SDN 中基于条件熵和 GHSOM 的 DDoS 攻击检测方法

田俊峰[1,2]，齐鎏岭[1,2]

（1. 河北大学网络空间安全与计算机学院，河北 保定 071002；2. 河北省高可信信息系统重点实验室，河北 保定 071002）

摘 要：软件定义网络（SDN, software defined networking）简化了网络结构，但同时控制器也面临着“单点失效”的安全威胁。攻击者可以发送大量交换机流表中并不存在的伪造数据流，影响网络正常性能。为了准确检测这种攻击的存在，提出了基于条件熵和 GHSOM（growing hierarchical SOM）神经网络的 DDoS 攻击检测方法 MBCE&G 。首先，依据此 DDoS 的阶段性特征，定位了网络中的受损交换机以发现可疑攻击流；然后，依据可疑攻击流种类的多样性特征，以条件熵的形式提取了四元组特征向量，将其作为神经网络的输入特征进行更加精确的分析；最后，搭建了实验环境完成验证。实验结果显示，MBCE&G 检测方法可以有效检测 SDN 中的 DDoS 攻击。

关键词：软件定义网络；条件熵；神经网络；DDoS 攻击

中图分类号：TP309

文献标识码：A

doi: 10.11959/j.issn.1000−436x.2018140

DDoS attack detection method based on conditional entropy and GHSOM in SDN

TIAN Junfeng[1,2], QI Liuling[1,2]

1. School of Cyber Security and Computer, Hebei University, Baoding 071002, China
2. Key Lab on High Trusted Information System in Hebei Province, Baoding 071002, China

Abstract: Software defined networking (SDN) simplifies the network architecture, while the controller is also faced with a security threat of “single point of failure”. Attackers can send a large number of forged data flows that do not exist in the flow tables of the switches, affecting the normal performance of the network. In order to detect the existence of this kind of attack, the DDoS attack detection method based on conditional entropy and GHSOM in SDN (MBCE&G) was presented. Firstly, according to the phased features of DDoS, the damaged switch in the network was located to find the suspect attack flows. Then, according to the diversity characteristics of the suspected attack flow, the quaternion feature vector was extracted in the form of conditional entropy, as the input features of the neural network for more accurate analysis. Finally, the experimental environment was built to complete the verification. The experimental results show that MBCE&G detection method can effectively detect DDoS attacks in SDN network.

Key words: software defined networking, conditional entropy, neural network, DDoS attack

1 引言

软件定义网络作为新型网络架构给传统网络带来了巨大的变革和提升。在云计算环境中，随着网络规模的不断扩大，以 TCP/IP 架构为基础的传统网络架构在网络优化时遇到了诸多问题[1]，主要表现为：1) 由于控制平面和数据平面难以分离，使网络更新变得十分烦琐，一旦确定了转发

收稿日期：2017−09−08；**修回日期**：2018−07−03

基金项目：国家自然科学基金资助项目（No.61170254）；河北省自然科学基金资助项目（No.F2016201244）

Foundation Items: The National Natural Science Foundation of China (No.61170254), The Natural Science Foundation of Hebei Province (No.F2016201244)

策略，如果后期需要对策略进行调整，只能通过对设备配置进行大规模更改才能实现；2) 大规模的网络设备使管理员对这些设备的管理也变得十分困难。

SDN 作为新型网络架构，将传统网络架构解耦为数据平面、控制平面和应用平面，简化了网络结构，使网络控制变得更为灵活和集中。其开放性和可编程性，令 SDN 在网络虚拟化、云数据中心网络、无线局域网和云计算中得到了大规模的应用[2]。

由于 SDN 的广泛应用，其本身的安全问题日益突出[1]。其中，DDoS 攻击由于可以造成控制器过载，对 SDN 的威胁巨大。Shin 等[3]首次论述了在 SDN 中存在 DDoS 攻击的可能性，并且进行了实验论证。Neelam 等[4]将传统网络中的各种 DDoS 攻击形式在 SDN 中逐一进行了测试，总结出不同种类的 DDoS 攻击对 SDN 的影响。

OpenFlow 是 SDN 中的南向接口协议，定义了 SDN 交换机和 SDN 控制器之间的通信规则。在 OpenFlow 中，如果交换机遇到无法匹配的流请求信息，会利用 packet_in 数据帧对其封装并发送至控制器，由控制器为其提供相关的应答策略，这种工作模式极大地增加了控制器遭受 DDoS 攻击的可能性。攻击者可以制造大量的在交换机流表中并不存在的恶意流请求信息，进而交换机将大量的 packet_in 数据帧发送至控制器，使控制器资源被耗尽。

为了准确分析和精确检测这种 DDoS 攻击，本文提出将信息熵和 GHSOM 神经网络进行结合，将 DDoS 流量的攻击特征用条件熵进行量化，再将其输入 GHSOM 网络中进行攻击检测。并且在进行流量特征提取之前，首先定位受损交换机，实现对可疑流量的精确定位。

2　相关工作

Shin 等[3]首次通过实验论述了在 SDN 中存在 DDoS 攻击的可能性，之后的大量工作也同样指出了在 SDN 中存在针对控制器的 DDoS 攻击隐患。Chen 等[5]以不同速率向 SDN 中发送伪造的数据流，观察合法数据流受到的影响。实验结果表明：当攻击流强度超过 300 flows/s 时，便会出现合法数据流的分组丢失情况；当攻击流强度达到 1 200 flows/s 时，超过 60%的合法数据流匹配失败。Neelam 等[4]分析了传统 DDoS 攻击对 SDN 的影响，其中，HTTP 和 TCP_SYN 洪泛攻击并不需要大规模的流量冲击，即可对控制器造成 90%的 table_miss 攻击效果，且其对控制器造成的危害是不可逆的。一些针对 OpenFlow 协议的研究[6-7]提出 DDoS 攻击同样可以对 OpenFlow 协议造成影响，进而影响交换机和控制器之间的通信情况。

