文章编号：1008-1542(2019)05-0404-10

基于 Web 访问路径的应用层 DDoS 攻击防御检测模型

任　皓[1]，许向阳[1]，马金龙[1]，张志浩[2]

(1.河北科技大学信息科学与工程学院，河北石家庄　050018；2.北华航天工业学院电子与控制工程学院，河北廊坊　065000)

摘　要：为了提高防御应用层分布式拒绝服务攻击的有效性、时效性和准确性，对应用层 DDoS 攻击的演化、模式，以及攻击者的攻击路径和攻击行为进行深入研究。提出一种基于 Web 访问路径的防御检测模型，根据访问路径轨迹、攻击行为特点和网站链接规则，建立请求路径、请求分布、路径循环、行为时隙和路径长度 5 种异常检测模型。通过计算合法用户访问网站时的正常值以及具有攻击行为用户的实时异常值偏离程度，可判定是否遭到应用层 DDoS 攻击。防御模块依据用户非法值大小选取最佳防御策略，抵御应用层 DDoS 攻击，实现网站数据安全与计算机安全。实验采用真实日志数据进行训练，向实验网站发动 5 种不同类型的应用层 DDoS 攻击。结果表明，防御检测模型能在短时间内准确辨别具有攻击行为的用户，并联合防御模块抵抗针对 Web 服务器的 DDoS 攻击，能够实现实时检测、实时防御，有效降低误报率。所提出的检测模型可以对路径长度进行监控，提升了异常判定的准确性和可靠性，有效提高了 Web 网站防御 DDoS 攻击的能力。

关键词：数据安全与计算机安全；应用层；分布式拒绝服务；访问路径；异常检测；攻击行为

中图分类号：TP393.08　　文献标志码：A　　**doi**：10.7535/hbkd.2019yx05006

Application-layer DDoS attack defense detection model based on Web access path

REN Hao[1], XU Xiangyang[1], MA Jinlong[1], ZHANG Zhihao[2]

(1. School of Information Science and Engineering, Hebei University of Science and Technology, Shijiazhuang, Hebei 050018, China; 2. School of Electrical and Control Engineering, North China Institute of Aerospace Engineering, Langfang, Hebei 065000, China)

Abstract: In order to improve the effectiveness and timeliness of defense against distributed denial of service (DDoS) attacks in application layer, the evolution and mode of application-layer DDoS attacks, as well as the attack path and behavior of attackers

收稿日期：2019-04-01；修回日期：2019-08-29；责任编辑：陈书欣

基金项目：教育部人文社科基金(19YJAZH069)

第一作者简介：任　皓(1993—)，男，河北石家庄人，硕士研究生，主要从事网络安全、复杂网络方面的研究。

通信作者：许向阳副教授。E-mail：Xxy@hebust.edu.cn

任皓，许向阳，马金龙，等.基于 Web 访问路径的应用层 DDoS 攻击防御检测模型[J].河北科技大学学报，2019，40(5)：404-413.

REN Hao，XU Xiangyang，MA Jinlong，et al.Application-layer DDoS attack defense detection model based on Web access path[J].Journal of Hebei University of Science and Technology，2019，40(5)：404-413.

are explored in depth. A defense detection model based on Web access path is proposed, according to access path trajectory, attack behavior characteristics and website link rules, five anomaly detection models including request path, request distribution, path loop, behavior slot and path length are established. By calculating the normal value of legitimate users accessing websites and the deviation degree of real-time outliers of users with aggressive behavior, it can determine whether it is attacked by application-layer DDoS. In order to improve the accuracy of detection, the defense module chooses the best defense strategy according to the size of user's illegal value, resists application-layer DDoS attacks, and achieves website data security and computer security. The experiment is trained with real log data, launch five different types of application-layer DDoS attacks on experimental website, the result show that the defense detection model can accurately identify users with aggressive behavior in a short time, it combines with defense module to defend the application-layer DDoS to a specific website, realizes real-time detection and real-time defense, and the false alarm rate is significantly reduced.

Keywords: data security and computer security; application layer; distributed denial of service; access path; anomaly detection; attack behaviors

在当代网络中,分布式拒绝服务(distributed denial of service,DDoS)攻击依然能够产生强大的威胁并给许多国家带来严重影响,它利用目标系统网络服务功能缺陷,以为合法用户提供服务作为枪靶,直接消耗受害者的服务器资源,最终导致服务器瘫痪[1-3]。

应用层面的 DDoS 攻击向受害服务器发送大量合法数据包,但入站流量不足以使受害服务器的带宽饱和,且出站流量亦可实现[4],由于大量用户同时访问站点是一种合法行为,因而应用层 DDoS 攻击与突发访问相似,辨别它们有一定的难度。随着 Web 应用重要性的不断提高,其安全风险与日俱增,目标主机或服务器遭受应用层 DDoS 攻击往往会产生一定的网络攻击行为,即存在防御检测的切入点[5-6]。为了检测应用层 DDoS 攻击,国内外专家提出了一系列防御检测策略。李锦玲等[7]将应用层检测映射到网络流量层面,使用卡尔曼滤波算法检测单位时间内每个用户的字节数,当检测出异常时,该源端口熵值超出预设值,则判定为 DDoS 攻击。ZHAO 等[8]为了识别 DDoS 攻击,构造联合熵向量的映射矩阵,对熵向量坐标判别图中边界和熵向量所处区域进行判别,能够有效区分应用层 DDoS 攻击的类别。王风宇等[9]对主干网中用户访问行为剖析,用户产生的外联行为特点为区分突发访问和攻击行为提供了有效辨别依据,但场景条件苛刻,检测难度大。康松林等[10]利用通信中套接字的参数进行检测,分别计算出源地址、源端口号、目的端口号和标识符字段(PRS)各自变量的熵值,在统计时隙内,统计各个变量的熵值并进行分析,能够得出有效区分和判定网络协议层面 DDoS 与应用层面 DDoS 的方法。

