基于多模态深度神经网络的应用层DDoS攻击检测模型

周奕涛[1,2],张 斌[1,2],刘自豪[3]
(1. 战略支援部队信息工程大学,河南郑州 450001; 2. 河南省信息安全重点实验室,河南郑州450001; 3. 61660部队,北京 100080)

摘 要: 为进一步提升应用层DDoS攻击检测准确率,提出一种将流量与用户行为特征相结合且模型参数可高效更新的应用层DDoS攻击检测模型. 为统一处理流量与用户行为特征的异源数据,利用多模态深度(Multimodal Deep Learning,MDL)神经网络从数据流量与网页日志中提取流量与用户行为深层特征后输入汇聚深度神经网络进行检测. 为减少MDL神经网络参数更新时的灾难性遗忘现象,在模型参数更新过程中基于弹性权重保持(Elastic Weight Consolidation,EWC)算法为重要模型参数增加惩罚项,保持对初始训练数据集检测准确率的同时,提升对新数据集的检测性能. 最后,基于K-Means算法获得模型初始训练数据集聚类,并筛选出新数据集中聚类外数据进行模型参数更新,防止EWC算法因数据相关性过高而失效. 实验表明,所提应用层DDoS检测模型检测准确率可达98.2%,且相对MLP_Whole方法模型参数更新性能较好.

关键词: 应用层DDoS攻击; 攻击检测模型; 多模态深度神经网络; 弹性权重保持算法; 参数更新
中图分类号: TP393 **文献标识码:** A **文章编号:** 0372-2112(2022)02-0508-05
电子学报URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20210009

Application Layer DDoS Detection Model Based on Multimodal Deep Learning Neural Network

ZHOU Yi-tao[1,2],ZHANG Bin[1,2],LIU Zi-hao[3]
(1. *SSF Information Engineering University*, *Zhengzhou*, *Henan* 450001, *China*;
2. *Key Laboratory of Information Security*, *Zhengzhou*, *Henan* 450001, *China*; 3. *No.*61660 *Troop*, *Beijing* 100080, *China*)

Abstract: To further improve the accuracy of application-layer DDoS attack detection, an application-layer DDoS attack detection model is proposed to combine traffic and user behavior features and to update model parameters efficiently. To integrate the heterogeneous data of traffic and user behavior characteristics, a multimodal deep learning(MDL) neural network is applied to extract the deep features of traffic and user behavior, which are employed for detection. To alleviate catastrophic forgetting in the update process of the MDL neural network, a penalty item is added to the important parameters based on the elastic weight consolidation(EWC) algorithm. The detection performance on the new dataset is improved while maintaining the detection accuracy of the initial training dataset. Based on the K-Means algorithm, the clusters of the initial training dataset are calculated. To prevent the EWC algorithm from failing due to high data correlation, the data outside the clusters are used to update model parameters. Experiments show that the detection accuracy of the proposed application layer DDoS detection model reaches 98.2 %, and it has better model update performance than the MLP_Whole method.

Key words: application layer DDoS attack; attack detection model; multimodal deep learning network; elastic weight consolidation algorithm; parameter update

1 引言

应用层DDoS攻击具有隐蔽性强、使用合法连接、与正常用户集中访问(Flash Crowd)事件相似等特点,近年来得到快速发展,已逐步成为主流的DDoS攻击方式[1]. 近年来,研究者通过分析应用层DDoS攻击与正常访问之间的流量特征或用户行为特征差异进行攻击

收稿日期:2020-12-22;修回日期:2021-11-12;责任编辑:李勇锋
基金项目:信息保障技术重点实验室开放基金(No.KJ-15-109);信息工程大学新兴科研方向培育基金(No.2016604703);信息工程大学科研项目(No.2019f3303)

检测,该方向已成为研究热点.

