基于改进自适应灰色模型的网络安全态势预测

陈 雷[1] 司志刚[1] 鹤荣育[1] 周 飞[2]
(信息工程大学密码工程学院 郑州 450001)[1] (河南省产品质量监督检查院 郑州 450004)[2]

摘 要 对网络安全态势的研究是网络安全领域的热门话题。现在的研究更多关注的是评估当前态势,而对未来态势的预测讨论较少。为实现对未来网络安全态势的准确预测,改进了现有的自适应灰色分析模型。同时,为提高预测精度,针对传统分析模型的不足,提出了自适应灰色参数和等维灰色填充方法。实验结果表明,这个模型是有效的。

关键词 态势感知,态势预测,灰色模型,灰色参数
中图法分类号 TP393 **文献标识码** A

Network Security Situation Prediction Based on Improved Adaptive Grey Model

CHEN Lei[1] SI Zhi-gang[1] HE Rong-yu[1] ZHOU Fei[2]
(Institute of Cipher Engineering, Information Engineering University, Zhengzhou 450001, China)[1]
(Institute of Product Quality Supervision and Inspection in Henan Province, Zhengzhou 450004, China)[2]

Abstract Network security situation is a hot research topic in the field of network security. The now-existing research is more concerned about the assessment of the current situation, but has little discussion about the future trend predictions. An improved grey Verhulst model was put forward to predict network security situation accurately in the future. Aiming at the shortages in the prediction based on traditional Verhulst model, the adaptive grey parameters and equaldimensions grey filling methods were proposed to improve the precision. The results show that the model is valid.

Keywords Situational awareness, Situational prediction, Grey model, Grey parameters

1 介绍

作为研究的一个热点话题,态势感知(Situation Awareness,SA)这一概念源于航天飞行中的人因研究[1]。但到目前为止,对网络态势感知依然没有给出统一的、权威的定义。Endsley对SA进行了定义,"在一定的时空条件下,对环境因素的获取、理解以及对未来状态的预测[2]。"可以看出他把整个SA分为3个阶段:当前态势感知、态势评估及未来态势预测[3]。1998年,为了应对和处理越来越紧迫的网络安全问题,Bass将对SA的研究引入到网络安全领域,并提出了网络空间态势感知[4]。

对SA的研究重点是对当前态势的评估以及对未来态势的预测。然而,过去的研究仅仅关注对当前态势的评估,却很少讨论对未来态势的预测。蜜网项目采用统计分析手段,使用被称作3DMA(3天动态平均)的方法来预测第二天即将到来的安全事件,然而,预测只讨论了单一类型的安全事件,并且预测信息过于含糊。Ren等[5]研究认为态势值具有非线性时间序列的特点。根据神经网络在处理非线性的模糊数据上的优势,他们提出基于径向基函数神经网络的网络安全态势预测模型。而且,Lai等[6]也研究了相关的问题,把灰色理论引进态势预测。由BIT-ISA Lab(北京技术信息安全和对立实验室)开发的网络安全态势评估系统为用户提供了态势评估和态势预测功能。这个系统使用态势曲线和预警组件来完成态势分析和威胁预测的态势可视化。为了适应态势的多样性,这个系统采用线性回归、多项式回归和指数过滤。但是,不同预测结果之间存在很多认知的差异,直接导致了使用者在选择最佳应对措施的时候存在困难。

针对基于传统GM(1,1)模型和灰色预测分析模型存在的不足,本文提出了改进的灰色分析模型,采用自适应灰色参数和等维灰色填充(Equal Dimensions Grey Filling,EDGF)措施来提高评估和预测精度。

2 基本概念

1982年,Deng第一次提出了灰色理论。灰色模型利用通过累加生成操作(Accumulated Generating Operation,AGO)产生的序列来建立模型,以削弱原始序列中存在的随机性。该措施可以很容易地发现在序列中存在的变化规律,并利用该规律来进行预测。目前,灰色模型被广泛地用于优先等级,因为它在微小抽样时具有主导权,在短期预测方面有更好的精度[7]。