SDN 中的 DDoS 攻击检测方法主要分为以下 3 类：基于策略的检测方法、基于统计信息的检测方法和基于机器学习的检测方法[8]。基于策略的检测方法类似于防火墙机制，通过允许合法请求、拒绝非法请求的形式，达到维护网络安全的目的。基于统计信息的检测方法通过信息熵可以良好地体现被测序列的随机性。Mousavi 等[9]提出了一种针对 SDN 控制器的 DDoS 攻击的早期检测算法，该算法基于数据分组目的地址的分布概率计算熵值，通过与阈值进行比较，可以在 250 个数据分组之内判断是否发生 DDoS 攻击，实现早期检测。该检测方法仅对数据分组的目的地址进行了熵值的计算，而对攻击特征的描述较为单一。Dong 等[10]提出了一种针对 SDN 控制器的 DoS 攻击的检测方法，该方法通过检测 low-traffic 确定了 low-traffic 事件，并且通过序贯概率比测试控制了攻击检测的漏报率和误报率。在基于机器学习的检测方法中，神经网络通过对训练样本的学习后，可以实现对攻击流量的高效检测。Braga 等[11]利用 SDN 集中管控的特点，提出利用自组织神经网络 SOM 对信息流进行分类，通过设立的六元组特征（信息流中数据分组的平均数量、信息流的平均字节数、信息流的平均持续时间、配对信息流的百分比、单信息流的增长率、不同端口的增长率）对 DDoS 攻击进行检测，然而并没有考虑到 DDoS 攻击对 SDN 控制器的影响。除此之外，SOM 神经网络神经元的排列形式较为固定，对攻击检测的实时性有一定影响。姚琳元等[12]同样使用了基于神经网络的检测方法，首先提取了基于目的 IP 地址的检测七元组，通过 GHSOM 神经网络对 DDoS 攻击进行了检测。相较 SOM 神经网络，GHSOM 神经网络具有生长、分层的特性，结构更加灵活，数据处理更加高效。

在传统网络中，研究者对 GHSOM 神经网络也进行了一定的研究工作。杨雅辉等[13]针对传统的 GHSOM 网络只能处理数字型样本的缺陷，对 GHSOM 网络进行改进，提出一种可以

混合处理数字型和字符型样本的改进 GHSOM 入侵检测算法。阳时来等[14]基于半监督 GHSOM 的入侵检测方法，将传统无监督的 GHSOM 模型进行了改进。

通过上述分析可知，虽然在传统网络中 DDoS 的检测算法多种多样，但是目前对 SDN 中 DDoS 的攻击检测主要还是基于信息熵和神经网络技术，并且 GHSOM 神经网络在传统网络中的入侵检测方面已有较成熟的研究。信息熵能够良好地体现被测序列的随机性，通过配置相应参数的阈值，可以对 DDoS 攻击进行检测。然而单纯使用基于熵的检测算法若存在待检测数据维数较多的情况，则难以对 SDN 中的大量数据进行精确刻画。神经网络具有较强的抽象和概括能力，可以对多维数据进行高效处理。

因此，结合 SDN 中针对控制器的 DDoS 攻击特征，提出了基于条件熵和 GHSOM 神经网络的 DDoS 攻击检测方法 MBCE&G。MBCE&G 具有如下特点。

1) 针对攻击的阶段性特征，通过确定 SDN 中的受损交换机来判断可疑的 DDoS 攻击流量。

2) 针对攻击种类的多样性特征，提出使用基于目的地址和目的端口的条件熵作为神经网络的输入向量，更加准确地表现了 DDoS 攻击流量中数据分组地址间多对一的映射关系。

3) 利用神经网络强大的分类能力，对 DDoS 攻击进行精确检测。

3 预备知识

3.1 条件熵

信息熵[15]用对数函数度量随机变量的期望，定量地表示变量的随机性，变量的随机性越大，其熵值越大，反之亦然。

假设某个随机变量 x，其取值集合为 $X=\{x_1,x_2,\cdots,x_n\}$，取值的概率分布为 $P=\{p_1,p_2,\cdots,p_n\}$，且为独立分布，其中，$\sum_{i=1}^{n}p_i=1$，$0\leqslant p_i\leqslant 1$，$i\in(1,\cdots,n)$，变量 x 的信息熵为

$$H=-\sum_{i=1}^{n}p_i\mathrm{lb}p_i \tag{1}$$

根据式(1)得到关于条件熵的定义，对于随机变量 X 和 Y，X 关于 Y 的条件熵可以表示为

$$\begin{aligned}&H(X|Y)\sum_{y}p(y)H(X|Y=y)\\&=-\sum_{y}p(y)\sum_{x}p(x|y)\mathrm{lb}\,p(x|y)\end{aligned} \tag{2}$$

条件熵可以更加精确地表示当某一个随机变量为定值时，另一个随机变量的随机性分布，可以很好地表现 SDN 中针对控制器的 DDoS 攻击流量的特征。

3.2 GHSOM 神经网络

GHSOM 网络为多层树状结构，如图 1 所示，每一层结构中包含多个独立的 SOM 网络，每个子网都遵循 SOM 过程。和 SOM 网络相比，GHSOM 网络增加了横向和纵向扩展方向，并且每个方向都有自己的扩展条件。