目前,应用层DDoS攻击检测主要有三种方法.第一种为基于人机测试的检测方法.这种方法通过向服务请求者发送人机测试问题,根据服务请求者的回答判断身份,进而实现应用层DDoS攻击检测.例如CAPTCHA方法[2],给予服务请求者人工容易解决而机器却难以识别的问题,进而识别攻击者;以及Speak-up方法[3],通过要求服务请求者提高访问速率以识别访问速率达到上限的攻击者.这种方法简单有效,但是需要消耗额外的服务器资源,并且会降低服务质量[4].第二种为基于流量特征的检测方法.这种方法基于应用层DDoS攻击对网络流量的统计特征改变以实现攻击检测,通过提取区分度较高的流量特征,识别攻击流量与正常流量,如基于机器学习的检测方法[5],以及基于数学统计学的检测方法[6].这种方法检测速率快,支持实时检测,但是难以区分Flash Crowd事件与HTTP Flood攻击[7].第三种为基于用户行为特征的检测方法,这种方法基于正常用户与攻击用户访问行为的差异性进行攻击检测,如基于访问顺序的检测方法[8]以及基于用户行为特征聚类的检测方法[9].这种方法检测准确率较高,对Flash Crowd事件与HTTP Flood攻击具有一定的区分能力,但是建模复杂,检测速率较慢,对用户访问记录较少的Slow DDoS攻击检测准确率较低[7].

为进一步提升应用层DDoS攻击检测准确率,提出一种基于MDL(Multimodal Deep Learning)神经网络结合流量与用户行为特征的应用层DDoS攻击检测模型.有效区分Flash Crowd事件与HTTP Flood攻击,同时对Slow DDoS攻击同样具备良好的检测能力,实现更为全面的应用层DDoS攻击检测.同时,为减少MDL神经网络在网络参数更新过程中的灾难性遗忘现象[10],实现检测模型参数的高效更新,针对MDL神经网络提出一种基于EWC(Elastic Weight Consolidation)算法[11]的模型参数高效更新方法EWC-UD,无需存储原始数据集,即可在不丢失原有知识的前提下,获得对新数据的检测能力,提升检测模型的更新性能.

2　基于MDL神经网络的应用层DDoS攻击检测模型

2.1　应用层DDoS攻击检测流程

为获得流量与用户行为特征,分别从网络关键节点处部署嗅探器以收集网络数据包Pcap文件,以及从网页服务器主机上收集服务器日志.Pcap文件主要记录流经网络关键节点处的数据包信息,通过统计Pcap文件信息,可以得到反映网络数据流的流量大小、流量速率、数据包分片信息等流量特征,而网页服务器日志主要记录用户请求网页的内容、访问顺序、请求状态码等信息,通过统计网页服务器日志,可以得到用户访问请求成功率、阅读时间等用户行为特征.

流量与用户行为特征经过提取后输入多模态深度神经网络.多模态深度神经网络由多个深度神经网络构成,流量特征向量与用户行为特征向量分别输入独立的流量深度神经网络(Traffic Deep Neural Network,T-DNN)与用户行为深度神经网络(User-behavior Deep Neural Network,U-DNN),T-DNN与U-DNN通过全连接的多层神经元将流量与用户行为特征向量转换为深层抽象特征,以相同形式表示流量与用户行为特征,从而实现特征结合.随后,通过一个汇聚神经网络(Merge Deep Neural Network,M-DNN)将流量与用户行为的深层抽象特征进行汇聚并输出最终结果.

2.2　基于EWC算法的模型参数更新方法

MDL神经网络作为连接性神经网络,在参数更新过程中存在灾难性遗忘问题,即对新数据集的学习可能导致神经网络遗忘初始数据集知识,使得对初始数据集所反映攻击类型的检测效果下降.因此,针对MDL神经网络,设计了一种基于EWC算法的模型参数更新方法EWC-UD,无需存储全部原始数据,仅需原始数据的少量抽样以及现有的神经网络参数,即可实现模型更新.

设N_{Sample}为费雪信息矩阵计算样本数,$\theta_i(i\in\{1,2,\cdots,N\})$为MDL神经网络参数,其中$N$为MDL神经网络参数数目,$\boldsymbol{F}$为费雪信息矩阵,$L_{\text{B}}(\theta)$为仅训练新数据集训练过程的损失函数,$L_{\text{EWC-UD}}(\theta)$为EWC-UD方法训练过程中的损失函数,$y_p(x)$为MDL神经网络预测输出.

