文章编号：1006-3080(2019)03-0478-08 DOI：10.14135/j.cnki.1006-3080.20180326002

基于注意力机制的 DGA 域名检测算法

陈立皇， 程 华， 房一泉
（华东理工大学信息科学与工程学院，上海 200237）

摘要：DGA 域名（Domain Generation Algorithm）检测是恶意 C&C 通信检测的关键技术之一。已有的检测方法通常基于域名构成的随机性进行检测，存在误报率高等问题，对于低随机性 DGA 域名的检测准确率较低，主要是因为此类方法未能有效提取低随机性 DGA 域名中的部分高随机性，为此提出了域名的多字符随机性提取方法。采用门控循环单元（GRU）实现多字符组合编码及其随机性提取；引入注意力机制，加强域名中部分高随机性特征。构建了基于注意力机制的循环神经网络的 DGA 域名检测算法（ATT-GRU），提升了低随机性 DGA 域名识别的有效性。实验结果表明，ATT-GRU 算法在检测 DGA 域名上取得了比传统方法更高的检测精确率和更低的误报率。

关键词：域名检测；注意力机制；门控循环单元 (GRU)
中图分类号：TP309.1 **文献标志码：**A

2017 年前后各种勒索软件[1] 爆发，对政府、能源、制造业等关键信息基础设施造成了重大损失。勒索软件等恶意软件采用随机生成大量 DGA（Domain Generation Algorithm）域名方式，与 C&C（Command and Control）服务器建立通信，躲避安全检测。文献 [2] 分析和评估了 43 个恶意软件，其中有 23 个采用 DGA 作为唯一的通信方式，其余的把 DGA 域名作为硬编码域名外的主要通信方式，因此 DGA 域名检测在应对此类问题时具有重要意义。

传统的 DGA 域名检测算法提取网络流量上下文信息等特征，利用机器学习区分正常域名和恶意域名。特征有统计特征，如域名的 unigram 和 bigram 特征[3-4]、域名长度[5]、熵、字符频率、数字及有意义字符比[6] 等；DNS 信息，如应答时间[7]、查询率[8]、whois 变更信息等；另外还有流量特征，如域名 TTL[9] 等。这些检测方法都存在着误报率较高、整体检测率低[10] 等问题。

文献 [11-12] 分别利用长短期记忆网络 (Long Short-Term Memory, LSTM)、门控循环单元 (Gated Recurrent Unit, GRU) 检测恶意 DGA 域名。此类循环神经网络方法对随机性高的 DGA 域名检测准确率高，但对随机性低的 DGA 域名识别率低，导致对正常域名产生较高的误报。误报是安全防御系统的一个重要指标，频繁的误报会降低安全检测的可信度。

DGA 算法通常采取随机策略生成域名，但部分新出现的 DGA 算法模拟正常域名的取名策略，如 suppobox 算法采用单词拼接，生成的域名看起来像正常的域名，字符组合的随机性比传统 DGA 域名的随机性低得多，本文称为低随机性 DGA。这类低随机性 DGA 域名中仍然存在高随机性的字符组合，如 suppobox 类域名中的单词拼接处的字符组合。

在图像分类[13] 中，注意力机制模拟人眼的注意模式，提高对图像中显著局部的关注，提高分类精确性。同样对于低随机性 DGA，研究其域名序列中的高随机性字符组合并加以关注，用于提高 DGA 域名的识别。

本文对 DGA 建立随机模型，采用 GRU 结合注意力机制，提出了 DGA 域名检测算法（ATT-GRU），采取双向 GRU 提取域名多字符组合编码信息，通过注意力机制学习字符组合在域名中的重要程度，提高域名中高随机性字符组合的关注度，从而在识别

收稿日期：2018-03-26
作者简介：陈立皇 (1994-)，男，硕士生，研究方向为网络安全。E-mail: chenlihuang1994@163.com
通信联系人：程 华，E-mail: hcheng@ecust.edu.cn
引用本文：陈立皇，程 华，房一泉. 基于注意力机制的 DGA 域名检测算法 [J]. 华东理工大学学报（自然科学版），2019, 45(3): 478-485.
Citation: CHEN Lihuang, CHENG Hua, FANG Yiquan. Detecting Domain Generation Algorithm Based on Attention Mechanism[J]. Journal of East China University of Science and Technology, 2019, 45(3): 478-485.

