计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2018,27(8):198−202 [doi: 10.15888/j.cnki.csa.006466]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于 GRU 型循环神经网络的随机域名检测①

陈立国[1,2,3], 张跃冬[3], 耿光刚[3], 延志伟[3]

[1](中国科学院 计算机网络信息中心, 北京 100190)
[2](中国科学院大学, 北京 100049)
[3](中国互联网络信息中心, 北京 100190)
通讯作者: 张跃冬, E-mail: zhangyuedong@cnnic.cn

摘　要: 随机域名是指由随机域名算法生成的域名, 被针对计算机网络系统的恶意软件广泛使用, 随机域名的检测任务是域名系统过滤攻击流量的基础性工作. 传统方法对随机域名的检测效果不理想, 精确率与召回率较低, 导致过滤攻击流量时会出现较多的误判. 本文提出和实现了一种基于 GRU 型循环神经网络的随机域名检测模型, 该模型首先将域名转换成向量, 然后借助 GRU 自动学习域名向量的特征, 最后通过神经网络计算分类. 相比于传统方法, 该模型不再需要人工提取特征的过程, 减少了特征提取的时间. 且经过算法生成数据与真实场景数据的实验验证, 该方法在随机域名检测任务中相比传统模型表现更加出色.

关键词: 随机域名; GRU; 循环神经网络; 域名系统; 流量过滤

引用格式: 陈立国,张跃冬,耿光刚,延志伟.基于 GRU 型循环神经网络的随机域名检测.计算机系统应用,2018,27(8):198–202. http://www.c-s-a.org.cn/1003-3254/6466.html

Detection of Random Generated Names Using Recurrent Neural Network with Gated Recurrent Unit

CHEN Li-Guo[1,2,3], ZHANG Yue-Dong[3], GENG Guang-Gang[3], YAN Zhi-Wei[3]

[1](Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China)
[2](University of Chinese Academy of Sciences, Beijing 100049, China)
[3](China Internet Network Information Center, Beijing 100190, China)

Abstract: Random domain names refer to the names generated by domain generation algorithms, which are widely used by the malware of the computer network system. The detection of random domain names is the basic work of the traffic filtering operation of the domain name system. The traditional method of detecting random domain name is not ideal and the precision and recall are low which will lead to erroneous judgement in attack traffic filtering. In this paper, the random names detection model are built based on recurrent neural network with gated recurrent unit. In this model, domain names are converted to vectors at first, then GRU are adopted to learn features automatically and which will be taken by the neural network to compute the class scores. Compared to traditional methods, this method is able to extract features without human help and which will reduce the time cost of feature extraction. This method performs better in the experiments of the algorithm generated data and the real world data than traditional models.

Key words: random names; GRU; recurrent neural network; domain name system; traffic filtering

随机域名是指由随机域名算法生成的域名, 在针对域名系统[1]发起的分布式拒绝服务攻击中被普遍使

① 基金项目: 国家自然科学基金 (61375039)
Foundation item: National Natural Science Foundation of China (61375039)
收稿时间: 2017-12-05; 修改时间: 2017-12-27; 采用时间: 2018-01-08; csa 在线出版时间: 2018-07-28

用, 随机域名检测是过滤拒绝服务攻击流量的前提. 随机域名的检测工作主要基于随机域名的随机性特点来进行. 随机域名由随机域名算法生成, 各字符出现的概率相同且字符之间没有依赖关系. 基于随机域名的随机性特点, 现有的随机域名检测方法主要包括基于域名分布相似度的检测方法与基于机器学习的检测方法两类, 这两类方法将在第 1 章中详细介绍.

随机域名检测是域名系统 DDoS 流量过滤、僵尸网络通信检测等域名安全任务的基础工作, 在检测实时性、检测准确率等方面都有较高的要求. 传统的随机域名检测方法检测准确率较低, 误报率较高, 且多数方法在流量过滤等任务中不能满足实时性的需求.

