基于支持向量机的P2P网络DoS攻击检测

吴 敏[1],王汝传[1],王治平[2]
(1. 南京邮电大学 计算机学院,江苏 南京 210003;
2. 中兴通讯股份有限公司,江苏 南京 210012)

摘 要:对等网络技术近年来发展迅速,但其安全性问题一直是P2P网络进一步发展和应用急待解决的重要问题之一。由于对等网络的松散性,基于洪泛式请求的拒绝服务(DoS)攻击已经成为主要威胁之一。文章首先介绍和分析了DoS攻击在对等网络下的特点,然后提出了一种基于支持向量机的P2P环境下DoS攻击的检测模型,该模型能够通过以离线的方式对发生DoS攻击时流的统计特性进行特征提取,并能实时识别攻击的发生。实验证明,这种模型具有较高的检测率和较低的误检率。

关键词:对等网络;拒绝服务;支持向量机
中图分类号:TP391 **文献标识码**:A **文章编号**:1673-629X(2009)11-0151-04

Detection Mechanism of DoS Attacks in Peer-to-Peer Networks Based on Support Vector Machine

WU Min[1], WANG Ru-chuan[1], WANG Zhi-ping[2]
(1. College of Computer, Nanjing University of Posts and Telecommunications, Nanjing 210003, China;
2. ZTE Corporation, Nanjing 210012, China)

Abstract: Peer-to-peer network technology has improved quickly in recent years, however the security problem is always the most important issue that impede the further improvement of P2P application which should be solved immediately. Among all these security issues the denial-of-service(DoS) attacks are becoming one of the major threats in peer-to-peer networks which are loosely connected. First describe the characters of DoS attack in P2P environment, then propose a DoS attack detection model based on support vector machine mechanism. The model can extract flow statistics feature when DoS attack occurs offline and can identify the event online. The result of the experiment shows that this model has high detection rate and low false reject rate in detecting DoS attack in P2P environment.
Key words: P2P network; denial-of-service; support vector machine

0 引 言

随着分布式计算技术的快速发展和计算环境的日趋复杂化,对等计算也就成为一种商业需求,而且可能成为影响未来互联网发展最重要的技术之一[1~3]。P2P网络环境的灵活和开放性决定了其不安全性,从而也带来了很多问题和挑战。在P2P的资源安全中,拒绝服务(Denial of Service,简称DoS)攻击是对资源可用性的主要威胁之一[4~7]。最近研究状况表明黑客正越来越多地利用P2P网络欺骗一些PC对其它计算机发动攻击,即利用了P2P协议设计时的漏洞,远程控制P2P网络中大量计算机,形成一个用于发动拒绝服务攻击的"僵尸网络",如果产生的流量足够大,就会导致目标计算机瘫痪,即使目标计算机并不属于P2P网络的一部分,如DNS服务器。

目前P2P网络的研究尚处于起步阶段,研究热点又集中在如何提高资源查找效率以及如何基于系统构建更为复杂的应用系统等方面,其安全方面的研究并不是十分充分。现有的P2P网络软件设计存在缺陷,从而使得攻击者可以轻松地发起庞大的拒绝服务攻击,导致互联网网站的崩溃。而目前网络的DDos攻

收稿日期:2009-03-02;**修回日期**:2009-06-13
基金项目:国家自然科学基金(60573141,60773041);江苏省自然科学基金(BK2008451);国家高科技863项目(2007AA01Z404,2007AA01Z478);南京市高科技项目(2007软资127);现代通信国家重点实验室基金(9140C1105040805);江苏高校科技创新计划项目(CX08B-085Z,CX08B-086Z);中兴通讯高校合作基金
作者简介:吴 敏(1976-),女,江苏泰州人,讲师,博士研究生,研究方向为移动代理技术、分布式计算、计算机密码学和网格计算;王汝传,教授,博士生导师,研究方向为计算机软件、计算机网络和网格、对等计算、信息安全、无线传感器网络、移动代理和虚拟现实技术。