过程中降低误报率。实验结果表明, 考虑字符组合随机性的域名序列编码可以显著提高检测准确性, 与文献 [11-12] 相比, ATT-GRU 具有更低的误报率。

1　域名局部特性

1.1　单字符随机性分析

域名单字符的随机性通过计算字符的相对熵 $(\overline{H})$ 得到:

$$\overline{H}=\frac{H}{H_{\max}}=\frac{-\sum_{i=1}^{n}p(x_i)\log_2 p(x_i)}{-\log_2\frac{1}{n}} \tag{1}$$

其中: p 为字符 x_i 出现的概率; n 为出现的字符种类总数; H 为字符熵; $H_{\max}$ 为 n 字符的最大熵。

不同种类域名单字符的相对熵如表 1 所示。corebot 类、ramnit 类等 DGA 域名的相对熵接近 1, 单字符分布近似完全随机。banjori 类、suppobox[2] 类和 gozi 类(基于词典)等低随机性 DGA 域名以及 legit 类正常域名的相对熵为 0.9 左右, 体现了域名单字符随机性低的特点。其中, banjori 类域名的算法只随机改变域名前 4 个字符, 后面的字符串为种子字符串。基于词典的 gozi 类和 suppobox 类 DGA 域名的平均相对熵为 0.869, 正常域名 legit 类的相对熵与基于词典的 DGA 域名相近, 为 0.861。legit 类中大部分域名的构成为单词组合, 而基于词典 DGA 域名也是采用单词拼接的域名构建策略, 如正常域名中 'facebook' 由单词 face 和 book 拼接, suppobox 中的 'forgetshort' 域名由 forget 和 short 组成。

表 1　域名单字符的相对熵

Table 1　Relative entropy of domain single-character

Family	$\overline{H}$	Type	Example(Second-level domain)
legit	0.861	Normal	google; youtube
corebot	0.999	Arithmetic-based DGA	at367lsnux1n1vg; kl3xe0gf3pmditc
kraken	0.995	Arithmetic-based DGA	xfdvisu; gopqfnsmjb
locky	0.999	Arithmetic-based DGA	oewvdjhwxkwdr; xwkcxka
ramnit	0.999	Arithmetic-based DGA	fhafbkfdjpw; kxoggojma
banjori	0.930	Arithmetic-based DGA*	pdtmstring; umfpstring
gozi	0.857	wordlist-baesd DGA	commithissunt; suntpotestatem
suppobox	0.880	wordlist-baesd DGA	forgetopinion; forgetsupply

基于词典的 DGA 域名和正常域名在整体随机性上的区分度低, 导致相应的基于统计特征的机器学习方法、循环神经网络方法检测率低。

1.2　低随机性 DGA 域名的分析

1.2.1　基于词典 DGA 域名　基于词典的 DGA域名构造方法使域名中单字符的分布不再随机, 其单字符分布近似正常域名。构造 DGA 域名时, 从词典中选取单词的行为是随机的, 这种随机性体现在单词的拼接部分, 因此, 拼接部分中 2 个字符的共现分布低于域名中单词内字符分布。

基于词典 DGA 域名拼接部分的字符分布低于组成单词部分的字符, 拼接部分体现了 DGA 域名的随机程度。表 2 所示为 suppobox 域名 'forgetshort' 的 bigram 分布 (f), 域名拼接部分 'ts' 的分布低于组成单词的部分, 仅为 0.28%, 低于组成单词的字符平均分布值(0.89%)。

1.2.2　半随机域名 banjori 类　banjori 类 DGA 域名中的种子字符串部分导致域名字符分布的半随机性, 这类域名的随机性集中在字符有随机变化的前 4 个字符。

banjori 类域名字符的前 4 个字符组合的分布低于种子字符串中字符的分布, 反映了 DGA 的随机性。表 3 所示为域名 'umfpstring' 的 bigram 分布, 前 4 个字符相关 bigram 的分布约为 0.06%, 低于种子字符串中字符平均分布值 0.86%。

1.3　多字符随机性分析

域名双字符分布的不同对应域名随机性的差异, DGA 域名自身的随机性高于正常域名, 表明字符随机性有助于区分 DGA 域名。域名双字符的随机性通过计算双字符的相对熵 $\overline{H}^2$ 得到:

$$\overline{H}^2=\frac{H}{H_{\max}}=\frac{-\sum_{i=1}^{n}p(xx_i)\log_2 p(xx_i)}{-\log_2\frac{1}{n}} \tag{2}$$

表2 “forgetshort”的bigram分布值

Table 2 Bigram distribution of “forgetshort”

