

基于尖点突变模型的联动网络流量异常检测方法

邱 卫 杨英杰

(解放军信息工程大学 郑州 450001) (河南省信息安全重点实验室 郑州 450001)

摘 要 针对现有方法没有考虑联动网络流量的非线性动力学特性,以及不能有效区分正常联动业务流量和异常攻击流量的问题,提出了一种基于尖点突变模型的联动流量异常检测方法。通过对联动网络流量非线性动力学特征参数的分析与提取,建立正常流量的尖点突变模型,利用模型的平衡曲面来描述网络流量系统的行为,构造正常网络流量行为的平衡曲面;并以网络流量行为相对于正常平衡曲面的偏离程度作为异常检测的依据。实验结果表明,所提方法具有较高的检测率和较低的误报率。

关键词 尖点突变,联动,流量异常,非线性动力学,平衡曲面

中图分类号 TP393.08 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2016.3.031

Interaction Network Traffic Anomaly Detection Method Based on Cusp Catastrophic Model

QIU Wei YANG Ying-jie

(The PLA Information Engineering University, Zhengzhou 450001, China)

(Henan Province Key Laboratory of Information Security, Zhengzhou 450001, China)

Abstract As the exiting methods do not consider the nonlinear dynamics feature of interaction network traffic, and cannot distinguish between normal interaction traffic and abnormal attack traffic effectively, we proposed an interaction traffic anomaly detection method based on cusp catastrophe. The normal traffic cusp catastrophe model is established on the nonlinear dynamics parameters of interaction network traffic, and the equilibrium surface is used to describe the behavior of network traffic system and the balance surface of normal network traffic behavior is structured. Then the deviation of normal balance surface is taken as basis to detect anomaly. Experimental results show that this method gets higher detection rate and lower false alarm rate.

Keywords Cusp catastrophe, Interaction, Traffic anomaly, Nonlinear dynamics, Equilibrium surface

网络安全联动^[1,2]是指通过利用相关的安全技术,使得网络中分散、独立的安全产品通过有机集成达到协同保护网络安全的目的;安全联动行为则是联动系统中各安全产品为应对安全事件进行的信息交互和决策过程。保护安全联动系统的正常运行是确保网络安全的基础,而联动系统中诸如网络故障、配置错误、网络病毒和网络攻击等都会对安全产品之间的联动行为造成威胁并引起联动流量参数的异常变化^[3,4]。通过对联动流量行为的分析,能够及时检测出联动过程中的流量异常,提高联动系统的可靠性和安全性,并为联动行为的安全增强提供依据。

流量异常检测是网络异常分析与定位的关键技术,按流量特性的不同,可分为短相关性和长相关性两类^[5,6]。短相关性是指网络流量在短时间序列上的关联性,主要有 Poisson 模型和 Markov 模型等;长相关性是指对业务流量自相似性的刻画,主要有 FARIMA 模型和分形布朗模型等。这些模型都是基于网络流量线性平稳性的假设,且只考虑流量的某一方面的特性,但网络是一个复杂的系统,网络流量的变化受网络状况、拓扑结构、网络配置等许多因素的影响,现有研究^[7,8]表明网络流量不仅具有长相关性,还具有突变性、非平

稳性、重尾性等非线性动力学特性。传统的基于线性平稳性假设的方法不能准确刻画网络流量行为,在应用上具有很大的局限性。针对这些问题,研究者提出了许多基于非线性动力学的新方法。Wei Xiong^[9]通过协同神经网络和突变理论来描述网络流量状态的变化,以状态偏离程度来检测异常。Yang Yue^[10]针对网络流量的非线性动力学特性,利用蝴蝶突变级数模型对网络流量进行建模,通过突变级数的跳变来检测流量的突变。Jianren Lin 等人^[11]利用尖点突变模型对流量的正常和异常数据建模,取得了一定效果,但该模型的属性统计特征参数不能有效刻画流量特性。本文在此基础上,以流量不同属性抽象出的特征参数作为模型变量进行建模,并且将其应用于联动网络流量的分析中。联动系统是一种特殊的通信网络,联动流量不仅具有典型的非线性动力学特性,它的变化还受背景通信网络状况的影响,当背景通信网络异常时,联动系统为应对安全事件产生大量的联动行为,联动流量发生正常跳变;而当联动系统本身受到扰乱或攻击时,联动流量会发生异常跳变。现有方法没有考虑联动系统的实际特性,不能有效区分联动系统中正常的业务流量突变与异常的攻击流量突变。

