

工业以太网网络流量的特性分析与建模研究

刘 亮, 江汉红, 王 洁, 芮万智

(海军工程大学 电气与信息工程学院, 湖北 武汉 430033)

摘 要: 针对一个特定工业以太网系统提出了一种有效的网络流量建模思路. 首先选取系统中一条流量最大的关键链路作为研究对象, 采用 Wireshark 软件捕获流经该链路的的不同协议数据包, 并根据数据包大小和到达时间计算出每种协议的流量. 然后分别对其中主体协议(CIP 协议和 TCP 协议)的流量进行时频特性和小波特性分析, 发现 CIP 协议具有平稳的周期性, 而 TCP 协议则体现出明显的趋势特性; 针对这两种协议的不同特性, 提出采用 SVR 模型对网络流量进行建模, 利用 Matlab 中的 libsvm 工具箱进行仿真分析, 结果表明了算法的有效性.

关键词: 工业以太网; 流量建模; 短时傅里叶变换; 小波变换; 支持向量回归机

中图分类号: TP273

文献标识码: A

文章编号: 1000-7180(2015)03-0001-05

Study on Characteristics and Modeling of an Industrial Ethernet Traffic

LIU Liang, JIANG Han-hong, WANG Jie, RUI Wan-zhi

(National Key Laboratory for VIPS Technology, Naval University of Engineering, Wuhan 430033, China)

Abstract: Aiming at a practical Industrial Ethernet system, this paper proposed an effective method for Traffic Modeling. First a link of maximum traffic was selected as a typical one, all packets of different protocols passing through the link is obtained by Wireshark, and then the traffic of each protocol can be calculated according to the size and arrival time of packets. It is found that CIP traffic is of relatively smooth periodicity while TCP traffic is of obvious trend after analysis of the characteristics of main protocols traffic (mainly CIP and TCP traffic) by STFT and wavelet transform. Hence, a traffic model using SVR method is proposed due to the different characteristics of the two protocols and verified by libsvm toolbox in Matlab, and the effectiveness of the method is demonstrated by the simulation results.

Key words: industrial ethernet; traffic modeling; STFT; wavelet transform; SVR

1 引言

工业以太网主要优点在于与传统以太网兼容, 可以为大多数编程语言所支持, 适合搭建工业环境下支持多种类型控制器的工业测控网络系统. 正常稳定的工业测控网络系统, 在其运行期间必须保持合理的网络流量水平. 一旦网络流量出现长期高负荷运行, 或者网络流量出现了异常的减小, 有可能是遭到病毒攻击或系统设备发生故障等情况, 将会影响系统正常运行, 甚至导致整个系统的瘫痪^[1-2].

目前常用的网络流量建模方法有 Poisson 过程模型、Markov 过程模型和自回归过程模型等传统方法^[3], 以及分形布朗运动模型、带重尾分布的开关源叠加模型、FARIMA 模型、M/G/∞ 排队模型和小波模型等基于网络流量自相似和长相关等特性的现代模型^[4-7], 每种模型有其各自的适用范围, 如果流量模型选择与实际系统流量的特性不一致, 就无法准确地刻画网络流量的真实特性. 从理论上来说, 最直观最有效的网络流量建模方法就是机理建模. 流量机理建模是从微观上考虑每条流量的发送、传

收稿日期: 2014-06-18; 修回日期: 2014-08-11

基金项目: 国家重点基础研究发展计划(“九七三”计划)项目(2012CB215103); 国家自然科学基金项目(51377167)

输、中转和接收,通过每条流量的汇集得到网络流量,但是这种建模方式涉及到交换机的端口服务优先级、多播流量限制、时延、功能协议开销、转发方式、数据的传输编码方式、TCP 协议流控机制,以及网络上各种应用软件内部定义的通讯机制等多种不确定因素,限制条件较多,模型稳定性也较差.因此本文拟采用试验分析-假设推理-模型验证的建模方式,在通过分析网络流量特性的基础上,对网络流量进行建模.