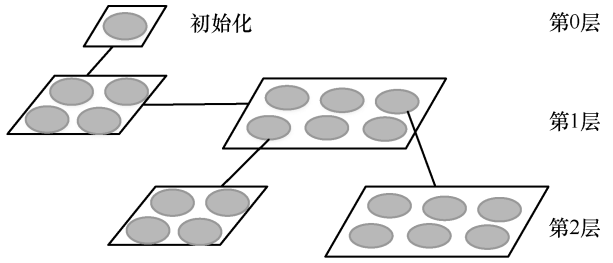


图 1 GHSOM 网络示意

第 0 层初始化：第 0 层只有一个神经元，在训练过程中，全部待训数据将落在该神经元上，并且将待训数据的平均值作为该神经元的初始权值。

训练过程：在训练开始之后，第 0 层的神经元作为父神经元会产生一个新的 SOM 子网，该子网共包含 4 个神经元，之后会从输入数据中取样执行 SOM 过程，进行进一步的细分。

在每个 SOM 中，根据最小欧氏距离作为竞争标准，得到最后的获胜神经元。$e_i(x)=\underset{j}{arg\min}\|e_i-x\|$，$e_i$ 表示获胜神经元，为 e_j 中的某一个，x 表示输入数据；权重调整遵循 $e_i(t+1)=e_i(t)+\alpha(t)\beta_i(t)\left[x(t)-e_i(t)\right]$，$t$ 表示迭代次数，α 表示学习率参数，β_i 表示获胜神经元 i 周围的邻域函数。重复上述竞争过程和权重调整过程，直到不再发生明显变化，则视为 SOM 过程结束。

GHSOM 算法采用量化误差（qe）和平均量化误差（mqe）来决定模型的生长。神经元 i 的量化误差 $qe_i=\sum_{X_j\in C_i}\|W_i-X_j\|$，其中，$W_i$ 表示神经元 i 的权值向量，C_i 表示映射到神经元 i 的所有输入模式

向量的集合。

横向扩展：计算子网中所有神经元的平均量化误差（mqe），并且每隔一定的训练次数后都会重新计算。如果 $mqe_i > qe_{\text{father}} \cdot \mu_1$（$qe_{\text{father}}$ 表示该子网对应的父神经元的量化误差，μ_1 表示横向扩展系数），则说明该层子网没有稳定，需要向子网中插入一行或一列神经元，然后继续进行 SOM 过程训练，并且判断上述条件。当不再满足上述条件时，进行纵向扩展。

纵向扩展：如果某个神经元 $qe > qe_0 \cdot \mu_2$（qe_0 表示第 0 层神经元的量化误差，μ_2 表示纵向扩展系数），则表示该神经元需要扩展出新的子网，此时重新构造一个 2×2 规模的子网，进行 SOM 过程训练，并且进行横向扩展条件判定。当不满足条件时，表示网络趋于稳定，过程结束。

4　基于条件熵和 GHSOM 的 DDoS 攻击检测方法

4.1　攻击特征

在针对控制器的 DDoS 中，攻击过程被分为 2 个阶段：攻击者发送大量伪造数据流到达 SDN 交换机，在交换机内部出现流表不匹配情况；交换机发送大量 packet_in 数据帧到达控制器。和传统网络中的 DDoS 攻击不同，攻击者不需要知道控制器的 IP 地址与 IP 端口，就可以发动针对控制器的 DDoS 攻击，属于盲 DDoS 攻击[16]。

然而，SDN 更多表现的是一种思想，即网络的自动化运维，具体表现为数据和控制分离。OpenFlow 协议是实现 SDN 数控分离的一种手段，就 OpenFlow 协议本身而言，它是一种网络控制器和网络转发设备之间交换信息和下发命令的协议，主要是为了实现二层交换和三层路由。因此，虽然 SDN 改变了传统网络路由设备的转发方式，但并没有改变数据分组的封装结构。所以在网络的数据传输中，目的 IP 地址仍然是数据分组到达目的地的重要依据。

所以，这种针对 SDN 控制器的 DDoS 攻击既有传统网络中 DDoS 攻击的一般特征，结合 SDN 的分层结构又有一些新的表现形式，具体如下。

1) 阶段性。此类攻击具有明显的阶段特征，攻击者发送攻击流量到达交换机，交换机产生 packet_in 数据帧发送至控制器，所以攻击流量并没有直接到达控制器。

2) 多样性。在传统网络中的 DDoS 攻击中，任何可以产生大量 packet_in 消息的攻击形式都可用来发动针对控制器的 DDoS 攻击。其中，文献[4]已经通过实验论证，HTTP 洪泛攻击和 TCP_SYN 洪泛攻击并不需要大规模流量，即可对控制器造成不可恢复的攻击效果。

3) 多对一映射。攻击者从多个攻击源对目标发起攻击，因此对于攻击目标而言，攻击流量中的源地址相对于目的地址为多对一映射，而正常流量则具有多对一、一对一、一对多这 3 种映射形式；除此之外，合法用户在一定时间段内请求的服务比较单一，而攻击者为了快速消耗目标资源，通常会请求尽可能多的服务，因此，目的端口与目的地址之间也存在着多对一的映射关系；最后，正常流量和攻击流量在数据分组长度上也有很大的区别。

基于上述特征，MBCE&G 首先通过分析受损交换机的存在，判断出 SDN 中确实存在流量突发情况；然后通过受损交换机的可疑流量，利用条件熵提取攻击特征；最后通过 GHSOM 神经网络对可疑流量进行分析，判断 DDoS 攻击。检测流程如图 2 所示。