关于 Web DDoS 检测,孙未等[11]提出为每个 Web 访问用户设定忠实度,根据用户行为忠实度和访问频次忠实度进行联合判定,将忠实度低于阈值的用户过滤,从而实现应用层 DDoS 防御。KEDJAR 等[12]提出了一种混合模型,解决了 Web 服务安全访问控制问题,然而,该模型缺乏灵活性,某些固定的阈值必须由管理员进行配置。KANDULA 等[13]提出攻击由程序生成,不具有人类智能型,当网络资源消耗超出预设阈值时,产生简单问题并进行验证连接是否为攻击源,该方法能够有效区分攻击源与正常用户,且会损失自身资源,并对正常用户体验和访问造成影响。肖军等[14]分析 DDoS 与正常访问行为的不同,提出行为异常度属性和 Session 异常度模型,依据平均异常度的值来判断是否为合法请求,该方法忽视网站链接规则和 Session 访问行为的影响。刘泽宇等[15]分析 Web 用户访问行为,提出 4 种方式检测应用层 DDoS 攻击异常属性,若用户异常值超出异常值均值或方差值时,则判定为遭受应用层 DDoS 攻击,但未对用户会话长度进行检测,容易导致 DDoS 绕过该检测模型进行攻击。

针对上述不足,本文主要针对网站结构链接性特点、行为概率分布特点和 Web 用户访问路径特点,提出一种基于 Web 访问路径的 DDoS 攻击防御模型(UAP)。该模型部署在 Web 服务器与应用服务器的中间地带,路径节点数据依次传入检测模块内,对请求路径、请求分布、路径循环、行为时隙和路径长度执行异常检测,检测结果能够准确辨别 DDoS 攻击和正常用户突发流访问。防御模块根据检测结果做出有效防御,依据用户非法值选取最佳防御策略,避免小概率事件的发生。该防御模型不受客户端环境差异性影响,不依赖主干网络,采用自身数据作为检测基础,有效减少了数据收集时间。对网站链接规则和节点信息分布以及路径特性都进行充分的考虑,有效提高了 Web 网站 DDoS 攻击的防御能力。

1 模型结构

UAP模型是由访问路径检测模块和防火墙过滤转发组成的防御检测模型，如图1所示。用户访问请求经过防火墙过滤后传送至Web服务器，用于异常用户过滤的黑名单过滤机制架设在应用层，防火墙策略部署在传输层。异常检测与应用服务器接受请求并行执行，路径数据实时传送至检测模型进行异常判定，判定结果发送到防火墙。若判定结果为用户异常，根据路径数据中的客户端IP，SessionId和会话信息中的浏览器标识符等特征，将异常用户加入到黑名单中，并为其累加非法值，根据非法值大小部署最佳防御策略。在预设时间范围内，防火墙对黑名单用户发起的请求进行滤除。当出现误判时，可通过输入正确验证信息等安全措施，从黑名单内移除，并清除上次非法值的累加。

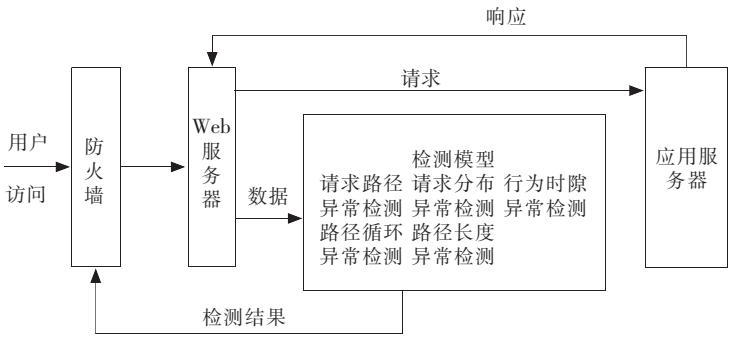


图1 防御模型结构示意图

Fig.1 Structure diagrammatic sketch of defense model

该模型的核心部分为异常检测模型，对用户的一次访问行为进行异常检测，并对用户路径生命周期内的同种异常检测值作全局判定，只要有一种异常检测超出正常范围则可判定为异常用户。异常判定具有全面性和准确性，可做到访问与检测并行执行，实现实时检测、实时防御部署。同时，可以有效降低防御模块的误报率。

1.1 用户路径

网站是指在因特网上根据一定规则，用于展示特定内容的相关网页的集合[16]。每个网站中包含许多网页，页面由若干个嵌入对象组成，若将页面和嵌入对象看作是网站的内嵌对象，一个网站就由无数个内嵌对象编制成一张大型网络[17]。将每个页面和嵌入对象看作网络的节点，网络中所有节点的集合记为 p；相邻页面间的链接则作为连接节点的边，记为 l；若边的形成不包含网络图中的任意边，则为非法边，反之为合法边，记为 l_r。节点和边构成了一张网络图，由网络图中一系列边和节点组成的用户路径，记为 L。

1.2 基本参数

每个节点 p_i 都拥有6个属性值(Id，SessionId，ObjectId，Time，ReferId，Count)，Id表示该节点的唯一标识符，采用自增的整数设定；SessionId表示用户访问网站建立一个会话的唯一标识符；ObjectId表示网站内部所包含页面节点和所有内嵌对象节点的唯一标识符；Time表示用户访问节点的时刻；ReferId表示用户访问节点的来源节点的标识符；Count表示访问用户一次会话内访问Web服务的次数[15]。每个属性值都能反映路径信息。

