文章编号：2095－6835（2019）07－0001－02

分布式环境下基于机器学习的 DDoS 攻击检测的研究与发现*

唐思均
（宜宾职业技术学院，四川 宜宾 644000）

摘　要：随着计算技术的快速发展，分布式拒绝服务（Distriibuted Denital of Service，DDoS）攻击已经成为目前信息网络不稳定的重要原因，由于僵尸网络的盛行，DDoS 已经越来越严重，每次攻击都会产生严重的影响。主要围绕分布式环境下 DDoS 攻击原理和 DDoS 检测方法进行分析，探讨分布式环境下 DDoS 攻击系统设计与实现的有效方式，从而为相关领域提供丰富的理论基础。

关键词：分布式环境；机器学习；DDoS；攻击检测
中图分类号：TP393.08　　**文献标识码：**A　　DOI：10.15913/j.cnki.kjycx.2019.07.001

1　DDoS 攻击原理及攻击工具

1.1　DDoS 攻击原理的概述

DDoS 攻击通常也被称为分布式拒绝服务攻击，这是一种通过控制网络上的主机产生大量的流量来制造巨大规模数据，流损耗主机网络带宽使主机无法正常工作的攻击方式。DDoS 的攻击方式主要有攻击网络带宽资源和攻击系统资源两种。由于互联网中路由器、服务机、交换机设备在带宽和数据包解析方面能力有限，因此如果大量的网络数据同时发过来，会导致网络出现堵塞，无法提供正常的服务。DDoS 攻击利用消耗网络带宽的原理，产生大量不必要的数据包，给被攻击的网络设备带来巨大的数据流量，使被攻击的主机无法正常响应。消耗网络资源的 DDoS 攻击主要分为直流洪水攻击和反射放大攻击。

其中直流洪水攻击主要是通过控制一定数量的僵尸主机，制造大量的网络数据包，同时发送给被攻击的主机网络，使被攻击的主机因为网络带宽被占满而无法正常运行。通常情况下，直流洪水攻击有 ICMP 洪水攻击和 UDP 洪水攻击两种。反射放大攻击跟直流洪水攻击相比，原理较为复杂，由于直流洪水攻击的攻击效果并不理想，非常容易被查到攻击源头，而反射放大攻击则完美地解决了直流洪水攻击存在的缺点。反射攻击主要通过路由器、服务器等网络设备产生响应来发动攻击，反射式攻击不仅能够有效地减轻僵尸主机的运行负担，并且攻击源头也不容易被查找。常见的反射式攻击包括 ACK 反射攻击和 DNS 反射攻击等。

1.2　DDoS 攻击工具的概述

根据目前数据统计，DDoS 的攻击工具种类很多，最常见的攻击工具有 TFN2K、SynK4、LetDown、Hyenae 等。其中 TFN2K 的最早版本是 TFN，这个程序支持多种攻击方式，经过不断完善，已经成为目前最常见的 DDoS 攻击方式。Synk4 攻击工具是由 C 语言编写的，不仅可以在 UNIX 使用，同时还能够与 GNU 兼容，在没有保护设备的网络环境中，攻击效果十分明显。LetDown 可以发动非常强大的洪水攻击，可以在 TCP 的任何阶段切断网路，攻击方式非常复杂，很多网络入侵检测系统都无法有效地拦截。Hyenae 是一种有效的网络数据伪装包发生器，可以在 IPv4 和 IPv6 环境中发动各种形式的 DDoS 攻击，并且 Hyenae 可以独立运行，不受网络设备的干扰，甚至可以绕过防火墙对主机造成严重的影响。

2　DDoS 检测方法

目前，DDoS 的攻击方式越来越复杂，对 DDoS 攻击进行检测也变得越来越困难，在过去的几年里，很多网络安全研究者研发了多种检测 DDoS 攻击的有效工具，这些检测工具和方法有的是分布式的，有的则是集中式的。DDoS 攻击检测系统主要是通过检测干扰系统服务信号为主机提供网络保护。

2.1　基于网络流量的 DDoS 检测方法

当主机遭到 DDoS 攻击时，被攻击的主机网络中会瞬间出现大量的数据包，网络流量数据会出现明显的异常，因此，基于网络流量的 DDoS 检测方法最直接也最有效。基于网络流量的 DDoS 检测方法包括基于全网流量变化检测、TCP 数据包变化比例检测、子网流量变化检测三种方式。其中基于全网流量变化的检测是通过对网络中源 IP 和目的 IP 之间的流量进行检测，根据网络攻击流的相关性，可以建立相应的流量矩阵，最终检测出 DDoS 的攻击方式。TCP 数据包变化

