A Stacking Approach to Objectionable-Related Domain Names Identification by Passive DNS Traffic (Short Paper)

Chen Zhao[1,2], Yongzheng Zhang[1,2(✉)], Tianning Zang[1,2], Zhizhou Liang[1,2], and Yipeng Wang[1,2]

[1] School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China
{zhaochen,zhangyongzheng,zangtianning,wangyipeng}@iie.ac.cn
[2] Institute of Information Engineering, CAS, Beijing 100093, China

Abstract. Domain name classification is an important issue in the field of cyber security. Notice that objectionable-related domain names are one category of domain names that serve services such as gambling, pornography, etc. They are classified and even forbidden in some areas, some of these domain names may defraud visitors privacy and property. Timely and accurate identification of these domain names is significant for Internet content censorship and users security. In this work, we analyze the behavior of objectionable-related domain names from the real-world DNS traffic, finding that there exist evidently differences between objectionable-related domain names and none-objectionable ones. In this paper, we propose a stacking approach to objectionable-related domain names identification, VisSensor, that automatically extracts name features and latent visiting patterns of domain names from the DNS traffic and distinguishes objectionable-related ones. We integrate convolutional neural networks with fully-connected neural networks to collaborate features of different dimensions and improve experimental results. The accuracy of VisSensor is 88.48% with a false positive rate of 9.11%. We also compared VisSensor with a public domain name tagging system, and our VisSensor performed better than the tagging system on the identification task of the objectionable-related domain names.

Keywords: Objectionable-related domain name · Traffic analysis · Convolutional neural network

1 Introduction

1.1 Background of Objectionable-Related Domain Names

Domain name system (DNS) is a bridge between the resources on the Internet and the Internet users. The classification of domain names are important

H. Gao et al. (Eds.): CollaborateCom 2018, LNICST 268, pp. 284–294, 2019.
https://doi.org/10.1007/978-3-030-12981-1_20

in the field of cyber security. Many researchers have paid their attentions to this area [5,7,8]. In this paper, we concerned the issue of objectionable-related domain names identification. Objectionable-related domain names are one kind of domain names that related to the objectionable contents such as gambling (e.g. Fig. 1), pornography (e.g. Fig. 2) and other services associated with them (e.g. in Fig. 3, the domain name www.80dytt.com offers pirate medias to attract visitor, and show promotions of ① gambling and ② pornography in its media). The contents of these domain names are harmful for teenager's mental health, and some of these domain names even try to steal users' privacy and property. Current practices on objectionable-related domain names highly rely on manual efforts. However, manual efforts lack of timeliness and cannot fully cover all the active objectionable-related domain names in practice.

Fig. 1. An illegal gambling domain name (01kjz.com).

Fig. 2. A pornography domain name.

Fig. 3. A pirate media platform.

1.2 Contributions

In this paper, we propose VisSensor, a stacking based approach to objectionable-related domain names identification. VisSensor collects the DNS answering traffic from the resolver and transforms the traffic into visiting features and name features of domain names, and automatically classifies the domain names appeared in the traffic into objectionable-related ones and none-objectionable ones. Our approach is based on the key insight that the periodical variations of DNS querying traffic are the embodiments of overall visitor behaviors which strongly indicate the services offered by domain names. We leveraged this characteristic for the identification of objectionable-related domain names.

The key novelty of VisSensor lies in the stacking of convolutional neural networks (CNN) and fully-connected neural networks (NN). This combination enables the collaborate of data with different orders of magnitude. VisSensor integrates the identification results based on DNS querying sequences with the results based on the name features. Moreover, VisSensor has outperformed the domain name tagging works aforementioned on the timeliness and completeness using passive DNS traffic.

The key contribution of our work are listed below:

- We propose a stacking based method that can integrate data with diverse orders of magnitude by stacking convolutional neural networks and fully-connected neural networks together. We apply CNN on high dimensional data and fcNN on simple data. And this collaboration evidently improves the overall classification result than any separate sub modules.
- We propose a stacking based approach of objectionable-related domain names identification, VisSensor, which automatically extracts the latent visiting patterns of domain names from the DNS answering traffic and identifies the objectionable-related domain names from the normal ones. VisSensor consists of five parts: data preprocessing module, training module, stacking module, filter and classification module. We build a prototype of VisSensor based on our design, train and test VisSensor on a real-world DNS traffic. The best sub-model of VisSensor achieves an accuracy of 87.47% and the overall results of VisSensor reach an accuracy of 88.48%.
- We compare our VisSensor with the public accessible URL tagging system of McAfee, trustedsource.org, on the task of identifying objectionable-related domain names. The recall and precision of our VisSensor is 85.07% and 90.89% higher while that of Trusted Source is 4.19% and 10.92% which evidently shown the effectiveness of the VisSensor over the state of arts labeling method.