Character	f/%
fo	0.65
or	1.67
rg	0.31
Ge	0.92
et	1.25
ts	0.28
sh	0.32
ho	0.66
or	1.67
rt	0.56

表3 “umfpstring”的bigram分布

Table 3 Bigram distribution of “umfpstring”

Character	f/%
um	0.08
mf	0.06
fp	0.06
ps	0.06
st	0.49
tr	0.74
ri	0.76
in	1.17
ng	1.15

其中:p为双字符xx_i出现的概率;n为出现的双字符种类总数;H为字符熵;$H_{\max}$为n个双字符的最大熵。

在低随机域名中,双字符的随机性与单字符随机性有较大差别,如表4所示,低随机DGA双字符的平均熵为0.876,高于正常legit类域名。

低随机性DGA域名和正常域名的单字符随机程度近似,但双字符分布不同。如表5所示,共现字符串‘or’在正常域名bigram分布中为1.07%,而‘or’

表4 低随机性域名双字符的相对熵

Table 4 Relative entropy of low-random domain double-character

Family	$\overline{H}^2$
legit	0.813
banjori	0.916
gozi	0.853
suppobox	0.861

表5 部分bigram分布

Table 5 Partial bigram distribution

Character	f/%	
	legit	suppobox
oo	0.43	0.21
op	0.33	0.05
oq	0.01	0.00
or	1.07	1.76

在suppobox的bigram分布中为1.76%,这是因为基于词典suppobox类域名使用的单词中很多由‘or’组成,如‘storm’、‘therefore’和‘effort’等。故本文提出提取域名中更长字符的组合,突出随机性高的部分来检测DGA域名。

2 基于注意力机制的DGA检测模型

由上文可知,若在DGA检测中提高部分域名字符的随机性权重,可以提高域名识别的精度。为此本文构建DGA域名随机模型,用循环神经网络获取域名序列变长字符组合及编码,设计注意力机制捕捉域名中高随机性的字符组合,构建DGA域名的检测方法。

2.1 域名随机模型

由域名生成算法及其局部特性分析,构建DGA域名的随机模型D_n:

$$D_n = c_1 c_2 \cdots c_{n-1} c_n,\ c_i \in C \tag{3}$$

C为候选集:

$$C\begin{cases} c_{\mathrm{ch}} \Rightarrow \hat{}[a-z][0-9\backslash\_]\$ \\ c_{\mathrm{st}} \Rightarrow \hat{}[a-z0-9\_]\{1,\}\$ \end{cases} \tag{4}$$

其中:c_{ch}为单字符集,传统的候选集是纯单字符c_{ch}的候选集;c_{st}为包含一个以上字母、数字或下划线的字符串集。基于词典与半随机DGA域名可以视为一个变异的候选集,即候选集中包括单字符c_{ch}以及字符串集c_{st}。

在候选集C中使用随机概率P选取元素c_i组成域名。DGA域名从候选集中随机选取元素拼接时,即选取概率$P(c_i \in C|D_n)$在所有i上都相同。当c_{st}作为元素拼接时,拼接的首尾部定义为$\mathrm{edge}(c_{\mathrm{st}})$。

2.2 多字符组合随机性的提取方法

基于GRU[14]提取多字符组合,获得域名中多字符组合编码。对字符编码计算注意力权重及其分布,权重值反映了多字符组合的随机性,域名中高随机字符组合的权重值大。

2.2.1　域名序列的多字符组合编码　图 1 所示为 GRU 实现变长字符组合的编码。当前字符的编码 g_t 依赖于当前字符输入 e_t 与上一时刻的字符编码 g_{t-1}。

$$g_t = f(g_{t-1}, e_t) \tag{5}$$

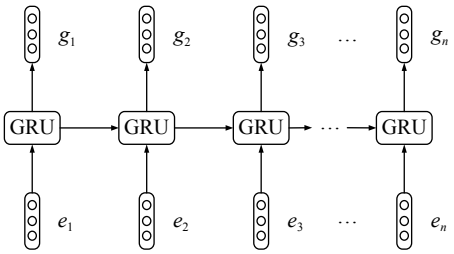

图 1　GRU 编码图

Fig. 1　GRU encoding

GRU 编码通过训练得到序列下一个字符在序列上的概率分布 $P(c_n|D_{n-1})$。序列中的字符编码结合了长期依赖的字符信息, 多于两字符组合的信息。域名序列中, 当前字符编码考虑了之前每一个字符的输入信息, 输入信息的取舍决定于 GRU 隐藏状态中的两个门。

