

基于免疫和灰色系统理论的网络认知研究

王 伟¹, 王 辉², 张 潇¹

(1 河南科技大学 电子信息工程学院, 河南 洛阳 471003)

2 河南科技大学 现代教育技术与信息中心, 河南 洛阳 471003)

摘 要: 针对传统 TCP 控制协议在异构网络中效率低下问题, 提出一种基于免疫和灰色系统理论的网络认知研究. 该研究利用人工免疫系统分析各参数的特性和利用灰色关联度建立该网络各参数与当前网络性能之间的相关性, 通过一定的数据分析, 对网络进行准确、高效的认知. 该研究采用传输控制协议 TCPL 进行仿真实验. 仿真实验证明: 该算法能够在不同背景流的条件下稳定发送窗口, 减少拥塞发生并提高吞吐量, 改善了网络性能.

关键词: 异构网络; 人工免疫系统; 灰色系统理论; 瓶颈链路队列长度; 网络认知

中图分类号: TN915.04

文献标识码: A

文章编号: 1000-7180(2013)06-0117-05

Research of Network Cognitive Based on Immune and Gray System Theory

WANG Wei¹, WANG Hui², ZHANG Xiao¹

(1 School of Electronic Information Engineering, Henan University of Science and Technology, Luoyang 471003, China;

2 Modern Educational Technology and Information Center, Henan University of Science and Technology, Luoyang 471003, China)

Abstract: In view of the traditional TCP control protocol in low efficiency in heterogeneous network, research of network cognitive based on immune and gray system theory is proposed. The research utilizes artificial immune system analysis parameters characteristics and the use of grey correlation to build the performance of the correlation between network parameters and the current network, through some data analysis to accurate and efficient cognition network. This research uses Transmission Control Protocol (TCPL) to finish simulation experiments. Simulation show: Under different background flow conditions, the algorithm can stabilize sending window, reduce congestion and improve throughput and transmission efficiency.

Key words: heterogeneous network; artificial immune system; gray system theory; the bottleneck link queue's length; network cognitive

1 引言

为了提高异构网络传输性能, 目前有三种基本的解决方案^[1-5]. 链路层解决方法虽然具有灵活的操作手段, 但由于该层协议与高层协议都有独立的差错控制功能, 有一定的重复性, 相互竞争会降低无线信道的利用率. 分段连接在基站处加入代理功能, 虽然对 TCP 发送方完全屏蔽了无线链路, 但该方法需要基站大量的缓存和处理能力. 端到端方法一般是基于 TCP 协议的改进, TCP-Reno、TCP-Vegas 以

及 TCP-Westwood 等都是比较常用的典型代表, 然而这类协议都比较片面的针对某一个参数或一种情况做出了改进, 并不能够全面和准确地解决异构问题.

目前方法的关键在于如何清楚正确地对异构网络状态进行认知. 在传输层, 对当前网络的真实状态进行认知并不简单, 需要根据参数的特性, 分别分析该参数对网络真实状态反应的程度, 即得到其相关性, 然后综合各参数的相关性结合其当前状态, 最终得到网络当前的状态.

收稿日期: 2012-08-30; 修回日期: 2012-10-22

基金项目: 国家自然科学基金(61070247); 河南省教育厅自然科学研究计划项目(2010A520017); 河南省科技攻关项目(112102210186)

该文将目标数据流在瓶颈链路的队列长度量化为当前网络的真实状态,该数据为一个模糊集合,同样提取的参数也量化为一个模糊集合,在计算两个模糊集的相关性中,提出使用灰色系统理论的方法来解决该问题.该方法易于计算,得到各参数的相关性准确,对当前异构网络状态有较好的认知.通过对网络状态认知建立的数学模型,从而达到对网络流量控制的目的.

2 网络参数选择

免疫系统^[6]是一个分布式、自组织和具有动态平衡能力的自适应复杂系统.它对外界入侵的抗原,可有分布全身的不同种类的淋巴系统产生相应的抗体,其目标是尽可能保证整个生物系统的基本生理功能得到正常运转.人工免疫系统就是研究、借鉴、利用生物免疫系统的原理、机制而发展起来的各种信息处理技术、计算技术及其在工程和科学中的应用而产生的多种智能系统的统称.免疫系统作为中枢神经系统的感受器官,感知体内环境的化学和生物性的动态变化,该文拟利用人工免疫系统的特性来感知网络内部的参数变化和网络状态,提取网络参数并分析各参数的特性.