Our arrangement of this paper is listed as followings: in Sect. 2, we are going to talk about our findings in the study of real world domain name visiting traffic; in Sect. 3, we will describe the features we use in the VisSensor; in Sect. 4, we will introduce the classifiers in VisSensor and show the overall design; Sect. 5 will illustrate the experimental results of VisSensor on a real-world DNS data, and compare the results with one of the state of arts applications in domain name tagging; in the last section, we will discuss about the limitation and application of the VisSensor, and provide our opinions on further study of objectionable domain names.

2 Observations

In this section, we provide an intuitive overview on the different visiting patterns of objectionable-related domain names and none-objectionable domain names. Note that we use DNS queries to refer to the DNS querying packets sent by clients that passively recorded on the recursive resolver side. Motived by the previous works [4,6], we design a new way of visualizing DNS queries. Explicitly, we count the queries for every five-minute span, and we illustrate the relative count of each span by the illumination of its corresponding black and white pixel point.

Given a domain name d, assume its five-minute counts in w days are $P = \{p_1, p_2, \ldots, p_{w\times 288}\}$, then the illumination of point p_i can be denoted as:

$$I_i = \lfloor \frac{p_i}{max(P)} \times 255 \rfloor$$

From this equation, we can say that lager queries counts have lager I_is and consequently have brighter pixel points.

We fill all the $I = \{I_1, I_2, \ldots, I_{w\times 288}\}$ into an image. Along the width of an image is 24 h of a day, and along the height represents 14 days of our sampling time. For example, we choose two typical illustrations of domain names to demonstrate the differences between objectionable-related and none-objectionable. As shown in Figs. 4 and 5: *Zompim.com* is the domain name of a live chat software solution company; *5303008.com* is a gambling website and is reported to have potentially harmful software by the Google Chrome. As mentioned above, brighter pixel points indicate higher queries counts. We can see the bright points of normal domain name *zompim.com* gather around the areas that represent the work time (around 8 a.m. to 6 p.m.), while points of gambling domain name are gradually get brighter after working hours and reach the brightest area at around 9 p.m.

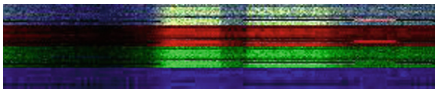

Fig. 4. Zompim.com, a normal commercial domain name.

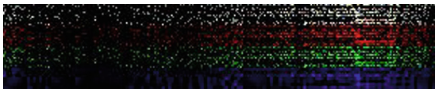

Fig. 5. 5303008.com, a gambling related domain name.

3 Features

3.1 Visiting Features

To quantify a domain name's time sequential accessions, we count the visiting features which compose three kinds of count numbers measured in five-minute grained spans to maintain the visiting details in the passive DNS traffic:

- Query counts denote how many times this domain name is queried;
- Client counts denote how many clients have queried this domain name;
- Network counts denote how many networks that querying clients come from.

We rearrange the arrays of three features into three 14×288 matrices. In matrix form, every time span is located between the five-minute span before and latter in the same row, and the time spans on the same column is the same time on different days. And we stacked the three matrices together, making each matrix as one channel of a $3 \times 14 \times 288$ sized domain name visiting features sample.

3.2 Name Features

In our research, we also find that the objectionable-related domain names appear following characteristics: 1. unreadable; 2. the proportion of numbers in the registrable part of domain names (for example, the registrable part of '12345foo.com' is '12345foo', and numbers take up $\frac{5}{8} = 62.5\%$). This finding

motivates us to impose the naming features that represent how much a domain name matches with these characteristics. Meanwhile, previous works have shown that objectionable-related domain names distribute unevenly among TLDs (top level domains, the right most label of a domain name; all effective domain name should be registered under a TLD) due to the different regulations of TLDs. For these reasons, we profile the name features in two aspects:

– The percentage of numbers in a domain name's registrable part;
– The index of TLD in the one-hot form.