4.2　对 SDN 交换机的分析

将攻击者发送的大量伪造数据流定义为新流

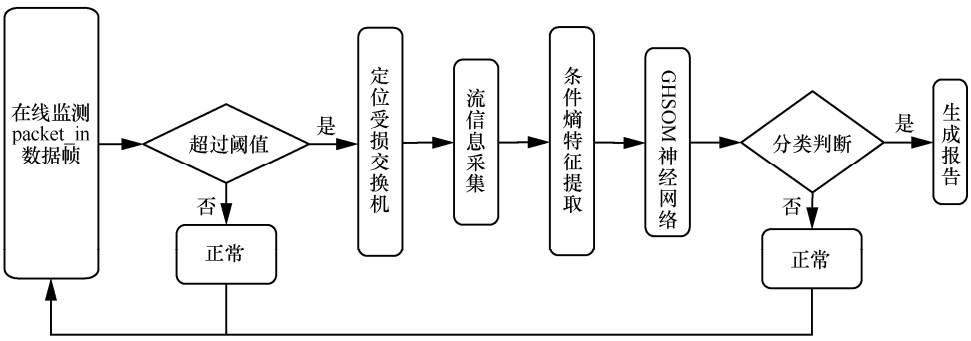


图 2　检测流程

（new_flow），那么检测算法的第一步就是确认这些 new_flow 的存在。在 OpenFlow 协议中，packet_in 数据帧可以侧面反映出交换机是否遇到了无法匹配的数据流。另一方面，由于交换机是攻击流到达的第一个 SDN 实体，因此首先通过监测 packet_in 数据帧定位受损交换机，进而确认 new_flow 的存在。

通过监测一段时间内 packet_in 数据帧的个数来表示交换机的流请求速率，进而说明有流量突发情况的存在。

将整个检测时长定义为 T，共有 m 个检测时隙。每个检测时隙的检测时长为 w，和控制器设置的 idle_timeout 相等，通常为 5 s。SDN 为了提高交换机的匹配效率，设置了 2 种超时时间：一种是软超时时间 idle_tiemout，表示一条流表项不再匹配数据分组时能持续的最大时间；另一种是硬超时时间 hard_timeout, 表示一条流表项在交换机流表中的最大生存时间。一旦达到时间期限，交换机就会自动删除相应的表项，此时交换机内的流表项就会更新。通过上述分析，可知 idle_time≤hard_time，即最少在一个 idle_time 时间内，交换机内的流表项是不变的，将 idle_time 的时间长度设置为每个检测时隙的检测时长，可在每个 idle_time 的时间长度内，检测 packet_in 数据帧的个数。

将每个检测时隙定义为 t，在每个检测时隙中，检查每个 packet_in 消息的源地址，用 PIM_i^t 表示交换机 i 在时隙 t 内发送的 packet_in 数据帧的集合，用 $num(PIM_i^t)$ 表示集合中 packet_in 数据帧的数量。这样，交换机在每个时隙的流请求速率定义为

$$R_i(t)=\frac{num(PIM_i^t)}{w} \tag{3}$$

在 m 个检测时隙中，关于交换机 i 的流请求速率的集合可以表示为

$$R_i=\left\{R_i(t)\,|\,t=1,2,\cdots,m\right\} \tag{4}$$

至此，网络中每个交换机的流请求速率的集合都已得到。在实验过程中，通过观测 packet_in 数据帧的流量变化，定义相关阈值为 R_i^α。当 $R_i>R_i^\alpha$ 时，认为网络中在交换机 i 处有流量突发情况产生，受损交换机定位完成，进行下一步更细致的检测。基于此，提出一种基于交换机流请求速率确定受损交换机的检测算法，通过监控交换机的流请求速率，检测出可疑攻击流，具体如下所示。

算法 交换机流请求速率算法

输入 packet_in 数据帧

输出 受损交换机

$R_i=\varnothing$;

for t = 1; $t\leqslant m$; t++ do

通过监测 packet_in 数据帧个数，计算 $R_i(t)$；

$R_i=R_i\cup R_i(t)$；

if $R_i>R_i^\alpha$

则 R_i 为受损交换机；

end if

end for

4.3 利用条件熵提取攻击特征

当检测到网络中交换机的流请求速率超过阈值时，并不能判断发生 DDoS 攻击，因为存在着合法用户的突发流量。因此，上述过程只能判断流量突发情况确实存在，不能作为判断攻击发生的依据，还需要更加精细地检测，本阶段对通过受损交换机的数据分组的头信息进行分析。

针对攻击特征，提出使用基于目的地址和目的端口的条件熵作为神经网络的输入向量。当确定网络中的受损交换机后，对通过受损交换机的流量进行特征提取，得到以下多维条件熵特征：{H(sip|dip), H(dport|dip), H(sip|dport), H(psize|dip)}。

1) H(sip|dip)：源 IP 地址关于目的 IP 地址的条件熵。对于 DDoS 攻击而言，攻击流量中源地址相对于目的地址具有明显的多对一的映射关系，但是正常流量间具有多对一、一对多与一对一等多种映射关系。所以，当源 IP 地址和目的 IP 地址之间存在多对一的映射关系时，H(sip|dip)的值会显著增加。

2) H(dport|dip)：目的端口关于目的 IP 地址的条件熵。对于 DDoS 攻击而言，攻击者通常会向目标主机请求尽可能多的服务，目的端口和目的地址之间具有多对一映射关系，而合法用户通常在一段时间内请求的服务较为单一。H(dport|dip)可用来描述目的端口和目的地址间的多对一映射关系。

3) H(sip|dport)：源 IP 地址关于目的端口的条件熵。针对某一特定服务的 DDoS 攻击而言，大量主机会向某固定端口请求服务，在这个过程中，同样有很大概率会造成交换机中流表不匹配，源地址和目的端口间存在多对一的映射关系。H(sip|dport)可用来描述源地址和目的端口之间的多对一关系。