首先,介绍一些基本概念[7,8]。1837年,根据生物繁殖和物种变化的饱和度,德国生物学家Verhulst修改了传统的Malthus模型。

定义1(Verhulst模型) 原始模型引入了一个动态限制,Verhulst模型是非线性的微分方程

$$\frac{dp(t)}{dt}=rp(t)-up^2(t) \tag{1}$$

陈 雷(1986—),男,硕士生,主要研究方向为网络与信息安全,E-mail:460986065@qq.com;司志刚(1965—),男,硕士,教授,主要研究方向为信息安全、计算机工程;鹤荣育(1966—),男,硕士,教授,主要研究方向为信息安全;周 飞(1980—),女,硕士生,主要研究方向为信息安全。

其中，r,u 是模型参数，保持不变，对方程式(1)进行求解得

$$p(t)=\frac{r}{u}\{1+[\frac{r}{up(t_0)}-1]e^{-r(t-t_0)}\}^{-1} \quad (2)$$

其中，t_0 是起始时间，$p(t_0)$是序列中的原始值。

假设 1　假设原始数据序列为

$x^{(0)}=\{x^{(0)}(1),x^{(0)}(2),\cdots,x^{(0)}(n)\}$

将 1-AGO 代入原始公式，累积的数据序列可表示为

$x^{(1)}=\{x^{(1)}(1),x^{(1)}(2),\cdots,x^{(1)}(n)\}$

在 $x^{(1)}$ 中

$$x^{(1)}(k)=\sum_{i=1}^{k}x^{(0)}(i),x^{(0)}(i)\geqslant 0,k=1,2,\cdots,n \quad (3)$$

$x^{(1)}$ 被称为 $x^{(0)}$ 的 1-AGO，从方程式(3)中可知，显然 $x^{(0)}$ 可用 $x^{(1)}$ 表示，即

$$x^{(0)}(k)=\begin{cases}x^{(1)}(k)-x^{(1)}(k-1), & k\neq 1\\ x^{(1)}(k), & k=1\end{cases}$$

其中，$x^{(1)}(k)\in x^{(1)}$，这个操作被称为一阶逆累加生成操作(Inverse Accumulated Generation Operation，IAGO)。Deng 试图用 Verhulst 模型来预测 S 类曲线，并改良 Verhulst 模型。灰色 Verhulst 模型变为

$$\frac{dx^{(1)}(k)}{dt}+ax^{(1)}(k)=b[x^{(1)}(k)]^2 \quad (4)$$

其中，a 是动态系数，b 是灰色输入。两个参数的值都通过构造均值类和使用最小二乘法来解决。最后，我们可以得到灰色 Verhulst 模型的 1-AGO 时间反应式如下：

$$\hat{x}^{(1)}(k+1)=\frac{ax^{(0)}(1)}{bx^{(0)}(1)+[a-bx^{(0)}(1)]e^{ak}} \quad (5)$$

在方程(5)中，$x^{(0)}(1)=x^{(1)}(1)$，假设选择 n 维数据序列以适应这个模型。当 $k\geqslant n$ 时，这个自适应模型可以用来预测未来值，如：

$$\hat{x}^{(0)}(k+1)=\hat{x}^{(1)}(k+1)-\hat{x}^{(1)}(k) \quad (6)$$

其中，$\hat{x}^{(0)}(1),\hat{x}^{(0)}(2),\cdots,\hat{x}^{(0)}(n)$被称为灰色 Verhulst 自适应序列，$\hat{x}^{(0)}(n+1),\overline{x}^{(0)}(n+2),\cdots,\hat{x}^{(0)}(n+t)$被称为灰色 Verhulst 预测值，构造的灰色 Verhulst 模型被称作自适应模型。

3　提出预测模型

由于网络安全态势(Network Security Situation，NSS)多变的特点，NSS 预测系统和灰色系统很接近。因此，灰色模型通常用来预测未来的安全态势。传统的 GM(1,1)模型适用于具有指数规律的预测，这个模型只可用来描述单一变化的情况。在它的变化中存在一定的随意性[9]。过去的研究表明，NSS 是由网络风险状况决定的，并且变化过程存在一定的随机性，风险指数往往随时间变化。其通常表现为以下过程：

首先，网络面临着大量恶意性的攻击。尽管许多攻击的规模很大，攻击严重程度却较低。因此，风险指数的累积曲线增长缓慢。但在这个阶段过后，侵入者开始认识到目标的脆弱性和安全级别。针对目标的安全水平和脆弱性，采取更强烈的攻击。这种入侵行为使得风险指数的累积曲线增长更快，曲线变得陡峭。在接下来的过程中，风险指数达到其极限，之后的曲线又变得较为平滑。