空间权值 S 是指网络图中任意两节点之间最短路径。对于任意边 $l=(p_j,p_k)$，设 $S(l)=s_{jk}$，则 s_{jk} 为 l 上的空间权值。例如节点元素 a,b,c，即 $a,b,c\in p$；a 可以访问相邻节点元素 b 和 c，则边为 $l_1=(a,b)$和 $l_2=(a,c)$，它们的空间权值 $S(l_1)=S(l_2)=1$。

时间权值 t 是指用户离开节点前所驻留的时间。对于任意节点 p_j,p_k，设 $t(p_j,p_k)=t_{jk}$，则 t_{jk} 为节点 p_j 上的时间权值。

请求次数 G 是指网络模型中任意节点被合法请求次数。对于任意节点 p_i，设 $G(p_i)=g_i$，则 g_i 为节点 p_i 上的请求次数。

1.3 路径分析

用户路径在网络图中具有以下5种特征。

1）依据节点间链接原则，路径中任意相邻两节点间的连接，即 $l_r=l=(p_j,p_k)$，$S(l)=1$。当用户破坏链接规则时，当前节点的下一跳节点[18]不属于该点的邻接节点，即 $l_r\neq l=(p_j,p_k)$，$S(d)>1$。

2）对训练集中节点请求次数进行统计，得出节点请求频次与节点次数位序之间满足 Zipf 分布[19]，如图 2 所示。Zipf 定律反映的是规模与位序之间的一种幂律关系，大约 20％的节点请求总数占全部节点请求总数的 80％，由此可看出访问发生在少数节点上。因页面之间的兴趣关联性，用户路径上节点兴趣相似[20]，故用户路径 $G(p_i)$ 的轨迹应趋于稳定。若攻击发生，恶意程序无法按照兴趣关联性智能访问网站节点，与真实用户路径相差较大，导致用户路径 $G(p_i)$ 的轨迹有较大波动。

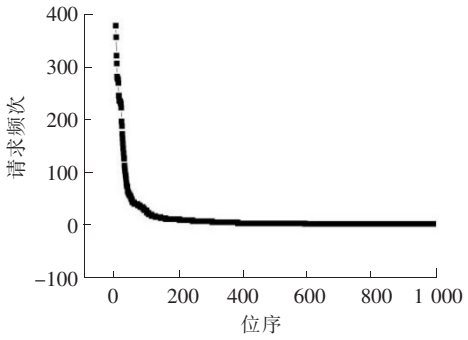


图 2　节点请求频次与节点位序的关系

Fig.2　Relationship between frequency of node request and node order

3）用户对新页面有新鲜感，故用户不会反复循环请求一个或多个页面。当用户路径在网络图中出现路径闭环时，说明存在攻击行为。

4）用户在节点的驻留时间 t_i 表明用户对节点内容的思考和关注度，节点 p_i 与驻留时间 t_i 是相关的。攻击者为了能够提高攻击力度，往往会锐减节点驻留时间，由此造成节点的时间权值出现较小值，导致出现行为时隙异常。

5）路径分析对路径信息的有效性需求较高，为了避免攻击者使用相对攻击模式绕过路径分析，需要对路径的有效长度加以检测，避免检测方法受到挟制。

1.4　数据预处理与采集

Web 日志是指客户端与 Web 服务器交互信息过程中产生的日志，客户端提交服务请求过程中，可能会途经代理服务器和防火墙，最后到 Web 应用服务器[21]。为了更高效地利用训练集的日志数据，需要对日志中的错误数据和重复记录进行处理，例如非成功请求（主要是指出现 403，404 等状态码的请求）；将页面加载内嵌对象请求也看作一个单独请求的网页页面请求，对于用户的请求是没有意义的，应予以删除。

将处理后的日志数据进行采集识别处理，对相同 IP 的用户数据进行归类，按时间大小排序。根据设定的 Session 失效时间为每个用户计算响应的 SessionId，一次会话过程中的 SessionId 相同，页面驻留时间 Time 根据一次会话中前后 ObjectId 请求时间差值来获取。

1.5　基于 Web 用户的路径异常检测

UAP 模型的核心检测模块为访问路径异常检测，检测模块分为请求路径异常、请求分布异常、路径循环异常、行为时隙异常和路径长度异常 5 个方面。

1.5.1　请求路径异常

在任意用户的访问路径中，若存在大量的子路径 $d=(p_j,p_k)$，$S(d)>1$，则访问路径出现请求路径异常，记为 f_{rpa}。

在给定的网站中，每个节点所拥有的邻接节点集合 p_{adj} 是依据网站页面间链接原则选定的，即 p_i 的邻接节点集合为 p_{iadj}。边的有向性由节点指向其邻接节点，如图 3 所示节点链接关系，节点 $a,b,c,e,h\in p$，节点 a 的邻接节点为 b,c，节点 e 为 b 的邻接节点。当 a 下一跳节点为 b 时，构成一条子路径，$d=(a,b)$，$S(d)=1$，属于正常访问。在节点 a 访问节点 e 的路径中，未出现中间节点 b，则可能出现非法访问，该路径中节点 a 请求路径异常。用户每次访问下一跳节点时，遍历当前节点的邻接节点集合，若出现异常，请求路径异常点集合：

