

基于混沌理论的网络流量性能评估

杨 谈 崔毅东 金跃辉 程时端

(北京邮电大学网络与交换技术国家重点实验室 北京 100876)

摘 要: 该文基于混沌理论提出了一种使用海量网络流量数据对大规模网络性能进行有效评估的方法。在长期链路利用率数据呈现出明显的周期性行为, 和短期链路利用率数据具有混沌特征的前提下, 选取最大 Lyapunov 指数作为一项性能评估参数来评估网络性能。分析结果表明最大 Lyapunov 指数较常见统计量如数学期望、方差等更能有效反映流量的行为趋势。

关键词: 网络性能评估; 混沌; 链路利用率; 最大 Lyapunov 指数

中图分类号: TP393

文献标识码: A

文章编号: 1009-5896(2009)12-2993-04

Performance Evaluation of Network Traffic Based on Chaos Theory

Yang Tan Cui Yi-dong Jin Yue-hui Cheng Shi-duan

(State Key Laboratory of Networking and Switching Technology,

Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: A method based on chaos theory is presented in this paper, to evaluate large-scale network performance using massive traffic measurement. As the periodicity of long-term link utilization measurement and the chaotic nature of short-term link utilization measurement, the largest Lyapunov exponent can be selected as a performance evaluation parameter to represent the performance of network. Analysis results show that the largest Lyapunov exponent can achieves better results than commonly used statistics, such as mathematical expectation and variance.

Key words: Performance evaluation of network; Chaos; Link utilization; Largest Lyapunov exponent

1 引言

如何使用海量的网络流量数据对网络性能进行有效评估是网络性能研究中有待解决的一个问题。常用的数学期望、方差等统计值不足以全面有效的描述和量化网络特征。例如, IP 网络中流量和链路利用率参数信息是网络运维的主要参考信息。目前网络运维人员普遍采用将 QoS 与平均链路利用率相关联的方式, 若平均链路利用率未超出预先设定的阈值(如 50%), 则认为网络的 QoS 得到保证。总的来说, 预先设定平均链路利用率阈值的方式不能真实反映用户流量的实际 QoS 需求; 同时, 链路利用率的平均值和方差等仅能有效反映较小时间粒度内的数值的统计特征, 不能反映链路利用率的变化趋势和行为特征, 进而对整网进行有效性能评估。

Internet 作为一个远离平衡态的开放系统, 实质上是一个耗散系统。网络环境下的海量流量信息组成的动力系统呈现了丰富的非线性动力学特征。

Lelad 等^[1]首次发现以太网上流量具有自相似特征; Internet 中 WWW, TCP 拥塞控制和文件大小重尾分布等是造成流量自相似和长相关的主要原因。网络流量的混沌研究与传统的网络流量自相似研究紧密相关, 从本质上来说, 混沌是产生时空结构的物质非线性运动, 而这种具有自相似特性的结构本身就是分形。

已有研究表明大规模网络流量呈现明显的周期性行为。而文献[2,3] 通过计算短时间粒度内的离散流量时间序列的最大 Lyapunov 值, 对 Internet 流量的混沌特性进行了定量分析, 但其中只涉及了对单个网络接口的出/入流量的混沌特征的讨论。结合以上两个方面, 本文中选取最大 Lyapunov 指数作为刻画流量性能的参数, 提出了一种评估大规模网络在较长时间内的流量性能趋势方法。首先, 分析流量的周期性行为。然后计算每个周期内链路利用率数据组成的离散时间序列的最大 Lyapunov 指数, 根据计算结果得到表征单条链路或多条链路乃至全网流量趋势的性能向量, 在此基础上对部分网络区域乃至整网进行评估。

2008-12-08 收到, 2009-05-07 改回

国家 863 计划项目(2006AA01Z235, 2007AA01Z206, 2009AA01Z210)
资助课题

2 基于混沌理论的网络流量性能评估方法

2.1 混沌理论介绍

混沌是指确定性非线性动力复杂系统中存在着内在随机性的现象。如果一个系统表现出非线性、确定性、无序中的有序、“蝴蝶效应”等特征时，称该系统是混沌的。混沌系统中在其相空间(即状态空间)中经过一定时间的演化后，最终会产生规则有形的运动轨迹，称之为混沌吸引子。混沌吸引子是系统状态点在相空间中的集合，它具有固定的复杂几何结构，经过拉伸扭曲操作后，在时序上表现出无序和自相似的特点。