态势变化使累积的风险时间曲线的增长像 S 型曲线一样，而不是简单的指数曲线。因此，用灰色 Verhulst 模型对网络安全态势进行预测效果更好。

3.1　现状评估

为了确切地预测未来网络态势的变化趋势，应当首先确定当前态势的评估方案。在评估方案中，网络被分为 3 个层面：网络层、主机层、服务层。当前风险值的获取采用层次分析法(Analytic Hierarchy Proces，AHP)来实现，过程描述如下：

假设 2　一个已知的网络系统包括几台主机，而且这些主机提供 p 种服务。这些服务表示为 $S_i(1\leqslant i\leqslant p)$，$A_j$ 表示对服务 S_i 所发起的攻击，而 T_{A_j} 则表示 A_j 攻击的严重程度。C 表示攻击类别的数值，且 j 满足 $1\leqslant j\leqslant c$，时间间隔用 τ 表示，时间间隔的重要性用 ω_τ 表示，在时间间隔 τ 中，N_{A_j} 是攻击 A_j 的检测数值。最后，C_T 表示划分评估时间间隔的数值。

根据上述假设，I_{S_i}(服务器 S_i 的风险指数)是 NSS 的基本指数，数学描述为：

$$I_{S_i}=\sum_{\tau=1}^{C_T}\omega_\tau\sum_{j=1}^{C}10^{T_{A_j}}N_{A_j} \quad (7)$$

基于式(7)，在评估期间，NSS 数学描述可表示为：

$$NSS=\sum_{i=1}^{p}W_{S_i}I_{S_i}=\sum_{i=1}^{p}W_{S_i}\sum_{\tau=1}^{C_T}\omega_\tau\sum_{j=1}^{C}10^{T_jN_j} \quad (8)$$

W_{S_i} 对应的是在整个层次模型中的服务 S_i 总量。

3.2　自适应灰色参数

灰色参数 a,b 的确定是保证灰色 Verhulst 模型精度的关键。a,b 的值可通过构建生成序列 $Z^{(1)}$ 并使用最小二乘法获得。通常生成序列可通过以下方式实现：

$$z^{(1)}(k+1)=\frac{1}{2}[x^{(1)}(k+1)+x^{(1)}(k)]$$

而 1-AGO 曲线在很短的时间间隔内是平滑变化的。以 1 秒内生成的序列来估算参数是比较适合的，否则，生成的序列会使预测产生提前或延后的误差，从而使得模型的精度降低，且使灰色 Verhulst 模型的应用范围缩小。在文中，采用自适应的方法确定灰色参数以保证模型精度。根据函数的 S 型曲线和存在的风险指数极限，可以确定函数的 AGO 曲线为：

$$X^{(1)}(t)=\frac{1}{\alpha e^{\beta t}}$$

曲线从 k 到 $k+1$，可得到以下方程式：

$$\left.\begin{aligned}x^{(1)}(k)&=\frac{1}{ae^{bk}}\\ x^{(1)}(k+1)&=\frac{1}{ae^{b(k+1)}}\end{aligned}\right\} \quad (9)$$

解方程(9)可得 α,β 的值为

$\alpha=x^{(1)}(k+1)/[x^{(1)}(k)]^{k+1}$

$\beta=\ln[x^{(1)}(k)]-\ln[x^{(1)}(k+1)]$

生成序列可以通过在区间$(k,k+1)$上整合 $Z^{(1)}$ 得到：

$$\left.\begin{aligned}z^{(1)}(k+1)&=\int_k^{k+1}\frac{1}{\alpha e^{\beta t}}dt\\ \frac{1}{\alpha}\int_k^{k+1}e^{-\beta t}dt&=-\frac{1}{\alpha\beta}[e^{-\beta(k+1)}-e^{-\beta k}]\end{aligned}\right. \quad (10)$$

代入方程式(10)，根据曲线的变化可用下面的式子来确定生成序列：

$$z^{(1)}(k+1)=\frac{x^{(1)}(k)-x^{(1)}(k+1)}{\ln[x^{(1)}(k)]-\ln[x^{(1)}(k+1)]} \quad (11)$$