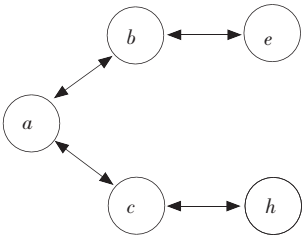


图 3　相邻节点示意图

Fig.3　Schematic diagram of adjacent nodes

$$P_{abn}=\{p_i \mid S(p_i,p_{i+1})>1,\ p_i\in p,p_{i+1}\notin p_{iadj}\},\tag{1}$$

由此得出用户第 i 次访问的请求路径异常值：

$$f_{rpa}=\frac{P_{abn}}{|L|}。\tag{2}$$

1.5.2 请求分布异常

因攻击程序节点的选取不具备智能化，若任意用户在访问路径中的路径行为概率均值波动较大，呈现出异常趋势，则访问路径出现请求分布异常，记为 f_{red}。

网站内每一个节点的 g_i 与位序呈现 Zipf 分布，按照节点的请求次数从大到小进行排序，将节点按照 g_i 大小划分为 M 类，相同类的节点 g_i 是相等的，其分类结果为

$$A=\{a_1,a_2,\cdots,a_M\},\tag{3}$$

$$\mathrm{NUM}=\{n_1,n_2,\cdots,n_M\},\tag{4}$$

式中：A 表示划分为 M 类；n_i 表示在 a_i 类中所包含 p_i 的个数。由于符合 Zipf 分布，依据 Zipf 二八原则可以分为两大类，当满足：

$$\sum_{i=1}^{n}a_in_i=0.8\times\sum_{i=1}^{M}a_in_i,\tag{5}$$

可划分为高频 E 类和低频 F 类：

$$E=\{a_1,a_2,\cdots,a_n\},\tag{6}$$

$$F=\{a_n,a_{n+1},\cdots,a_M\}。\tag{7}$$

用户点击第 m 个类的节点行为概率为

$$OP=a\times a_mn_m/\sum_{m=1}^{M}a_mn_m,\tag{8}$$

式中 a 参数的引入将高频类和低频类之间的界限划分得更明确。由此得出用户第 i 次访问的请求分布异常值：

$$f_{red}=\sum_{n=1}^{i}OP_n/|L|。\tag{9}$$

1.5.3 路径循环异常

在任意用户的访问路径中，如果访问路径中的子路径片段重复出现，并呈现路径闭环形态，则访问路径出现路径循环异常，记为 f_{pci}。

定义访问路径中子路径循环的次数为 c，子路径的长度 $|d|$，第 i 次访问的节点所处子路径的位置为 d'，例如 Session 的访问路径为 b,a,w,e,r,c,a,w,e,r,a，由此可知 $d=\{a,w,e,r\}$，$|d|=4$，$|D|=10$，循环次数为 2，路径末尾节点的位置标识 $d'=1$。对于用户第 i 次访问的路径循环异常值：

$$f_{pci}=\frac{|d|c+d'}{|L|}。\tag{10}$$

1.5.4 行为时隙异常

在任意用户的访问路径中，如果用户访问时隙与访问节点的时间权值集合的中值差异较大，则访问路径出现行为时隙异常，记为 f_{bet}。

每个节点对于用户的吸引程度不同，节点驻留时间也会有所差异，因节点提供的内容不变，故节点驻留时间的差异性较小。对用户行为时隙进行监控，用户访问节点的实际驻留时间为该节点时间权值 t_i，即节点时间权值集合 T_i。用户往往会按照兴趣或内容相关度寻找下一个节点，驻留时间不会太短[21]，将节点时间权值集合按照数值大小依次排序，以 t_{im} 表示节点 T_i 的中位数。用户一次访问路径中的行为时隙异常节点集合为

$$T_{abn}=\{p_i\mid t_i-t_{im}<0,\ p_i\in p,t_{im}\in T_i\}。\tag{11}$$

访问路径中，用户第 i 次访问的行为时隙异常值：

$$f_{bet}=\frac{|T_{abn}|}{|L|}。\tag{12}$$

1.5.5 路径长度异常

为了避免攻击者在路径较短时丢弃当前会话，会对以上 4 种异常检测算法产生抑制效果，提出路径长度异常，记为 f_{ple}。

采用训练集对全局会话长度进行统计,计算其均值 $\hat{L}_n$ 和方差 $\hat{\sigma}_n$。但现实中均值与方差都会随时间的变化而变化[22],文中使用切比雪夫不等式构造自适应偏移变量。

首先设置两者初始值:

$$\hat{L}_n=\frac{1}{N}\sum_{n=1}^{N}L_n, \tag{13}$$

$$\hat{\sigma}_n=\sqrt{\sum_{n=1}^{N}(L_n-\hat{L}_n)^2/(N-1)}。 \tag{14}$$

当未检测出异常时,均值 $\hat{L}_n$ 和方差 $\hat{\sigma}_n$ 的计算公式为

$$\hat{L}_n=(1-\omega)\times\hat{L}_{n-1}+\omega\times\hat{L}_n, \tag{15}$$

$$\hat{\sigma}_n=\sqrt{\sum_{n=1}^{N}(L_n-L_{n-1})^2/(N-1)}, \tag{16}$$

式中,ω 为平滑因子,是对路径长度均值的估计值与真值的进一步预测,本文中 ω 取 0.12。当检测到用户非法时,$\hat{L}_n$ 和 $\hat{\sigma}_n$ 不变,即 $\hat{L}_n=\hat{L}_{n-1}$,$\hat{\sigma}_n=\hat{\sigma}_{n-1}$。