“蝴蝶效应”即系统演化轨迹的初值敏感性。混沌系统中，两个初值很靠近的轨道会因为细小的差别随时间推移按指数速度分离。而 Lyapunov 指数谱 λ 正是定量描述这一现象的量，它计算了系统在多个状态子空间中每次迭代所引起的指数分离中的指数的平均值。有

$$\lambda = \lim_{t \rightarrow \infty} \frac{1}{t} \ln \frac{|\delta Z(t)|}{|\delta Z(0)|} \quad (1)$$

其中 $\delta Z(t)$ 表示 t 时刻的偏离量， $|\cdot|$ 表示向量的模。 λ 中的最大值即被称为最大 Lyapunov 指数。为简单起见，下文中的 λ 均指最大 Lyapunov 指数。

$\lambda > 0$ 是判断系统为混沌系统的最常用方法之一。 λ 值的估算是在重构的相空间中进行的。相空间重构是指把系统的观测值(通常是时间序列)嵌入到高维的状态空间中，在状态空间中获取该动力系统的特征。相空间重构需要使用两个重要的参数，时间延迟 τ 和嵌入维数 m 。下文中将提到单变量时间序列和多变量时间序列相空间重构的区别。

2.2 性能评估问题描述

H 条链路在 N 个时间间隔内的链路利用率统计值组成 H 维时间序列 $\mathbf{L} = \{\mathbf{U}(i), i = 1, 2, \dots, N\}$ 。若 $H = 1$ ，则 \mathbf{L} 是单变量时间序列， $\mathbf{U}(i)$ 是单条链路上第 i 个时间段的测量值；若 $H > 1$ ，则 \mathbf{L} 是多变量时间序列， $\mathbf{U}(i)$ 是多条链路上第 i 个时间段的测量值，即 $\mathbf{U}(i) = \{l_1(i), l_2(i), \dots, l_H(i)\} (i = 1, 2, \dots, H)$ 。

将 \mathbf{L} 按 k 个时间段划分为 M 个子序列，即 $M = \lfloor N/k \rfloor$ ，得

$$\left. \begin{aligned} \mathbf{L}_1 &= \{\mathbf{U}(1), \mathbf{U}(2), \dots, \mathbf{U}(k)\}, \\ \mathbf{L}_2 &= \{\mathbf{U}(k+1), \mathbf{U}(k+2), \dots, \mathbf{U}(2k)\} \dots \\ \mathbf{L}_j &= \{\mathbf{U}[k(j-1)+1], \mathbf{U}[k(j-1)+2], \dots, \mathbf{U}(kj)\} \dots \\ \mathbf{L}_M &= \{\mathbf{U}[k(M-1)+1], \mathbf{U}[k(M-1)+2], \dots, \mathbf{U}(N)\} \end{aligned} \right\} \quad (2)$$

将每个子序列 $\mathbf{L}_j (j = 1, 2, \dots, M)$ 映射到性能评估值 $\psi_j (j = 1, 2, \dots, M)$ ，即

$$\psi_j = f(\mathbf{L}_j) \quad (3)$$

最后得到性能评估向量 $\bar{\psi} = \{\psi_1, \psi_2, \dots, \psi_M\}$ 。

需要解决的两个问题包括：(1)确定子序列的粒度，即 k 值的大小；(2)构造性能映射函数 $f(\cdot)$ ，以及根据得到的向量 $\bar{\psi}$ 进行评估。

2.3 基于混沌理论的网络流量性能评估方法

首先，需要解决子序列粒度 k 的取值问题。已有研究表明网络流量存在着周期性行为，故本文中选取流量周期长度为 k 的取值。同时，由于短期网络流量存在着混沌特性，而最大 Lyapunov 指数量化了系统中状态点之间的变化速度，它表征了混沌运动的基本特点，故选取最大 Lyapunov 指数 λ 作为性能评估值 ψ_j 。离散时间序列 \mathbf{L} 的性能评估方法如下：

(1) 使用快速傅里叶变换 (Fast Fourier Transform, FFT) 分别估计离散时间序列 \mathbf{L} 的 H 个子序列的功率谱，得该子序列的周期长度 $k_i (i = 1, 2, \dots, H)$ ，最后取 $k = \min\{k_i\}$ ；

