基于递归率REC特征的网络流量相空间重构监测

谢胜军[1] 殷 锋[1] 周绪川[2]
(西南民族大学网络中心 成都 610041)[1] (西南民族大学计算机科学与技术学院 成都 610041)[2]

摘 要 传统上对网络流时间序列分析多采用线性分析方法,没有充分利用到网络流客观存在的非线性特征信息,从而使数据分析能力受限。提出了基于定量递归分析递归率REC特征的网络流量相空间重构监测模型,基于相空间重构和递归图分析,设计了网络流量的REC递归率的定量递归特征作为网络流量序列分析的数据支撑。使用平均互信息算法和虚假最近邻点算法求取流量序列的相空间重构的关键参数,利用递归图中有规律的点线检验网络总出口流量的确定性和可预测性,利用REC特征监测网络流量序列的异常流量和特性进行分析。仿真实验表明,网络流量序列的定量递归特征具有较强的稳定性和自相似性,精度较传统特征统计方法提高19%以上,采用REC递归率特征对异常流量序列的预测预报监测准确率为99.7%,比采用传统的其它非线性递归特征提高了13.2%,展示了算法在网络流量和非平稳数据序列分析中的优越性能。

关键词 网络流量,递归率,相空间重构,递归图
中图法分类号 TP393 **文献标识码** A

Testing of Network Traffic Series in Reconstructed Phase Space Based on Recurrence Rate Feature

XIE Sheng-jun[1] YIN Feng[1] ZHOU Xu-chuan[2]
(Network Center, Southwest University for Nationalities, Chengdu 610041, China)[1]
(Collage of Computer Science and Technology, Southwest University for Nationalities, Chengdu 610041, China)[2]

Abstract Traditional linear analysis method takes the network traffic time series analysis, and does not fully use the existed non-linear feature information. The data analysis performance is limited as result. An improved testing model of network traffic with phase space reconstruction based on recurrence quantitative analysis(RQA) was proposed, and on the basis of phase space reconstruction, the recurrence rate called REC feature was designed as the sequence data support. The average mutual information algorithm and false nearest neighbors algorithm were proposed in calculating the key parameters for phase space reconstruction. The determinism and prediction property of network traffic were tested based on the regular points and lines, and the abnormal flow feature was detected based on REC recurrence feature. Simulation result shows that the series has strong stability and self-similar characteristics, and the feature detection performance is stable and precise, and the monitoring precision is improved by 19% with REC feature, and the abnormal traffic detection precision is 99.7% with the improvement of 13.2%. It shows good performance in analysis of network traffic and non-stationary data sequence.

Keywords Network traffic, Recurrence rate, Phase space reconstruction, Recurrence plot

1 引言

随着互联网的普及和广泛应用,网络已成为社会各行各业、生产生活中必不可少的工具,然而,在生产工作中,或许会出现网络拥堵,造成系统瘫痪、贻误商机、行政不作为等,甚至将造成人民财产和经济利益的巨大损失,因此研究网络流量时间序列的统计特性和特征分析,从而预测性地分析和监测网络流,使客户端有针对性地开展工作,避免拥堵,对为营运商宏观调剂、填补漏洞、避免网络攻击提供数据参考、采取宏观决策具有重大的现实意义。

随着网络规模的扩大和网络特性的变化,以及人们对网络流量时间序列分析的不断深入,人们发现,网络流量序列更多的是呈现一种非线性特征,使用传统的线性时间序列分析的方法去预测网络流量已经不符合事物之本质实际,他们已经开始尝试使用非线性时间序列分析的方法去分析网络流量时间序列,利用其定量递归特性分析网络流量的特征,实现预测评估和监测。文献[1]采用递归图分析的方法检验时间序列的确定性,为递归图定性分析时间序列奠定了基础。然而要定量分析时间序列的内部特征,需要建立一种定量递归分析模型并提取对应的非线性特征,这是本文需要完成的内容。文献[2,3]主要采用以Marlov模型算法、Poisson模型算法对网络流量序列特征进行分析的方法,建立了线性特征提取的

到稿日期:2013-01-20 返修日期:2013-04-25 本文受中央高校基本科研业务费专项基金项目(12NZYQN27)资助。
谢胜军(1977—),男,硕士,工程师,主要研究领域为网络运行管理,E-mail:xieshj@163.com;殷 锋(1972—)男,博士,教授,主要研究领域为计算机应用软件测试;周绪川(1972—),男,博士,副教授,主要研究领域为软件工程。