首先,利用原数据的少量样本,与现有的MDL神经网络参数计算MDL神经网络输出关于神经网络参数的一阶导数,以获得反映现有神经网络参数对初始数据集重要程度的费雪信息矩阵,如下式所示:

$$\boldsymbol{F}_{i,j}=\begin{cases}\sum\limits_{k=1}^{N_{\text{sample}}}\left(\dfrac{\partial\log(y_p(x_k))}{\partial\theta_i}\right), & i=j\\ 0, & i\neq j\end{cases}\tag{1}$$

式(1)中,i,j分别为费雪信息矩阵的行标与列标.在之后的更新过程中,根据式(1)得到MDL神经网络参数对应的重要程度参数,即费雪信息矩阵的对角元素后,修改更新过程的损失函数,为重要的MDL神经网络参数增加惩罚项,使其在更新过程中难以改变,如下式所示:

$$L_{\text{EWC-UD}}(\theta)=L_{\text{B}}(\theta)+\frac{1}{2}\sum_i\boldsymbol{F}_i\cdot(\theta_i-\theta^*_{\text{A},i})^2\tag{2}$$

2.3　基于K-Means算法的数据筛选方法

记初始数据集A、新数据集B为D_{A}、D_{B},θ_{A}、θ_{B}、$\theta_{\text{A,B}}$

分别为MLP神经网络分别在D_A、D_B以及联合D_A与D_B下训练后的模型参数，设θ_{EWC-UD}为经过EWC-UD更新后的模型参数.

EWC-UD方法正确性证明需要将神经网络训练过程看作条件概率的寻优过程，从而证明$\theta_{A,B}$与θ_{EWC-UD}近似相等. 而在联合数据集D_A、D_B下训练神经网络的过程可表示为：$\theta_{A,B}=\mathrm{argmax}_\theta(\log(P(\theta|D_A,D_B)))$. 若数据集$D_A$、$D_B$具有较强独立性，根据条件概率公式与贝叶斯公式，$\log(P(\theta|D_A,D_B)$可以拆分为下式：

$$\begin{aligned}\log(P(\theta|D_A,D_B))&=\log\left(\frac{P(\theta,D_A,D_B)}{P(D_A,D_B)}\right)\\&=\log P(D_B|\theta)+\log P(\theta|D_A)-\log P(D_B)\end{aligned}\tag{3}$$

其中，$\log P(D_B|\theta)$可看作为训练过程中的交叉熵函数$L_B(\theta)$，$\log P(\theta|D_A)$经过拉普拉斯对角近似后，忽略3次方及以上项后可表示为

$$\log P(\theta_A|D_A)-\frac{1}{2}\frac{\partial^2\log P(\theta|D_A)}{\partial\theta^2}(\theta-\theta_A)^2$$

$\log P(D_B)$为常数. 综上对式(3)改写后，对参数θ进行寻优可得

$$\theta_{EWC-UD}=\arg\min_\theta(L_B(\theta)+\frac{1}{2}\sum_i\boldsymbol{F}_i^A\cdot(\theta_i-\theta_A)^2)\approx\theta_{A,B}$$

综上，EWC-UD更新方法的正确性前提是要求初始数据集与新数据集具有较强的独立性. 因此，设计一种基于K-Means聚类方法的数据筛选方法，如图1所示.

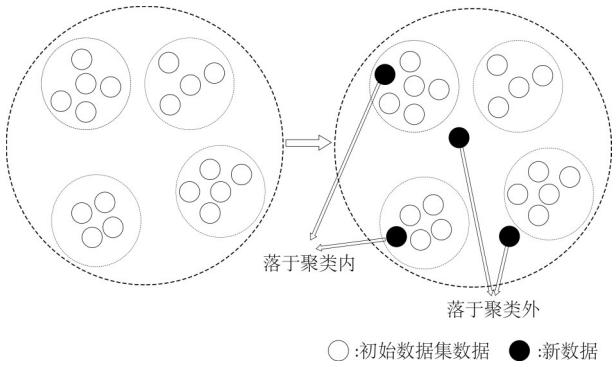


图1 K-Means数据筛选方法图示

图1中，空白节点表示初始数据集数据，黑色节点表示新数据. 假设聚类簇划分为$(C_1,C_2,C_3,\cdots,C_k)$，每个类簇存在一个质心$\mu_i=\frac{1}{|C_i|}\sum_{x\in C_i}x$，K-Means算法通过最小化平方误差$E=\sum_i^k\sum_{x\in C_i}\|x-\mu_i\|^2$，让类簇内的点尽可能密集，从而实现聚类效果. 采用该类方法，可以有效防止因初始数据集与新数据集相关性过高而导致的更新失败问题，同时过滤了重复数据，减少模型参数更新过程的计算量，进一步提升模型参数更新效率.