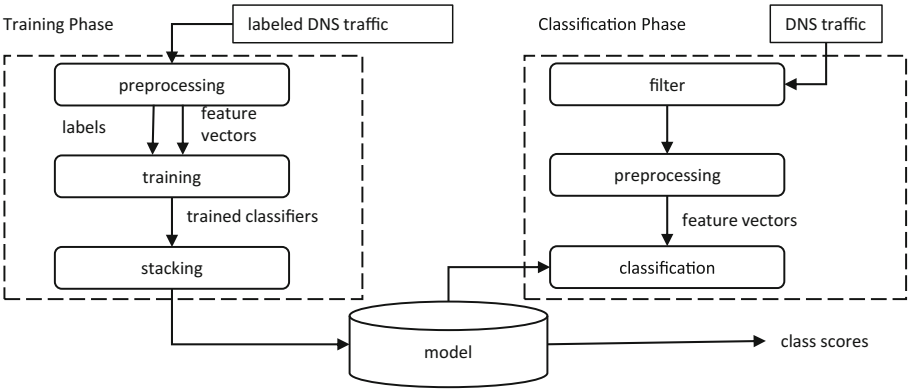

Fig. 6. The overview of VisSensor.

4 The VisSensor

In this section, we present the VisSensor, an ensemble system for classifying domain names into objectionable-related and none-objectionable based on their visiting and name features. We will introduce the core classification model of VisSensor which composed by four classifiers in Sect. 4.1; and we are going to introduce the two phases of VisSensor, the training phase and the classification phase, explicitly in Sects. 4.2 and 4.3. Figure 6 provides an overview of VisSensors.

4.1 Classifiers

In our work, we apply two kinds of neural network classifiers, specifically, fully-connected Neural Network (fcNN) and Convolutional Neural Network (CNN) classifiers. The labels we used are objectionable-related (positive) and none-objectionable (negative), they are known for training purposes.

fcNN Classifiers. FcNN classifiers in VisSensor give objectionable scores according to name features. A fcNN classifier consists of hidden layers with neurons that have learnable weights and biases; each neuron links to all neurons on the previous layer, performs dot product and has an optional non-linearity operation. The fcNN transformed instances on the input end to class scores at the output end.

CNN Classifiers. CNN classifiers aim to extract latent visiting patterns from visiting features and map them to objectionable scores. A CNN classifier is similar to fcNN in the overall structure, but it modified some of the hidden layers into convolutional layers with multi-dimensional neurons that only connect to a small region of the upper layer. The CNN proposed by [9] is able to deal with large scale and high dimensional inputs. Note that we impose dilated convolutional layers [10] in our CNN classifiers to measure the weekly querying characteristics in additional to regular convolutional layers.

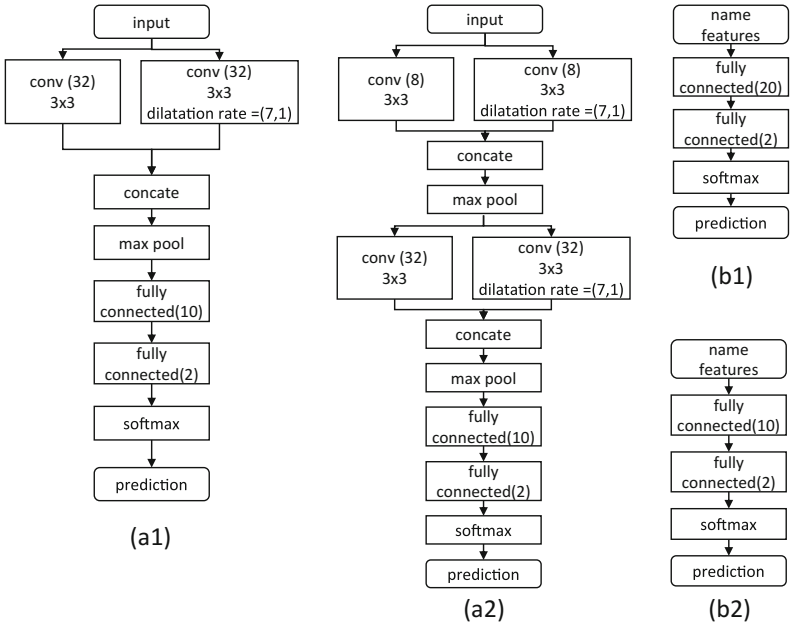

Fig. 7. (a1) the structure of CNN1; (a2) the structure of CNN2; (b1) the structure of fcNN1; (b2) the structure of fcNN2.

4.2 Training Phase

In this phase, we aim to train a stacked model composes the fcNN and CNN classifiers. The model is able to tell the possibility of a domain name to be objectionable-related by giving class scores. First of all, the preprocessing module transforms the labeled DNS traffic into visiting features and name features defined in Sect. 3. Then the training module directs the classifiers to automatically extract the most distinguishing patterns of objectionable-related domain names from visiting and name features. After the classifiers finish training separately, stacking module integrates them together and build an integral classifying model as the output of training phase.