近年来, 深度学习在图像、语音、自然语言处理等领域都取得了重大突破[2,3], 基于深度学习的应用越来越多[4–6], 应用深度学习的模型解决传统研究领域的问题也成为了研究的热点. 本文将 GRU(Gated Recurrent Unit) 型循环神经网络模型[7]应用于随机域名的检测工作中, 经过算法生成数据与真实场景数据的实验验证, 该模型相比传统检测模型拥有更好的检测性能, 且检测过程中不需要对域名提取特征.

1 传统的随机域名检测方法

目前, 已有的随机域名检测方法主要分为两类, 一类是基于域名分布相似度的检测方法, 另一类是基于机器学习的检测方法.

1.1 基于域名分布相似度的检测方法

Sandeep 等[8]提出了一种基于分布相似度判定域名是否遭受随机域名攻击的方法.

首先, 对于给定的单一域名 d, 该方法生成其对应的 n-$gram$, 如式 (1), 其中 $i \in \{0,1,\cdots,len(d)-n\}, n \geqslant 1$. $len(d)$ 为域名 d 包含的字符数, n 在该文章中取值为 1, 2.

$$n-gram = \{g | g = d_i d_{i+1},\cdots,d_{i+n-1}\} \quad (1)$$

然后, 该方法分别统计正常域名与随机域名的 n-$gram$ 分布. 得到上述两种分布后, 文章选取了三种分布相似度的评价指标: 相对熵、杰卡德距离与编辑距离. 最后对每一个 IP 地址或权威域名计算其所对应域名的 n-$gram$ 分布分别与正常域名 n-$gram$ 分布和随机域名 n-$gram$ 分布的相似度, 根据相似度的大小判断当前 IP 地址或权威域名是否遭受了随机域名攻击.

文章中验证了 n 的不同取值以及不同分布相似度时实验的效果. 实验结果显示, 当 n 取 1, 相似度选用相对熵时, 达到 100% 检出率, 2.5% 假阳率, 实验结果最优.

该方法优势是实现简单, 仅需要对比待检测的权威域名或 IP 地址对应的域名 n-$gram$ 分布与正常域名、随机域名的分布计算相似度即可. 但是单一域名由于字符数有限, 字符分布不具有统计特性, 基于分布相似度的方法难以做出类别判断.

1.2 基于机器学习的检测方法

基于机器学习的检测方法将随机域名检测作为二分类任务处理, 通过最小化代价函数的方式提升检测准确率. 这类检测方法由于具有检出率高、误报率小、可实时检测等优势, 是近年来的研究热点.

Davuth 等[9]提出了一种基于支持向量机的随机域名检测方法. 该方法将域名的 bi-gram 作为特征, 通过人工阈值的方式过滤出现频率较低的 bi-gram, 实现特征选择. 文章中采用支持向量机分类器. 在第 3 章中, 本文将该方法与基于 GRU 型循环神经网络的方法进行了对比实验.

王红凯等[10]提出了一种基于随机森林的随机域名检测方法. 该方法通过人工提取域名特征来构建随机森林模型训练分类, 实现对随机域名的检测. 但人工特征的构造过程较为繁琐, 而且由于所使用的统计特征过多, 特征抽取过程也较为耗时.

章思宇[11]提出了一种基于 DNS 图挖掘的随机域名检测算法, 该方法根据 DNS 查询日志将域名与主机的集合构建成图模型, 然后应用置信传播算法进行声望推断, 据此进行随机域名的检测. 该方法不仅可以检测出随机域名, 也可以分析出网络中的控制服务器和受害主机, 但检测准确率较低, 且检测实时性较差.

2 基于 GRU 型循环神经网络的随机域名检测方法

近年来, 深度循环神经网络模型在众多自然语言处理的任务中表现出色[7,12,13], 由于常规域名通常也保留一定的语言特性[14], 注册域名中也通常包含汉语拼音、单词等子串. 因此, 本文尝试将神经网络模型应用在随机域名的检测工作中, 通过挖掘域名语言特性区分常规域名与随机域名.