利用切比雪夫不等式计算:

$$Pf(|X-\hat{L}_n|\geqslant k\hat{\sigma}_n)\leqslant\frac{1}{k^2}, \tag{17}$$

式中:$k\hat{\sigma}_n$ 代表路径长度异常的阈值;$Pf(|X-\hat{L}_n|\geqslant k\hat{\sigma}_n)$表示小于路径长度阈值的比例;$k$ 为固定值。

路径长度异常值为

$$f_{ple}=\begin{cases}1, & Pf(|L_{n+1}-\hat{L}_n|\geqslant k\hat{\sigma}_n)-\dfrac{1}{k^2}>0,\\ 0, & Pf(|L_{n+1}-\hat{L}_n|\geqslant k\hat{\sigma}_n)-\dfrac{1}{k^2}\leqslant 0。\end{cases} \tag{18}$$

1.6　路径异常度

用户路径的异常包含请求路径异常 f_{rpa},请求分布异常 f_{red},路径循环异常 f_{pci},行为时隙异常 f_{bet}和路径长度异常 f_{ple}。基于上述 5 种异常检测,在检测过程中,为前 4 个异常值各分配一个权重,对应的权重分别为 ε,λ,φ,γ,满足:

$$\varepsilon+\lambda+\varphi+\gamma=1。 \tag{19}$$

网站结构和内容影响异常值的分布,制约着各异常权重的大小,可根据各异常分布占比大小确定各异常权重值,异常检测中各异常值占比越大,异常情况越明显,赋予其对应比例的权重值。例如,就此实验网站而言,对日志数据集进行异常检测,计算各异常值大小,并统计超出正常范围的异常值数量,以及其中各异常所占的比例,若请求路径异常占比为 60%,请求分布异常、路径循环异常、行为时隙异常占比均为 10%,则令 $\varepsilon=0.6$,$\lambda=\varphi=\gamma=0.1$。在异常检测中,因路径长度易被挟制,容易绕过异常检测对网站进行侵害,故对路径长度异常检测判定赋予“一票否决”属性。当检测到用户路径长度出现异常时,要求该用户再次发送请求进行信息验证,验证失败则限制访问。

根据以上异常检测,可以计算得出用户路径中第 i 次访问异常度为

$$F_{fin}=\varepsilon f_{rpa}+\lambda f_{red}+\varphi f_{pci}+\gamma f_{bet}。 \tag{20}$$

当用户被判定为异常用户时,异常度则累加到用户的非法值里,依据用户非法值大小实施相应措施,防御应用层 DDoS 攻击。

2　实验仿真

2.1　模型搭建

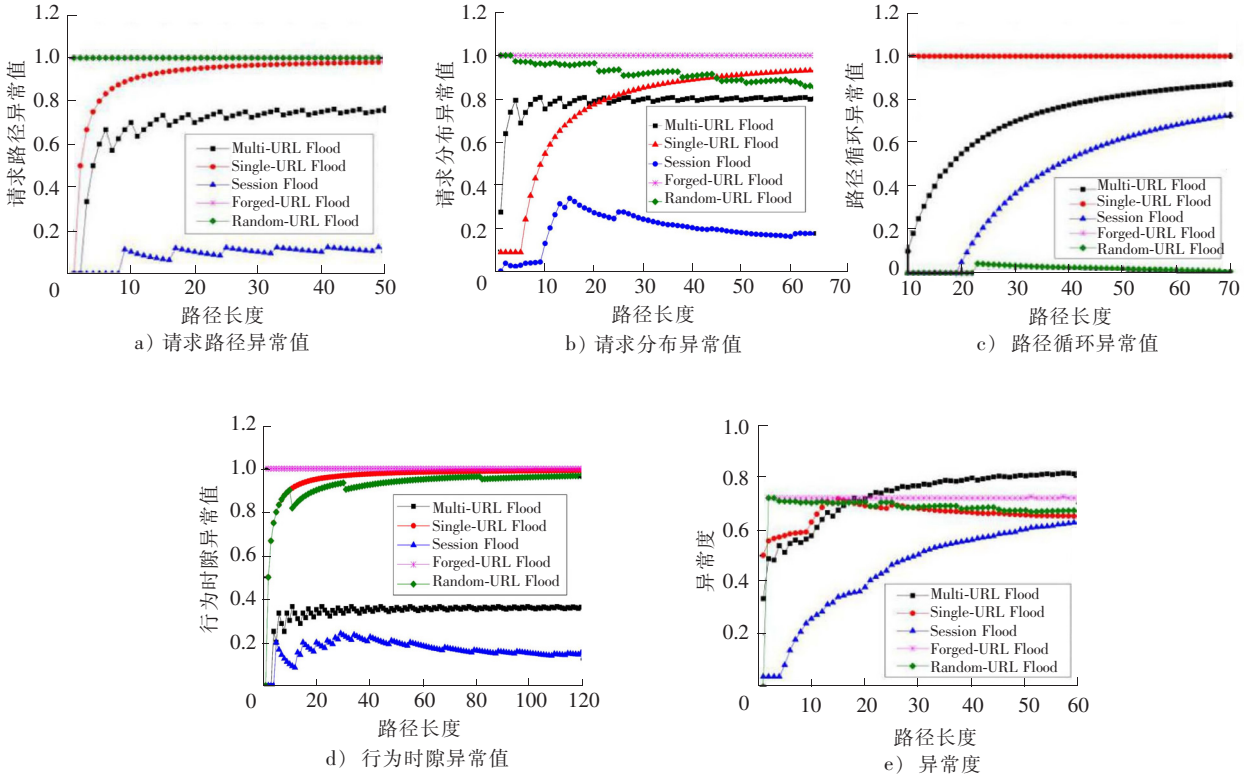
UAP 模型运行在 Ubuntu 16.04 系统下,采用 Iptables 作为包过滤防火墙,Web 服务器使用 NGINX 实现,Tomcat 当作应用服务器。使用 Forged-URL Flood,Random-URL Flood,Single-URL Flood,Multi-URL Flood 和 Session Flood 5 类应用层 DDoS 攻击访问 Web 服务器,分别记为 A,B,C,D,E 5 类攻击。对 Web 服务器产生的路径数据进行异常检测,当访问路径任意异常值超出对应均值一个方差范围时,则判定用户为异常用户。检测结果交由防火墙作防御规则调整,当异常用户访问时,限制其访问。