3 实验结果与分析

实验主要验证多模态深度神经网络的检测性能、EWC-UD方法的更新性能、K-Means数据筛选方法性能. 共使用3个应用层DDoS攻击数据集，包括CIC 2017 App-DDoS Dataset[12]、CES-CIC-IDS2018-AWS[13]以及CIC DDoS 2019[14]. 提取CES-CIC-IDS2018-AWS中的DDoS攻击部分，与CIC 2017 App-DDoS Dataset结合成为融合数据集(简称为$\mathrm{Dataset_{old}}$). 使用CIC DDoS 2019作为新数据集(简称为$\mathrm{Dataset_{new}}$).

3.1 基于MDL神经网络的应用层DDoS攻击检测性能验证

为验证流量与用户行为特征分别对于HTTP Flood攻击与Slow DDoS攻击的检测能力，并验证结合流量与用户行为特征的检测性能，以检测准确率(Accuracy)、F1分数(F1_Score)、以及ROC曲线与X轴所围面积AUC作为评价参数.

基于流量特征、用户行为特征以及结合流量与用户行为特征的应用层DDoS攻击检测结果如图2所示.

实验表明，基于流量特征与基于用户行为特征对HTTP Flood与Slow DDoS攻击具有不同的检测性能. HTTP Flood攻击使用合法数据包，流量统计特征与Flash Crowd事件较为相似，但是请求分布随机，规律性较差，与正常用户访问行为存在较大差异，因此使用用户行为特征可以取得较好的检测效果，而使用流量特征则检测效果欠佳. 而Slow DDoS攻击通常采用少量而持续时间长的请求消耗服务器资源，其数据包分段信息等流量统计特征会发生较大改变，但是导致用户日志记录较少，因此使用用户行为特征对这类攻击难以检测，而基于流量特征则取得比较好的检测效果. 而所提模型结合流量与用户行为特征，对HTTP Flood攻击与Slow DDoS攻击都取得了更好的检测效果，在整体检测中表现同样为最优.

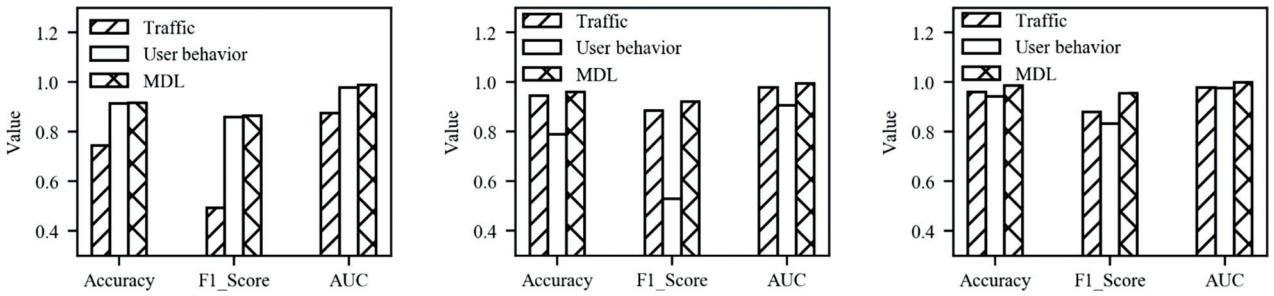
3.2 EWC-UD模型参数更新方法性能验证

为验证所提EWC-UD方法的模型参数更新性能，设置以下两类场景：

场景1：模型参数训练数据集为$\mathrm{Dataset_{old}}$，模型参数更新数据集为$\mathrm{Dataset_{new}}$. 该场景模拟检测模型在一段时间内进行计划内更新的情况.

场景2：模型参数训练数据集为$\mathrm{Dataset_{old}}$，模型参数更新数据集为CIC DDoS 2019中的Portmap攻击数据，该场景模拟检测模型在发现新型应用层DDoS攻击后进行紧急更新的场景.