到稿日期:2015-07-06 返修日期:2015-10-25 本文受国家 863 计划项目(2012AA012704),国家 973 计划项目(2011CB311801),郑州市科技领军人才项目(131PLJRC644)资助。

邱 卫(1990—),男,硕士生,主要研究方向为网络安全、数据分析;杨英杰(1971—),男,副教授,硕士生导师,主要研究方向为网络安全、数据分析。

针对上述问题,本文在充分分析网络流量不同特性的基础上,建立联动网络流量的尖点突变模型,对流量进行分析与检测。首先,根据流量的不同性质将其抽象为不同的特征参数,选出能够显著反映联动网络流量非线性动力学特性的参数作为模型的控制变量和状态变量;其次,在特征参数的基础上建立尖点突变模型,并基于滑动时间窗和 Hoeffding 假设检验规则对流量进行检测;最终,通过实验验证了本文方法的有效性。

1 相关理论

突变理论^[12,13]是数学家勒内·托姆于 1972 年提出的,描述非线性动力学系统随外界条件变化,系统本身状态发生突变的过程。它从系统整体的角度分析外界控制变量和系统内部状态之间的复杂关系,并能通过系统外界控制变量的连续变化来刻画系统状态变量的不连续突变。尖点突变模型是突变理论的一种,已被广泛应用于复杂系统行为的分析与预测中^[14],具体参数如下。

1. 势函数

势函数在突变理论中表示整个系统的变化趋势,由系统状态变量和外部控制变量决定。尖点突变模型的势函数用 $E(x) = x^4 + aux^2 + bvx$ 表示,其中, u 和 v 表示控制变量,对应外部影响因素; x 表示状态变量,对应网络流量本身; a 和 b 为系数。

2. 平衡曲面

平衡曲面是势函数中所有极值点的集合,表示系统的所有可能状态,以平衡曲面为参考,能够确定系统状态是否发生突变。尖点突变模型的平衡曲面用 $E'(x) = 4x^3 + 2aux + bv = 0$ 表示,几何结构如图 1 所示。

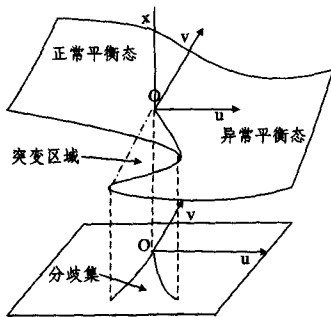


图 1 尖点突变模型几何结构

3. 奇点集

奇点集是系统可能发生突变的位置集合,能够表示系统是否处于稳定状态,用 $E''(x) = 6x^2 + au = 0$ 表示,其中,

$$\begin{cases} E''(x) > 0, & \text{表示系统处于稳定状态} \\ E''(x) = 0, & \text{表示系统处于分歧点} \\ E''(x) < 0, & \text{表示系统处于不稳定状态} \end{cases}$$

4. 分歧集

奇点集在控制变量平面上的映射轨迹就是分歧集,突变理论通过分歧集对突变现象进行界定和控制。分歧集由

$$\begin{cases} E'(x) = 0 \\ E''(x) = 0 \end{cases} \text{联合确定,用 } 8a^3u^3 + 27b^2v^2 = 0 \text{ 表示。}$$