(2) 根据 k 的长度将 \mathbf{L} 分为 M 个子序列， $M = \lfloor N/k \rfloor$ 。然后计算每个子序列的最大 Lyapunov 指数。最大 Lyapunov 指数的计算首先需要重构相空间^[4]。以子序列 \mathbf{L}_j 为例， $\mathbf{L}_j = \{\mathbf{U}(t), t = k(j-1)+1, k(j-1)+2, \dots, kj\}$ 。

若 \mathbf{L}_j 是单变量时间序列 ($H = 1$)，则使用常见的虚假最近邻点法^[5]和互信息法^[6]分别确定最小嵌入维数 m 和时间延迟 τ ，将 \mathbf{L}_j 重构为 m 维状态向量

$$\{l(t), l(t-\tau), \dots, l(t-(m-1)\tau)\} \in \mathbf{R}^m \quad (4)$$

其中 $t = k(j-1) + (m-1)\tau + 1, k(j-1) + (m-1)\tau + 2, \dots, kj$ 。

若 \mathbf{L}_j 是多变量时间序列，重构相空间时需要确保子变量间松耦合。通过计算子变量数据的 Pearson 相关系数，根据计算结果抽取出 H' ($H' \leq H$) 个松耦合的子变量，组成新的 H' 维的时间序列 \mathbf{L}'_j 。然后使用虚假最近邻点法和互信息法分别计算 \mathbf{L}'_j 中各个子变量序列的最小嵌入维数 $m_i (i = 1, 2, \dots, H')$ 和时间延迟 $\tau_i (i = 1, 2, \dots, H')$ ，重构相空间^[7]为

$$\left. \begin{aligned} & \{l_1(t), l_1(t-\tau_1), \dots, l_1[t-(m_1-1)\tau_1], \\ & l_2(t), l_2(t-\tau_2), \dots, l_2[t-(m_2-1)\tau_2], \dots, \in \mathbf{R}^{m_1+m_2+\dots+m_{H'}} \\ & l_{H'}(t), l_{H'}(t-\tau_{H'}), \dots, l_{H'}[t-(m_{H'}-1)\tau_{H'}] \end{aligned} \right\} \quad (5)$$

其中 $t = k(j-1) + 1 + N_0, k(j-1) + 2 + N_0, \dots, kj$ ， $N_0 = \max_{1 \leq i \leq H'} \{(m_i-1)\tau_i + 1\}$ 。

(3) 相空间重构后，选取常见的 Wolf 算法^[8]分别计算 M 个重构序列的最大 Lyapunov 指数，最终得性能评估向量 $\bar{\psi} = \{\psi_1, \psi_2, \dots, \psi_M\}$ 。

3 方法验证

3.1 Abilene 网络性能评估

本文中使用的 Abilene 网络上链路利用率数据^[9]对该评估方法进行验证。Abilene 网络是 Internet2 的高性能主干网络，由 11 个结点和 30 条链路组成，其拓扑结构如图 1 所示。该数据包括了 2004 年中持续 24 周的测量值，其中每条链路每 5 min 对应一个测量值，一共包括了 $288 \times 7 \times 24 = 48,384$ 个时间段的测量值。这些测量值从空间和时序的角度呈现了网络中流量的性能趋势。

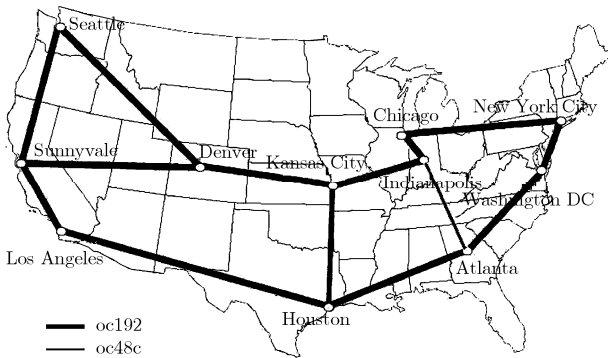


图 1 Abilene 网络拓扑图

首先，分析链路上流量的周期特征。以 Washington DC 到 Atlanta 的链路利用率数据为例，使用 FFT 变换得其功率谱如图 2 所示。从图中可以看到，频率 168 和 337 处有明显峰值，根据频率和周期的倒数关系，有 $48,384 \times 5 \text{ min} / 168 = 60 \text{ min} \times 24$ 以及 $48,384 \times 5 \text{ min} / 336 = 60 \text{ min} \times 12$ ，即链路利用率数据呈现明显的 24 h 和 12 h 周期特征，24 h 周期性尤其明显。经过分析，其他链路与该链路也有类似特征，故选取 $k=12 \times 24=288$ ，即将链路利用率数据按天分成多个子序列分别计算其最大 Lyapunov 指数。