设置以下两种MLP神经网络更新方式作为EWC-



(a) HTTP Flood攻击检测　(b) Slow DDoS攻击检测　(c) 整体检测

图2 流量与用户行为特征对各类应用层DDoS攻击检测性能比较

UD方法的对照组:

MLP_New:以 $Dataset_{old}$ 作为模型原参数训练数据集,以 $Dataset_{new}$ 作为模型参数更新所用数据集.

MLP_Whole:以 $Dataset_{old}$ 作为模型原参数训练数据集,以 $Dataset_{old}$ 与 $Dataset_{new}$ 的融合数据集作为模型参数更新所用数据集.

EWC-UD模型参数更新方法更新性能验证实验结果如表1所示.

实验表明,EWC-UD方法在场景1与场景2下都具备优秀的模型参数更新性能,相较于MLP_Whole方法,检测性能几乎无差别,但时间与空间开销显著下降.

表1 各类检测模型参数更新方法性能

	方法	Dataset_Old			Dataset_New			更新开销	
		Accuracy/%	AUC	F1	Accuracy/%	AUC	F1	Time/s	Memory/MB
场	MLP_Whole	97.01	0.978 9	0.907 8	98.66	0.987 1	0.993 1	718.53	11 874.22
景	MLP_New	96.68	0.977 5	0.899 9	99.43	0.999 1	0.996 8	128.28	7 467.34
1	EWC-UD	97.05	0.979 8	0.909 5	99.54	0.999 2	0.997 5	143.32	8 541.24
场	MLP_Whole	97.02	0.984 4	0.908 9	98.51	0.879 4	0.992 3	161.93	8 330.96
景	MLP_New	54.75	0.932 0	0.428 7	99.78	0.999 5	0.998 9	36.16	3 771.53
2	EWC-UD	95.32	0.976 6	0.865 0	99.73	0.999 3	0.998 6	38.45	5 697.93

3.3 K-Means数据筛选方法性能验证

以 $Dataset_{old}$ 为模型参数训练数据集,以 $Dataset_{new}$ 为更新数据集,分别随机抽样初始数据集的1%、2%、3%、4%、5%的数据样本,并与新数据集进行合并. 以初始数据集与新数据集的融合数据集作为测试数据集,进行5次重复实验,计算经过EWC-UD方法进行模型参数更新后,检测模型的检测准确率、F1分数、召回率、精确率的实验平均值,并与经过K-Means数据筛选后再经过EWC-UD方法进行模型参数更新后的检测模型进行对比,实验结果如表2所示.

实验表明,随着初始数据集与新数据集相关程度的增加,EWC-UD方法的更新性能显著下降,而经过数据筛选,删除与聚类中心点过近的重复数据后,EWC-UD方法的更新性能始终保持在较高水平,表明K-Means数据筛选方法可有效改善EWC-UD方法难以处理相关性过高数据集的问题.

4 总结

提出了一种基于MDL神经网络的应用层DDoS攻击检测模型,结合流量与用户行为特征对应用层DDoS攻击进行检测. 同时设计了一类高效的模型更新方法EWC-UD,并针对EWC算法对初始数据集与新数据集独立性要求较高的问题,提出一种K-Means数据筛选方

表2 K-Means数据筛选方法性能验证

抽样比例	平均准确率/%		F1-Score		Recall/%		Precision/%	
	K-Means	Normal	K-Means	Normal	K-Means	Normal	K-Means	Normal
1%	**95.89**	90.99	**0.940 1**	0.892 7	**92.62**	80.77	**95.45**	83.85
2%	**96.13**	86.80	**0.925 9**	0.757 6	**93.04**	82.22	**92.14**	74.54
3%	**92.79**	87.82	**0.876 7**	0.704 8	**93.30**	83.98	**83.56**	73.54
4%	**96.93**	68.67	**0.922 7**	0.367 7	**92.62**	44.27	**95.45**	31.71
5%	**91.07**	66.56	**0.844 2**	0.283 9	**93.26**	24.86	**77.15**	20.01

法,实现了应用层DDoS攻击模型的高效更新.实验表明,所提模型具备良好的应用层DDoS攻击检测性能,且具有较高的模型参数更新效率.