2.1 均值抖动积 MDP

均值抖动积 MDP 是基于传输往返时延(Round-trip time, RTT)计算得出的. RTT 是网络传输控制的一个最常用的参数,它反映了数据包从发送到接收 ACK 所用到的总体时间.它对网络传输过程中因为瓶颈链路而导致的拥塞丢包事件反映非常敏感,但是由于 TCP 应答机制(接收方无论是否受到一个数据包都会定时发送数据包应答 ACK) RTT 值无法确切反映出因为外界干扰而产生的丢包事件.同时由于无线网络的传输环境复杂性, RTT 值的抖动幅度比较大,干扰了实际网络传输情况的判断,即一次 RTT 绝对值的突然变化不能说明网络必然出现拥塞状况,有国外学者基于此认为

RTT 不能明确反映网络传输状态.因此我们提出基于往返时延 RTT 的参数均值抖动积 MDP,旨在克服这些缺陷.

$$MDP = \theta \times e^{\delta} \quad (1)$$

式中, θ 表示的是 RTT 的短期均值, δ 表示的是短期标准差. a 是调节系数,调节 RTT 抖动因素占的贡献比例.

2.2 短期吞吐量 STT

短期吞吐量是基于拥塞窗口(Congestion Window, cwnd) 计算得出的.拥塞窗口反映了发送方发送数据包的速率.它可以间接反映网络实时传输速率,如果没有发生拥塞,则传输速率保持平稳或者稳步增长.我们提出的短期吞吐量概念是为了能敏感反映网络实时传输特征,定义如下:

$$STT = \frac{(\sum_{i=1}^T cwnd_i) \times packet_size}{T} \quad (2)$$

式中, $cwnd_i$ 表示拥塞窗口在为每个数据包的大小, $packet_size$ 为每个数据包的大小, T 为统计周期.

2.3 瓶颈链路队列长队的估计值 L

瓶颈链路:对于一个数据流传输路径,在整个网络系统中,最易产生拥塞的节点和下一段剩余带宽最小的链路.一般为多个数据流交汇的小缓存节点和下一段窄带宽链路.

L 变量为瓶颈链路队列长队的估计值:瓶颈链路中节点缓冲区数据包个数和下一段链路数据包个数的总和.

$$L = L_1 + L_2 - P,$$

式中, L_1 为瓶颈节点缓冲区数据包个数, L_2 为瓶颈节点下一段链路数据包个数.定义 C_1 为节点缓冲区大小,故有 $L_1 \leq C_1$;定义 C_2 为下一段链路的时延带宽积,则下一段链路中最大数据包个数由这一段的 C_2 决定,故有 $L_2 \leq C_2$.这里将 C_2 换算成数据包个数,单位为 packets, P 为瓶颈链路丢失包数,包括误码丢包和拥塞丢包.

$$L_1(t) = \begin{cases} \sum_{i=1}^{i=n} \int_{t_i}^{t_{i+1}} (\frac{SentData}{ActualRTT} + g - TR) dt & L_1 + \Delta L_1 > 0 \\ 0 & L_1 + \Delta L_1 \leq 0 \end{cases}$$

$$= \begin{cases} \sum_{i=1}^{i=n} \frac{(\frac{SentData}{ActualRTT} + g)}{n} - TR \times t + cons1 & L_1 + \Delta L_1 > 0 \\ 0 & L_1 + \Delta L_1 \leq 0 \end{cases}$$

$$L_2(t) = \begin{cases} \sum_{i=1}^{i=n} \int_{t_i}^{t_{i+1}} (\frac{SentData}{ActualRTT} + g - \frac{L_2(t)}{basetime}) dt & L_1 = 0 \\ \int_{t_1}^{t_n} (TR - \frac{L_2(t)}{basetime}) dt & L_1 > 0 \end{cases}$$

$$= \begin{cases} cons2 \times \sum_{i=2}^n e^{\frac{-(t_i - t_{i-1})}{basetime}} + basetime \times ((\frac{SentData}{actualRTT} + g)_n - (\frac{SentData}{actualRTT} + g)_1) & L_1 = 0 \\ cons3 \times e^{\frac{-t}{basetime}} + basetime \times TR & L_1 > 0 \end{cases}$$