4) H(psize|dip)：数据分组大小关于目的 IP 地

址的条件熵。上述条件熵并不能很好地区分“合法突发流量”和 DDoS 攻击流量之间的区别。对于合法突发流量，对目标发送的数据分组大小往往是无规律的。而 DDoS 攻击流量往往具有固定大小的数据分组，因此相对于正常值，异常状况下该条件熵将下降。

4.4　使用 GHSOM 神经网络对流量进行分析

在使用神经网络进行攻击检测之前，需要利用训练样本对神经网络进行训练，目的是得到一个稳定的网络结构。

训练过程：训练过程就是构造 GHSOM 网络的过程。根据 GHSOM 算法的流程，主要有以下 4 个步骤。1) 初始第 0 层神经元，神经元的权值向量为所有输入模式向量的平均值，计算 0 层平均量化误差；2) 采用自顶向下的方式开始训练，将第 0 层神经元扩展为第 1 层 2×2 结构的 SOM 子网，此 SOM 子网采用传统的 SOM 学习方法；3) 将最新层的获胜神经元的平均量化误差和 0 层平均量化误差进行比较，判断横向扩展条件，若满足条件，在具有最大量化误差的误差神经元和距离它距离最远的神经元之间添加一行或一列，继续进行 SOM 学习，直到不能再进行横向扩展；4) 当 SOM 子网稳定不再进行扩展后，若某个神经元满足纵向扩展条件，则从这个神经元扩展出一个新的 2×2 结构子网。对于新的子网，重复上述训练过程，直到神经元的数量和层次不再增加，视为训练结束。

检测过程：从流量中提取出 DDoS 攻击的条件熵特征，输入神经网络，和之前的训练结果进行对比，判断流量是否为 DDoS 攻击流量。

生成报告：生成检测报告，由管理员决定下一步动作。

5　实验结果及分析

5.1　实验环境

本文选取 POX 控制器作为 SDN 的控制器。POX 作为一种轻量级的 OpenFlow 控制器平台，它的前身是 NOX 控制器，具有良好的可编程性。交换机使用支持 OpenFlow 协议的华为千兆交换机，型号为 S5720-32C-HI-24S-AC。主机配置为 CPU 3.30 GHz，4 GB 内存，Windows 10 操作系统。

网络拓扑如图 3 所示，POX 控制器连接 3 台交换机。由于本文所研究的针对控制器的 DDoS 攻击

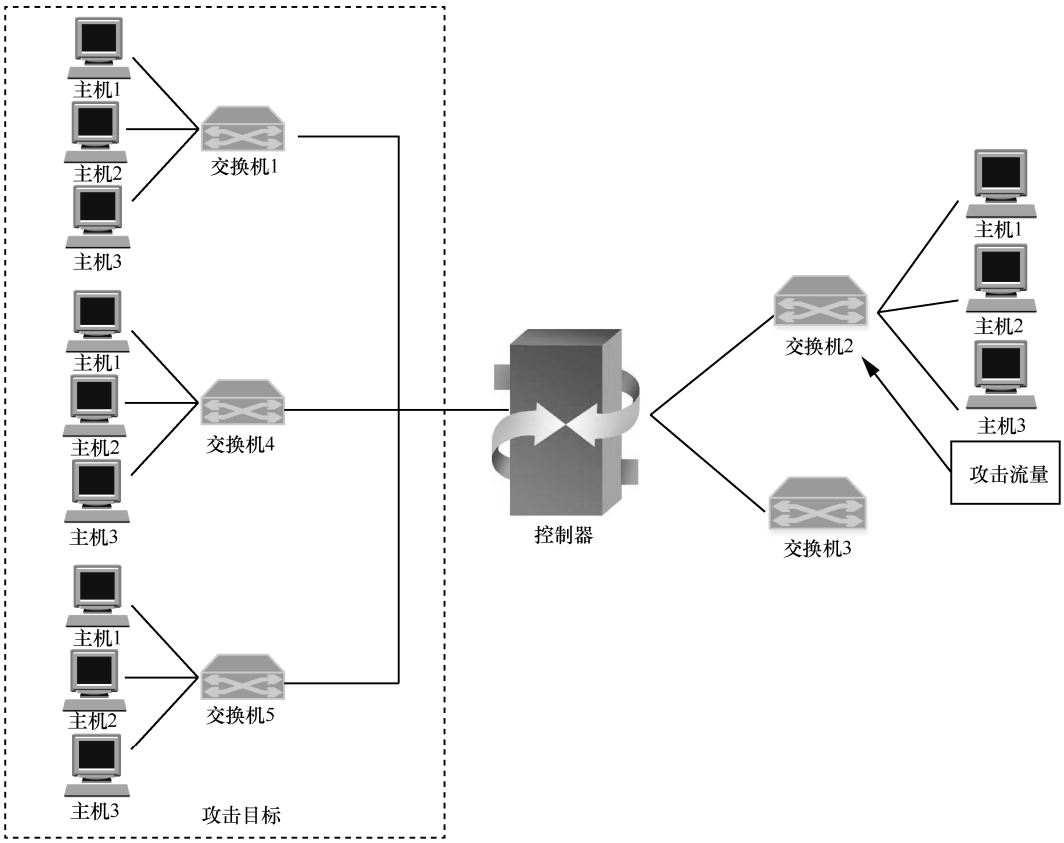


图 3　网络拓扑

仍然需要目的地址作为数据分组的目的依据，因此仍然需要目标主机作为攻击目标。交换机 1、交换机 4 和交换机 5 分别连接 3 台受攻击主机，交换机 2 连接的主机模拟发送攻击流量，交换机 3 作为正常交换机提供正常的数据转发。

5.2 实验数据集

实验使用 TFN2K 攻击软件获得攻击流量，用于训练和检测。为了不失一般性，取林肯实验室的 DARPA 1999 数据集中第 1、3 周中不包含攻击流量的正常流量数据集作为背景流量，并且依据文献[17]中提到的网络流量中 3 种常见网络协议（ICMP、TCP、UDP）的比例，5%为 ICMP 流量，85%为 TCP 流量，10%为 UDP 流量。