3.3 模型精度的修改

当自适应模型的精度不能满足要求时，利用剩余序列来修正自适应模型。参照文献[10]，可采取以下方式。

假设 3　给定一个误差序列 $\varepsilon^{(0)}(k)$，并且 $\varepsilon^{(0)}(k)$ 满足 $\varepsilon^{(0)}(k)=x^{(1)}(k)-\hat{x}^{(1)}(k)$。

假设 4　存在 $q(q\in N)$，$\varepsilon^{(0)}(q)>0$，对于任意 q，都有 $\varepsilon^{(0)}(q)<\varepsilon^{(0)}(k)$。

根据上述两个假设，选择整数 k，满足 $|k>\varepsilon^{(0)}(q)|$，即原始数据的误差序列可转化为 $\varepsilon^{(0)}(k)+k, k\in N$，将 1-AGO 应用于变化的误差序列并构建 GM(1,1)模型，可得到在 1-AGO 模拟误差的时间响应式 $\hat{\varepsilon}^{(1)}(k)$，将 1-IAGO 应用于 $\hat{\varepsilon}^{(1)}(k)$，可得到模拟残余初始序列 $\hat{\varepsilon}^{(0)}(k)$。最后将 $\hat{\varepsilon}^{(0)}(k)$ 代入 1-AGO 灰色 Verhulst 自适应序列并计算修改后的 1-AGO 模拟序列。

3.4 等维灰色填充

一般来说，自适应灰色 Verhulst 模型是随连续时间而变化的函数，在未来可拓展到随机时间。但实际上，未知的干扰会影响模型的精度。

因此，这里采用 EDGF 方法来构造 EGDF Verhulst 模型[11]，从而改善灰色 Verhulst 模型，以实现 NSS 准确预测的目标。一旦获得一个新的预测结果，原序列中的第一个元素将被删除，新的预测将会位于数据序列的末端。随着嵌合序列的维数保持稳定，新生成自适应序列总是包含最新的态势变化信息，灰色 Verhulst 模型在新生成的自适应序列基础上进行重建。一直重复上述过程直到实现预定目标，该 EDGF 方法流程如图 1 所示，其中 M 表示预测步骤，N 表示原始序列的维数。

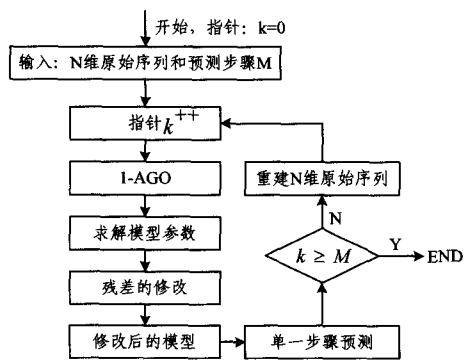

图 1　等维灰色填充方法流程

4　实验结果

评估数据集是在实验中所建立的局域网警报数据。实验分析了警报数据并评估了网络安全状况。采用 3.1 节提出的方案计算 NSS 风险值，并获得了持续 14 小时对未来评估的安全态势值。前 10 小时的值被用于拟合灰色 Verhuls 模型，前 10 小时的安全态势值分别是：3.019，4.254。由累积的态势值，得到了 1-AGO 序列，并使用该序列来构建灰色 Verhulst 模型。

根据自适应模型参数的解决方案，得到了参数值 $\alpha=0.5067$，$\beta=0.0029$。将得到的值代入 1-AGO 时间响应公式，得到：

$$\hat{x}(k+1)=\frac{1}{0.0049+0.4014e^{-0.5067k}}, k=0,1,\cdots \tag{12}$$

根据方程式(12)可以得到模拟 1-AGO 前 10 小时的态势值，通过与实际值比较来检查自适应模型的精确度，检查误差如表 1 所列。

表 1　相应时间间隔的模拟误差检查表

时间/小时	初始序列	1-AGO 序列		
		实际值	模拟值	误差
2	4.432	7.675	5.3651	2.3099
3	4.531	12.236	10.0142	2.2218
4	14.101	27.013	18.1543	8.8587
5	15.397	42.689	29.0565	13.6325
6	21.403	63.012	45.5269	17.4851
7	26.489	88.977	71.5244	17.4526
8	28.169	116.617	98.2547	18.3623
9	27.736	143.774	124.3578	19.4162
10	19.052	165.012	149.7546	15.2574