式中, $SentData$ 为一个 RTT 周期内发送的数据包个数, TR 为瓶颈链路节点最大包转发速率, $basetime$ 为瓶颈链路最小时延, $ActualRTT$ 为实时的 RTT, g 为背景流, 换算成数据包个数. $\Delta L_1, \Delta L_2$ 为 Δt 时间内 L_1 和 L_2 的改变量. $cons1, cons3$ 为固定常数, 由网络的初始状态决定, 可以通过带入初始值通过分析瓶颈链路队列长度的真实值反推出来. 事实上, 当网络链路初试状态为空时, $cons1 = 0 = cons3 = -1$.

假设丢包概率稳定不变, 值为 p , 则 t 时刻丢包数:

$$P = \int_{t_1}^{t_n} p dt = pt$$

3 灰色关联性分析

因素序列集: 每一个 X_i 因素为对网络拥塞进行预测的参数, 例如 MDP , 实时的 MDP 序列构成的一个时间序列. 选取参数构成因素集合为

$$X^{(0)} = (x_1(0), x_2(0), x_3(0))$$

式中, $X_1^{(0)} = MDP$, 为均值抖动积; $X_2^{(0)} = STT$, 为短期吞吐量; $X_3^{(0)} = L$, 为瓶颈链路队列长度的估计值.

目标序列: 所有正确传输的数据包都要经过瓶颈链路, 所以瓶颈链路队列长度能够实时地反映一个 TCP 流的拥塞情况, 也就是说可以选取瓶颈链路队列长度作为观察拥塞程度等性能的目标量^[7], 瓶颈链路队列长度真实值 $X_0^{(0)}$.

灰色关联度计算步骤如下:

(1) 求各序列的初值像:

$$X_1^{(1)} = \frac{X^{(0)}}{X_i^{(0)}(1)}; i = 0, 1, 2, 3;$$

(2) 求差序列:

$$\Delta_i(k) = |x_0^{(1)}(k) - x_i^{(1)}(k)|; i = 1, 2, 3; k = 1, 2, \dots, n;$$

(3) 计算两值最大差与最小差:

$$\begin{cases} M = \max_i \max_k \Delta_i(k) \\ M = \min_i \min_k \Delta_i(k) \end{cases}; i = 1, 2, 3; k = 1, 2, \dots, n$$

(4) 求关联系数:

$$\gamma_i(k) = \frac{m + \rho M}{\Delta_i(k) + \rho M}; i = 1, 2, 3; k = 1, 2, \dots, n$$

(5) 计算关联度:

$$R(i) = \frac{1}{n} \sum_{k=1}^n \gamma_i(k); i = 1, 2, 3$$

当网络传输稳定后, 提取相应的网络参数, 运用上述灰色关联度的方法分别计算出它们与瓶颈链路队列长度真实值的关联度. 表 1 为 MDP 、 STT 、 L 与瓶颈链路队列长度真实值在不同链路时延情况下贴近度比较.

表 1 贴近度比较

因素 \ 延时	MDP	STT	L
10	0.771 3	0.813 5	0.902 2
30	0.752 5	0.794 5	0.881 6
50	0.736 8	0.775 6	0.864 5

从上表中可以看出在不同的网络环境下, L 对瓶颈链路队列长度真实值的关联度最高, 也即 L 对网络性能的反映程度最大.

4 仿真实验

通过上述灰色关联度分析可知瓶颈链路队列长度的估计值 L 对网络性能的反映程度最大, 因此利用瓶颈链路队列长度的估计值 L 给出的数学模型, 对网络进行定量的分析.

当有 ACK 到达时执行以下四步:

(1) 执行 $L_1(t)$;

(2) 执行 $L_2(t)$;

(3) $p = \begin{cases} 1 & \text{收到三个重复 ACK} \\ 0 & \text{其他情况} \end{cases}$;

(4) 执行 $L = L_1 + L_2 - P$;

为了验证数学模型有效性, 本文在传统 TCP 协议的基础上, 提出了一种新的适用于异构网络的网络传输控制协议 TCPL, 该协议在传统 TCP 协议中加入上述描述的瓶颈链路队列长度估计值数学模型

进行仿真,并与其他算法进行吞吐量对比.其核心思想为:尽可能使得队列长度 L 保持在 $[a, b]$ 区间内,当队列长度小于 a 时,增大发送速率;当队列长度大于 b 时减少发送速率.当发生丢包,查看队列长度 L ,如果队列长度 L 大于 b ,为拥塞丢包,否则为误码丢包.