实验包括训练和检测 2 个阶段，均使用连续的 1 500 个样本。在训练阶段，确定交换机流请求速率检测阈值以及 GHSOM 神经网络的各个参数。在检测阶段，利用测试样本对 MBCE&G 方法进行测试，并且和基于 GHSOM 神经网络的七元组检测方法、基于 SOM 神经网络的式元组检测方法、基于目的地址的熵检测方法进行了对比。

5.3 实验结果分析

对背景流量和攻击流量进行混合采样用于 MBCE&G 检测方法的训练，在训练过程中将取样时间设置为 100 s，采样周期为 5 s，此处采样周期和 4.2 节的 idle_timeout 对应，T 为 100 s。其中，有 100 s 的连续正常流量，在 30~60 s 之间加入了攻击流量。实验确定交换机流请求速率检测阈值和 GHSOM 神经网络的各个参数，并且对选取的四元组条件熵特征进行了可行性分析。

5.3.1 交换机流请求速率分析

对交换机流请求速率进行分析。在数据采集过程中，对交换机发送的 packet_in 数据帧进行了采样，图 4 显示了 5 台交换机的 packrt_in 数据帧的流量速率随时间的变化曲线。可以看到，在 0~10 s 这段时间内，5 台交换机发送的 packet_in 数据帧都有所上升，此时因为随着未知流的进入，交换机需要向控制器询问流表规则；控制器下发规则之后，10~30 s 时间的 packet_in 速率趋于平缓；在 30 s 之后的一段时间内，交换机 1、交换机 4、交换机 5 的 packet_in 数据帧的速率骤升，但是期间达到的峰值并不相同，并且维持了一段时间，整个过程持续了 30 s。在此过程结束之后，packet_in 数据帧的速率和正常交换机的速率大致相等。

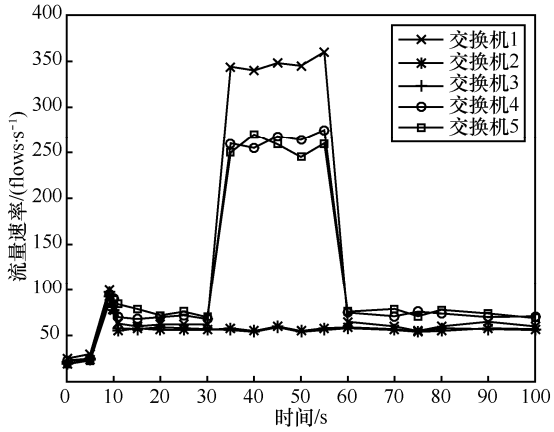


图 4 packet_in 数据帧流量变化曲线

根据在采样过程中统计的 packet_in 数据帧的数量，依据式(3)得到了各交换机的流请求速率，5 台交换机的流请求速率随时间的变化曲线如图 5 所示。

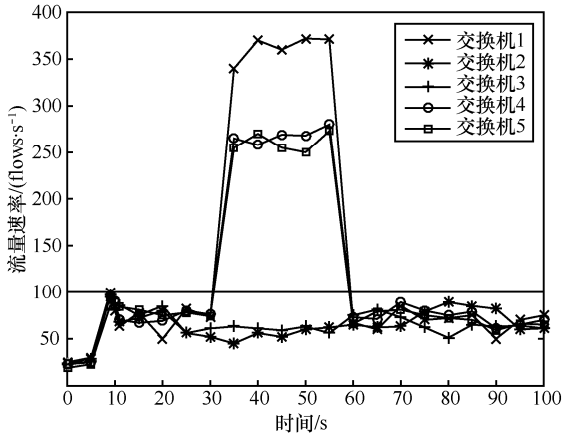


图 5 交换机流请求速率变化曲线

通过图 4 和图 5 的对比可以看出，随着 packet_in 数据帧请求速率的上升，交换机的流请求速率也有相应的提高。在 0~10 s，将交换机流请求速率随时间的上升视为攻击者的攻击侦测阶段，其目的是探寻 SDN 中的潜在攻击目标，将此阶段达到的流请求速率峰值 100 flows/s 作为判断受损交换机的阈值 R_i^{α}；在之后的攻击时间，交换机 1、交换机 4、交换机 5 的流请求速率持续超过阈值 100 flows/s，由于存在合法用户流量突发的情况，此时将这 3 台交换机视为受损交换机，而通过它们的流量则视为可疑攻击流。