流量估计，它对流量预测一定的效果，但模型没有考虑网络流量序列的非线性特性，评估效果具有受到限制。文献[4]提出使用定量递归分析的方法分析校园网特性，采用相空间重构和递归图分析的方法研究和校园网流量的混沌特性，并构建了异常检测模型，为本文方法的引出提供了原始基础。但文献[5]的定量递归特征的选取受到阈值大小的限制，对小数据量信息流具有一定的效果，但对大数据网络流信息效果欠佳。

2 网络流量序列分析及相空间重构

利用非线性时间序列分析网络流量序列并对其进行异常信号检测、特征提取以及预测预报的第一步即最重要的一步，是进行有效的相空间重构。相空间重构是非线性时间序列分析的基础，也是本文构建基于递归率 REC 特征下网络流量相空间重构监测系统模型的基础。

2.1 网络流量序列的相空间重构

相空间重构的原理是来自于 Takens. F 和 R. Mane 的延迟嵌入定理，基于 Takens 嵌入定理的相空间重构建立了实际监测的信号与系统时间波动反映到多维混沌相空间中的动力系统空间特征之间的桥梁[6]，为分析网络流量时间序列和特征监测奠定了基础。相空间重构的方法描述为：一个组成的某事件或系统可由一非线性差分方程表示，该非线性差分方程为：

$$z_{n+1}=F(z_n) \tag{1}$$

上式表示的差分方程为一个离散系统，微分方差表达式为：

$$\frac{\mathrm{d}z(t)}{\mathrm{d}t}=F(z) \tag{2}$$

一组实际观测检测的时间序列$\{x_n\}$作为输出表示为：

$$x_n=h(z_n)+\omega_n \tag{3}$$

当然，对于连续系统，表示为：

$$x_n=x(t_0+n\Delta t)=h[z(t_0+n\Delta t)]+\omega_n \tag{4}$$

式中，$h(.)$表示函数；ω_n表示误差。

设 M 是 d 维的紧流形，F 表示为一矢量场，具有光滑性，对于 $\Phi:M\to R^{2d+1}$，有：

$$\Phi(z)=(h(z),h(\varphi_1(z)),\cdots,h(\varphi_{2d}(z)))^{\mathrm{T}} \tag{5}$$

当 $\omega_n=0$ 时，研究的网络流量序列$\{x_n\}_{n=1}^{N}$以如下向量在相空间中形成新的映射：

$$x_n=(x_n,x_{n-\tau},\cdots,x_{n-(m-1)\tau}) \tag{6}$$

在有干扰和噪声的情形下，T. Sauer 把上述 Takens 延迟嵌入定理推广如下。对于时间序列$\{x(t_0+i\Delta t)\}$，$i=0,1,\cdots,N-1$，它的相空间重构轨迹为：

$$X=[x(t_0),x(t_0+\Delta t),\cdots,x(t_0+(K-1)\Delta t)] \tag{7}$$

式中，$x(t)$表示嵌入的相空间中的一组空间状态矢量，m 是嵌入维数，Δt 是抽样时间间隔，$K=N-(m-1)J$，τ 为相空间重构的时延参数，$\tau=J\Delta t$，时间窗为：

$$\tau_w=(m-1)\tau=(m-1)J\Delta t \tag{8}$$

在重构相空间过程中，选择 m、τ 两个参数最为关键，本文采用平均互信息算法估计最佳时延参数 τ，采用虚假最近邻点算法计算最小嵌入维数 m。

2.2 平均互信息量估计相空间最佳时延

相空间重构中，选择合适的重构嵌入时间延迟，能够有效实现相空间轨迹的合理展开，既不会出现重叠也不会有冗余，如果实现相空间重构的嵌入时延太小，其冗余度就会增大，导致相空间中包含的原吸引子信息量偏少。与之相对应，如果选择的相空间重构时延太大，那么得到的相空间中的轨迹就会彼此分离，信息出现丢失，相互之间出现不相关性[7,8]。