基于对模拟和实际的 1-AGO 时序的对比发现，参与序列从第三个时间点误差变得越来越大，这个精度不可能被接受。因此，采用误差修正的方法来提高精度，从第三个时间点开始截取误差序列：

$\varepsilon^{(0)}(k)=\{2.2218,8.8587,13.6325,17.4851,17.4526,18.3623,19.4162,15.2574\}$

采用 1-AGO 得到的误差序列为：

$\varepsilon^{(1)}(k)=\{4.5317,13.3904,27.0229,44.5080,61.9606,80.3229,99.7391,114.9965\}$

按照 1-AGO 误差序列构建 GM(1,1)模型。最后得到模拟误差序列的 1-AGO 时间响应式为：

$$\hat{\varepsilon}^{(1)}(k+1)=205.6745e^{0.0540k}-197.1526 \tag{13}$$

使用 1-AGO 时间响应式，得到模拟误差序列的时间响应式为：

$$\hat{\varepsilon}^{(0)}(k+1)=14.687e^{0.0541k}, k\in N \tag{14}$$

联合方程式(12)－式(14)得到 1-AGO 修改后的态势模型。如下所示：

$$\hat{x}^{(1)}(k+1)=\begin{cases}\dfrac{1}{0.0049+0.4014e^{-0.5067k}}, & k<2\\ \dfrac{1}{0.0049+0.4014e^{-0.5067k}}+14.687e^{-0.0541(k-2)}, & k\geqslant 2\end{cases} \tag{15}$$

根据改进模型得出方程式(15)，执行单步预测得到 11 小时以内的结果，使用新预测结果代替前 1 小时的原始数据序列元素，重复单一步骤直到得出 3 个小时的预测结果。

为突出新提出模型的优点，将该新模型的预测结果和传统灰色模型预测方法所得的预测结果进行比较。

方案 1 为 GM(1,1)传统模型，方案 2 为灰色 Verhulst 传统模型，方案 3 为新建模型。

预测结果的比较如表 2 所列。

表 2　不同预测方案的比较

时间/小时	态势值			
	实际值	方案 1	方案 2	方案 3
11	9.814	33.6714	26.3672	15.3014
12	7.985	35.2367	14.0184	7.2675
13	7.124	44.0139	9.1543	5.8553
14	4.248	48.6809	6.0523	2.9676

如表 2 所列，不同方案的精度不同：

方案 1 采用传统的 GM(1,1)模型进行预测，1-AGO 预测

结果误差依然保持快速上升的趋势且偏离实际较多。方案2采用灰色Verhulst传统模型，预测结果同样偏离实际值较多，误差依然很大，新建的模型最贴近实际值，误差小，精度高。

结束语 自适应灰色参数克服了使用传统模型在决定参数系数和提高精度上的不足。通过使用新的预测结果替换原始序列中的第一个元素，这有益于所提出的预测模型及时辨别曲线变化趋势，而使传统序列的维度更加稳定，这个方法使得新建模型能更准确预测曲线的趋势和提高预测精度。

本文虽已分析和讨论了单峰态势的变化情况，并且改进了灰色Verhulst模型，但需要注意的是更为常见的多峰态势变化。在未来的研究中，将更加注重研究分析多峰态势变化以及预测复杂态势变化。

参 考 文 献

[1] 王慧强. 网络安全态势感知研究新进展[J]. 大庆师范学院学报，2010，30(3)：1-8

[2] 李颖. NSSA中网络安全态势预测研究[J]. 技术与市场，2010，17(12)：43-44

[3] 赵嶷飞，史永亮. 基于模糊综合评价的航路交通态势评估[J]. 中国民航大学学报，2011，29(1)：5-8

[4] Bass T. Intrusion detection systems and multi. Sensor data fusion：Creating cyberspace situational awareness [J]. Communications of the ACM，2000，43(4)：99-105

[5] Ren W，Jiang X H，Sun T F. Rbfnn. based prediction of networks security situation [J]. Computer engineering and Applications，2006，42(31)：136-139

[6] Lai J B，Wang H Q，Zhao L. Study of network security situation awareness model based on simple additive weight and grey theory[C]//Proceedings of 2006 International Conference on Computational Intelligence and Security. Hangzhou：IEEE Press，2006：1545-1548

