection can result in high false positives

Company	Associated Keywords	exposeascam	realscam	scambaits	scamfound	scamvictimsunited
C1	fee,scam				*	
C2	fee,scam				*	
C3	fee,fraud,scam,transaction			*	*	*
C4	charge,scam,transaction,unauthorise	*	*		*	*
C5	bill,charge,fee,scam,unfair	*			*	
C6	bill,charge,fee,illegal,refund,scam,unfair	*	*		*	
C7	fee,scam				*	*
C8	fee,scam				*	
C9	fee,fraud,scam			*	*	*
C10	fee,scam,unethical	*	*		*	*
C11	fee,fraud,scam	*		*	*	*
C12	fee,fraud,scam				*	
C13	fee,fraud,scam,steal		*		*	
C14	fee,scam				*	
C15	bill,fee,refund,rip,scam				*	
C16	fee,fraud,scam	*			*	*
C17	charge,fee,scam,unauthorize				*	
C18	fee,porn,scam,unauthorize,steal	*			*	*
C19	charge,fee,fraud,scam,transaction	*	*	*	*	*
C20	fee,illegal,scam				*	
C21	scam		*			

TABLE VI: Detected Individual Company and Associated Keywords in Post Contents, and Evidence of Presence of Detected Companies in Forums

when the detection feature is applied to generic online source. Therefore, we investigated the different keyword combinations and showed that the identification and selection of the top 10 keyword combinations is sufficient to support 65.60% to 93.55% coverage of the posts in the forums. We also evaluated the coverage of companies in the forums and investigated the association of keywords with each company. Our results showed that the merchant coverage in the forums is sufficiently wide for the identified popular companies. Based on our findings, we proposed some important recommendations to improve and enhance the applicability of forums and online sources in raising public awareness to new scams and fraud, and the early detection and prevention of such incidents to new potential victims.

REFERENCES

[1] J. Zhang, M. S. Ackerman, and L. Adamic, "Expertise networks in online communities: structure and algorithms," in *WWW Conference*, 2007, pp. 221–230.

[2] N. Glance, M. Hurst, K. Nigam, M. Siegler, R. Stockton, and T. Tomokiyo, "Deriving marketing intelligence from online discussion," in *ACM SIGKDD International Conference on Knowledge discovery in Data Mining*, 2005, pp. 419–428.

[3] B. Bickart and R. M. Schindler, "Internet forums as influential sources of consumer information," vol. 15, no. 3, pp. 31–40, 2001.

[4] S. Hariharan, R. Srimathi, M. Sivasubramanian, and S. Pavithra, "Opinion mining and summarization of reviews in web forums," in *ACM Bangalore Conference*, 2010.

[5] Y. Zhang, S. Zeng, L. Fan, Y. Dang, C. A. Larson, and H. Chen, "Dark web forums portal: searching and analyzing jihadist forums," in *IEEE International Conference on Intelligence and Security Informatics*, 2009, pp. 71–76.

[6] Y. Zhou, J. Qin, G. Lai, and H. Chen, "Collection of u.s. extremist online forums: A web mining approach," in *Annual Hawaii International Conference on System Sciences*, 2007, p. 70.

[7] A. Abbasi and H. Chen, "Affect intensity analysis of dark web forums," in *IEEE International Conference on Intelligence and Security Informatics*, 2007, pp. 282–288.

[8] A. Abbasi, H. Chen, and A. Salem, "Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums," vol. 26, no. 3, 2008.

[9] C. Yang, X. Tang, and B. Thuraisingham, "An analysis of user influence ranking algorithms on dark web forums," in *ACM SIGKDD Workshop on Intelligence and Security Informatics*, 2010.

[10] R. Cai, J.-M. Yang, W. Lai, Y. Wang, and L. Zhang, "iRobot: An intelligent crawler for web forums," in *WWW Conference*, 2008, pp. 447–456.

[11] Y. Wang, J.-M. Yang, W. Lai, R. Cai, L. Zhang, and W.-Y. Ma, "Exploring traversal strategy for web forum crawling," in *ACM SIGIR International Conference on Research and Development in Information Retrieval*, 2008, pp. 459–466.

[12] J.-M. Yang, R. Cai, C. Wang, H. Huang, L. Zhang, and W.-Y. Ma, "A threadwise strategy for incremental crawling of web forums," in *WWW Conference*, 2009.

[13] A. Sachan, W.-Y. Lim, and V. L. L. Thing, "A generalized links and text properties based forum crawling," in *IEEE/WIC/ACM International Conference on Web Intelligence*, 2012, pp. 113–120.

[14] H.-M. Ying and V. L. L. Thing, "An enhanced intelligent forum crawler," *IEEE Symposium on Computational Intelligence for Security and Defence Applications*, pp. 1–8, 2012.

[15] S. Li, L. Tang, J. Hu, and Z. Chen, "Automatic data extraction from web discussion forums," in *Proceedings of the 2009 Fourth International Conference on Frontier of Computer Science and Technology*, 2009, pp. 219–225.

[16] S. Pretzsch, K. Muthmann, and A. Schill, "Fodex—towards generic data extraction from web forums," in *Advanced Information Networking and Applications Workshops (WAINA), 2012 26th International Conference on*. IEEE, 2012, pp. 821–826.

[17] J.-M. Yang, R. Cai, Y. Wang, J. Zhu, L. Zhang, and W.-Y. Ma, "Incorporating site-level knowledge to extract structured data from web forums," in *WWW Conference*, 2009, pp. 181–190.

[18] W.-Y. Lim, A. Sachan, and V. L. L. Thing, "A lightweight algorithm for automated forum information processing," in *IEEE/WIC/ACM International Conference on Web Intelligence*, 2013, pp. 121–126.

[19] W.-Y. Lim, V. Raja, and V. L. L. Thing, "Generalized and lightweight algorithms for automated web forum content extraction," in *IEEE International Conference on Computational Intelligence and Computing Research*, 2013.

[20] Exposeascam, <http://www.exposeascam.com>.

[21] Realscam, <http://www.realscam.com>.

[22] Scambaits, <http://www.scambaits.net>.

[23] Scambook, <http://www.scambook.com>.

[24] Scamfound, <http://www.scamfound.com>.

[25] Scamvictimsunited, <http://www.scamvictimsunited.com>.