图 2 所示为 GRU 门控循环单元示意图。

GRU 单元中更新门 z 决定了前一位置信息保留的程度, 重置门 r 控制当前字符信息的输入程度, 其更新公式如下:

$$r_t = \sigma(\boldsymbol{W}_{\mathrm{r}} \cdot [g_{t-1}, e_t] + b_{\mathrm{r}}) \tag{6}$$

$$z_t = \sigma(\boldsymbol{W}_{\mathrm{z}} \cdot [g_{t-1}, e_t] + b_{\mathrm{z}}) \tag{7}$$

$$g_t' = \tanh(\boldsymbol{W}_{\mathrm{h}} \cdot [r_t g_{t-1}, e_t] + b_{\mathrm{h}}) \tag{8}$$

$$g_t = (1 - z_t) g_{t-1} + z_t g_t' \tag{9}$$

其中 σ 和 tanh 为 sigmoid 和双曲正切函数; $\boldsymbol{W}_{\mathrm{r}}$、$\boldsymbol{W}_{\mathrm{z}}$ 和 $\boldsymbol{W}_{\mathrm{h}}$ 为参数矩阵; b_{r}、b_{z} 和 b_{h} 为偏置参数。

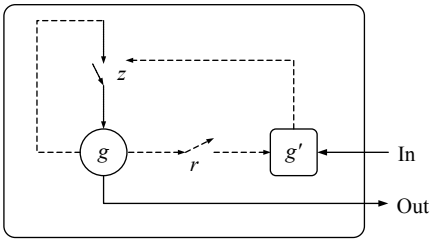

图 2　GRU 门控循环单元

Fig. 2　GRU unit

2.2.2　多字符随机性提取方法　研究域名中不同字符的组合对域名检测的贡献。当 DGA 域名候选集仅由单字符集 c_{ch} 组成时, 域名中不同单字符及其多字符组合的随机性相同。

对于域名候选集包括 c_{ch} 和 c_{st} 的域名, 由于 c_{st} 由单字符组成, 而随机策略 P 不能影响到 c_{st} 的组成, 因此, 算法应当区别对待元素 c 之间的随机分布与组成 c_{st} 字符之间的随机程度。

在域名字符组合编码上, 注意力函数 $f_{\mathrm{aat}}(\boldsymbol{G}_t, \boldsymbol{u}_w)$ 计算得到各字符属于 c_{ch} 的概率或属于 $\mathrm{edge}(c_{\mathrm{st}})$ 的概率:

$$P[c_i \ \text{in} \ (c_{\mathrm{ch}} \| \mathrm{edge}(c_{\mathrm{st}}))] \tag{10}$$

注意力机制通过分配不同权重 α_t, 表现出域名中多字符组合的不同随机性。用 softmax 归一化权重:

$$\alpha_t = \mathrm{soft\,max}(f_{\mathrm{aat}}(\boldsymbol{g}_t, \boldsymbol{u}_w)) \tag{11}$$

2.3　ATT-GRU 模型的检测方法

图 3 示出了 ATT-GRU 模型结构, 形成了基于 ATT-GRU 模型的 DGA 检测方法, 由 GRU 字符编码器、注意力机制、softmax 分类器组成, 分类器输出待判定域名的检测结果。