模拟实验采用 blog-http 真实数据从 2013-09-18 T 06:49:18 到 2013-09-18 T 22:14:51 作为训练集，进行异常检测模型训练。采用 2013-09-18 T 22:16:47 到 2013-09-19 T 06:26:36 的真实日志数据作为验证数据集检测模型的有效性。依据数据集的来源和请求页面的映射关系建设模拟网站，根据网站结构为每个异常值分配权重 $\varepsilon=0.08$，$\lambda=0.42$，$\varphi=0.28$ 和 $\gamma=0.22$，通过训练集计算异常检测的均值和方差，即 $\hat{f}_{rpa}$，$\hat{f}_{red}$，$\hat{f}_{pci}$，$\hat{f}_{bet}$，$\hat{L}_n$，$\hat{\sigma}(f_{rpa})$，$\hat{\sigma}(f_{red})$，$\hat{\sigma}(f_{pci})$，$\hat{\sigma}(f_{bet})$，$\hat{\sigma}_n$，计算结果如式(21)所示。前 4 种检测异常值超出均值一个方差时，则判定为用户异常，路径长度异常的用户需要通过验证解除用户的异常判定。

$$\begin{cases}\hat{f}_{rpa}=0.034\ 17, & \hat{\sigma}(f_{rpa})=0.051\ 87,\\ \hat{f}_{pci}=0.054\ 29, & \hat{\sigma}(f_{pci})=0.098\ 73,\\ \hat{f}_{bet}=0.145\ 35, & \hat{\sigma}(f_{bet})=0.091\ 96,\\ \hat{f}_{red}=0.190\ 76, & \hat{\sigma}(f_{red})=0.127\ 40,\\ \hat{L}_n=13.576\ 53, & \hat{\sigma}_n=14.180\ 09。\end{cases} \tag{21}$$

2.2 实验结果

5 种应用层 DDoS 攻击的异常值和异常度随路径变化的关系如图 4 所示。



a) 请求路径异常值　b) 请求分布异常值　c) 路径循环异常值

d) 行为时隙异常值　e) 异常度

图 4　5 种攻击异常值和异常度

Fig.4　Outlier and abnormal values from five attacks

如图 4 a)可知，由于 A 类和 B 类两种攻击在路径节点的选择上具有很强的随机性，而 A 类攻击使用伪造的超长 URL 进行攻击，使请求路径异常值稳定在最大值，即 $f_{rpa}=1$。C 类攻击对单个节点持续访问，异常值逐渐增大，随着时间推移异常值趋于最大值。D 类攻击中攻击路径中节点的连接出现非法衔接时，轨迹出现抖动，抖动最大值逐渐增大，直到轨迹稳定并超出正常范围。E 类攻击始终处于较小值，由于路径选取历史记录里真实用户的访问片段，导致截取路径的重复使用出现轨迹抖动。

图 4 b) 中 A 类攻击请求分布异常值 $f_{red}=1$。B 类攻击的节点选取随机性高，高频节点所占比例低选取性低，异常值始终较大。C 类攻击对单一节点持续访问次数超过正常用户单一节点访问次数最大值时，异常值持续增加，逐渐趋于最大值。D 类攻击异常值轨迹出现小幅抖动，最终稳定在较大值。E 类攻击由于真实用户路径中个别节点的频次较低，轨迹出现较大峰值但仍小于阈值，随后持续走低。

由图 4 c)可知，D 类和 E 类攻击都是对路径的重复使用，当路径第 1 次形成一个闭环时，异常值逐渐持

续增大,趋于较大值。A 类攻击和 B 类攻击的路径循环异常值为 0,C 类攻击的路径循环异常值 $f_{pci}=1$。

如图 4 d)所示,A 类、B 类和 C 类攻击行为时隙较小,异常值上升较快,趋于最大值。D 类攻击由于路径和攻击时隙固定,攻击反复执行,其中单一节点时隙异常判断值较小,轨迹出现上下波动,异常值稳定在阈值以上。E 类攻击由于截取正常用户访问片段,访问时隙处于正常时隙,异常值处于较低值。

如图 4 e)所示,根据所分配的权重,对用户的 4 种异常检测值做综合性异常判定,A 类、B 类、C 类和 D 类异常判定所需的路径长度较短,E 类攻击需要较长路径才能做出准确判定。将用户的综合性判定值作为防御策略部署的依据。