2 获取流量

现有一工业控制系统,其网络结构为多环形拓扑结构,如图 1 所示.在系统开机运行后,各个底层控制器向状态数据服务器发送周期性的状态数据,该服务器通过状态数据判断系统是否正常,当系统正常时可以通过管理面板下达各种指令进行试验.在每次试验完成后,底层控制器向事件数据服务器发送本次试验采集到的波形,事件数据服务器通过各个底层控制器的波形来分析本次试验的结果数据,并运用该结果判断系统是否可以进行下一步试验.

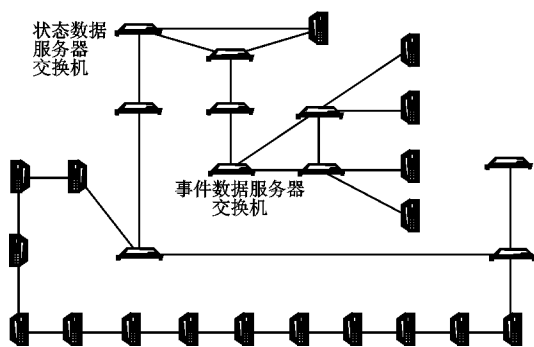


图 1 某工业控制系统的网络拓扑结构

由于所有的试验波形都通过交换机汇集到事件数据服务器,在传输试验波形的时候该服务器和交换机相连的端口数据流量最大,此外,还与底层控制器之间通过 CIP 协议交换大部分状态和参数信息,是网络主要流量的汇聚端口,故可以通过分析事件数据服务器与主干网交换机相连的链路的网络流量来评估整个控制系统的网络运行状态,并作为该控制系统网络带宽需求的一个上界评估参数.

在确定了研究目标后,本文采用 wireshark 软件对选定的链路进行抓包分析,该软件可以捕获数据包的到达时间、大小、源地址、目标地址、协议类型等各种信息,并按照需求导出相关信息,捕获数据包的时间精度为 $1 \mu\text{s}$.为了从一个孤立的数据包得

到连续的网络流量 s ,可以使用如下近似算法:

$$s(t) \approx \frac{1}{\Delta t} \sum_{\tau=t}^{t+\Delta t} L(\tau) \quad (1)$$

式中, $L(t)$ 为 t 时刻捕获的数据包大小, f_s 为采样频率, Δt 为采样时间间隔, $\Delta t \gg 1 \mu\text{s}$, $s(t)$ 为 t 时刻的网络流量.该近似算法中,对于设定不同的采样频率就可以得到不同精度的流量序列.在折中频率分辨率的要求和计算量的限制以后,本文取采样频率为 50 Hz ,即采样时间间隔为 0.02 s .通过式(1)处理得到的网络流量如图 2 所示.

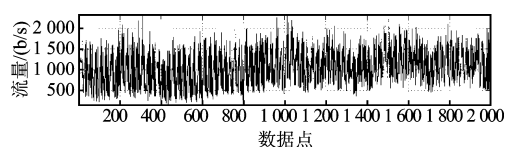


图 2 事件数据服务器采集到的网络流量波形

如果要得到其中一种或几种特定协议、特定 IP 设备的网络数据流量,可以对抓取的数据包的协议类型、源地址或者目标地址进行过滤,提取相应的数据包大小和时间信息后,再按式(1)得到该协议或者该 IP 的网络流量.

通过对抓取的数据包进行初步分析发现,该链路中的网络流量主要包括 CIP 协议和 TCP 协议.以字节计算,CIP 和 TCP 这两项协议占网络流量总体比重的 99% 以上,因此本文主要针对系统中这两种协议的网络流量特性进行研究.

3 流量的特性分析

由于本文控制系统的网络流量表现出较为复杂的特性,其中 CIP 和 TCP 协议流量具有明显不同的特点.为了建立网络流量的模型,采用试验分析-假设推理-模型验证的建模方式,先从流量的频率特性和小波特性上进行分析,根据分析结果寻找适合本系统的网络流量建模方法.