击研究主要是针对一般网络的，P2P 网络和传统的 C/S 模式是有着非常不同的特点，因此研究适合 P2P 网络[7]的 DDos 攻击特点及其检测防御机制对于保障 P2P 网络中的安全有非常重要的意义。

1 典型的 P2P 环境下的 DDos 攻击

图 1 是一个典型的利用 P2P 网络特点发布虚假消息进行 DDos 攻击示意图，使用协议为 Bittorrent，步骤如下：

(1) 恶意节点解析. torrent 文件，向 tracker 发布伪造的虚假信息，例如宣称受害节点拥有某资源；

(2) 当客户主机向 tracker 请求文件拥有节点列表时，tracker 返回信息中包含了受害主机的 IP 地址和端口号；

(3) BT 网络中大量客户主机试图向受害主机发起连接，请求文件资源，形成 DDos 攻击。

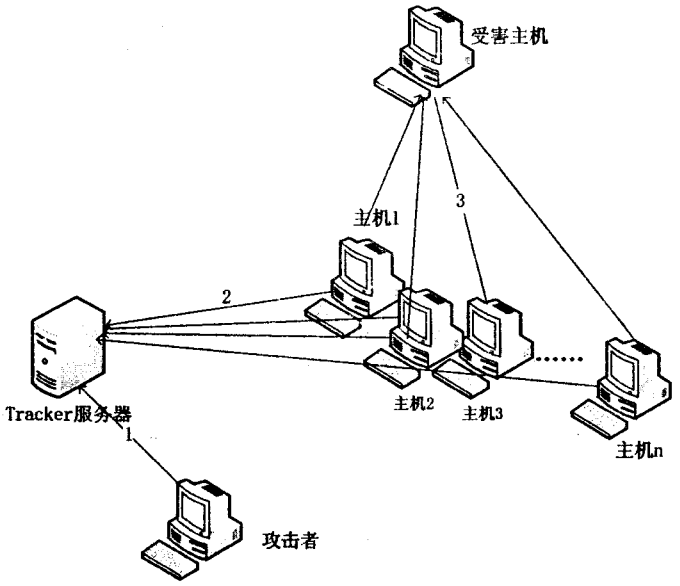

图 1 P2P 网络中拒绝服务攻击示意图

分析可知，在分布式 P2P 环境下，系统中的攻击者冒充攻击目标节点向索引服务器发布虚假资源拥有消息，将某个热门资源与受害主机或者是 P2P 网络中不存在的主机绑定，从而造成资源定位指针列表的污染。然后那些请求资源或服务的节点纷纷与这个节点联系，如果这个节点存在，则大量请求涌向这个节点，而由于目标节点上并没有该资源，因此受害主机会忽略这个请求，挂起连接，而连接资源是有限的，大量挂起的连接会使目标节点的连接资源耗尽，无法响应其他的连接请求，导致服务崩溃。

2 基于支持向量机的 P2P 下 DDoS 攻击检测模型

2.1 支持向量机技术

20 世纪 80 年代后期以来，V. Vapnik 等人一直致力于统计学习理论的研究，针对传统模式识别的问题，提出了一种小样本的学习方法—支持向量机[8,9]。这是根据统计学习理论，以结构风险最小化原则为理论基础的一种新的机器学习方法，对有限样本情况下模式识别中的一些根本性问题进行了系统的理论研究，很大程度上解决了模型选择与过学习问题、非线性、维数灾难、局部极小点等问题。

DDos 攻击检测实际上也是一个分类问题，也就是对数据流量进行分类，分辨什么样的数据是正常的，什么样的数据是异常的。从本质上讲属于机器学习中模式识别的范畴，即通过对已知样本的学习，找到规律用于对将来未知样本的预测。其中训练样本是已知的网络流量或系统审计记等，输出是类别标号，所以基于支持向量机的方法可以应用到 P2P 网络中 DoS 攻击是否发生的检测中。