在网络流量时间序列数据分布的区间上创建流量序列的概率分布曲线。用 p_i 表示 $x(t)$ 出现在分布曲线的区间 i 的概率，$p_{ij}(\tau)$表示信号 $x(t)$出现在区域 i 以及出现在延迟了一定时间量 τ 后的延迟信号 $x(t+\tau)$出现在区域 j 的联合概率。延迟时间互信息为：

$$I(\tau)=-\sum_{ij}p_{ij}(\tau)\ln\frac{p_{ij}(\tau)}{p_ip_j} \tag{9}$$

取互信息曲线 $I(\tau)$的第一个最小值时对应的横坐标值即为最佳相空间重构时间延迟参数选定值。

2.3 虚假最近邻点法估计最小嵌入维数

令 $x(n)$为有限数据长度的网络流量时间序列，τ 为通过上述方法取得的相空间重构时间延迟。根据 Takens 定理，在 m 维相空间中形成的 m 维矢量为：

$$X(n)=\{x(n),x(n+\tau),\cdots,x(n+(m-1)\tau)\},n=1,2,\cdots,N \tag{10}$$

对于重构相空间中的任意一点 X_n，它的最近邻点表示为 $X_{\gamma(n)}$，定义 R_m 为 X_n 与 $X_{\gamma(n)}$ 两点之间的距离，用欧式距离表示为：

$$R_m=\|X_{\gamma(n)}-X_n\|_2^{(m)}=\min_{j=N_0,\cdots,N,j\neq n}\|X_n-X_j\|_2 \tag{11}$$

随着 m 增加到 $m+1$，两点之间的距离也相应为：

$$R_{(m+1)n}=\|X_{\gamma(n)}-X_n\|_2^{(m+1)} \tag{12}$$

$R_{(m+1)n}$比R_m大得多的情况被认为是由于高维吸引子中两个不相邻的点投影在低维轨线上形成相邻的两点造成的，这种邻点是虚假的，称作虚假最近邻点。当满足以下准则，则称 $X_{\gamma(n)}$为 X_n 的虚假最近邻点。

$$\frac{|x_{\gamma(n)+m\tau}-x_{n+m\tau}|}{\|X_{\gamma(n)}-X_n\|_2^{(m+1)}}\geqslant R_{tol} \tag{13}$$

式中，R_{tol}为阈值，可取 15 左右。计算虚假最近邻点占的比例曲线，随着 m 增加，虚假最近邻点的比例减少，最终趋于收敛，但当虚假最近邻点的比例小于 5%或虚假最近邻点的比例趋于稳定不再递减时，可以得到混沌吸引子几何结构，从而得到有效拓扑，此时的 m 为所求的相空间重构最小嵌入维数[9]。

3 流量序列递归图与定量递归分析

递归图是分析时间序列周期性、非线性以及非平稳性的一个重要方法，如果网络流量时间序列中有确定性成分存在，则递归图中的图像点线分布具有规律性，例如在对角线周围有规律线段，表明该时间序列具有一定的识别和预测能力，可以对其作非线性定量递归分析研究，寻找有效的特征量，对网络流量进行异常检测和评估预测[10]。递归图的算法描述具体步骤描述为：

以 i 为横坐标、j 为纵坐标的坐标系，绘制出 $R(i,j)$，其中 i,j 分别称为时间序列标号 i 和 j，所得到的图为递归图。其中 $R(i,j)$为递归值，表达为：

$$R(i,j)=H(\varepsilon_i-\|d_{ij}\|),i,j=1,2,\cdots,N \tag{14}$$

式中，ε_i 为截止距离，可取固定值，或随 i 改变，使得半径为 ε_i 的球包含一定邻域数；d_{ij}为网络流量序列重构相空间后的网络流量序列重组空间中第 i 点 x_i 和第 j 点 x_j 的距离，表示

为：

$$d_{ij}=\| x_i-x_j \| \tag{15}$$

式中，$i=1,2,\cdots,N-(m-1)\tau$；$j=1,2,\cdots,N-(m-1)\tau$。

据式(15)可知$R(i,j)$的值在0和1之间变化，由此，在反映递归图的坐标系中，对于$R(i,j)$取值为0的点，不进行打点，对取值为1的点，进行打点，最终递归图将通过黑白相间的点线反映出系统的内在规律性和特征性，从而在此基础上提取特征参数，进行网络流量序列的异常检测分析。