2.1 域名向量化

由于循环神经网络模型接受的输入 x 为定长的向量, 而域名本身为字符串, 因此需要引入域名向量化步骤, 将字符串转换成循环神经网络输入.

域名向量化过程首先统计在所有域名中出现的字符集合, 假设该集合中字符数为 n, 则将域名中的每个字符编码成长度为 n 的 one-hot 向量, 最后将该域名中的所有字符对应的 one-hot 向量按照各个字符在域名中的顺序拼接得到循环神经网络的输入 x.

2.2 GRU 型循环神经网络结构

GRU 型神经网络是一种循环神经网络, 隐层节点相互连接, 呈现出循环的特性. 不同的是该网络隐层节点内部结构并非单一激活函数, 而是采用 GRU 结构. 图 1 为 GRU 内部结构示意图, 其中 h_{t-1} 为 t–1 时刻隐藏层输出, X_t 为 t 时刻输入, h_t 为 t 时刻隐藏层输出, r_t 表示重置门, z_t 表示更新门, $\tilde{h}_t$ 表示 t 时刻隐层节点的候选值, 其他元素操作如右侧图例所示, 用公式描述图 1 如下:

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \tag{2}$$

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \tag{3}$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \tag{4}$$

$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t]) \tag{5}$$

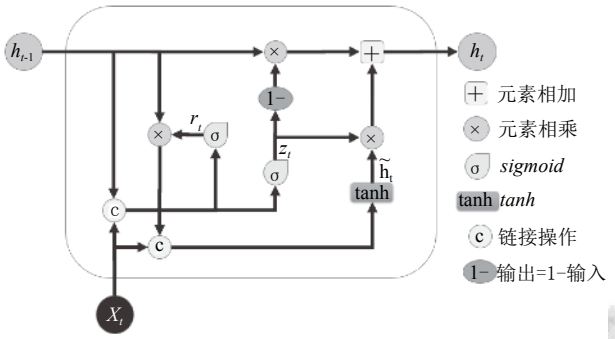

图 1 GRU 内部结构

实验所采用的循环神经网络由输入层、隐含层、全连接层和输出层组成, 其结构如图 2 所示. 其中 l 为域名的截断长度, 当域名长度超出 l 则对域名进行截断, 当域名长度小于 l 则在域名向量后拼接零向量, 直到域名向量长度达到 l. x_t 为当前输入域名中的第 t 个字符, 采用 one-hot 编码. h_t 为当前时刻第 t 个字符对应的隐含状态. drop_out 为随机丢弃层, 用于防止网络的过拟合[15], 经过实验测试, 当随机丢弃概率为 0.5 时, 训练得到的模型分类性能最优. full_connected 为全连接层, 全连接层的节点数设置为 32. Y 为输出的类别, 取值为 0 或 1, 分别表示当前输入是否为随机域名. 实验中, 模型的超参数设置将在 3.2 节中详细介绍.

