

基于独立子空间算法与集成策略的仪表微小故障诊断方法

胡吉晨¹, 黄国勇^{1*}, 邵宗凯^{1,2}, 王晓东^{1,2}, 邹金慧^{1,2}

(1. 昆明理工大学 信息工程与自动化学院, 昆明 650500; 2. 云南省矿物管道输送工程技术研究中心, 昆明 650500)

(* 通信作者电子邮箱 42427566@qq.com)

摘要:针对流程工业中多仪表微小故障难以检测的问题,利用独立元分析(ICA)提取仪表变量的独立元信息,根据独立元贡献度矩阵构建独立元子空间,并分别在每个独立元子空间上根据不同的贡献率选择独立元个数,得出三个统计量及其控制限,建立故障检测模型。再综合所有子空间故障检测模型的检测结果,根据实际需求制定集成故障检测策略,最后通过贡献度算法对故障源进行识别和分离。对 Tennessee Eastman 过程数据的仿真实验结果表明独立子空间算法提高了微小故障的检测精度,在流程工业中多仪表故障诊断中配合不同的集成故障检测策略在应用中更具有灵活性。

关键词:仪表故障检测;独立元子空间;独立元贡献度;集成策略

中图分类号:TP277 **文献标志码:**A

Small fault detection method of instruments based on independent component subspace algorithm and ensemble strategy

HU Jichen¹, HUANG Guoyong^{1*}, SHAO Zongkai^{1,2}, WANG Xiaodong^{1,2}, ZOU Jinhui^{1,2}

(1. Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming Yunnan 650500, China;

2. Yunnan Engineering Research Center for Mineral Pipeline Transportation, Kunming Yunnan 650500, China)

Abstract: To solve the problem of small fault detection of instruments in process industry, independent components were extracted by Independent Component Analysis (ICA) from instruments recorded data. And independent component subspaces were established according to the contribution matrix. Fault detection model was constructed in each independent component subspace with statistical variables. A proper ensemble strategy was chosen by combining all the fault detection results. Finally, the instrument with fault was located by contribution algorithm. The simulation results with TE (Tennessee Eastman) process show that this method has higher precision on small fault detection and more flexibility with proper ensemble strategy.

Key words: instrument fault detection; independent component subspace; independent component contribution; ensemble strategy

0 引言

在流程工业中,为了保证产品质量以及提高生产过程的安全性,流程工业中的仪表是否正常工作显得尤为重要。当前仪表故障诊断方法可分为三类,即基于定量模型的方法、基于知识经验的方法和基于过程历史数据驱动的方法。其中通过仪表数据处理分析的数据驱动方法是目前仪表故障诊断的研究重点,但大部分研究成果基本上都是基于单个仪表的故障检测,无法对流程中的所有仪表状态进行同步实时监控。

近年来,基于多元统计理论的多元过程控制(Multivariate Statistical Process Control, MSCP)故障诊断技术,如主元分析(Principal Component Analysis, PCA)、偏最小二乘(Partial Least Squares, PLS)及其拓展方法已经有了一定的研究成果^[1-5],然而传统的多元统计方法都是假定数据服从高斯分布,而实际工业数据并不符合这一假设,因此容易发生错报、漏报等现象。

独立元分析(Independent Component Analysis, ICA)是处

理数据中非高斯信息的一种方法。近年来,ICA 与其改进方法在故障诊断领域得到了较好的应用^[6-10]。然而这些方法都是对整体全部变量建立统一的单个模型进行故障检测,尤其对于仪表的微小故障检测中,故障往往体现在单个变量上,采用全部变量建立的检测模型会掩盖细节信息,不能反映局部变化特征,致使微小故障的检测效果不佳。

针对这一情况,首先根据独立元(Independent Component, IC)贡献度将全部变量划分为若干个子空间,然后在每一个子空间里分别建立检测模型。为补偿由于独立元数量选择不适所造成的影响,在原有的统计量基础上增加了统计量 I^2 ,构成更为全面的故障诊断模型,改善了微小故障的诊断效果。综合所有子空间故障检测模型的结果,根据不同的实际故障检测需要提供一种灵活的集成策略。