3.1 流量的时频特性分析

时频表示是一种为了描述非平稳信号的频谱随时间而变化的特点,而引入某种时间和频率的联合函数来表示信号的方法.基于时频表示的信号分析方法统称时频分析法,其中满足线性迭加定理的时频表示称为线性时频表示,主要有短时傅里叶变换、Gabor 变换、小波变换等.与传统的傅里叶分析相比,时频分析是一种对信号的局部频谱进行分析,除了可以用于分析流量的频谱特性外,还可以分析流量的平稳性.

定义 $\gamma(t) \in L^2(R)$, $\hat{\gamma}(\omega) \in L^2(R)$, $t\gamma(t) \in L^2(R)$, $\omega\hat{\gamma}(\omega) \in L^2(R)$, 则对于 $x(t)$ 有

$$STFT_x(t, \omega) = \int_{-\infty}^{\infty} [x(\tau)\gamma^*(\tau-t)]e^{-i\omega\tau} d\tau,$$

$$STFT_x(t, f) = \int_{-\infty}^{\infty} [x(\tau)\gamma^*(\tau-t)]e^{-i2\pi f\tau} d\tau.$$

式中, $STFT_x(t, f)$ 即为信号 $x(t)$ 在时间 t 附近的一个“窗口”内的局部信号段的频谱, 邻近的程度决定于窗函数 $\gamma(t)$ 的宽窄. 时窗越窄, STFT 的时间分辨率越高, 而频率分辨率则越低; 若时窗变宽则频率分辨率提高, 而时间分辨率降低, 可以根据需要选择合适的时窗宽度. 在实际应用中, 一般采用离散信号的短时傅里叶变换, 其定义为

$$STFT_x(n, \omega) = \sum_{m=-\infty}^{\infty} x(m) \gamma(n-m) e^{-i\omega m}$$

采用第2节中的流量算法分别计算该链路同一时段内 CIP 和 TCP 协议的网络流量, 然后使用 STFT 变换分别对其进行时频特性分析, 窗口函数选择高斯窗函数, 窗口半径设备总数数据长度的 0.05, 得到的结果如图3和图4所示.

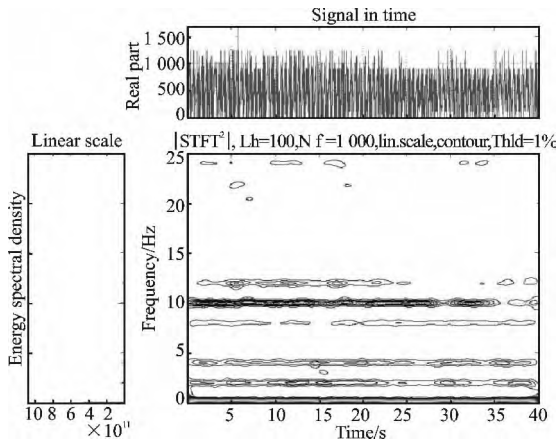


图3 CIP协议流量的STFT频谱(10%阈值)

图3为CIP协议的网络流量的STFT频谱图, 图中右上部为CIP协议流量的时域波形, 图左边为频域波形, 图右下部为该段流量时频特性的等高线表示. CIP协议流量虽然在时域上随机性较强, 但在频域上呈现出较为平稳的周期特性, 从图中可以看出CIP协议流量呈现出平稳的周期性, 主要能量集中在低频段, 其他频率点在2 Hz、4 Hz、8 Hz、10 Hz和12 Hz, 并且几乎不随时间变化, 其中10 Hz处的幅值最大, 这与控制系统中使用CIP协议的实际设定的100 ms通信周期情况相符. 因此在较大的时间尺度上, CIP协议的流量可以做平滑处理; 在较小的细节尺度上, 可以近似使

用周期流量模型进行建模.