然后，根据待评估网络区域的链路利用率测量

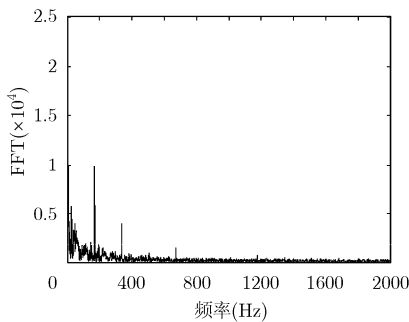


图 2 链路利用率数据功率谱

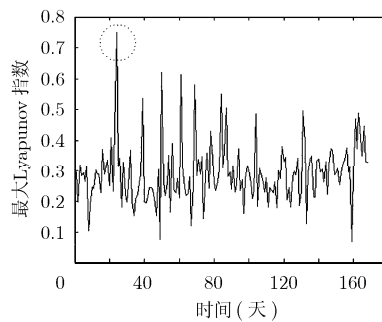


图 3 最大 Lyapunov 指数反映的网络性能趋势

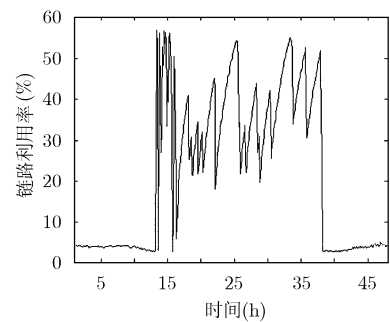


图 4 Indianapolis 到 Kansas City 链路利用率

值组成的单变量或多变量时间序列，对区域性能趋势进行评估。

以整个 Abilene 网络为例，计算其性能评估向量 $\bar{\psi}$ 。全网的链路利用率测量值组成一个多变量时间序列。首先，通过计算 Pearson 相关系数，发现 Los Angeles 到 Chicago、Chicago 到 Log Angeles, Chicago 到 Washington DC, Los Angeles 到 Washington DC 以及 Washington DC 到 Houston 总共 5 条路径上的链路的测量数据分别强相关。故从这 5 条路径上分别挑选 1 条链路，和其余链路数据组成一个新的多变量时间序列。分别计算得到以天为单位的 168 个子序列的最大 Lyapunov 指数，其结果均为正，说明 Abilene 网络的流量数据组成的多变量时间序列呈现以天为单位的混沌特征。计算得到的 168 个最大 Lyapunov 指数组成了性能评估向量 $\bar{\psi}$ ，如图 3 所示。从图中可以看到， $\bar{\psi}$ 中的值大部分落在 0.002 和 0.004 之间，反映了 Abilene 网络的性能的大致稳定性；但序列中的一些明显抖动也反映了其相应时间区间内的流量事件，如存在明显流量突发性等。例如，图 3 中第 23 天(2004 年 4 月 10 日，星期六)和 24 天(2004 年 4 月 11 日，星期日)对应了两个明显的峰值，通过对原始测量值的分析，发现从第 23 天的下午到第 24 天的下午，受大流影响，Chicago 到 Los Angeles 路径上链路利用率在小于 10% 和大于 50% 的区间内长时间波动，该路径上 Indianapolis 到 Kansas City 链路在这两条内的测量值如图 4 所示。

3.2 与常见统计量比较

常见的统计量如数学期望、方差等，能反映测量值在较短时间粒度内的性能状况，而最大 Lyapunov 指数由于刻画了系统状态的持续变化趋势，故能更好地反映其在较长时间粒度内的性能变化。

以 2004 年 6 月 21 日 Washington DC 到 Atlanta

链路上的测量值为例,如图 5 所示,可以看到 21:00 到 22:00 之间有一个明显的峰值,该峰值导致该天数据的方差计算结果比平时明显增大,而最大 Lyapunov 指数的计算结果和平时并没有明显区别。