2.2 基于支持向量机的 P2P 网络 DDos 攻击检测模型

图 2 是基于支持向量机的 P2P 网络中 DDos 攻击检测模型。其中，数据采集模块使用某种数据包嗅探器从网络中捕获当前网络中的数据包信息，提取网络连接的特征数据信息。数据预处理模块是对大量的审计数据进行处理，由于支持向量机的分类器只能对维数相同的数字向量进行分类，所以必须将量化后数据转换为支持向量机能够识别的数字向量形式。如果是处于基于支持向量机的 DDos 特征选取模块，则选取网络中 DDos 攻击可能相关特征库，按照相应的步骤确立P2P环境

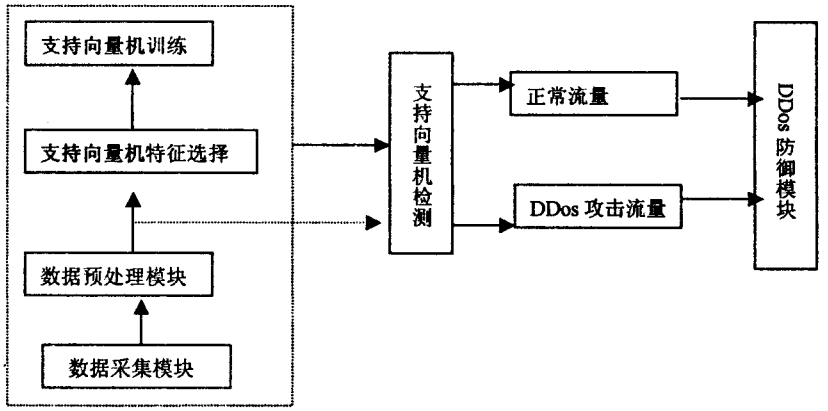

图 2 基于支持向量机的 P2P 网络中 DDos 攻击检测模型

中的 DDos 攻击特征库;如果处于 SVM 训练状态,则训练 SVM,并将训练后的结果,即若干个支持向量存入 SVM 支持向量库;如果处于 SVM 检测状态,则由 SVM 检测模块对输入向量进行检测;根据检测结果进行相应的 DDos 攻击防御操作。

2.3　基于支持向量机技术的 P2P 网络中 DDos 攻击特征选取模块

由于 P2P 网络的分布性、动态性,以及攻击者采用伪造、随机变化消息源地址等方法,使得 P2P 网络下的 DoS 攻击特征不同于传统网络中的 DoS 攻击特征,而攻击特征的难以提取,将给攻击防御带来很大的困难。因此检测的第一步就是研究 P2P 网络下 DoS 良好的表示特征,使用支持向量机以特征约减方法,去除冗余特征,分析提取 P2P 网络下 DoS 攻击的重要特征,为下一步建立进行防御做好准备。

(1) P2P 网络下与 DoS 攻击可能相关特征库。

将每个网络连接与攻击相关的属性分为四类,即基本属性集、内容属性集、流量属性集和主机流量属性集。

(2)基于 SVM 的 DoS 攻击特征筛选方法。

特征筛选是一个最优化问题,我们提出一个对特征筛选方案,即在确保检测精度和降低误判的前提下,去除非重要特征,选择关键特征,从而提高核参数选取,降低 DoS 攻击检测系统的处理时间,降低运算量,节约系统资源。该方法是利用经验数据来估计每个属性对于检测结果的贡献,提出了一种方法每次只删除掉一个属性,保留其他全部属性,观察检测结果的变化,从而确定该属性的贡献。删除结果的影响主要表现在训练时间、测试时间、检测精度三个方面。删除某个属性后,检测结果存在三种可能的变化增加、下降、持平,可以按照表 1 规则判别。

表 1　属性重要程度判断表

规则	检测精度	训练时间	测试时间	属性判断
1	下降	增加	下降	重要
2	下降	增加	增加	重要
3	下降	下降	增加	重要
4	基本无变化	增加	增加	重要
5	基本无变化	下降	增加	辅助特征
6	基本无变化	增加	下降	辅助特征
7	基本无变化	下降	下降	不重要
8	增加	增加	下降	辅助特征
9	增加	下降	增加	辅助特征
10	增加	下降	下降	不重要