4.3 Classification Phase

In the classification phase, VisSensor classifies arbitrary domain names into objectionable-related and none-objectionable based on the integral model built in the training phase. Firstly, the filter module receives the raw DNS traffic and removes the domain names that do not have enough queries or popular domain names that is irrelevant to objectionable contents (such as Alexa top domain names). The preprocessing module accepts the purified DNS traffic and transforms it into features described in Sect. 3. After that, the previous trained model performs the classification on these features and generates objectionable-related scores for the domain names.

5 Experimental Analysis

5.1 Data Set

Data Collection. The data we use in our research is the passively sampled DNS traffic which is first proposed by Weimer [2], and it became a significant analytic data source of DNS-associated security issues since then [3]. The passive DNS traffic is often collected on the level of resolvers, and it is generated by consecutively sampling the DNS queries and answers between clients and the resolver. Monitoring objectionable-related domain names through passive DNS traffic can significantly improve the timeliness and discover newly appeared objectionable-related domain names when they are visited.

We collect domain name samples by consecutively counting DNS querying answers from a provincial backbone resolver of a major ISP for 14 days (4th August, 2017 to 17th August, 2017). We select the domain names which were queried around 10^3 to 10^7 times in the two weeks, discard the domain names that are either very popular or lack of visiting.

Data Tagging. To label the domain names, we refer to the URL Ticketing System called Trusted Source [1] of McAfee on the Sep. 2017 at first. We refer the categories of domain names given by the Trusted Source rather than the risk

levels to ensure the accuracy of tagging. Due to the websites that our targeting domain names hosted, they should be labeled either Malicious, Pornography, Gambling or PUPs (potentially unwanted programs) by the Trusted Source [1]. But we find that a large portion of these domain names are ticketed as Forum/Bulletin Boards or Public Information which might be confused with normal domain names. To guarantee the reliability of our dataset, we manually label 5460 normal domain names and 5661 objectionable-related domain names, and partition them into three sets for training, testing and validation purposes, as shown in Table 1.

Table 1. Domain name samples and partition

Partition	Normal	Objectionable	Total
Training set	2730	2830	5560
Validation set	1365	1416	2781
Testing set	1365	1415	2780
Total	5460	5661	11121

5.2 Experiment Results

With the visiting features and name features of domain names, we separately train four classifiers for the two kinds of features. Specific structures of all classifiers are shown in Fig. 7(b1), (b2), (c1), (c2). For visiting features, we build two CNN classifiers to learn the visiting patterns of domain names. And we use two fcNN classifiers to learn the objectionable-scores from name features. The results of four classifiers are shown as Table 2.

Table 2. Model accuracies

Model	CNN 1	CNN 2	fcNN 1	fcNN 2	Stacked
Valid set	87.03%	86.75%	87.10%	87.25%	-
Test set	87.03%	86.10%	86.66%	87.47%	88.48%

5.3 Comparisons with Trusted Source

The Trusted Source [1] is a URL ticketing system of McAfee that provides the category and risk of a site, it also manually verifies the categories of websites that reported by its users before updating to its databases.

We compare the results of VisSensor with the labels tagged by Trusted Source on the domain names of the testing set, and summarize in the Table 3. We mark the domain names related to gambling, pornography and pirate media as

objectionable, and the Trusted Source tickets 4.19% (true positive) those domain names with a total accuracy of 34.37%. While VisSensor can figure out 85.07% objectionable domain names with a false positive rate of 8.23%, which shows a significant improvement upon the state of art of objectionable-related domain name tagging implementation.

Table 3. Comparisons with Trusted Source on the test set

Manual label	Trusted Source		VisSensor		Total
	Normal	Abnormal	Negative	Positive	
Normal	895 (32.31%)	465 (16.79%)	1294 (46.71%)	116 (4.19%)	1410
Abnormal	1353 (48.84%)	57 (2.06%)	203 (7.33%)	1157 (41.77%)	1360
Total	2248	522	1497	1273	2770

6 Related Works

The most common method of domain name classification is domain names tagging, and the typical ways of tagging are blacklists, whitelist and tagging systems. Although the importance of blacklists and whitelists are acknowledged widely in the domain names classification field, many researchers also found that the reliability of these lists are limited. Sinha et al. [11] found that blacklists shown high false positive and false negative rates; Sheng et al. [12] pointed out that blacklists' updating is sometimes not timely, and their coverages varied a lot; Kührer et al. [13] found that 15 public blacklists failed to cover more than 80% of the malicious domain names queried by malwares. Some researchers tried to improve the accuracy of blacklists: Kheir et al. [14] proposed methods that filter legal domain names from blacklists to reduce the false positive rates. Stevanovic et al. built a semi-manual labeling method which tracks the domain names with frequently changed IP addresses and relates the domain names with the reputations of these IP addresses in the blacklists. These works show that the effectiveness of blacklists is questionable.