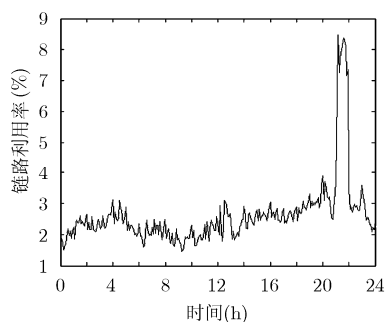


图 5 2004 年 6 月 21 日 Washington DC 到 Atlanta 链路利用率

又以 2004 年 9 月 10 日该链路上的测量值为例,如图 6 所示,可以看到该天内流量波动较大,而该天数据的数学期望计算结果和平时并没有明显区别,而最大 Lyapunov 指数的计算结果比平时明显增大,较好地反映了链路上流量的波动情况。

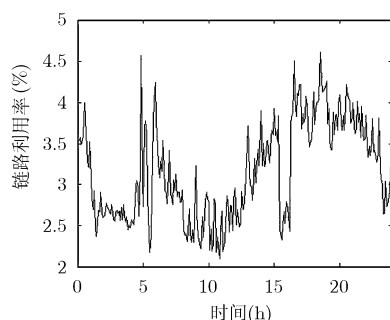


图 6 2004 年 9 月 10 日 Washington DC 到 Atlanta 链路利用率

4 结论

本文在链路利用率数据呈现明显的周期性行为,同时单个周期内链路利用率数据组成的离散时间序列呈现明显的混沌特征的前提下,选取最大 Lyapunov 指数作为一项性能评估参数来评估网络性能。通过对实际网络数据进行分析,验证了该基于混沌理论的网络流量性能评估方法的有效性,并且最大 Lyapunov 指数较常见统计量如数学期望、方差等更能有效刻画流量的行为趋势。

参考文献

- [1] Leland W E, Taqqu M S, and Willinger W, *et al.* On the self-similar nature of Ethernet traffic. ACM Sigcomm'93, San Francisco, CA, USA, 1993: 183-193.
- [2] 徐喆, 王培荣, 付冲等. Internet 流量的混沌动力学模型研究. 通信学报, 2006, 27(11A): 199-203.
Xu Z, Wang P R, and Fu C, *et al.* Chaotic behavior decision of complex system and fractal based chaotic time series modeling. *Journal on Communication*, 2006, 27(11A): 199-203.
- [3] 陆锦军, 王执铨. 一种基于混沌特性的网络流量改进预测算法. 兵工学报, 2007, 28(11): 1346-1350.
Lu J J and Wang Z Q. An improved prediction method of network traffic flow base on chaos characteristics. *Acta Armamentarii*, 2007, 28(11): 1346-1350.
- [4] 秦奕青, 蔡卫, 杨炳儒. 非线性时间序列的相空间重构技术研究. 系统仿真学报, 2008, 20(11): 2969-2973.
Qin Y Q, Cai W D, and Yang B R. Research on phase space reconstruction of nonlinear time series. *Journal of System Simulation*, 2008, 20(11): 2969-2973.
- [5] Kennel M, Brown R, and Abarbanel H. Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Physical Review A*, 1992, 45: 3403-3411.
- [6] Fraser A. Information and entropy in strange attractors. *IEEE Transactions on Information Theory*, 1989, 35(2): 245-262.
- [7] 黎敏, 徐金梧, 阳建宏等. 基于多变量相重构的混沌时间序列预测. 北京科技大学学报, 2008, 30(11): 208-212.
Li M, Xu J W, and Yang J H, *et al.* Prediction for chaotic time series based on phase reconstruction of multivariate time series. *Journal of University of Science and Technology Beijing*, 2008, 30(11): 208-212.
- [8] Wolf A, Swift J, and Swinney H, *et al.* Determining Lyapunov exponents from a time series. *Physica D*, 1985, 16: 285-317.
- [9] Zhang Y. Abilene traffic metrics. <http://www.cs.utexas.edu/~yzhang/research/AbileneTM/>, 2004.

杨 谈: 女, 1982 年生, 博士生, 研究方向为网络测量与性能评估。

崔毅东: 男, 1976 年生, 讲师, 研究方向为网络测量与性能评估。

金跃辉: 女, 1965 年生, 副教授, 研究方向为网络测量技术与体系结构、IP 网络服务质量与性能评估体系。