4 定量递归 REC 特征的提出和计算

递归图只能从形象上判断信号内一些内部规律性特征，本文需要对网络流量时间序列进行定量分析。定量的特征分析作为流量监测的数据支撑，需要对递归图内部特征进行分析，实现特征提取和信号特征的检测研究。本文为了更加精确地研究网络流量序列递归图上反映出来的一些内在规律性特征，提出使用递归图平面中递归点占总点数的比例，即定义为递归率(REC)，来反映网络流量序列的特征。在递归图中，对角线方向线段的递归点占总递归点数的百分比，表示为确定性(DET)，表达式为：

$$DET=\frac{Nl}{Nr}=\frac{\sum_{l=l_{\min}}^{N-1} l\cdot p(l)}{\sum_{i=1}^{N}\sum_{j=1}^{N}R_{ij}} \tag{16}$$

式中，Nl是包含在与对角线平行线段中的递归点的个数，l为递归图中沿对角线方向的线段长度。定义最大对角线长度($L_{\max}$)，即除主对角线外，对角线方向线段长度的最大值：

$$L_{\max}=\max(l_i),i=1,2,\cdots,N-1 \tag{17}$$

计算递归图的递归率(REC)，其反映在递归图平面中递归点占总点数的比例：

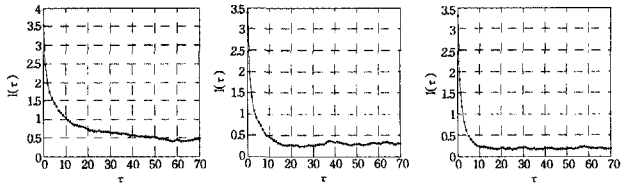
$$REC=\frac{Nr}{Nt}=\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}R_{ij} \tag{18}$$

式中，Nr表示递归图中打上黑点的递归点的数目，Nt表示所有平面上可能点的数目。物理上，REC表明了在m维相空间中彼此靠近的相空间点矢量所占的比例。

5 仿真实验与结果分析

5.1 实验数据采集

仿真实验中，研究的网络流量数据采集于某内部网络中心的监测数据，在中心交换机上进行采集分析，采集时间为2013年6月20日至22日的流量信息，采集样本每天为一段，作为一组样本实验集。采集方法是等时间间隔监测网络流量的数据包个数和数据量信息，采集采样时间间隔为1min，组成一组时间序列。流量监测的数据包括用户进行网页浏览、下载传输、文件传送等与产生流量相关的一切信息流。实验中，首先需要对采集的数据样本建立网络流量序列滑动窗口数学模型，作为网络流量序列重组空间异构分析的输入模型，相空间重构分析中，采用互信息算法和虚假最近邻点法求得网络流量序列重构相空间的关键性参数。3天3组样本的重构最佳试验参数计算仿真图如图1所示，取曲线第1个极小值点对应的横坐标为最佳异构时延参数，该样本序列计算得到的最佳时延参数分别为$\tau=16,18$和15。采用虚假最近邻点算法计算得到网络流量序列最佳嵌入维数的计算曲线，如图2所示，最近邻点的比例趋于稳定不再递减时，可以得到非线性时间序列吸引子几何结构，从而得到有效拓扑，由此得到最优化网络流量序列重组空间异构嵌入维数m的值，得到3天3组网络流量序列样本的最小也即最佳的相空间重构嵌入维数为$m=7$、5和5。



(a) 20日采集样本　(b) 21日采集样本　(c) 22日采集样本

图1 平均互信息算法计算相空间最佳时延参数的3组样本平均互信息曲线

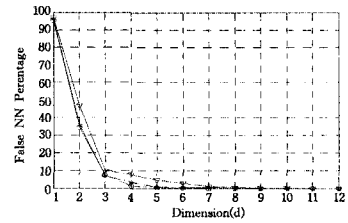
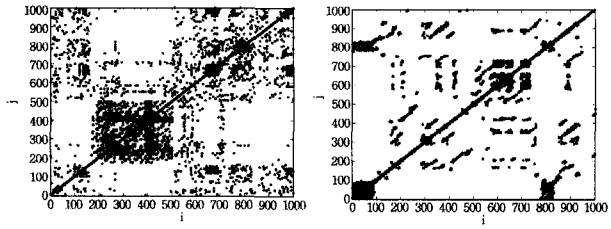
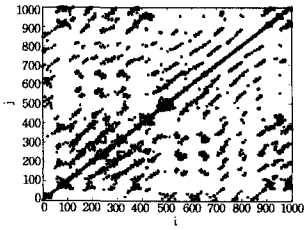