2 基于尖点突变模型的联动网络流量异常检测

联动网络的流量变化受网络拓扑、网络攻击和背景通信

网络状况等诸多因素的影响,流量状态随着时间的推移,在外界条件的变化下,不断从一个平衡态转变到另一个平衡态,这个状态转变过程是一个非平稳、非连续的变化过程,具有典型的非线性动力学特性^[15]。传统的基于线性平稳性假设的方法没有考虑联动流量的非线性动力学特性,并且不能有效区分联动系统中正常的业务流量突变和异常的攻击流量突变。而尖点突变模型能够从系统整体的角度,通过外部控制变量的连续变化分析系统内部状态的不连续变化,从本质上揭示联动系统流量在正常和异常两种状态之间的跳变行为。

尖点突变模型具有 3 种可能状态,通过对特征参数的建模,两种平衡态能够分别刻画联动流量的正常和异常状态,一种非平衡态能描述系统可能发生突变的区域。尖点突变模型通过分析流量系统势函数的临界点来研究系统状态的突变现象,当势函数的取值唯一时,系统处于平稳状态;当外部控制变量处于突变区域,即控制变量在分歧集内取值时,系统的势函数取值不唯一,系统处于不稳定状态;在外部控制变量的作用下,系统可能偏离原有的平衡态,向分歧集靠拢,当偏离程度超过一定阈值时,系统会由正常的平衡状态跃变为异常的平衡状态,发生流量异常突变。

针对传统的基于线性平稳性假设的流量异常检测方法没有充分考虑实际联动网络流量的非线性动力学特征,不能有效区分联动系统中正常的业务流量突变和异常的攻击流量突变的问题,本文提出一种基于尖点突变模型的联动系统流量异常检测新方法。通过对联动流量各属性特征参数的筛选,选出能够显著反映网络流量非线性、长相关性和突变性等特征的参数作为模型的变量;然后,基于训练数据建立正常联动流量的尖点突变模型;最后,在滑动时间窗口模型下,以当前流量系统状态相对于正常平衡状态的偏离度作为异常检测的依据,对联动流量进行实时检测,具体处理流程如图 2 所示。

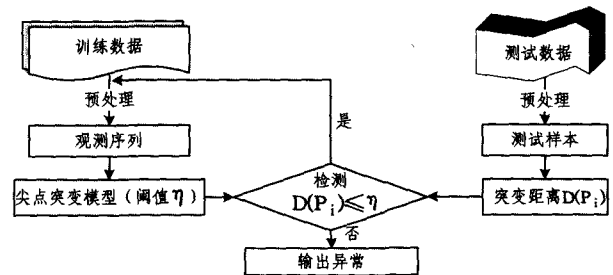


图 2 流量异常检测流程

2.1 特征参数选择

联动网络是一个复杂的通信系统,联动流量的变化受诸多因素的影响,具有非线性、长相关性和非平稳性等非线性动力学特征。尖点突变理论基于流量的特征参数从不同的侧面、宏观上分析联动流量行为的变化。已有研究^[16,17]表明,根据网络流量不同的动力学特性,可以将网络特征参数分为自相似特征类、动力学结构类和统计特征类 3 种。自相似特征是网络业务流量长相关性的体现,从整体反映网络流量的内在变化,包括 Hurst 指数和 Lyapunov 指数;动力学结构特征从短相关性上反映网络流量在动力学上的差异,反映流量的局部变化,包括 Q 因子;统计类特征从表征上反映网络流量的统计情况,直观地描述流量的当前变化,包括均值、方差和变差系数等。这 3 类特征所包含的具体参数如表 1 所列。

基于本文设计的实验,利用 TCPDUMP 工具采集实验网络在 3 周内的数据。

第 1 周,背景通信网络正常时,系统产生联动行为的网络流量,其中包括所有 PC、路由器、交换机、防火墙及入侵检测系统与 CS-MARS 之间的正常通信网络流量。在思科交换机上采集流量数据,共 400 次,每次采集 2 分钟的流量数据,形成训练数据集 Data1。

第 2 周,背景通信网络异常时,系统产生大量联动行为时的联动业务流量。通过外部主机对网络内部主机、交换机等关键节点进行攻击,由于实验环境的限制,主要模拟了 ICMP FLOOD、TCP FLOOD 和 UDP FLOOD 攻击,每种攻击分别进行 400 次,每次持续 2 分钟,离散分布在 1 周时间内。在思科交换机上采集流量,并基于 IP 地址过滤攻击流量数据,将真正的联动业务流量作为训练数据集 Data2。