(3)特征选取步骤。

1)对于具有全部属性的数据集 Z,记录下测试数据;

2)删除数据集 Z 的第 $i(1\leqslant i\leqslant 41)$ 个属性,记录测试结果;

3)重复 3),直到删掉最后一个属性,并记录测试结果;

4)通过对最后测试结果的分析,将属性判断不为重要的属性删除,剩余的特征属性重新组成一个新的训练集;

5)在这个新训练集上进行步长为 2^2 的粗网格搜索,得到核参数(C_k, r_k);

6)在上一步中得到的最优参数(q, Yk) 旁正负 22,以 $2o.zs$ 为步长进行一次更精细 的网格搜索,得到最优参数$(C_{\text{final}}, r_{\text{final}})$;

7)用第 6 步得出的适合于此分类问题的最好参数$(C_{\text{final}}, r_{\text{final}})$ 训练整个训练集;

8)在数据集上进行测试,得出该分类问题的分类精确率;

9)根据分类精度确定最佳 Dos 攻击判别特征,并确立相应的最佳训练模型。

2.4　基于支持向量机的 DDos 攻击训练模块

该模块对预先选定的训练数据集进行训练,训练数据集中的数据是从数据预处理模块中得到的。训练有两种方式:一种是监督学习,就是对训练数据集中的每条数据都给出其类别信息,即训练样本是由 (x, y) 成对给出;另一种是非监督学习,就是训练数据集中的每条数据不给出其类别信息,经训练之后将得到一组支持向量并存入 SVM 支持向量库,这组支持向量也就是训练后得到的模型。

2.5　基于支持向量机的 DDos 攻击检测模块

该模块也是整个系统最重要的部分,是该模型的核心部分。它利用 SVM 训练模块得到的 SVM 支持向量组对实际需要检测的网络连接记录(这些记录也是从数据预处理模块中得到的)进行预测,预测的结果即输出值为 1 表示正常,即未发生 DDos 攻击行为;-1 表示异常,即发生 DDos 攻击行为。对于正常流量放行,对于发生 DDos 攻击的流量则交由相应的 DDos 攻击防御模块进行防御,该部分可以参考传统网络中防御 DDos 的方式。

3　实验验证

采用三个 P2P 网络环境测试和验证基于支持向量机的拒绝服务攻击,包括 bttorrent、Pplive、Emule。验证指标包括检测准确率和误检率,如图 3,图 4 所示。