图2 虚假最近邻点算法计算3组样本相空间重构最小嵌入维

在上述取得的网络流量序列相空间重构的基础上，得到3组样本的递归图，如图3所示。从图中可以看出递归图有规律性表现，点线分布具有规律性，在对角线周围有规律线段，表明该时间序列具有一定的识别和预测能力，得出网络流量序列具有确定性和可预测性。



(a) 20日采集样本　(b) 21日采集样本



(c) 22日采集样本

图3 3组网络流量序列样本的递归图

在上述对网络流量递归图定性分析的基础上，采用定量递归分析的方法，对流量序列进行定量分析和定量特征参数提取的仿真实验。采用传统的确定性(DET)、最大对角线长度($L_{\max}$)特征量和本文提出的递归率(REC)特征量，计算得到3天的取样区间下每天的流量信息的定量递归特征结果，如图4所示。从图中计算结果对比可知，网络流量序列的定量递归特征具有较强的稳定性和自相似性，采用传统的最大对角线长度$L_{\max}$和DET特征分析网络流量时间序列，发现在各组样本之间，特征值出现上下浮动和偏差，计算标准差分别为0.23和0.19，而采用本文提出的递归率REC分析特征，样本监测统计标准偏差为0，证明本文提出的递归率特征监测网络流量序列比较稳定准确，精度较传统特征统计方法提高

19%以上。

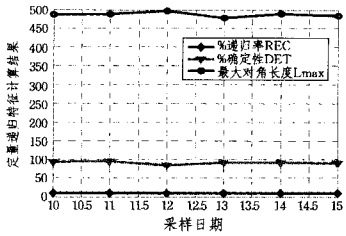


图 4 样本序列网络总出口量序列的定量递归特征提取值

采用 REC 定量递归分析特征提取的方法，在多维相空间中，对网络流量序列进行流量异常监测，全段样本网络流量序列时域抽取波形如图 5 所示，引入 REC 定量递归分析特征的网络流量异常谱递归图结果如图 6 所示。从图 6 可以看出，网络流量在流量序列出现异常特征时将会以异常光谱特征形式的点显示于递归图中，计算异常谱的点递归率等定量递归分析特征，得到准确的异常数据监测结果。从计算流程可知，采用本文方法无需引入网络流量序列的初始特征和先验信息，尤其适合于大数据量时间序列和非平稳数据的检测和分析。仿真结果表明，本文提出的网络流量监测算法和模型能有效检测出网络流量序列的隐藏异常流量特征，异常特征监测频谱清晰可见，采用蒙特卡洛算法计算得到的对异常特征预测预报监测的准确率为 99.7%，比采用传统的其它非线性递归特征提高了 13.2%，展示了算法在网络流量和非平稳数据序列分析中的优越性能。

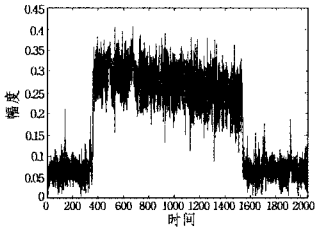


图 5 流量序列样本时域抽取波形图

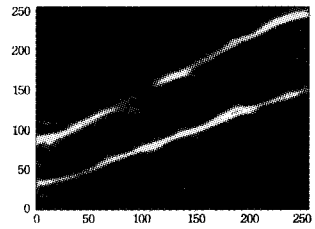


图 6 基于 REC 特征分析的网络流量序列特征监测

结束语 本文使用非线性时间序列分析的方法对网络流量序列进行相空间重构和定量递归 REC 特征提取及异常监测研究，利用递归图和定量递归分析的方法，检测出网络总出口流量的确定性，通过递归图确定性分析和检验，证明网络流量序列是可预测评估的；提出了 REC 递归率这一新的非线性定量递归分析特征量，并以改特征量作为监测网络流量序列的数据支撑。仿真分析得到结论：(1)网络流量序列的定量递归特征具有较强的稳定性和自相似性，本文提出的递归率特征监测网络流量序列比较稳定准确，精度较传统特征统计方法提高 19%以上；(2)本文提出的时间序列分析和监测方法无需引入网络流量序列的初始特征和先验信息，尤其适合于大数据量时间序列和非平稳数据的检测和分析；(3)采用 REC 递归率特征对异常流量序列预测预报监测的准确率为 99.7%，比采用传统的其它非线性递归特征提高了 13.2%。研究成果展示了算法在网络流量和非平稳数据序列分析中的优越性能，能有效应用于网络数据处理和网络安全管理领域。