1 独立元子空间算法

1.1 独立元检测方法

ICA 算法的本质是将某个观测变量分解为若干个之间相

收稿日期:2013-01-17;修回日期:2013-02-17。

基金项目:国家自然科学基金资助项目(51169007);云南省科技计划项目(2010DH004,2011DA005,2011FZ036);云南省中青年学术和技术带头人后备人才培养计划项目(2011CJ017);云南省教育厅基金资助项目(2011Y386)。

作者简介:胡吉晨(1988-),男,安徽马鞍山人,硕士研究生,主要研究方向:基于数据驱动的仪表故障诊断;黄国勇(1977-),男,湖北天门人,副教授,博士,主要研究方向:工业过程故障诊断、工业过程智能控制;邵宗凯(1974-),男,云南保山人,副教授,博士,主要研究方向:智能交通联网监控、电机控制。

互独立的源信息的线性组合^[11],在多元变量统计中,可表示为:

$$X = AS + E \quad (1)$$

其中: $X = [x_1, x_2, \dots, x_m]^T \in \mathbf{R}^{m \times n}$ 为观测变量矩阵, m 为观测变量个数, n 为样本数量; $A \in \mathbf{R}^{m \times r}$ ($r \leq m$) 为混合矩阵; $S = [s_1, s_2, \dots, s_r]^T \in \mathbf{R}^{r \times n}$ 为独立元矩阵; $E \in \mathbf{R}^{m \times n}$ 为残差矩阵。ICA 算法在独立元矩阵 S 和混合矩阵 A 未知的情况下,得出分离矩阵 W ,从观测变量矩阵中分离出独立元,即:

$$\hat{S} = WX = W \cdot A \cdot S \approx S \quad (2)$$

对观测变量矩阵进行白化和标准化预处理,分离矩阵 W 可采用快速不动点 ICA 算法^[12] 得出。由于在快速不动点 ICA 算法中,假定了独立元个数 r 与变量个数 m 相等,所以没有残差矩阵。

1.2 独立元贡献度和贡献度矩阵

$$\text{设混合矩阵 } A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1r} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mr} \end{bmatrix}, \text{独立元 } s_j \text{ 在变量}$$

x_i 上的贡献度 η_{ij} 和贡献度矩阵 η 可如式(3)^[13] 定义:

$$\eta_{ij} = \frac{|a_{ij}|}{|a_{i1}| + |a_{i2}| + \cdots + |a_{ir}|} \quad (3)$$

其中: $i = 1, \dots, m; j = 1, \dots, r$ 。

$$\eta = \begin{bmatrix} \eta_{11} & \eta_{12} & \cdots & \eta_{1r} \\ \eta_{21} & \eta_{22} & \cdots & \eta_{2r} \\ \vdots & \vdots & & \vdots \\ \eta_{m1} & \eta_{m2} & \cdots & \eta_{mr} \end{bmatrix} \quad (4)$$

η_{ij} 越大,说明独立元 s_j 对变量 x_i 的贡献越大,体现出各独立元的特征信息,采用这样的变量建模可以更好地体现出独立元 s_j 的内在变化。根据贡献度矩阵 η 可挑选合适的独立元构造独立元子空间。

1.3 独立元子空间构建

由独立元的性质可知独立元 s_1, s_2, \dots, s_r 彼此之间独立且差别最大,在构造子空间时应让不同的子空间体现出不同独立元的特征信息,使子空间呈现出较大的差异性并保证有效性。

对贡献度矩阵 η 第 j 列向量 $(\eta_{1j}, \eta_{2j}, \dots, \eta_{mj})^T$ 中所有元素按照从大到小排列为:

$$\eta_{h_{1j}} > \eta_{h_{2j}} > \cdots > \eta_{h_{Kj}} \quad (5)$$

采用 K 最大值法挑选相对应的 K 个变量 $x_{h_1}, x_{h_2}, \dots, x_{h_K}$ 建立第 j 个子空间模型^[14]。 K 值可根据在不同的子空间内,前 K 个变量的贡献度 η_{h_j} 之和达到 80% 以上来确定,以此保证具有贡献度主导地位的变量被选取构建子空间。