攻击时隙选取为正常间隔时间,可以得到如下结论。

1)A 类攻击针对的是网络图中的非法节点,主要表现为请求路径、请求分布异常。

2)B 类攻击忽视网站链接规则和兴趣的关联性,随机访问节点,故请求路径异常和请求分布异常尤为突出。

3)C 类攻击对单一节点持续请求,偏离正常用户浏览路径,出现零长度路径,主要表现为请求分布异常和路径循环异常。

4)D 类攻击对多个固定节点循环请求,若节点选择符合链接规则和兴趣匹配,则表现为循环路径异常;不符合时,则表现为请求路径异常和请求分布异常。

5)E 类攻击使用真实 Session 路径的截取片段,截取片段较短时,表现为请求路径异常和路径循环异常,片段较长时,表现为路径循环异常。

6)若攻击时隙较短,则均表现为行为时隙异常。

5 种应用层 DDoS 攻击访问路径检测性能如表 1 所示。A 类、B 类和 C 类攻击的检测判定时间很快,异常行为因素突出,随着路径的增长异常值增长最快。D 类攻击时隙较小时,检测效率高,时隙增大时,异常行为种类减少,也能得到较快判定。E 类攻击片段为正常用户历史截取片段,检测较有难度,但是对于 UAP 检测模型来说,对 E 类攻击的检测效率相对较高。

表 1　不同攻击时隙的 5 种攻击性能检测

Tab.1　Performance test for the five attacks at different time slot

模拟攻击类型	不同攻击时隙的告警时间/s			
	0.01	0.1	1	10
Forged-URL Flood(A)	0.1	0.1	2.1	32
Random-URL Flood(B)	0.1	0.2	3.3	51
Single-URL Flood(C)	0.1	0.2	3.7	46
Multi-URL Flood(D)	0.1	0.3	7.2	79
Session Flood(E)	0.1	1.2	11.6	108

最后,分析判定误差随请求数的变化关系(与文献[14]工作相比较)如图 5 所示,与文献[14]中的 SSM 方法相比,本文提出的 UAP 模型误报率有明显降低,随着请求数的增加,误报率稳定值也低于 SSM 方法。因 UAP 模型中建立了网站规则检测和路径长度范围内异常综合判定模块,提高了检测准确性,防御模块具有信息验证功能,使得误报率大幅降低。

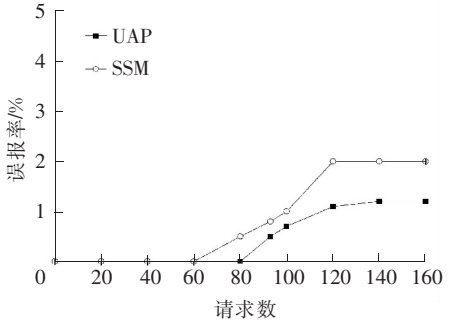


图 5　判定误差随请求数的变化关系

Fig.5　Relation between decision error and request number

2.3　性能分析

对于数据的来源,文献[9]采用收集计算网站和主干网络的数据作为检测的依据,而 UAP 模型采用网站自身产生的数据,在数据读取和处理时间上优于文献[9]所描述的方法,由于数据来源简便,因此适用范围更广泛。对于关键值的存取,文献[11]提出以用户忠实度判定用户异常,使用 Cookie 存储用户忠实度,而

Cookie存在一个安全性问题，被拦截后原样转发Cookie可以达到挟制目的，本文中用户非法值存储于服务端，直接由服务端布置相应防御策略，不受客户端影响，可靠性高。对于异常的检测，文献[15]只针对网站结构和访问时间间隔进行检测，当挟持较短路径长度发起会话攻击时，几乎完全避开了攻击检测。本文提出的UAP模型除上述两方面外，还增加了网站内容对用户行为产生的影响和自适应的路径长度异常检测，封堵了检测算法对路径长度依赖的漏洞，提高了异常检测的全面性。

3 结 语

针对网站结构，将行为概率分布特点、攻击行为特征和访问路径异常特征与Web访问路径相结合，建立异常检测模型，能综合判定路径长度范围内的异常值。将异常检测模型架设在应用服务器和Web服务器之间，能有效减少路径信息采集时间。配合过滤转发防火墙防御模块完成UAP模型的搭设，实现实时检测、实时防御应用层DDoS攻击，并有效降低误报率。通过统计网站正常用户历史访问信息中的节点驻留时间、节点访问频次和固有链接关系，定义网站固有节点 p_i、合法边 l_r、空间权值 S、时间权值 t 和请求频次 G。检测模型对路径长度的有效监控，提升了异常判定的准确性和可靠性。

服务器会对用户的每次访问请求进行记录，通过对记录信息的提取和处理，按照用户访问时间的先后顺序，将数据进行组合形成路径信息，并传入异常检测模块进行计算。计算异常值是否超过正常范围，从而判定用户是否为异常用户。同时非法值会记入用户的历史记录，为防御部署提供理论选择。由于各网站结构功能不同，异常检测所分配的权重值以及各参数值取值仍需进一步完善。为了能够与分布式框架相匹配，与分布式应用程序协调运作的策略需进一步研究。

参考文献/References：

[1] SHAFIEIAN S, ZULKERNINE M, HAQUE A. CloudZombie: Launching and detecting slow-read distributed denial of service attacks from the cloud[C]// 2015 IEEE International Conference on Computer and Information Technology. Liverpool: IEEE, 2015: 1733-1740.

[2] XIE Yinglian, YU Fang, KE Qifa, et al. Innocent by association: Early recognition of legitimate users[C]// Proceedings of the 2012 Acm Conference on Computer and Communications Security. New York: ACM, 2012: 353-364.

[3] 邹存强.基于ISP网络的DDoS攻击防御的研究及应用[D].北京:北京邮电大学,2010.
ZOU Cunqiang. Research and Application of Defense of DDoS Attacks Based on ISP Network[D]. Beijing:Beijing University of Posts and Telecommunications, 2010.

[4] BEITOLLAHI H, DECONINCK G. ConnectionScore: A statistical technique to resist application-layer DDoS attacks[J]. Journal of Ambient Intelligence and Humanized Computing, 2014, 5(3): 425-442.