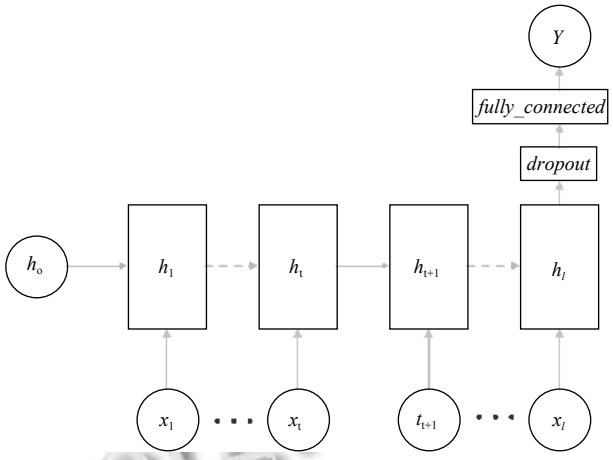

图 2 GRU 型循环神经网络随机域名检测模型结构

传统方法为了提升随机域名的检测准确率, 往往会抽取一定维度的域名特征[10], 统计当前一段时间内被请求域名的信息, 当 DNS 系统遭受 DDoS 攻击后, 解析请求量急速上升, 特征的提取需要更多的计算资源, 为了保证攻击流量过滤的实时性, 需要减少特征提取的时间, 降低特征提取所需的算力成本. 而 GRU 型循环神经网络随机域名检测模型是一个端到端的模型, 输入为域名向量, 输出为该域名对应的类别. 相比传统方法, 该模型不需要对该域名抽取统计特征, 节省了特征抽取的算力成本和时间成本, 使得模型能够及时对当前域名的类别做出判断并设置防火墙规则, 从而能够满足大规模攻击流量过滤任务的实时性需求.

相比于使用 *n-gram* 特征的方法[9], GRU 型神经网络的随机域名检测方法检测性能更好, 准确率更高, 在相同的实验环境下, 该模型拥有更低的误报率和漏报率. 具体实验结果将在 3.3 节中详细介绍.

3 实验分析

3.1 实验数据

本文中所使用的实验数据包含算法生成数据与真实场景数据两部分.

算法生成数据中, 负类样本由 5 种著名僵尸网络的随机域名算法生成: newGOZ[16]、Ramnit[17]、Shiotob[18]、Symmni[19]、Banjori[20]. 负类样本共计 50 万条, 每种随机算法产生的随机域名占比 20%, 且各个算法产生的随机域名在训练集中服从均匀分布. 正类样本为.cn 在正常服务状态下被查询的域名经过随机采样、反向查询过滤得到的域名集合, 共计 50 万条.

```
1,www.zhong5.cn
0,ip9o8pfg8nm9f.vrj.cn
1,www.zj3000.cn
0,uipoz4vs44jli.vwxq.cn
1,n.7k7kimg.cn
0,zyilb6bc7atbl.vmwv.cn
1,ouyang2013.hanbang.org.cn
0,u-i8xtss3-15n.dfl.cn
1,kaouthiusg.cn
0,f0xybr5zkt0kj.rqqi.cn
1,www.kaoxia.cn
0,qz0ci3yi7t7gn.hbo.cn
```

图 3 算法生成数据示例

真实场景数据来自.cn 域名服务在 2015 年 5 月 12 日遭受的一次随机子域名攻击事件. 负类样本为遭受攻击时被攻击域名的子域名去除已注册子域名得到的域名集合, 共计 50 万条. 正类样本为.cn 在正常服务状态下被查询的域名经过随机采样和过滤反向查询得到的域名集合, 共计 50 万条. 正负类样本合计 100 万条, 数据格式与算法生成数据相同.

3.2 实验设置与超参数

在随机域名检测任务中, 模型的分类性能是评价模型优劣的首要考虑因素, 实验的第一部分将考察模型的分类性能. 而神经网络模型通常存在容易过拟合[21]、收敛难度大等缺陷, 实验的第二部分将会比较逻辑回归模型与基于 GRU 型神经网络模型的收敛过程, 验证.

在分类性能实验中, 本文分别实现了基于逻辑回归与基于支持向量机的随机域名检测模型, 并分别使用算法生成数据和真实场景数据训练、测试两种模型. 在实验中, 训练集占 80%, 验证集占 10%, 测试集占 10%. 使用 5 种评估指标来评估模型效果, 评估结果在 3.3 小节中详细描述.

在收敛过程实验中, 本文将对比 GRU 型循环神经网络检测模型与基于逻辑回归的检测模型在每轮迭代过程中的 AUC, 比较两种模型的收敛过程, 为了更明显的效果对比, 两种模型的学习率都调整为分类性能实验的 0.2 倍.