5.3.2 四元组条件熵分析

图 6 中的(a)、(b)、(c)、(d)分别代表了通过受损交换机 1、受损交换机 4、受损交换机 5 的流量

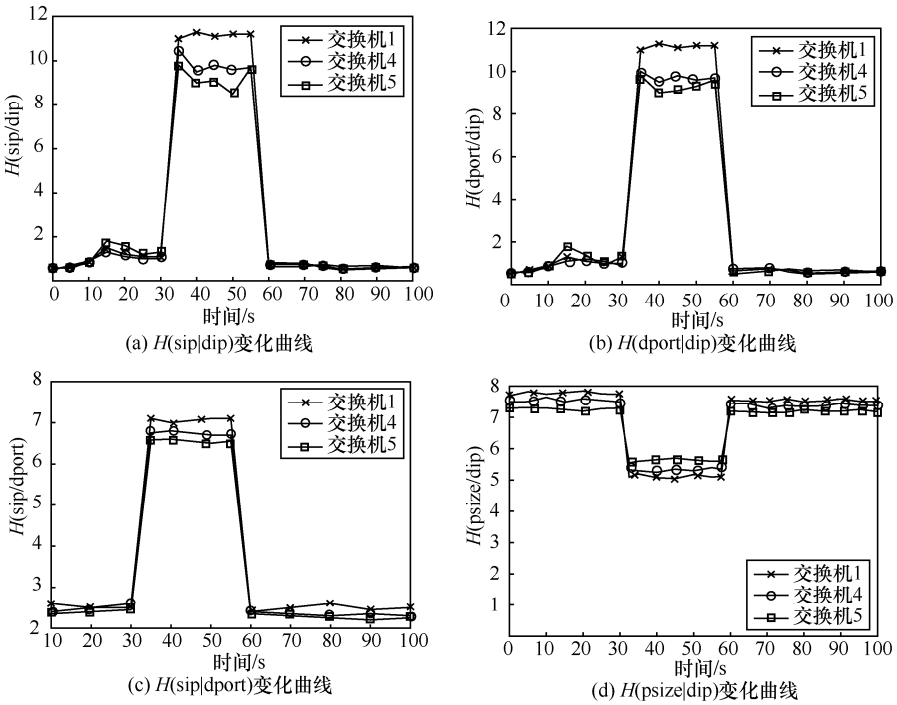


图 6　条件熵变化曲线对比图

的 H(sip|dip)，H(dport|dip)，H(sip|dport)，H(psize|dip)随时间的变化曲线。由于在 DDoS 攻击中前 3 种条件熵都表现出明显的多对一特征，而正常流量间具有多对一、一对多与一对一等多种映射关系，因此在图 6(a)、(b)和(c)中可以看出相应的条件熵都有明显的升高。从图 6(d)可以看出，合法数据分组对目标发送的数据分组的大小往往是无规律的，此时熵值较高，并且趋于平稳。当存在攻击流量时，由于数据分组通常具有固定长度，因此相对于正常值熵值有所下降。上述四元组特征在攻击前后都表现出了明显的差异性，可以用来检测 SDN 是否遭受到了 DDoS 攻击。

5.3.3　检测率分析

经过对实验数据的采集、训练，GHSOM 神经网络参数设置如下：横向扩展参数 $\mu_1 = 0.7$，纵向扩展参数 $\mu_2 = 0.035$，学习速率 $\alpha(t) = \dfrac{0.1}{1+0.001t}$。

在检测过程中，将本文方法分别和 GHSOM 方法、SOM 方法以及熵检测法进行了对比，并使用了相同的采样数据。依照文献[12]提取了七元组特征、依照文献[11]提取了六元组特征、依照文献[9]对目的地址求熵的方法分别进行了实验验证。在实验中，对正常流量和攻击流量分别进行了 3 700 次和 2 300 次训练，1 500 次和 2 000 次检测。

通过实际检测，得到了如下的检测结果，如表 1 所示。实验选取了较有代表性的 neptune 攻击、portsweep 攻击和 ipsweep 攻击，这 3 种 DDoS 攻击都可以对控制器产生较大影响。通过 TFN2K 攻击软件发送不同的攻击流量，将 4 种检测方法的检测率进行了对比。从表 1 可以看出，MBCE&G 检测方法和其他 3 种相比，都拥有较高的检测率。其中，针对 portsweep 的攻击，熵检测法的检测率较低。因为这种攻击是为了侦测网络中的更多主机，目的 IP 地址较为分散，所以熵检测法中基于目的 IP 地址的熵值会比较高，导致检测率较低。同理，熵检测法对 ipsweep 攻击的检测率也不理想。而本文所提出的 MBCE&G 检测方法，在检测之前确定了受损交换机，并且以目的地址和目的端口为基准提取了不同的条件熵。相比文献[12]的七元组特征、文献[11]的六元组特征、文献[9]的特征选取更加明确，更能突出针对控制器的 DDoS 攻击流量的多对一特征。

表 1　　检测率对比

攻击类型	MBCE&G	GHSOM	SOM	熵检测法
neptune	97.6%	96.8%	96.4%	94.3%
portsweep	95.7%	94.6%	81.3%	71.6%
ipsweep	95.5%	93.5%	82.1%	70.7%

5.3.4 检测性能及算法开销分析

为了评估算法的检测性能，针对 neptune、portsweep、ipsweep 攻击，将 MBCE&G 检测方法和 GHSOM 检测方法、SOM 检测方法以及熵检测法在检测时间上进行了比较，实验结果如表 2 所示。

表 2 检测时间对比

攻击类型	MBCE&G/ms	GHSOM/ms	SOM/ms	熵检测法/ms
neptune	517	430	314	212
portsweep	552	453	342	233
ipsweep	564	465	357	246

从检测时间的对比可以看出，MBCE&G 检测方法的检测时间比其他 3 种检测方法的检测时间要长，这主要是因为 MBCE&G 检测方法比其他 3 种检测方法更加复杂。和 GHSOM 检测方法相比，MBCE&G 增加了对受损交换机的判定；GHSOM 神经网络的扩展过程相比 SOM 神经网络具有更多的迭代运算；熵检测法由于只需对熵值和阈值进行对比，检测过程较为简单，所需时间较短。

除此之外，针对神经网络的训练过程，可独立进行线下训练，因此神经网络的训练过程对检测方法的影响可以忽略。

MBCE&G 检测方法在检测过程中的开销主要包括 2 个部分：受损交换机的判断和 GHSOM 神经网络的分类判断。受损交换机的判断阶段的计算复杂度主要由网络中交换机的数量和检测周期 T 内时隙的数量 m 来决定，在本实验中交换机数量为 5 台，时隙数量为 20，而在实际网络部署中此阶段的计算复杂度由网络中交换机的数量决定。GHSOM 神经网络的分类判断的计算复杂度主要由待检测样本的数量和算法本身的匹配时间来决定，在本实验中检测样本数量为 1 500 个连续样本，算法本身的匹配时间受神经元的匹配过程影响，单位小于待检测样本的数量，因此在实际网络部署中此阶段的计算复杂度由待检测的样本数量决定。