第 3 周,联动网络本身受到 DDOS 攻击时,产生异常的联动网络数据,此时,内部网络 2 的主机作为傀儡机通过 TFN2K 软件对 CS-MARS 发起 ICMP FLOOD、TCP FLOOD 和 UDP FLOOD 攻击,每种攻击分别进行 400 次,每次持续 2 分钟,离散分布在 1 周时间内。采集攻击发生的 4 分钟流量,将包含攻击流量的联动流量作为测试数据集 Data3。

3.2 结果及分析

在实验数据的基础上,首先,需要筛选出合适的特征组合参数作为模型的控制变量和状态变量,本文利用穷举的方法,对所有可能的特征组合参数建立模型并进行实验,根据实验结果的优劣选择出最合适的组合参数作为模型的变量。为方便实验比较,本文采用误报率 R_F 、检测率 R_T 和漏报率 R_L 作为检测评估的标准,其定义如下:

$$R_F = \frac{TF}{TN + TF} \quad (5)$$

$$R_T = \frac{FN}{FT + FN} \quad (6)$$

$$R_L = \frac{FT}{FT + FN} \quad (7)$$

其中, TN 表示检测为正常的正常数据; FN 表示检测为异常的异常数据; TF 表示被误检为异常的正常数据; FT 表示被漏检为正常的异常数据。

测试中,根据特征参数的选取原则,本文对 10 种所有可能的特征参数组合进行实验,实验中,置信区间 $\theta=0.95$,将模型的漏报率和误报率作为评判标准,实验结果如图 4 所示。

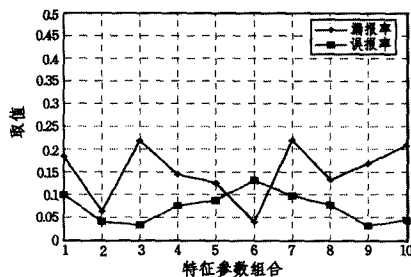


图 4 特征参数选择实验结果

其中特征参数组合如表 4 所列。

由实验结果可知,选取组合 2(Hurst 指数, Q 因子, 三阶矩)作为模型的参数变量时,模型能够取得较低的漏报率和误报率,因此本文在后续的实验中选择 (Hurst 指数, Q 因子,

三阶矩)这个组合特征作为模型变量。

表 4 特征参数组合

特征序号	特征组合	特征序号	特征组合
1	Hurst 指数, Q 因子, 四阶矩	6	Lyapunov 指数, Q 因子, 均值
2	Hurst 指数, Q 因子, 三阶矩	7	Lyapunov 指数, Q 因子, 方差
3	Hurst 指数, Q 因子, 变差系数	8	Lyapunov 指数, Q 因子, 变差系数
4	Hurst 指数, Q 因子, 方差	9	Lyapunov 指数, Q 因子, 三阶矩
5	Hurst 指数, Q 因子, 均值	10	Lyapunov 指数, Q 因子, 四阶矩

为进一步说明本文方法的有效性,提高研究价值,本文与文献[19]的基于自相关函数(ACF)的方法和文献[11]的方法进行比较,实验结果如图 5、图 6 所示。

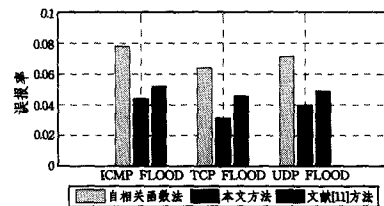


图 5 误报率实验结果

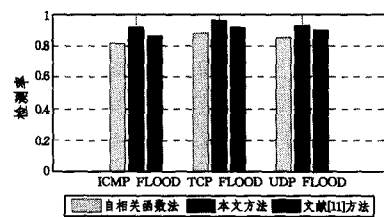


图 6 检测率实验结果

由结果可知,本文方法在检测率和误报率上都优于其它两种方法。当网络流量发生异常时,流量的属性统计特性(如自相关系数、均值等)变化可能十分微小,很难通过这些微小的变动来检测异常;而尖点突变模型利用流量的不同特征参数,建立流量平衡状态模型,当流量发生异常时,流量一定会从正常平衡状态跳变为异常平衡状态,脱离正常平衡曲面,因此本文方法能取得更优的结果;此外,同一检测模型中, TCP FLOOD 的检测率较高,这是因为在联动网络中,基于 TCP 的应用数据更多,优化了训练模型。