2.3.1　GRU 字符编码　ATT-GRU 的输入为以字符为单位的域名序列, 之后通过嵌入层 Embedding 训练[15], 得到域名在字符向量空间的向量 $\boldsymbol{E}$。

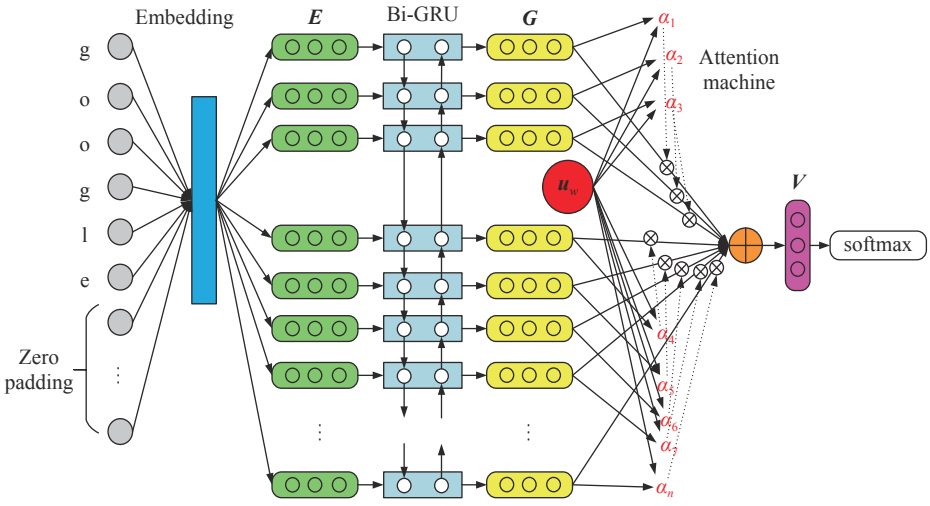

图 3　ATT-GRU 模型结构

Fig. 3　ATT-GRU model structure

$$\boldsymbol{E}=\{e_1,e_2,e_3,\cdots,e_n\} \tag{12}$$

普通的GRU模型只能得到域名中正向的字符组合编码,缺乏获得逆序字符组合的能力,因此本文在编码字符组合时采用Bi-GRU模型对域名向量$\boldsymbol{E}$进行编码。在Bi-GRU中,字符组合编码的输出为当前正向GRU和逆向GRU的输出连接而成,表示为

$$\boldsymbol{G}_t=[g_t,\ g_{n-t}] \tag{13}$$

2.3.2 注意力机制 将Bi-GRU得到的字符组合编码输入注意力模型中,计算编码$\boldsymbol{G}_t$的注意力权重α_t。权重采用加性注意力函数计算,α_t与字符编码和一个随机初始化的向量$\boldsymbol{u}_w$有关,由式(10)和式(11)得到。

$$\alpha_t=\frac{\exp(f_{\text{aat}}(\boldsymbol{G}_t,\boldsymbol{u}_w))}{\sum_{t=1}^{n}\exp(f_{\text{aat}}(\boldsymbol{G}_t,\boldsymbol{u}_w))} \tag{14}$$

$$f_{\text{aat}}(\boldsymbol{G}_t,\boldsymbol{u}_w)=\boldsymbol{u}_w^{\mathrm{T}}\tanh(\boldsymbol{W}_s\boldsymbol{G}_t+b_s) \tag{15}$$

其中:$\boldsymbol{W}_s$为模型的权重矩阵;b_s为模型的偏置,与$\boldsymbol{u}_w$一同作为训练模型过程中的参数,通过不断学习训练得到。使用α_t对$\boldsymbol{G}_t$进行加权,构造一个关注到相应字符组合随机程度的域名表示向量$\boldsymbol{v}$。

$$\boldsymbol{v}=\sum_{t=1}^{n}\alpha_t\boldsymbol{G}_t \tag{16}$$

2.3.3 域名分类 将$\boldsymbol{v}$输入到softmax层进行分类检测。域名分类得到域名的预测标签(DGA或者正常):

$$\partial=\text{soft}\max(\boldsymbol{W}_c\boldsymbol{v}+b_c) \tag{17}$$

定义训练损失loss为真实标签的负对数似然率,其中l是域名的真实标签:

$$\text{loss}=-\sum\lg p_{\partial,l} \tag{18}$$

3 仿真实验

3.1 数据集及训练策略

评估本文模型在12类DGA域名以及正常域名上的有效性。实验数据为二级域名,如表6所示,包括标签、数据集大小、域名最大和最小长度、包含字符数以及相对熵。80%数据用于训练模型,10%数据用于模型训练时验证,10%数据用于最终的测试。训练策略采取最大训练轮数为15;验证集上连续两轮准确率没有提升则结束模型训练;在测试集中测试。

3.2 实验结果与分析

3.2.1 与传统方法的比较 将ATT-GRU方法与LSTM方法[11]、GRU方法[12]、使用统计特征(包括域名长度,bigram,trigram)[3-4]的逻辑回归LR方法进行对比。另外,将ATT-GRU模型中的GRU编码器替换为LSTM单元构成ATT-LSTM方法,也作为对比方法。

表6 域名数据集

Table 6 Domain dataset

Family	Label	Size	Max length	Min length	CharSize	$\overline{H}$
legit	0	120 000	53	1	37	0.861
corebot	1	10 000	23	15	34	0.999
cryptolocker	1	10 000	31	8	25	0.999
dircrypt	1	10 000	20	8	26	0.999
kraken	1	10 000	12	7	26	0.995
locky	1	10 000	17	7	25	0.999
1akbot	1	10 000	25	8	26	0.999
ramdo	1	10 000	31	8	13	0.999
ramnit	1	10 000	19	8	25	0.999
simda	1	10 000	23	1	26	0.954
banjori	1	10 000	15	6	34	0.930
gozi	1	10 000	24	12	26	0.857
suppobox	1	10 000	20	9	26	0.880