在 NS2.34^[8] 上进行算法仿真验证,其拓扑结构如图 1 所示,采用哑铃式的结构,发送端为有线链路,接收端为无线链路,R-BS 为瓶颈链路,节点缓冲区最大存储包数为 10. S1-R1 建立 TCP 连接作为目标流,分别采用传统 TCP 协议和基于瓶颈链路长度控制机制的控制协议 TCPL 进行对比,在 20 s 后开始传输数据. S2-R2, S3-R3, S4-R4 建立 UDP 流,分别在 40 s, 60 s, 80 s 以稳定速率发送数据,用以提供阶梯变动的背景流.

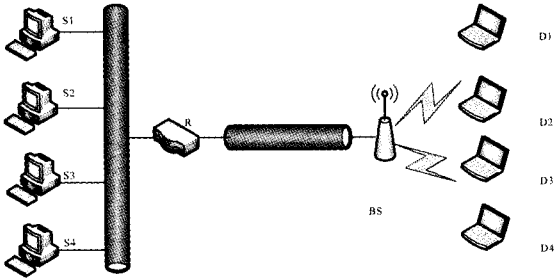


图 1 拓扑结构

根据拓扑结构和初试状态,对模型中各个参数进行初始化: $TR=250$, $g=0$, 两个门限值 $a=7$, $b=10$.

图 2 所示为上述拓扑结构中 MDP、STT、 L 与瓶颈链路队列长度真实值在不同时延情况下贴近期度比较.

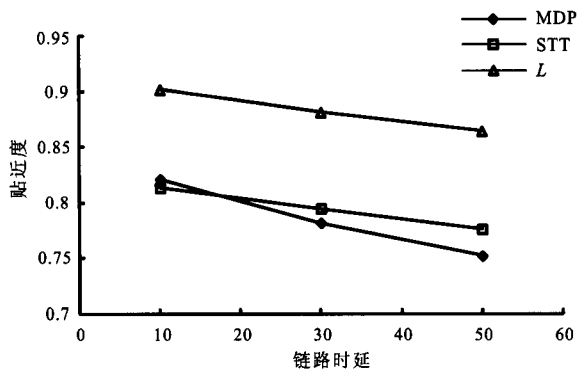


图 2 MDP、STT、 L 与瓶颈链路队列长度的真实值贴近期度

可以看出,贴近期度随链路延时的增大而减小.参数 MDP 递减得最快;STT 的贴近期度递减程度较少但贴近期度不高;反观参数 L ,不论从递减程度还是

贴近期度值,都是较好的,可以得出参数 L 在长延时的网络环境下与网络状态有着较好的相关性,能有效地反映当前网络性能.

图 3 所示为 TCP 和 TCPL 在上述拓扑结构下,吞吐量的对比.

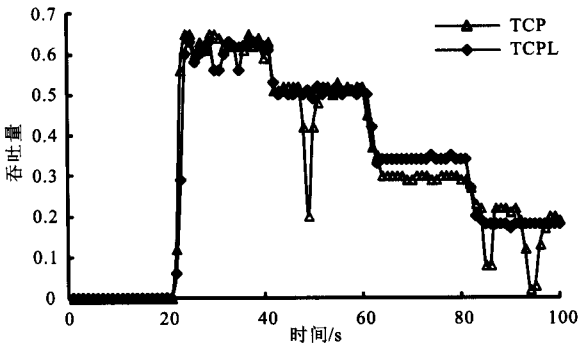


图 3 TCP 和 TCPL 协议吞吐量对比

明显可以看出,在 20~40 s 中没有背景流的情况下,它们取得了几乎相同的吞吐量,在 40~60 s 中加入了微小背景流,此时两种策略吞吐量也没有巨大差异,当出现了一次误码丢包时,由于 TCPL 策略能够区分丢包,所以仍然保持吞吐量不变,而 TCP 协议受影响较大,吞吐量减少.在 60~80 s 中再次加入背景流,TCPL 协议较 TCP 协议有明显的吞吐量提高.在 80~100 s 网络负载较重,频繁发生拥塞,TCP 协议发生剧烈抖动,而 TCPL 协议仍可以较低速率发送数据包,保持稳定吞吐量.