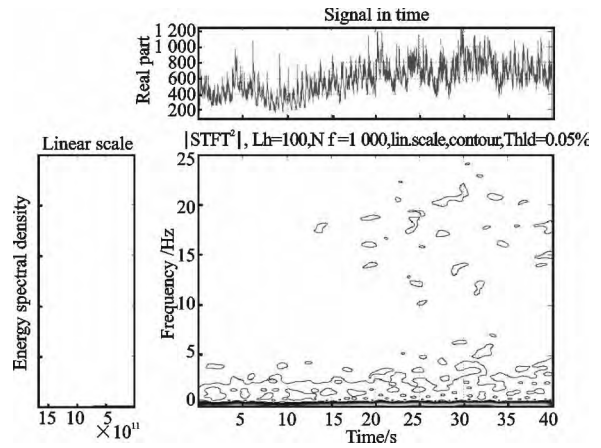


图4 TCP协议流量的STFT频谱(10%阈值)

图4为同一时间段内TCP协议的网络流量的STFT频谱. 从时域波形来看, 与CIP流量具有明显的周期特性相比, 本系统中TCP流量从频谱来看没有发现明显的频率分量, 主要能量也集中在低频段, 因此在较大的时间尺度上, TCP流量也可以做平滑处理, 但在较为精细的时间尺度上不适合采用周期流量模型进行建模.

通过对比图3和图4发现, 虽然系统中不同协议的网络流量具有不同的频率特性, 不过两者的共同特点在于低频谱段能量较为集中, 可以考虑采用小波变换分别提取网络流量的低频部分和高频部分进行分析和建模.

3.2 流量的小波特性的分析

小波变换是一种非线性序列分析方法, 可以把具有复杂特性的网络流量进行分解, 得到多个频率特性相对比较单一的时间信号, 便于分析网络流量在不同频段所表现的特性, 可以得到能够反映网络流量的总体趋势信号和多个细节信号, 从而能够方便地建立系统的网络流量模型.

一个网络流量信号 $x(t)$ 的小波变换可以表示为

$$W_x(a, b) = \langle x, \psi_{a,b} \rangle = \frac{1}{\sqrt{|a|}} \int_{\mathbb{R}} x(t) \overline{\psi(\frac{t-b}{a})} dt \quad (2)$$

式中, $\psi_{a,b}(t)$ 为小波函数, $\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi(\frac{t-b}{a})$,

$a, b \in \mathbb{R}, a \neq 0$.

重构公式为

$$x(t) = \frac{1}{\sqrt{|a|}} \int_{\mathbb{R}} W_x(a, b) \psi(\frac{t-b}{a}) db \quad (3)$$

多分辨分析是小波分析中一种对信号的空间分解方法. 其主要思想是把有限能量的信号空间 $L^2(R)$ 分解为一系列具有不同频率分辨率的子空间, 小波变换将原始流量信号 $x(t) \in L^2(R)$ 通过高通和低通两个滤波器逐一投影到各个子空间, 得到低频信号和各层高频信号的投影. 低频信号是网络流量的主体趋势项部分, 该部分反映的是网络流量整体的变化趋势特征; 高频信号是网络流量的随机性部分, 反映随机扰动等动态因素对网络流量的影响, 是对趋势项部分的细节补充.

小波分析中一般采用 Mallat 算法对流量信号 $x(t)$ 进行分解和重构:

$$\text{分解: } \begin{cases} a_{j+1} = Ha_j \\ d_{j+1} = Ga_j \end{cases}, j = 0, 1, 2, \dots, n \quad (4)$$

$$\text{重构: } a_{j-1} = H^* a_j + G^* d_j \quad (5)$$

式中 H 为一低通滤波器, G 为一高通滤波器, H^* 和 G^* 为 H 和 G 的对偶滤波器; a_j 为 $x(t)$ 经过小波分解得到的低频系数, d_j 为高频系数.

采用快速 Matlab 算法对采集到的链路总流量进行 5 层小波分析, 小波母函数选择 db4 小波, 得到各层趋势分量 $a_1 \sim a_5$ 和细节分量 $d_1 \sim d_5$ 如图 5 所示.