ATT-GRU、ATT-LSTM、GRU和LSTM模型的参数选择相同,包括嵌入层维度为128维,GRU和LSTM都选取128个单元,Dropout率为0.5。

对模型的整体评价标准为精确率(Precision)和召回率(Recall)以及$F1$值。精确率反映模型对正常域名的误报情况,召回率反映模型对DGA域名的检测率,$F1$值综合考虑模型的有效性,公式如下:

$$\text{Precision}=\frac{TP}{TP+FP} \tag{19}$$

$$\text{Recall}=\frac{TP}{TP+FN} \tag{20}$$

$$F1=\frac{\text{Precision}\times\text{Recall}\times 2}{\text{Precision}+\text{Recall}} \tag{21}$$

$$\text{Accuracy}=\frac{TP+TN}{TP+TN+FP+FN} \tag{22}$$

实验结果见表7。结果表明,ATT-GRU方法的$F1$值为0.993 3,是所有方法中最好的方法;ATT-GRU方法比传统应用统计特征的LR方法识别效果好,说明新出现的DGA域名规避了传统统计特征,LR方法不适合检测这类域名。

ATT-GRU方法的召回率和精确率比传统方法中表现最优的GRU分别提高了0.93%和4.06%,表明注意力机制可以更好地检测DGA域名,提升了整

表 7　不同方法的精确率、召回率、$F1$ 值

Table 7　Precision, recall and $F1$ score of different algorithms

Algorithm	Recall/%	Precision/%	$F1$
ATT-GRU	**99.26**	**99.40**	**0.993 3**
ATT-LSTM	99.22	**99.40**	0.993 0
GRU	98.33	95.34	0.968 1
LSTM	97.76	95.59	0.966 6
LR	90.51	88.46	0.894 7

体的分类效果。

3.2.2　误报率分析　在保证精确率的前提下, 误报率是安全系统考虑的重点。误报的发生会造成大量系统资源的误消耗, 长期的高误报会降低安全检测的可靠性。

分别对正常域名 legit、高随机性域名和低随机性域名进行检测。对不同种类 DGA 域名识别效果的评价标准为识别率(Detection rate), 公式如下:

$$\text{Detection rate} = \begin{cases} \dfrac{TP}{TP+FN}, & \text{legit} \\ \dfrac{TN_i}{FP_i+TN_i}, & \text{DGA}_i \end{cases} \tag{23}$$

将 DGA 域名根据单字符随机程度分为两类: 低随机性域名 DGA-low 和高随机性域名 DGA-high。其中 DGA-low 包括半随机 DGA 域名 banjori、两类基于词典的 DGA 域名 gozi 和 suppobox; DGA-high 包括表 6 中剩余的 DGA 域名, 实验结果见表 8。

表 8　不同种类域名检测结果比较

Table 8　Detecting results comparison of different types of domains

Algorithm	Detection rate /%		
	legit	DGA-high	DGA-low
ATT-GRU	**99.40**	**99.56**	**99.27**
ATT-LSTM	**99.40**	99.52	99.00
GRU	95.20	99.08	96.60
LSTM	95.50	98.82	95.40
LR	88.20	96.36	76.53

由表 8 可以看出, ATT-GRU 方法的检测效果在 3 类域名上比其他方法更有效。其中 GRU、LSTM 和 LR 方法对高随机性 DGA 域名的识别率高, 但不能很好地区分正常域名和低随机性 DGA 域名, 特别是 LR 方法的识别率只有 88.20% 和 76.53%, 这些方法对正常域名误报率较高。

ATT-GRU 和 ATT-LSTM 方法对低随机性域名的识别率均有提高, 比 GRU 和 LSTM 方法分别提高了 2.67% 和 3.6%。说明注意力机制解决了区分正常域名和低随机性 DGA 域名的问题, 从而降低了模型的误报率, 提升了模型可靠性。

3.2.3　域名字符组合随机性分析　图 4 示出了注意力机制获取的对应域名字符权值的分布, 颜色的深浅反映字符组合的随机性。如, banjori 类 DGA 域名前 4 个字符属于 c_{ch}, 是该类域名半随机性的体现, 被注意力机制完全关注到了; 基于词典的 suppobox 类 DGA 域名中 edge(c_{st})部分 'ts'、'gs'被注意力机制着重关注到, 并延伸到 4 个字符。