参 考 文 献

[1] 闫源江，胡光波，周勇. 舰船辐射噪声的非线性和确定性检验[J]. 舰船电子工程，2010，30(10)：150-153

[2] Stan C，Cristescu C P，Dimitriu D G. Analysis of the intermittent behavior in a low-temperature discharge plasma by recurrence plot quantification[J]. Physics of Plasmas，2010，17(4)：1-6

[3] lan T P，James W S，Sadasivam S，et al. Attractor structure discriminates sleep states：Recurrence plot analysis applied to infant breathing patterns [J]. IEEE Transactions on Biomedical Engineering，2010，57(5)：1108-1116

[4] 朱凡，吴敏. 基于定量递归分析的校园网流量特性分析[J]. 计算机应用与软件，2012，29(6)：275-281

[5] 李超顺，周建中，方仍存，等. 基于混沌优化的模糊聚类分析方法[J]. 系统仿真学报，2009，21(10)：2977-2980

[6] 汪中才，黎永碧. 基于数据挖掘的入侵检测系统研究[J]. 科技通报，2012，28(8)：150-152

[7] Kousik G，Basabi B，Roy C A. Using recurrence plot analysis to distinguish between endogenous and exogenous stock market crashes[J]. Physica A：Statistical Mechanics and its Applications，2010，389(9)：1874-1882

[8] 赵岩，何鹏. 网络流量的非线性组合预测模型应用研究[J]. 计算机仿真，2012，6(29)：140-144

[9] 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

[10] 丁玲，赵小刚. 改进 BP 神经网络用于入侵检测[J]. 微计算机信息，2012，28(3)：131-134

(上接第 47 页)

[4] 延志伟. 基于 MIPv6/PMIPv6 的移动性支持关键技术研究[D]. 北京：北京交通大学，2011

[5] 袁琦凯，张奇支. 基于瞬间绑定的 PMIPv6 协议的域间切换方案[J]. 华南师范大学学报：自然科学版，2012，44(2)：58-62

[6] 周华春，张宏科，秦雅娟. 一种基于代理移动 IPv6 的全局移动性管理结构和协议[J]. 电子与信息学报，2008，30(12)：2999-3004

[7] AL-Hashimi H N，Baka K A，Ghafoor K Z. Inter-domain Proxy Mobile IPv6 based Vehicular Network[J]. Network Protocols and Algorithms，2010，2(4)：74-85

[8] Lee J-H，Han B-J，et al. Network Mobility Basic Support within Proxy Mobile IPv6：scenarios and analysis[S]. IETF draft-jhlee-netlmm-nemo-scenarios-01. 2008

[9] IEEE，Std P802. 21，IEEE standard for local and metropolitan area networks media independent handover services[S]. 2008

[10] Im I，Jeong J. Security-Effective Fast Authentication Scheme for PMIPv6-based NEMO with Global Mobility Support[C]// ICDIPC. 2012：90-95

[11] Yokota H，Chowdhury K，Koodli R，et al. Fast Handover for Proxy Mobile IPv6[S]. IETF RFC 5949. 2010

[12] Koodli R，et al. Mobile IPv 6 Fast Handovers [S]. IETF RFC 5568. 2009

[13] Lee K，Seo W，Cho Y，et al. Inter-domain Handover Scheme Using an Intermediate Mobile Access Gateway for Seamless Service in Vehicular Network[J]. International Journal of Communication Systems，2010，23：1127-1144

[14] Mussabbir Q B，Yao Wen-bing，Niu Ze-yun，et al. Optimized FMIPv6 using IEEE 802. 21 MIH Services in Vehicular Networks [J]. IEEE Transactions on Vehicular Technology，Special Issue on Vehicular Communications Networks，2007，56(6)：3397-3407