结束语 针对联动系统流量所具有的非线性动力学特性以及传统流量异常检测方法不能有效区分正常的联动业务流量突变和异常的攻击流量突变的问题,本文提出一种基于尖点突变理论的联动流量异常检测方法。当流量发生异常时,网络系统会从正常的平衡态跳变为异常的平衡态,网络系统的行为也会发生变化。本文利用尖点突变模型的平衡曲面来描述网络流量的系统行为,构造正常的网络流量行为的平衡曲面,当网络流量行为相对于正常平衡曲面的偏离程度超过一定阈值时,认为发生突变异常。通过在实验室现有联动系统中的部署与测试,验证了本文方法的有效性。下一步,本文将细粒度地对联动流量特征进行分析,并着重研究突变理论中其它突变模型在流量异常检测中的运用。

(下转第 173 页)

Product proxy Re-Encryption with constant size ciphertext[C]// SCC'15. Singapore: ACM, 2015: 31-40

[9] Goyal V, Pandey O, Amit S, et al. Attribute-Based encryption for fine-grained access control of encrypted data[C]// 13th ACM Conference on Computer and Communications Security. Alexandria, VA, United states: ACM, 2006: 89-98

[10] Bethencourt J, Amit S, Brent W. Ciphertext-policy attribute-based encryption[C]// IEEE Symposium on Security and Privacy, 2007(SP '07). Berkeley, CA, United states, 2007: 321-334

[11] Cheung L, Newport C. Provably secure ciphertext policy ABE [C]// 14th ACM Conference on Computer and Communications

Security. Alexandria, Virginia, USA: ACM, 2007: 456-465

[12] Khader D. Attribute based search in encrypted data[C]// 2014 ACM Workshop on Information Sharing and Colla. Scottsdale, AZ, United States: ACM, 2014: 31-40

[13] Chase M, Chow S M. Improving privacy and security in multi-authority attribute-based encryption[C]// 16th ACM Conference on Computer and Communications. Chicago, IL, United States: ACM, 2009: 121-130

[14] De Caro A, Vincenzo I. Jpbc: Java pairing based cryptography [C]// 2011 IEEE Symposium on Computer and Communications. Kerkyra, Corfu, Greece, 2011: 850-855

(上接第 166 页)

参 考 文 献

[1] Zang Tian-ning, Yun Xiao-chun, Zhang Yong-zheng. A Model of Network Device Corrdinative Run[J]. Chinese Journal of Computers, 2011, 34(2), 216-228(in Chinese)
臧天宁, 云晓春, 张永铮. 网络设备协同联动模型[J]. 计算机学报, 2011, 34(2): 216-228

[2] Wang B Y, Yang H P, Zhang S M. Research on application of interaction firewall with IDS in distribution automation system [M]// Advances in Electronic Engineering, Connumication and management Vol 1. Springer, 2012: 527-532

[3] Zheng Li-ming, Zou Peng, Jia Yan. How to Extract and Train the Classifier in Traffic Anomaly Detection System[J]. Chinese Journal of Computers, 2012, 35(4): 719-729(in Chinese)
郑黎明, 邹鹏, 贾焰. 网络流量异常检测中分类器的提取与训练方法研究[J]. 计算机学报, 2012, 35(4): 719-729

[4] Ahmed M, Mahmood A N. Novel Approach for Network Traffic Pattern Analysis using Clustering-based Collective Anomaly Detection[J]. Annals of Data Science, Springer, 2015, 2(1): 111-130