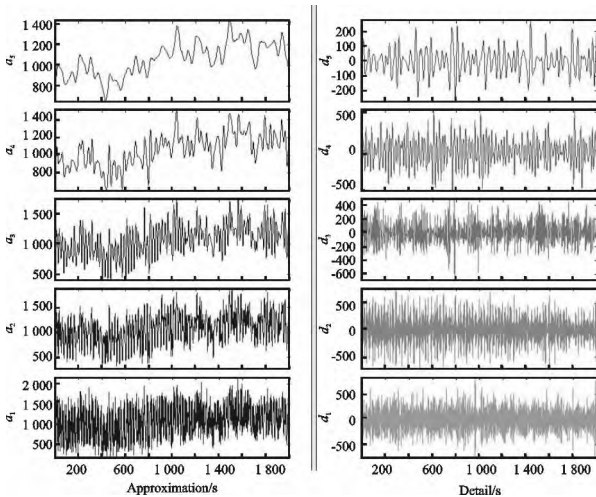


图 5 网络流量的小波趋势分析

从图中的趋势分量可以看出, 随着分解的层数越来越多, 趋势分量由 a_1 到 a_5 变得越来越平滑, 趋势也越来越明显; 而从各层细节分量的幅度来看, $d_1 \sim d_4$ 层都在 500 左右, 而 d_5 层的分量幅度(近似频率带宽为 1.6 ~ 3.1 Hz)有明显的减小, 由流量合成公式 $a_n = a_{n-1} + d_n$ 可知, a_4 与 a_5 层差别有所减小, 在精度要求不高的场合, a_4 层即能基本体现出了系统中网络流量的整体走向.

4 网络流量的 SVR 建模与仿真

前文通过对网络流量的特性进行研究, 发现了本系统的网络流量主要由 TCP 协议和 CIP 协议构成. 通过分别对两种协议流量的 STFT 分析发现两者具有不同的特性: CIP 流量具有较为平稳的通信周期, 而 TCP 流量没有明显的通信周期, 不过两者的主要能量都集中在低频段; 继而通过 db4 小波分析对系统流量进行 5 层分解, 对网络流量在低频段的表现进行分析, 发现整体流量表现出较为明显的趋势特性. 针对系统流量的这种特点, 本文初步选择 SVR 流量模型对系统的网络流量进行建模, 因为 SVR 模型具有可以使用较少的数据拟合周期项和趋势项的优点, 便于有效地简化流量模型.

4.1 支持向量回归机(SVR)模型

支持向量回归机的基本思想是寻找一个非线性映射 $\varphi(x): R \rightarrow F$, 将低维空间 R 映射到高维特征空间 F , 并在高维特征空间中对网络流量 $x(t)$ 进行估计:

$$x(t) = w \cdot \varphi(t) + b, w \in F \quad (6)$$

式中, w 为权值向量, b 为偏置项.

通过式(6)可以将一个低维空间内的非线性回归问题转化为一个高维特征空间的线性回归问题, 因此支持向量回归机模型可以表示为一个二次规划问题:

$$\min(\frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i^2 + \hat{\xi}_i^2) \quad (7)$$

$$\text{s. t. } w^T \varphi(x_i) + b - y_i \leq \varepsilon + \xi_i, i = 1, 2, \dots, n \quad (8)$$

$$y_i - (w^T \varphi(x_i) + b) \leq \varepsilon + \hat{\xi}_i, i = 1, 2, \dots, n \quad (9)$$

式中, ξ_i 和 $\hat{\xi}_i$ 为松弛变量; $\varepsilon > 0$ 称为不敏感参数, 其大小决定了对网络流量的拟合精度. 采用 Lagrange 乘子法对该二次规划问题进行求解可得

$$x(t) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(t_i, t) + b \quad (10)$$

式中, $K(t_i, t)$ 为核函数, α_i, α_i^* 为两个 Lagrange 乘子, 且满足 $\sum_{i=1}^n \alpha_i = \sum_{i=1}^n \alpha_i^*, 0 \leq \alpha_i, \alpha_i^* \leq C, \alpha_i - \alpha_i^*$ 等于 0 对应的数据点称为支持向量; $C > 0$ 称为惩罚系数, 用来平衡 $x(t)$ 的平滑程度和误差.