GRU 型循环神经网络检测模型的主要超参数包含: 隐层节点个数 *l*, 随机丢弃概率 *p_drop*, 全连接层节点个数 *m*, 学习率 *a*, 激活函数 *f*, 学习算法 *la*, 损失函数 *loss* 等. 基于逻辑回归的检测模型主要超参数包含: 学习率 *a*, 正则项 *penalty*, 正则化强度 *C*. 基于支持向量机的 *n-gram* 检测模型主要超参数有: 核函数 *kernel*, 惩罚因子 *C*. 两次实验的超参数如表 1 所示.

表 1 各模型超参数设置

模型	超参数	分类性能实验	收敛速度实验
GRU-RNN	*l*	64	64
	p_drop	0.5	0.5
	m	128	128
	a	0.0005	0.0001
	f	tanh	tanh
	la	adam	adam
	loss	cross entropy	cross entropy
LR	*a*	0.01	0.002
	penalty	L2	L2
	C	1.0	1.0
SVM	*kernel*	rbf	-
	C	1.0	-

3.3 实验结果分析

本文提出和实现了基于 GRU 型循环神经网络的随机域名检测模型, 为了对比传统模型的检测性能, 分别实现了基于 *n-gram* 特征的逻辑回归、支持向量机随机域名检测模型, 以上模型的分类性能在算法生成数据上的表现如表 2 所示, 在真实数据上的分类性能如表 3 所示. 根据实验可以得出, 基于 bi-gram 特征的模型分类性能普遍优于基于 uni-gram 特征的模型, 支持向量机模型普遍优于逻辑回归, 而 GRU 型循环神经网络模型在各项分类性能中都表现优异, 领先前两者.

表 2 各模型在算法生成数据上的分类性能

	精确率	召回率	F_1	AUC	log loss
lr_unigram	0.9566	0.9642	0.9604	0.9603	1.3729
lr_bigram	**1.0**	0.9845	0.9922	0.9922	0.2685
svm_unigram	**1.0**	0.9655	0.9824	0.9827	0.5957
svm_bigram	0.999	0.992	0.9955	0.9955	0.1554
rnn_gru	**1.0**	**0.9994**	**0.9997**	**0.9997**	**0.0104**

表 3 各模型在真实场景数据上的分类性能

	精确率	召回率	F_1	AUC	log loss
lr_unigram	0.8846	0.9504	0.9163	0.8873	3.5165
lr_bigram	0.9513	0.9798	0.9653	0.9543	1.4248
svm_unigram	0.9163	0.9698	0.9423	0.9069	2.6164
svm_bigram	0.998	0.97	0.9838	0.9834	0.6873
rnn_gru	**0.9997**	**0.9856**	**0.9925**	**0.9901**	**0.4213**

收敛过程实验结果如图 4 所示, 横轴为模型迭代次数, 纵轴为算法生成数据上的每一轮迭代后各模型的 AUC. 由图中可以看出, 基于 GRU 型循环神经网络

的检测模型收敛速度与逻辑回归模型相当且收敛过程较为平稳.