[7] 邓聚龙. 灰色预测与决策[M]. 武汉：华中科技大学出版社，2002

[8] 邓聚龙. 灰色系统基本方法[M]. 武汉：华中科技大学出版社，1987

[9] Liu S F，LIN Y. An introduction to grey systems theory [M]. Grove City：IIGSS Academic Publisher，1998

[10] Guo Z J，Song X Q，YE J. A Verhulst model on time series error corrected for port throughput forecasting[J]. Journal of the Eastern Asia Society or Transportation Studies，2005(6)：881-891

[11] 傅立. 灰色系统理论及其应用[M]. 北京：科学文献出版社，1992

(上接第240页)

[28] Ruj S，Cavenaghi M A，Huang Z，et al. On data-centric misbehavior detection in VANETs [C]//Vehicular Technology Conference (VTC Fall). San Francisco，CA，USA，IEEE，2011：1-5

[29] Lu Rong-xing，Li Xiao-dong，Luan T H，et al. Pseudonym changing at social spots：an effective strategy for location privacy in VANETs [J]. IEEE Transactions on Vehicular Technology，2012，61(1)：86-96

[30] Pan Yuan-yuan，Li Jian-qing，Feng Li，et al. An analytical model for random pseudonym change scheme in VANETs [J]. Cluster Computing，2013：1-9

[31] Zhou Tong，Choudhury R R，Ning Peng，et al. P2DAP—Sybil Attacks Detection in Vehicular Ad Hoc Networks [J]. IEEE Journal on Selected Areas in Communications，2011，29(3)：582-594

[32] Golle P，Greene D，Staddon J. Detecting and correcting malicious data in VANETs [C]//Proceedings of the 1st ACM international workshop on Vehicular ad hoc networks. Philadelphia，PA，USA，ACM，2004：29-37

[33] Grover J，Prajapati N K，Laxmi V，et al. Machine learning approach for multiple misbehavior detection in VANET [M]// Advances in Computing and Communications. Springer Berlin Heidelberg，2011：644-653

[34] Grover J，Gaur M S，Laxmi V，et al. A Sybil attack detection approach using neighboring vehicles in VANET [C]//Proceedings of the 4th International Conference on Security of Information and Networks. Sydney，Australia，ACM，2011：151-158

[35] Chen Chen，Wang X，Han Wei-li，et al. A robust detection of the Sybil attack in urban VANETs [C]//The 29th IEEE International Conference on Distributed Computing Systems Workshops(ICDCS 2009). Montreal，Quebec，Canada，IEEE，2009：270-276

[36] Chang Shan，Qi Yong，Zhu Hong-zi，et al. Footprint：detecting Sybil attacks in urban vehicular networks [J]. IEEE Transactions on Parallel and Distributed Systems，2012，23(6)：1103-1114

[37] Guette G，Ducourthial B. On the Sybil attack detection in VANET [C]// IEEE International Conference on Mobile Ad hoc and Sensor Systems(MASS 2007). Pisa，Italy，IEEE，2007：1-6

[38] Yu Bo，Xu Cheng-zhong，Xiao Bin. Detecting Sybil attacks in VANETs [J]. Journal of Parallel and Distributed Computing，2013：746-756

[39] Grover J，Kumar D，Sargurunathan M，et al. Performance evaluation and detection of Sybil attacks in vehicular ad-hoc networks [M]// Recent Trends in Network Security and Applications. Springer Berlin Heidelberg，2010：473-482

[40] Capun S，Hubaux J P. Secure positioning of wireless devices with application to sensor networks [C]// Proceeding of 24th Annual Joint Conference of the IEEE Computer and Communications Societies(INFOCOM 2005). Miami，USA，IEEE，2005：1917-1928

[41] Bouassida M S，Guette G，Shawky M，et al. Sybil nodes detection based on received signal strength variations within VANET [J]. IJ Network Security，2009，9(1)：22-33

[42] Leinmuller T，Schoch E，Kargl F. Position verification approaches for vehicular ad hoc networks [J]. IEEE Wireless Communications，2006，13(5)：16-21

[43] Hao Yong，Tang Jin，Cheng Yu. Cooperative Sybil attack detection for position based applications in privacy preserved VANETs[C]// Global Telecommunications Conference (GLOBECOM 2011). Houston，Texas，USA，IEEE，2011：1-5