参考文献

[1] 孙长华, 刘斌. 分布式拒绝服务攻击研究新进展综述[J]. 电子学报, 2009, 37(7): 1562-1570.
SUN Chang-hua, LIU Bin. Survey on new solutions against distributed denial of service attacks[J]. Acta Electronica Sinica, 2009, 37(7): 1562-1570. (in Chinese)

[2] SARAVANAN A, BAMA S, KADRY S, et al. A new framework to alleviate DDoS vulnerabilities in cloud computing[J]. International Journal of Electrical & Computer Engineering, 2019, 9(5): 4163-4175.

[3] GULIHAR P, GUPTA B B. Cooperative Mechanisms for Defending Distributed Denial of Service(DDoS) Attacks [M]//Handbook of Computer Networks and Cyber Security. Germany: Springer, 2020: 421-443.

[4] PRASEED A, THILAGAM P S. DDoS attacks at the application layer: Challenges and research perspectives for safeguarding Web applications[J]. IEEE Communications Surveys & Tutorials, 2018, 21(1): 661-685.

[5] 张斌, 刘自豪, 董书琴, 等. 基于偏二叉树SVM多分类算法的应用层DDoS检测方法[J]. 网络与信息安全学报, 2018, 4(3): 24-34.
ZHANG Bin, LIU Zi-hao, DONG Shu-qin, et al. App-DDoS detection method using partial binary tree based SVM algorithm[J]. Journal of Network and Information Security, 2018, 4(3): 24-34. (in Chinese)

[6] LIN H, CAO S, WU J, et al. Identifying application-layer DDoS attacks based on request rhythm matrices[J]. IEEE Access, 2019, 7: 164480-164491.

[7] JIANG J, YU Q, YU M, et al. ALDD: A hybrid traffic-user behavior detection method for application layer DDoS[C]// The 17th IEEE International Conference on Trust, Security and Privacy in Computing and Communications/12th IEEE International Conference on Big Data Science and Engineering. Piscataway, NJ: IEEE, 2018: 1565-1569.

[8] LI B, GAO M, MA L, et al. Web application-layer DDoS attack detection based on generalized Jaccard similarity and information entropy[C]//International Conference on Artificial Intelligence and Security. Germany: Springer, 2019: 576-585.

[9] 刘自豪, 张斌, 祝宁, 等. 基于改进AP聚类算法的自学习应用层DDoS检测方法[J]. 计算机研究与发展, 2018, 44(5): 729-736.
LIU Zi-hao, ZHANG Bin, ZHU Ning, et al. Adaptive app-DDoS detection method based on improved AP algorithm [J]. Journal of Computer Research and Development, 2018, 44(5): 729-736. (in Chinese)

[10] FRENCH R M. Catastrophic forgetting in connectionist networks[J]. Trends in Cognitive Sciences, 1999, 3(4): 128-135.

[11] KIRKPATRICK J, PASCANU R, RABINOWITZ N, et al. Overcoming catastrophic forgetting in neural networks [J]. Proceedings of the National Academy of Sciences, 2017, 114(13): 3521-3526.

[12] JAZI H H, GONZALEZ H, STAKHANOVA N, et al. Detecting HTTP-based application layer DoS attacks on web servers in the presence of sampling[J]. Computer Networks, 2017, 121(1): 25-36.

[13] SHARAFALDIN I, LASHKARI A H, GHORBANI A A. Toward generating a new intrusion detection dataset and intrusion traffic characterization[C]//The 4th International Conference on Information Systems Security and Privacy. Germany: Springer, 2018: 108-116.

[14] SHARAFALDIN I, LASHKARI A H, HAKAK S, et al. Developing realistic distributed denial of service(DDoS) attack dataset and taxonomy[C]//2019 International Carnahan Conference on Security Technology. Piscataway, NJ: IEEE, 2019: 1-8.

作者简介

周奕涛 男,1996年生,湖南怀化人,信息工程大学硕士.主要研究方向为应用层DDoS攻击检测.
E-mail: zyt1996715@163.com

张 斌 男,1969年生,河南郑州人.现为信息工程大学教授,博士生导师.主要研究方向为网络空间安全.