[5] 赵洋，单娟，宋超.复杂网络中的病毒传播机制研究[J].河北科技大学学报,2011,32(3):252-255.
ZHAO Yang, SHAN Juan, SONG Chao. Virus propagation mechanism of complex network[J]. Journal of Hebei University of Science and Technology, 2011, 32(3):252-255.

[6] 齐林，王静云，蔡凌云.SQL注入攻击检测与防御研究[J].河北科技大学学报,2012,33(6):530-533.
QI Lin, WANG Jingyun, CAI Lingyun. Detection of SQL injection attacks and the defense[J]. Journal of Hebei University of Science and Technology, 2012, 33(6):530-533.

[7] 李锦玲，汪斌强，张震.基于流量分析的App-DDoS攻击检测[J].计算机应用研究,2013,30(2):487-490.
LI Jinling, WANG Binqiang, ZHANG Zhen. Detecting App-DDoS attacks based on flow analysis[J]. Application Research of Computers, 2013, 30(2): 487-490.

[8] ZHAO Yuntao, ZHANG Wenbo, FENG Yongxin, et al. A classification detection algorithm based on joint entropy vector against application-layer DDoS attack[J]. Security and Communication Networks, 2018:9463653.

[9] 王风宇，曹首峰，肖军，等.一种基于Web群体外联行为的应用层DDoS检测方法[J].软件学报,2013,24(6): 1263-1273.
WANG Fengyu, CAO Shoufeng, XIAO Jun, et al. Method of detecting application-layer DDoS based on the out-linking behavior of Web community[J].Journal of Software, 2013,24(6): 1263-1273.

[10] 康松林，詹煜，樊晓平，等.基于蛋白质相互作用网络的DDoS攻击检测[J].小型微型计算机系统,2015, 36(6): 1283-1290.
KANG Songlin, ZHAN Yu, FAN Xiaoping, et al. DDoS detection which bases on protein interaction networks[J]. Journal of Chinese Computer Systems, 2015, 36(6): 1283-1290.

[11] 孙未，张亚平.基于用户忠实度的应用层DDoS防御模型[J].计算机工程与设计,2015,36(1): 93-97.

SUN Wei, ZHANG Yaping. Application layer DDoS defense model based on user loyalty[J]. Computer Engineering and Design, 2015,36(1): 93-97.

[12] KEDJAR S, TARI A. The hybrid model for web services security access control and information flow control[C]// 8th International Conference for Internet Technology and Secured Transactions. London: IEEE, 2013: 6750190.

[13] KANDULA S, KATABI D, JACOB M, et al. Botz-4-sale: Surviving organized DDoS attacks that mimic flash crowds[C]//Proceedings of the 2nd Conference on Symposium on Networked Systems Design and Implementation.[S.l.]:[s.n.], 2005: 287-300.

[14] 肖军, 云晓春, 张永铮.基于会话异常度模型的应用层分布式拒绝服务攻击过滤[J].计算机学报,2010, 33(9): 1713-1724.
XIAO Jun, YUN Xiaochun, ZHANG Yongzheng. Defend against application-layer distributed denial-of-service attacks based on session suspicion probability model[J]. Chinese Journal of Computers, 2010, 33(9): 1713-1724.

[15] 刘泽宇, 夏阳, 张义龙, 等.基于 Web 行为轨迹的应用层 DDoS 攻击防御模型[J].计算机应用,2017,37(1): 128-133.
LIU Zeyu, XIA Yang, ZHANG Yilong, et al. Application-layer DDoS defense model based on Web behavior trajectory[J]. Journal of Computer Applications, 2017,37(1): 128-133.

[16] 龚浩.西南地区民族中学校园网内容建设及影响因素研究[D].重庆:西南大学,2010.

[17] STOCCO A, LEOTTA M, RICCA F, et al. Clustering-aided page object generation for Web testing[C]// Proceedings of the 16th International Conference on Web Engineering. Switzerland: Springer International Publishing, 2016: 132-151.

[18] 岳永哲, 赵战民.网络通信信息传输效率控制仿真研究[J].计算机仿真,2017, 34(10): 269-272.
YUE Yongzhe, ZHAO Zhanmin. Simulation research on information transmission efficiency control of network communication[J]. Computer Simulation, 2017, 34(10): 269-272.

[19] 陈迪, 张鹏, 杨洁艳, 等.在线地图服务日志的大数据分析[J].小型微型计算机系统,2015, 36(1): 33-38.
CHEN Di, ZHANG Peng, YANG Jieyan, et al. Big data analysis of Web map service log[J]. Journal of Chinese Computer Systems, 2015, 36(1): 33-38.

[20] 李珊, 刘继超, 邵芬红.Web 日志与浏览行为结合下的用户浏览兴趣数据挖掘分析[J].现代电子技术,2017, 40(5): 22-25.
LI Shan, LIU Jichao, SHAO Fenhong. Analysis of user's browsing interest data mining combining Web log with user's browsing behavior [J]. Modern Electronics Technique, 2017, 40(5): 22-25.

[21] 张玺, 张学玲, 张洪欣.基于 Web 日志的数据预处理方法研究[J].滨州学院学报,2014,30(6): 98-104.
ZHANG Xi, ZHANG Xueling, ZHANG Hongxin. Research on data preparation based on Web log[J]. Journal of Binzhou University, 2014,30(6): 98-104.

[22] 何涛.基于 SIP 协议的攻击呼叫检测关键技术研究[D].郑州:解放军信息工程大学,2011.
HE Tao. Research on Key Technology of Attacking Call Detection Based on SIP[D]. Zhengzhou: The PLA Information Engineering University, 2011.