因此，MBCE&G 检测方法的计算复杂度为 $O(MN)$，其中，M 为交换机的数量，N 为待检测样本的数量。

综上所述，MBCE&G 的检测时延不足 0.6 s，远小于采样周期 5 s。在实际运行中，若对检测速度有较高的要求，可冗余设置多个检测模块，以流水线的形式并行处理检测任务，以此来弥补单一检测模块不足以应对海量数据的缺陷。综上所述，本文提出的 MBCE&G 检测方法是可行的。

6 结束语

本文介绍了 SDN 中一种针对控制器的 DDoS 攻击，分析了其攻击特征，并且针对这种攻击提出了 MBCE&G 检测算法。该检测算法在进行检测之前精确定位受损交换机，进而确定了可疑攻击流，之后以条件熵的形式对流量进行了特征提取，最后利用了神经网络强大的分类能力完成攻击检测。实验将 MBCE&G 检测算法与经典的熵检测法、基于 SOM 神经网络的六元组检测法和基于 GHSOM 神经网络的七元组检测法进行了比较，实验结果表明，MBCE&G 检测方法可以有效检测 SDN 中针对控制器的 DDoS 攻击。

参考文献：

[1] KREUTZ D, RAMOS F M V, ESTEVES V P, et al. Software-defined networking: a comprehensive survey[J]. Proceedings of the IEEE, 2014, 103(1):10-13.

[2] SEZER S, SCOTT H S, CHOUHAN P K, et al. Are we ready for SDN? implementation challenges for software-defined networks[J]. IEEE Communications Magazine, 2013, 51(7):36-43.

[3] SHIN S, GU G. Attacking software-defined networks: a first feasibility study[C]// ACM SIGCOMM Workshop on Hot Topics in Software Defined NETWORKING., 2013:165-166.

[4] NEELAM D, SHASHANK S. Analyzing behavior of DDoS attacks to identify DDoS detection features in SDN[C]//IEEE International Conference on Communication System and Networks (COMSNETS), 2017.

[5] CHEN K Y, JUNUTHULA A R, SIDDHRAU I K ,et al. SDNShiled: towards more comprehensive defense against DDoS attacks on SDN control plane[C]//IEEE Conference on Communications and Networks Security (CNS). 2016.

[6] KLOTI R, KOTRONIS V, SMITH P. OpenFlow: a security analysis[C]//IEEE International Conference on Network Protocols. 2013: 1-6.

[7] BENTON K, CAMP L J, SMALL C. OpenFlow vulnerability assessment[C]// ACM SIGCOMM Workshop on Hot Topics in Software Defined NETWORKING. 2013:151-152.

[8] DAYAL N, MAITY P, SRIVASTAVA S, et al. Research trends in security and DDoS in SDN[J]. Security & Communication Networks, 2016, 9.

[9] MOUSAVI S M, STHILAIRE M. Early detection of DDoS attacks against SDN controllers[C]// International Conference on Computing, NETWORKING and Communications. 2015:77-81.

[10] DONG P, DU X, ZHANG H, et al. A detection method for a novel

DDoS attack against SDN controllers by vast new low-traffic flows[C]//IEEE International Conference on Communications. 2016:1-6.

[11] BRAGA R, MOTA E, PASSITO A. Lightweight DDoS flooding attack detection using NOX/OpenFlow[C]// Conference on Local Computer Networks. 2010:408-415.

[12] 姚琳元, 董平, 张宏科. 基于对象特征的软件定义网络分布式拒绝服务攻击检测方法[J]. 电子与信息学报, 2017, 39(2): 381-388.
YAO L Y, DONG P, ZHANG H K. Distributed denial of service attack detection based on object character in software defined network[J]. Journal of Electronica & Information Technology, 2017, 39(2): 381-388.

[13] 杨雅辉, 姜电波, 沈晴霓, 等. 基于改进的 GHSOM 的入侵检测研究[J]. 通信学报, 2011, 32(1): 121-126.
YANG Y H, JIANG D B, SHEN Q N, et al. Research on intrusion detection based on an improved GHSOM[J]. Journal on Communications, 2011, 32(1): 121-126.

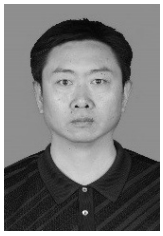
[14] 阳时来, 杨雅辉, 沈晴霓, 等. 一种基于半监督 GHSOM 的入侵检测方法[J]. 计算机研究与发展, 2013, 50(11): 2375-2382.
YANG S L, YANG Y H, SHEN Q N, et al. A method of intrusion detection based on semi-supervised GHSOM[J]. Journal of Computer Research and Development, 2013, 50(11): 2375-2382.

[15] SHANNON C E. A mathematical theory of communication[J]. ACM Sigmobile Mobile Computing & Communications Review, 1948, 27(4): 379-423.

[16] MA D, XU Z, LIN D. Defending blind DDoS attack on SDN based on moving target defense[C]//International Conference on Security and Privacy in Communication Systems. 2014: 463-480.

[17] BORGNAT P, DEWAELE G, FUKUDA K, et al. Seven years and one day: sketching the evolution of internet traffic[C]// INFOCOM. 2009: 711-719.

[作者简介]



田俊峰（1965−），男，河北保定人，河北大学教授、博士生导师，主要研究方向为信息安全与分布式计算。

齐鎏岭（1992−），男，河北保定人，河北大学硕士生，主要研究方向为信息安全与分布式计算。