C 和 ε 是支持向量回归机算法中两个需要调整的重要参数, 对流量预测模型的性能具有较大影响, 可以通过调整 C 和 ε 两个参数使得样本流量预测的

结构风险(即泛化误差上界)最小. 相比于人工神经网络的经验风险最小准则,其优化目标是即训练误差最小,容易陷入局部极小点,造成过拟合而影响模型的泛化能力,因此支持向量回归机在理论上具有更好的泛化性能. 一般情况下优先选择较小的 C 参数,这样可以减少支持向量的数量,从而有效地减少计算量.

4.2 仿真实验

为了验证模型的有效性,将系统网络流量进行归一化后输入到 libsvm 工具箱中进行建模仿真,然后对 SVR 模型的参数进行寻优:核函数选择径向基函数,并通过多次测试和调整,最后确定 SVR 模型的其他参数取 $\epsilon = 0.2$ 和 $C = 10$,径向基函数选取 $g = 0.5$. 建立支持向量回归机仿真并进行训练,得到网络流量的建模结果,并与原始流量进行对比,结果如图 6 所示.

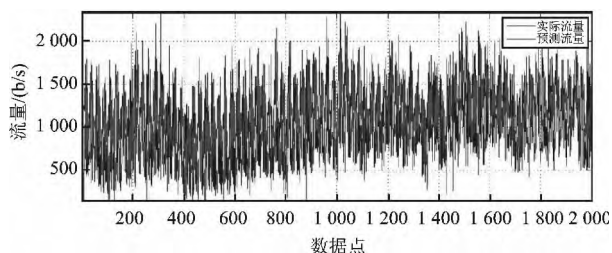


图6 SVR 仿真结果与实际网络流量的对比

从图中的对比结果可以看出,本模型对于系统网络流量主体趋势部分拟合得比较好,经过频谱分析发现 CIP 协议的频率特性部分也拟合得较好. 然而系统的流量存在较大的随机性,随着时间的增加,SVR 模型误差也逐渐有所增大,影响了建模精度. 不过流量的随机性起伏对实际系统性能影响没有趋势项大,在一定程度上可以忽略. 如果需要在此基础上进行进一步的精细刻画,可以考虑组合建模的方法.

5 结束语

本文以一个特定的工业控制系统的测控网络为研究对象,提出了一种有效的网络流量建模思路. 仿

真实验结果表明,SVR 模型可以有效地反映出系统网络流量的趋势特性和频率特性,然而对于网络流量的随机性拟合精度较差,导致整体流量建模精度不高.

参考文献:

- [1] Vonnahme E, Ruping S, Ruckbert U. Measurements in switched ethernet networks used for automation systems [C]// Proceedings of the 2000 IEEE International Workshop on Factory Communication Systems, Porto, Portugal, 2000; 231-238.
- [2] Ruan J H, Zhang W X. Identifying network communities with a high resolution [J]. Physical Review E, 2008, 77(1): 96-104.
- [3] 邹伯贤, 刘强. 基于 ARMA 模型的网络流量预测 [J]. 计算机研究与发展, 2002, 39(12): 1645-1652.
- [4] 罗赞骞, 夏靖波, 王焕彬. 混沌—支持向量机回归在流量预测中的应用研究 [J]. 计算机科学, 2009, 36(7): 244-245.
- [5] 李小航, 刘渊, 刘元珍. 基于小波多尺度分析的网络流量组合预测方法研究 [J]. 微电子学与计算机, 2008, 25(1): 130-133.
- [6] 刘渊, 王鹏. 融合小波变换与贝叶斯 LS-SVM 的网络流量预测 [J]. 计算机应用研究, 2009, 26(6): 2229-2231.
- [7] Ardhan S, Satsri S, Chutchavong V, et al. Improved model for traffic fluctuation prediction by neural network [C]// Proceedings of International Conference on Control Automation and Systems 2007. Korea; Seoul, 2007; 122-125.

作者简介:

刘亮 男, (1981-), 博士研究生. 研究方向为电力系统网络及自动化监控管理技术. E-mail: myquell@live.com.

江汉红 男, (1960-), 博士生导师. 研究方向为网络化系统、智能机器人等.

王洁 男, (1980-), 讲师. 研究方向为电力系统网络及自动化监控管理技术.

芮万智 男, (1984-), 讲师. 研究方向为网络化控制系统设计与故障诊断.