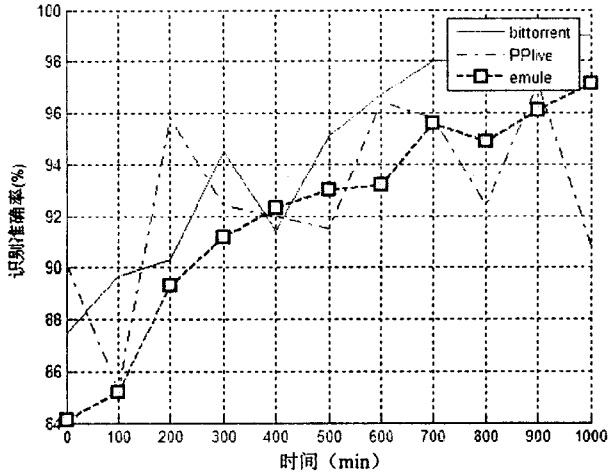

图 3 拒绝服务攻击流量检测准确率

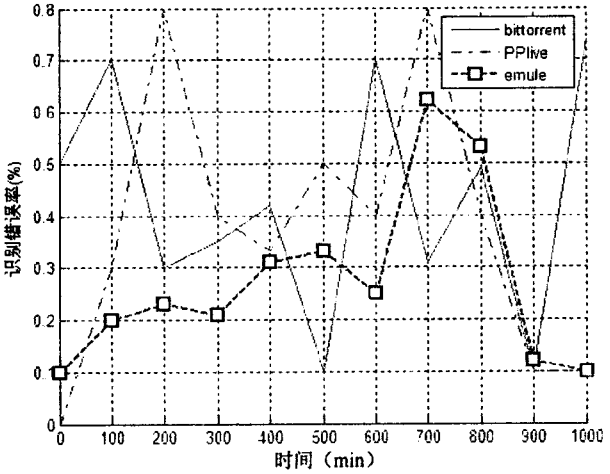

图 4 拒绝服务攻击流量识别错误率

从图 3 中可以看出,在基于 bittorrent 的 P2P 文件共享应用环境中 DoS 攻击检测准确率较高,其次是 Emule 和 Pplive,并且随着时间的增长,准确率逐渐上升,这是因为随着采集数据的增多,支持向量机能更为有效地分析流量规律,准确度随之上升;在图 4 中,它们错误率都相对比较低,其中 Pplive 环境中的攻击识别错误率比较高。

4 结束语

文中研究 P2P 网络中的 DDoS 攻击问题,并提出来一个基于支持向量机的 P2P 环境下 DDos 攻击检测模型,该模型能够筛选出 P2P 网络中的 DDos 攻击特征,建立 P2P 网络中的 DDos 攻击行为发生特征库,并进行基于支持向量机的训练,得到相应的支持向量后可以对预处理后的网络流量进行支持向量机检测,并根据检测的结果采取相应的防御措施。下一步的工作是优化算法性能,提高实时在线检测率。

参考文献:

[1] Sen S, Wang J. Analyzing Peer to Peer Traffic Across Large Networks [J]. ACM/IEEE Transactions on Networking, 2004,12(2):137-150.

[2] Azzouna B, Guillemin N. Impact of Peer to Peer Applications on Wide Area. Network Traffic: An Experimental Approach [C]// IEEE Globecom 2004. Dallas, USA:[s.n.],2004.

[3] 吴国庆. 对等网络技术研究[J]. 计算机技术与发展, 2008,18(7):100-104.

[4] Neil D, Garcia-Molina H, Beverly Y. Open Problems in Data-sharing Peer-to-peer Systems[C]// 9th International Conference on Database Theory (ICDT 2003). Siena, Italy: [s.n.],2003.

[5] Daswani N, Garcia-Molina H. Query-Flood DoS Attacks in Gnutella[C]// in the Proc. of CCS '02. Washington, US: [s.n.],2002:181-192.

[6] 蒋海明,张剑英,王青青,等. P2P 流量检测与分析[J]. 计算机技术与发展, 2008,18(7):74-76.

[7] 李俊青. 蚁群优化在 P2P 网络防范 DDoS 攻击中的应用研究[J]. 计算机应用研究, 2009,26(1):339-341.

[8] 瓦普尼克. 统计学习理论[M]. 许建华,等译. 北京:水利电力出版社,2004.

[9] 肖健华. 智能模式识别方法[M]. 广州:华南理工大学出版社, 2006.

(上接第 150 页)

Transactions on Parallel And Distributed Systems, 2008,19 (10):1325-1337.

[7] Takeda A, Hashimoto K, Kitagata G, et al. A New Authentication Method with Distributed Hash Table for P2P Network[C]//22nd International Conference on Advanced Information Networking and Applications-Workshops. Okinawa: [s.n.].2008:483-488.

[8] 左 敏, 张全海, 李建华. 对等网中一种基于 SAML 的委托认证方案[J]. 计算机工程, 2006,32(6):173-175.

[9] Merkle R C. Secure Communications over Insecure Channels [J]. Communications of the ACM,1978,21(4):294-299.

[10] Seo Dong Hwi, Sweeney P. Simple Authenticated Key Agreement Algorithm[J]. Electronics Letters,1999,35(13):1073-1074.

[11] Lin Iuon-Chang, Chang Chin-Chen, Hwang Min-Shiang. Security Enhancement for the Simple Authentication Key Agreement Algorithm[C]//24th Ann. Int. Computer Software and Application Conference. Taipei, Taiwan:[s.n.]. 2000:113-115.