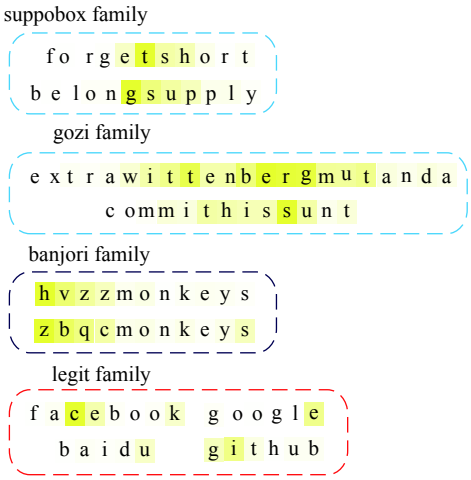

图 4　二级域名注意力权重

Fig. 4　Attention weight of second level domain

对于 gozi 类 DGA 域名, 域名构成为多个单词的拼接, 注意力机制关注到较长的单词拼接部分。由于单词首尾都与其他单词有拼接, 故在 GRU 编码上关注了高随机性的多个字符组合, 其中 edge(c_{st}) 的注意力权重相对 c_{st} 内部的字符组合仍较高。对于正常域名 legit 类, facebook 的单词拼接部分因为随机性低, 没有被关注, 但关注到其他字符组合部分'ce'。说明注意力机制有效地关注了域名中体现高随机性的多字符组合。

3.2.4　训练中准确率与损失函数值　不同方法在训练过程中的准确率变化与损失函数值的变化如图 5 所示, 其中的准确率和损失值都是在验证集上评估得到。

图 5 中, ATT-GRU在第 10 轮训练收敛, 比 ATT-LSTM 提早一轮。因为 GRU 结构相比 LSTM 减少了一个控制门的参数学习, 减少了模型的复杂度, 但两者最终的准确率相近。使用 LSTM 导致时间复杂度的提升并没有得到有效的准确率提升, 所以结合时间效率和准确率, 本文使用 GRU 进行多字符组合编码。

从准确率变化可以看出, 采取注意力机制的 ATT-GRU 和 ATT-LSTM 方法准确率收敛最快, 从第 3 轮

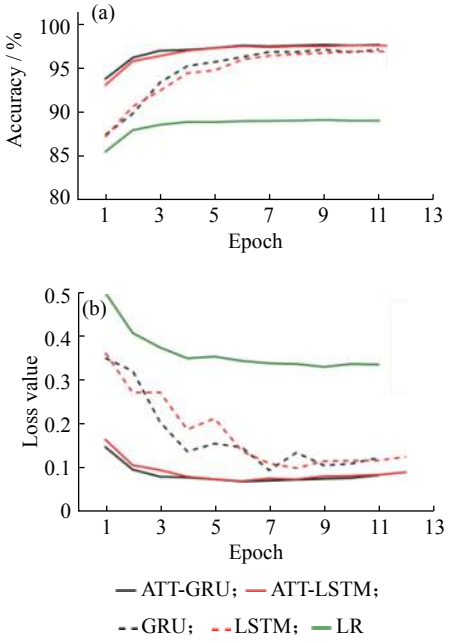

图5 训练准确率(a)与损失值(b)变化趋势

Fig. 5 Trends of training accuracy (a) and loss value (b)

训练开始准确率就提升至95%,此后的训练准确率缓慢提升至97.6%。而GRU和LSTM方法准确率收敛较慢,初始的准确率只有86%,经过8轮训练准确率才缓慢提升至95%,在这一过程中其准确率还有一次较大幅度的下降,其损失函数也有较大的反弹。说明注意力机制能更好地提取出正常域名和DGA域名的不同,有助于模型收敛。

4 结束语

本文提出了DGA域名的随机模型,将GRU神经网络与注意力机制结合提出了ATT-GRU模型,关注域名序列中随机性高的字符组合,进行DGA域名的检测。使用GRU对域名中变长字符组合编码;引入注意力机制,通过注意力权重强化了域名序列中高随机性的编码。

从实验结果可以看出,ATT-GRU模型与现有算法对比,提高了对低随机性DGA域名的识别,进而降低了对正常域名的误报率,整体模型的评价优于现有算法。

参考文献:

[1] KHARRAZ A, ROBERTSON W, BALZAROTTI D, *et al*. Cutting the gordian knot: A look under the hood of ransomware attacks[J]. Lecture Notes in Computer Science, 2016, 9148: 3-24.

[2] PLOHMANN D, YAKDAN K, KLATT M. A comprehensive measurement study of domain generating malware[C]// 25th USENIX Security Symposium. Austin: Usenix, 2016: 263-278.