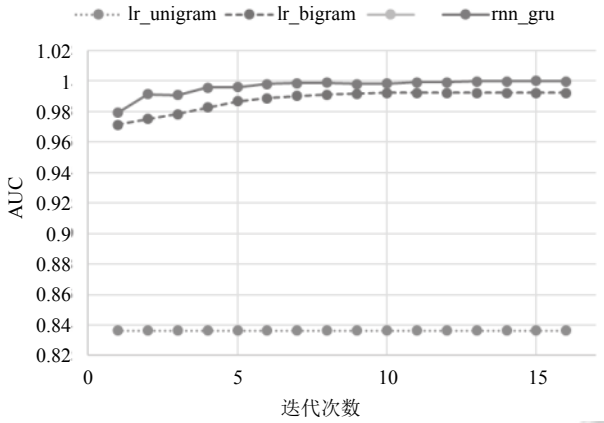

图 4 GRU 型循环神经网络检测模型与逻辑回归检测模型收敛过程比较

4 结论与展望

本文的主要工作是将 GRU 型循环神经网络应用到随机域名的检测工作中, 提出和实现了基于 GRU 型循环神经网络的随机域名检测方法, 并将该方法与传统检测方法进行对比分析. 经过实验验证, 该模型在算法生成数据和真实场景数据上都相比传统方法表现出色, 模型收敛速度快, 收敛过程平稳. 但目前该模型的检测对象仅限于随机域名, 对于可能包含语言特性的恶意域名检测效果尚未得到验证. 在接下来的工作中, 将研究如何将现有检测模型融入注册、解析等非语言本身的特征, 使得融合后的模型能够应对更加复杂的应用场景.

参考文献

1 颜挺进, 李淑英. 域名系统的应用. 计算机系统应用, 2001, 10(10): 26–27. [doi: 10.3969/j.issn.1003-3254.2001.10.008]

2 Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Communications of the ACM, 2017, 60(6): 84–90. [doi: 10.1145/3098997]

3 Lecun Y, Bengio Y, Hinton G. Deep learning. Nature, 2015, 521(7553): 436–444. [doi: 10.1038/nature14539]

4 Deng L, Hinton G, Kingsbury B. New types of deep neural network learning for speech recognition and related applications: An overview. Proceedings of 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. Vancouver, BC, Canada. 2013. 8599–8603.

5 Wan J, Wang DY, Hoi SCH, *et al*. Deep learning for content-based image retrieval: A comprehensive study. Proceedings of the 22nd ACM International Conference on Multimedia. Orlando, FL, USA. 2014. 157–166.

6 Deng L, Yu D. Deep learning: Methods and applications. Foundations and Trends in Signal Processing, 2014, 7(3-4): 197–387. [doi: 10.1561/2000000039]

7 Cho K, Van Merriënboer B, Gulcehre C, *et al*. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv: 1406.1078, 2014.

8 Yadav S, Reddy AKK, Reddy ALN, *et al*. Detecting algorithmically generated malicious domain names. Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement. Melbourne, Australia. 2010. 48–61.

9 Davuth N, KIM SR. Classification of malicious domain names using support vector machine and Bi-gram method. International Journal of Security and its Applications, 2013, 7(1): 51–58.

10 王红凯, 张旭东, 杨维永, 等. 基于随机森林的 DGA 域名检测方法: 中国, CN105577660A. 2016–05–11.

11 章思宇. 基于 DNS 流量的恶意软件域名挖掘[硕士学位论文]. 上海: 上海交通大学, 2014.

12 Sutskever I, Vinyals O, Le QV. Sequence to sequence learning with neural networks. arXiv: 1409.3215, 2014.

13 Mikolov T, Chen K, Corrado G, *et al*. Efficient estimation of word representations in vector space. arXiv: 1301.3781, 2013.

14 魏唯达. 域名中语言符号的探究. 现代语文 (语言研究), 2007, (7): 104–105.

15 Srivastava N, Hinton G, Krizhevsky A, *et al*. Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 2014, 15(1): 1929–1958.

16 Bader J. The dga of newgoz. https://www.johannesbader.ch/2014/12/the-dga-of-newgoz/, [2017-11-06].

17 Bader J. The dga of ramnit. https://johannesbader.ch/2014/12/the-dga-of-ramnit/, [2017-11-06].

18 Bader J. The dga of shiotob. https://www.johannesbader.ch/2015/01/the-dga-of-shiotob/, [2017-11-06].

19 Bader J. The dga of symmi. https://johannesbader.ch/2015/01/the-dga-of-symmi/, [2017-11-06].

20 Bader J. The dga of banjori. https://www.johannesbader.ch/2015/02/the-dga-of-banjori/, [2017-11-06].

21 Lawrence S, Giles CL, Tsoi AC. Lessons in neural network training: Overfitting may be harder than expected. Proceedings of the 14th National Conference on Artificial Intelligence and 9th Conference on Innovative Applications of Artificial intelligence. Providence, RI, USA. 1997. 540–545.