[5] Zhang Bin, Yang Jia-hai, Wu Jian-ping. Survey and Analysis on the Internet Traffic Model[J]. Journal of Software, 2011, 22(1): 115-131(in Chinese)
张宾, 杨家海, 吴建平. Internet 流量模型分析与评述[J]. 软件学报, 2011, 22(1): 115-131

[6] Alotibi G, Li F, Clarke N. Behavioral-Based Feature Abstraction from Network Traffic[C]// ICCWS 2015. 2015: 176-188

[7] Yang Xin-yu, Yang Shu-sen, Li Juan. A Flooding-Based DDoS Detection Algorithm Based on Non-Linear Preprocessing Network Traffic Predicted Method[J]. Chinese Journal of Computers, 2011, 34(2): 395-405(in Chinese)
杨新宇, 杨树森, 李娟. 基于非线性预处理网络流量预测方法的泛洪型 DDoS 攻击检测算法[J]. 计算机学报, 2011, 34(2): 395-405

[8] Hofleitner A, Herring R, Abbeel P. Learning the dynamics of arterial traffic from probe data using a dynamic Bayesian network[J]. IEEE Transactions on Intelligent Transportation System, 2012, 13(4): 1679-1693

[9] Wei Xiong, Hu Han-ping, Laurence T. Anomaly secure detection methods by analyzing dynamic characteristics of the network

traffic in cloud communications[J]. Information Sciences, 2014 (258): 403-415

[10] Yang Yue. Network Traffic Anomaly Detection Method Based on a Feature of Catastrophe Theory[J]. CHIN. PHYS. LETT., 2010, 27(6): 116-124

[11] Lin Jian-ren, Yang Xiao-long, Long Ke-ping, et al. Catastrophe Model Construction and Verification for Network Anomaly Detection[C]// SPIE Proceedings. Vol. 7137, 2008: 70-81

[12] Wei Xiong, Nai Xue-xiong, Laurence T. Yang. Network Traffic Anomaly Detection based on Catastrophe Theory[C]// 2010 IEEE GLOBECOM Workshops. 2010: 2070-2074

[13] Gu J, Chen S. Nonlinear Analysis on Traffic Flow Based on Catastrophe and Chaos Theory[J]. Discrete Dynamics in Nature and Society, 2014, 23(3): 253-264

[14] Yang Xiao-long, Zhang Min, Hu Wu-sheng. IP Network Anomalous Behaviors Detection Mechanism[J]. Journal of University of Electronic Science and Technology of China, 2011, 40(6): 892-897(in Chinese)
阳小龙, 张敏, 胡武生. 基于尖点突变模型的 IP 网络异常行为检测方法[J]. 电子科技大学学报, 2011, 40(6): 892-897

[15] Zhang Xian-jiang, Liu Xiao-qiang. Nonlinear Network Traffic Prediction Model Based on Parameters Joint Optimization[J]. Computer Engineering and Application, 2014, 50(6): 64-67(in Chinese)
张显江, 刘小强. 一种参数联合优化的网络流量非线性预测模型[J]. 计算机工程与应用, 2014, 50(6): 64-67

[16] Wen Xiang-xi, Meng Xiang-ru, Ma Zhi-qiang. The Chaotic Analysis and Trend Prediction on Small-Time Scale Network Traffic [J]. ACTA Electronica Sinica, 2012, 40(8): 1609-1616(in Chinese)
温祥西, 孟相如, 马志强. 小时间尺度网络流量混沌性分析及趋势预测[J]. 电子学报, 2012, 40(8): 1609-1616

[17] Kane J, Lawrence J, Farnon M. Analysis of network traffic: 883870[P].

[18] 胡晓洁. 正态分布及其扩展综述[J]. 数学学习与研究, 2014(3): 92-94

[19] Kolbusz J, Rozycki P, Korniak J. The Simulation of Malicious Traffic Using Self-similar Traffic Model[M]// Human-Computer Systems Interaction: Background and Applications 2. Springer, 2012: 327-341