[3] YADAV S, REDDY A K K, REDDY A L N, *et al*. Detecting algorithmically generated malicious domain names[C]// ACM SIGCOMM Conference on Internet Measurement. Australia: DBLP, 2010: 48-61.

[4] YADAV S, REDDY A K K, REDDY A L N, *et al*. Detecting algorithmically generated domain-flux attacks with DNS traffic analysis[J]. IEEE/ACM Transactions on Networking, 2012, 20(5): 1663-1677.

[5] MOWBRAY M, HAGEN J. Finding domain-generation algorithms by looking at length distribution[C]// IEEE International Symposium on Software Reliability Engineering Workshops. USA: IEEE, 2014: 395-400.

[6] SCHIAVONI S, MAGGI F, CAVALLARO L, *et al*. Phoenix: DGA-based botnet tracking and intelligence[C]// International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment. Cham: Springer, 2014: 192-211.

[7] BILGE L, SEN S, BALZAROTTI D, *et al*. Exposure: A passive DNS analysis service to detect and report malicious domains[J]. ACM Transactions on Information & System Security, 2014, 16(4): 1-28.

[8] VILLAMARIN-SALOMON R, BRUSTOLONI J C. Identifying botnets using anomaly detection techniques applied to DNS traffic[C]// Consumer Communications and Networking Conference. USA: IEEE, 2009: 476-481.

[9] 袁福祥, 刘粉林, 芦斌, 等. 基于历史数据的异常域名检测算法[J]. 通信学报, 2016, 37(10): 172-180.

[10] KRISHNAN S, TAYLOR T, MONROSE F, *et al*. Crossing the threshold: Detecting network malfeasance via sequential hypothesis testing[C]//2013 43rd Annual IEEE/IFIP International Conference on Dependable Systems and Networks(DSN). USA: IEEE, 2013: 1-12.

[11] WOODBRIDGE J, ANDERSON H S, AHUJA A, *et al*. Predicting domain generation algorithms with long short-term memory networks[EB/OL]. arXiv, 2016-11-2[1018-3-10]. https://arxiv.org/abs/1611.00791.

[12] LISON P, MAVROEIDIS V. Automatic detection of malware-generated domains with recurrent neural models[EB/OL]. arXiv, 2017-9-20[2018-3-10]. https://arxiv.org/abs/1709.07102.

[13] WANG F, JIANG M, QIAN C, *et al*. Residual attention network for image classification[C]//Computer Vision and Pattern Recognition. USA: IEEE, 2017: 6450-6458.

[14] JOZEFOWICZ R, ZAREMBA W, SUTSKEVER I. An empirical exploration of recurrent network architectures[C]// International Conference on International Conference on Machine Learning. USA: ACM, 2015: 2342-2350.

［15］ GAL Y, GHAHRAMANI Z. A theoretically grounded application of dropout in recurrent neural networks[C]// Advances in Neural Information Processing Systems. [s.l.]: [s.n.], 2016: 1019-1027.

Detecting Domain Generation Algorithm Based on Attention Mechanism

CHEN Lihuang, CHENG Hua, FANG Yiquan
（School of Information Science and Engineering, East China University of Science and Technology, Shanghai 200237, China）

Abstract: Domain generation algorithms (DGA) is one of the domain name detection key technologies for malicious C&C (command and control server) communication detection. Many existing detection methods, e.g., machine learning methods based on statistical features and deep learning methods based on recurrent neural networks, are usually based on the randomness of the domain name and have higher false positive rate and lower detection for these domain names with low random features. A main reason is that those methods cannot effectively extract some of the high randomness from the low-random domain names. This usually makes normal domain names be falsely reported as DGA domain names and increases the unnecessary consumption of the safety system and reduces its reliability. Aiming at the above shortcoming, this paper proposes a multi-character random extraction method for domain name. The gated recurrent unit (GRU) is utilized to encode multi-character combination and extract the randomness of the domain name. At the same time, the attention mechanism is introduced to extract the randomness of characters in the domain name and strengthen the high random features in the domain name. Besides, DGA domain name detection algorithm based on the attention-based recurrent neural network ATT-GRU is proposed to improve the identification validity on the low random DGA domain name. Finally, it is verified from experiments results that the ATT-GRU algorithm can achieve better accuracy and lower false positive rate than the traditional algorithm in detecting DGA domain name.

Key words: domain name detection**;** attention mechanism**;** gated recurrent unit (GRU)