*［基金项目］四川省教育厅科研项目“基于分布式拒绝服务攻击（DDoS）防御技术研究”（项目编号：18ZB0678）

比例检测主要是通过对网络环境中 TCP 数据包的标志位比例进行检测，一旦比例发生明显的变化，就会检测出相应的 DDoS 攻击方式。子网流量变化检测主要是通过网络中流量的变化情况，根据攻击原理，检测出 DDoS 攻击方式。

2.2 基于地址变化的 DDoS 检测方法

基于地址变化的 DDoS 检测方法是由源地址出现速率变化和源地址分布变化两部分组成。源地址出现速率变化的检测原理是，主机在受到 DDoS 攻击的时候，由于工具的数据包往往跟正常的数据包相同，但是被攻击的网络 IP 地址数量会急速增加，根据这个原理，使用机器学习的方式就能够快速有效地检测出 DDoS 的攻击方式。这种方式对源地址均匀分布的 DDoS 攻击检测效果非常显著。源地址分布变化的检测原理是，主机在正常运行的情况下，IP 地址分布比较稳定，一旦受到 DDoS 攻击之后，很多 IP 地址都是 DDoS 伪造的，因此，此时 IP 地址分布会变得非常不稳定，根据这个特点，使用滑动窗口将数据包进行分类，通过计算相邻滑动窗口的系数，就能够有效地检测出 DDoS 的攻击方式。

2.3 基于数据包头统计信息变化的检测方法

如果主机受到 DDoS 攻击的时候，由于攻击数据包头部信息很多都是随机产生的，因此，数据包信息的统计结果会发生明显的改变。利用相应的方法检验数据包头部的信息，并与正常情况的分布信息情况进行比较，可以快速有效地分析出响应的 DDoS 攻击方式。另外，还可以通过对数据包进行采样分析，从而降低计算的难度，进行高效的 DDoS 检验。目前常见的 DDoS 攻击方式主要是 SYN 洪水，由于攻击者在进行 DDoS 攻击的时候，会发送大量的 SYN 数据包，但是不会对被攻击主机的服务器做出相应的回应，因此，能够快速消耗被攻击主机的网络宽带，使其无法正常运行。利用被攻击主机受到的 SYN 数据包进行分析，可以检测出发动 SYN 攻击的主机网络位置。

3 分布式环境下 DDoS 攻击检测系统设计的有效方式

3.1 进行系统框架的设计

通常情况下，分布式环境 DDoS 攻击检测系统由四个部分组成，分别为数据采集系统、数据处理子系统、HDFS 分布式文件系统和数据分析子系统。其中数据采集子系统负责触发检测系统工作并进行源数据的获取，提取相关信息为后续的工作做准备。数据处理系统负责对数据进行处理加工，其中包括数据包分组信息、数据特征信息提取和统一归类操作等。HDFS 子系统是对海量的数据信息进行快速有效的储存，当进行数据处理的时候，能够将数据信息进行快速的提取。数据分析子系统主要负责对数据特征进行分析，检测出相应的 DDoS 攻击方式。

3.2 数据采集子系统的设计

数据采集子系统主要负责对数据包信息进行快速提取，在获取数据包的时候，数据采集子系统会根据相应的原理提取其中不同的特征信息，并快速发送给后面的系统进行相应的处理和检测工作。数据采集系统另一个重要的功能是能够快速启动 DDoS 攻击检测系统。在主机正常运行的情况下，不会对 DDoS 攻击进行检测，只有在数据采集子系统获取相应的数据包之后，才会启动这一功能。因此，数据采集子系统的设计对于检测 DDoS 有着非常重要的作用。

3.3 HDFS 分布式文件系统设计

在 DDoS 检测系统中，HDFS 分布式文件系统主要负责数据包的储存工作，数据采集系统会将获取的数据包传送给 HDFS 分布式文件系统，然后 HDFS 会以文本的形式进行保存。HDFS 系统有很多应有优势，不仅适应各种大数据的储存，同时也能够在性能较低的主机上运行，并且 HDFS 系统有着一定的容错机制，每份数据包信息都会进行多份储存。通常情况下，HDFS 系统是由客户端、Secondary NameNode 节点、NameNode 节点组成，数据被获取之后，系统会自动缓存一定的时间，且每过一定的时间，HDFS 客户端就会将文本文件储存到 HDFS 系统中，确保数据采集子系统提取的信息能够进行快速有效的存储。

3.4 数据处理子系统

数据处理子系统的主要作用是对数据进行一定的加工处理，主要包括分组、特征信息提取等。通常情况下，数据处理子系统都是利用 Java 语言进行编写，包括数据分组和数据提取两个部分。数据分组功能是将处理过的信息根据数据特征进行快速有效地分组，分组原理跟协议类型有所不同。数据提取是将分组后的数据根据特征进行提取，这个功能通常是利用 Spark 进行设计，具有较快的反应速度和较高的可拓展性。