图 4 所示为 TCP 和 TCPL 在上述拓扑结构下,发送窗口对比.

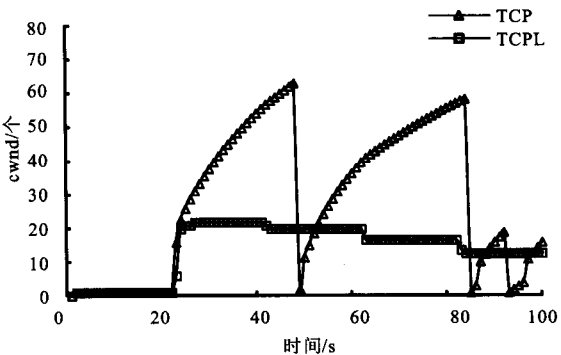


图 4 TCP 和 TCPL 协议发送窗口对比

结合图 3 明显可以发现,在网络负载较轻的 20~80 s 中,TCP 协议虽然把发送窗口提高的非常高,但是并没有提高吞吐量,反而使协议增大了产生拥塞的几率.在网络负载较重的时候,由于拥塞的频发发生,导致发送窗口频繁变化,TCPL 能够尽可能避免拥塞事件的发生,保持发送窗口和 RTT 稳定,提高了吞吐量.

5 结束语

该文应用人工免疫系统的特性来提取均值抖动积 MDP、短期吞吐量 STT、瓶颈链路队列长队估计值 L 这三个参数并利用灰色关联度建立 MDP、STT、 L 与当前网络性能瓶颈链路队列长度真实值的关联性,通过分析可知瓶颈链路队列长队估计值 L 与瓶颈链路队列长度真实值贴进度最高,也即是最能反映网络性能. 另一方面,在传统 TCP 协议的基础上,提出了一种新的适用于异构网络的网络传输控制协议 TCPL,该协议在传统 TCP 协议中加入上述描述的瓶颈链路队列长度估计值数学模型. 该算法能够稳定发送速率和瓶颈链路队列长度,在不同负荷的背景流下都能取得较良好的性能,尤其是在负载较重时性能更为突出.

参考文献:

[1] 王强,普杰信,刘伟. TCP 拥塞控制慢启动策略的研究[J]. 微电子学与计算机,2007,24(12):210-212.
[2] 王国栋,任勇毛,李俊. 多重慢启动 TCP 协议研究[J]. 微电子学与计算机,2011,28(3):178-181.
[3] 欧瑜枫,陈元琰,李鹏. TCPVegas-O:一种新的基于

延迟估计的 TCPVegas 改进算法[J]. 微电子学与计算机,2009,26(9):177-180.

[4] 叶进,王建新. 异构网络中丢包识别研究综述[J]. 计算机科学,2006,33(12):19-33.
[5] Osada S, Wang Hui, Yokohira T, et al. Throughput optimization in TCP with a performance enhancing proxy[J]. IEIC Technical Report,2006:145-150.
[6] Jerne N K. Towards a network of the immune system[J]. Annual Immunolog, 1974(125):373-389.
[7] 章森,吴建平,林闯. 互联网端到端拥塞控制研究综述[J]. 软件学报,2002,13(3):354-363.
[8] Information Sciences Institute. The ns-2 simulator [EB/OL]. [2012-08-30][2011-12-1]. <http://www.isi.edu/nsnam/ns/>.

作者简介:

王伟男,(1986—),硕士研究生. 研究方向为网络性能改善;
王辉女,(1966—),教授,硕士生导师. 研究方向为网络性能改善与数据挖掘;
张潇男,(1986—),硕士研究生. 研究方向为网络性能改善.

(上接第 116 页)

[7] 刘利枚,石彪. 一种结构化的多 Agent 任务分配机制[J]. 信息技术与信息化,2009(1):27-29.
[8] 刘淑华,张箭,吴洪岩,等. 基于群体智能的多机器人任务分配[J]. 吉林大学学报:工学版,2010, 40(1): 123-129.

作者简介:

陶雪丽女,(1974—),硕士,讲师. 研究方向为虚拟现实、多 Agent 协作.
郑延斌男,(1964—),博士,教授. 研究方向为人工智能、虚拟现实.