Meanwhile, the existing tagging systems have some limitations. For example, we have retrieved the data sets on the URL tagging system of McAfee (trustedsource.org) twice, one on Sep. 2017 and the other on July 2018 (shown in Table 4). The objectionable-related domain name should be tagged as 'gambling' or 'pornography' in the Trusted Source. And the identification true positive rates of Trusted Source decreased from 4.56% to 2.74%; some domain names that had been labeled in the 2017 became unverified in 2018. From these results, we can notice that Trusted Source keeps updating its label engines, but its identification of objectionable-related domain names is still need to be further processed.

Table 4. Comparisons with Trusted Source

Manual label	2017-09		2018-07			Total
	Normal	Objectionable	Normal	Objectionable	Unverified	
Normal	3706	1755	3820	1236	404	5460
Objectionable	5402	258	5373	155	133	5661
Total	9108	2013	9193	1391	537	11121

References

1. Customer URL Ticketing System. https://trustedsource.org/sources/index.pl. Accessed 12 July 2018
2. Weimer, F.: Passive DNS replication. In: FIRST Conference on Computer Security Incident, p. 98 (2005)
3. Zdrnja, B., Brownlee, N., Wessels, D.: Passive monitoring of DNS anomalies. In: M. Hämmerli, B., Sommer, R. (eds.) DIMVA 2007. LNCS, vol. 4579, pp. 129–139. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-73614-1_8
4. Antonakakis, M., Perdisci, R., Dagon, D., et al.: Building a dynamic reputation system for DNS. In: USENIX Security Symposium, pp. 273–290 (2010)
5. Bilge, L., Kirda, E., Kruegel, C., et al.: EXPOSURE: finding malicious domains using passive DNS analysis. In: NDSS (2011)
6. Antonakakis, M., Perdisci, R., Lee, W., et al.: Detecting malware domains at the upper DNS hierarchy. In: USENIX Security Symposium, pp. 1–16 (2011)
7. Rahbarinia, B., Perdisci, R., Antonakakis, M.: Segugio: efficient behavior-based tracking of malware-control domains in large ISP networks. In: 2015 45th Annual IEEE/IFIP International Conference on Dependable Systems and Networks, DSN, pp. 403–414. IEEE (2015)
8. Hao, S., Thomas, M., Paxson, V., et al.: Understanding the domain registration behavior of spammers. In: Proceedings of the 2013 Conference on Internet Measurement Conference, pp. 63–76. ACM (2013)
9. LeCun, Y., Jackel, L.D., Bottou, L., et al.: Learning algorithms for classification: a comparison on handwritten digit recognition. Neural Netw.: Stat. Mech. Perspect. **261**, 276 (1995)
10. Szegedy, C., Liu, W., Jia, Y., et al.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9 (2015)
11. Sinha, S., Bailey, M., Jahanian, F.: Shades of Grey: on the effectiveness of reputation-based "blacklists". In: 3rd International Conference on Malicious and Unwanted Software, MALWARE 2008, pp. 57–64. IEEE (2008)
12. Sheng, S., Wardman, B., Warner, G., et al.: An empirical analysis of phishing blacklists. In: Sixth Conference on Email and Anti-Spam, CEAS (2009)
13. Kührer, M., Rossow, C., Holz, T.: Paint it black: evaluating the effectiveness of malware blacklists. In: Stavrou, A., Bos, H., Portokalidis, G. (eds.) RAID 2014. LNCS, vol. 8688, pp. 1–21. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-11379-1_1

14. Kheir, N., Tran, F., Caron, P., Deschamps, N.: Mentor: positive DNS reputation to skim-off benign domains in botnet C&C blacklists. In: Cuppens-Boulahia, N., Cuppens, F., Jajodia, S., Abou El Kalam, A., Sans, T., et al. (eds.) SEC 2014. IFIPAICT, vol. 428, pp. 1–14. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-642-55415-5_1
15. Stevanovic, M., Pedersen, J.M., D'Alconzo, A., et al.: On the ground truth problem of malicious DNS traffic analysis. Comput. Secur. **55**, 142–158 (2015)