3.5 数据分析子系统的设计

数据分析子系统的功能主要是对检测的数据进行分析，获得相应的 DDoS 攻击的 IP 地址，其中 BP 神经网络是数据分析子系统的核心。在进行大量的样本分析的过程中，数据分析子系统会利用 BM-HJ-GSO 算法进行误差分析，从而获取 DDoS 攻击的引擎信息，对 DDPS 攻击检测系统输入相应的特征数据后，可以对 DDoS 进行全面的检测，并得出相应的分析结果。在进行 BM – HJ-GSO 计算的时候，通常需要使用 Spark 算法，这样可以快速有效地获取 DDoS 的初始值。通过使用 BP 神经网络的初始值，可以对 DDoS 进行全方位的检测工作。在对 BP 神经网络进行调试的过程中，需要获取相应的神经网络模型。将检测模型的特征信息转化成 RDD 信息，然后再对 DDoS 进行检测。

4 结语

随着计算机技术的快速发展，为了给用户的网络使用安全提供保障，应该完善分布式环境下基于机器学习的 DDoS 的检测技术，网络安全人员需要明确 DDoS 的攻击原理和常

（下转第 7 页）

4.2 加大了课堂教学信息量，丰富了课堂教学内容

传统的粉笔加黑板教学模式，教师用粉笔在黑板上画一个比较复杂的函数图像费时、费力且不够精确。利用 MATLAB 软件可精确地生成各种函数的图像，极大节约了时间，提高了教学效率，加大了课堂教学信息量。

4.3 帮助学生更深入地理解课程内容，提高了教学质量

高等数学的知识是从大量实践中抽象出来的，这一特点就决定了该课程内容很难理解，比如讲定积分的基本概念的时候，要从曲边梯形的面积说起，而曲面梯形的面积通过分割、近似、求和和取极限四个步骤得到。再讲这样的内容时候我们利用 MATLAB 软件编程，动态地演示不断分割的过程，能帮助学生更好地理解曲边梯形的面积，从而对定积分的概念有更深入的认识。在三年多的教学实践中引入 MATLAB 软件学生使得学生的考试成绩稳步提高，与同一专业之前的两个年级相比较及格率提高了 7.6%.

5 结束语

MATLAB 软件与高等数学课程深度融合，将抽象的问题变得形象、直观，提高了学生的学习兴趣，丰富了课堂内容，有利于学生理解数学知识，有利于学生掌握数学思想、方法。运用 MATLAB 软件对计算进行结果验证，增强学习自信心，培养学生的观察能力和创造能力。在以后的教学中我们将利用 MATLAB 软件搭建智能化高等数学的学习平台，让学生掌握数学知识的同时具备一定的编程素养，从而提高学生的创造力和综合应用能力。

参考文献：

[1] 杨大勇.线性规划灵敏度分析的一个应用[J].赤峰学院学报，2013（7）：6-7.
[2] 唐少芳.MATALB 在高等数学教学中的应用[J].亚太教育，2016（13）：120.
[3] 陈家颐.高职公共基础课改革的实践与思考[J].中国高教研究，2005（11）.
[4] 马丽娜，刘烁.MATLAB 数学软件在线性代数教学中的应用[J].产业与科技论坛，2011（10）.

〔编辑：严丽琴〕

（上接第 2 页）

见的 DDoS 攻击工具的特点，设计相应的 DDoS 检测系统，从而快速有效地检测出 DDoS 攻击方式，使用户的网络能够不受干扰。

参考文献：

[1] 谭森.分布式环境下基于机器学习的 DDoS 攻击检测的研究与实现[D].北京：北京邮电大学，2018.
[2] 贾斌.基于机器学习和统计分析的 DDoS 攻击检测技术研究[D].北京：北京邮电大学，2017.
[3] 刘运.DDoS Flooding 攻击检测技术研究[D].长沙：国防科学技术大学，2011.
[4] 孙永强.基于机器学习的分布式拒绝服务攻击检测方法研究[D].长沙：国防科学技术大学，2006.

作者简介：唐思均（1980—），男，四川宜宾人，研究生，讲师，研究方向为计算机科学。

〔编辑：严丽琴〕

（上接第 4 页）

2018，39（01）：80-86.
[3] 张振利，梁毓明，温如春.基于虚拟仪器的分布式温度测控系统设计[J].江西理工大学学报，2012，33（01）：44-47.
[4] 翁发禄，梁礼明，丁元春，等.随机地震波干扰作用下不确定结构系统主动控制研究[J].江西理工大学学报，2015，36（03）：78-84.
[5] 林春，方晓猛.胡天林投篮机器人气动式抛射机构的设计与分析[J].现代制造技术与装备，2017（02）：66-68，70.
[6] 帅超，廖贵超，张阳新，等.高压空气抛射角反射器过程内弹道研究[J].南京理工大学学报，2017（02）：152-158.

作者简介：范振（1999—），男，研究方向为控制理论与控制工程。

〔编辑：张思楠〕