通过以上方法将原始变量空间分解构造为 r 个子空间,构成一个子空间的变量还体现其他独立元的特征,对整体数据特征也有一定的描述能力,可保证每个子空间的准确性。

2 集成策略的子空间故障诊断

2.1 故障检测模型

首先对原始的观测变量矩阵 X 进行归一标准化和白化的预处理,再利用快速不动点算法得出分离矩阵 W 。再通过分离矩阵 W 每一行的 Euclidean 范数选取 Euclidean 范数百分比贡献度最大的 d 行构成 W_d ,获得主导独立元 $S_d = W_d X$,对第 k 个采样点,建立故障检测模型的统计量 I^2 和平方预测误差(Squared Prediction Error, SPE)统计量:

$$I^2(k) = s_d(k)^T \cdot s_d(k) \quad (6)$$

$$\text{SPE}(k) = e(k)^T \cdot e(k) = (\mathbf{x}(k) - \hat{\mathbf{x}}(k))^T \cdot (\mathbf{x}(k) - \hat{\mathbf{x}}(k)) \quad (7)$$

由于 Euclidean 范数百分比贡献度选取的不同,主导独立元数目选取也不同,从而可能造成诊断模型不能准确反映独立元内部变化。为补偿这种影响,在之前的基础上,对 W 的剩余部分 W_e ,获得非主导的独立元 $S_e = W_e X$,在第 k 个采样点增加统计量 I_e^2 :

$$I_e^2(k) = s_e(k)^T \cdot s_e(k) \quad (8)$$

由于数据并不服从某个特定的分布,所以每个统计量的控制限可以由核函数的方法确定^[15]。

2.2 集成策略的设计选取

根据独立元贡献度方法将原始变量空间构造为 r 个独立元子空间,在每一个独立元子空间内,采用 ICA 检测方法提取合适的子空间独立元建立故障检测模型。设 $p = 0$ 代表监测模型未检测到故障, $p = 1$ 代表检测模型检测到故障, r 个独立元子空间对应应有 r 个故障检测模型和 r 个检测结果,构成故障检测结果向量 $P = [p_1, p_2, \dots, p_r]$ 。

对于原始变量中所产生的微小故障,可能只影响到少数独立元信息,因此可能只被少数子空间故障检测模型检测到。设向量 P 中共有 l ($l \leq r$) 个元素值为 1,由此可设计如下的检测集成策略:

$$\varphi(x) = \begin{cases} 1, & i \geq l \\ 0, & i < l \end{cases} \quad (9)$$

其中 l ($1 \leq l \leq r$) 为根据实际需求所设定的值。当有 l 或 l 以上个子空间检测到故障时,即 $\varphi(x) = 1$ 表明有故障发生。根据不同的检测需要设定不同的 l 值,确定不同的集成策略:例如强调微小故障检测时,取较小的 l 值可制定激进检测策略;当只强调检测较大的故障时,取较大的 l 值制定保守检测策略。 l 值的选取也应当考虑不同种类的故障特征,以达到更好的故障检测效果^[13]。

2.3 故障源分离

在故障检测阶段,实时读取仪表 k 时刻的观测数据 $x_{\text{new}}(k) = [x_1, x_2, \dots, x_m]^T$,使用同样的归一化模型预处理,根据子空间的划分在各子空间内通过故障诊断模型得出 k 时刻的 I^2, I_e^2 , SPE 统计量,根据概率统计理论,当连续多个采样点统计量超过控制限时便可判断发生异常故障。

当子空间故障检测模型检测并判定发生异常故障时,需要对发生异常故障的仪表来源进行辨识和分离。由于通常的仪表故障是由于较少的仪表数据变量而产生的,因此可参考贡献度方法分离故障源仪表^[6]。

当子空间 I^2, I_e^2 , SPE 统计量未检测到故障时,不需实时计算统计量贡献度,减小系统运算量。当子空间的统计量检测到连续多点超过控制限时,得出该子空间内每个变量的贡献度,经过分析判断,辨识故障源仪表。

2.4 故障诊断过程

基于独立元子空间的仪表故障检测方法可归为以下几个步骤,其中步骤 1)~3) 为离线建模过程,步骤 4)~6) 为在线检测过程:

1) 通过正常历史数据,经过归一标准化和白化预处理,利用快速不动点 ICA 算法得出混合矩阵 A ;

2) 根据贡献度方法获取贡献度矩阵 η ,根据 K 最大值法选取变量,将整体变量空间划分为多个子空间;

3) 在每个子空间上确定各统计量控制限,建立故障检测模型;

4) 将在线采集的每个采样点数据采用相同的标准化模型处理;

5) 在不同的子空间上使用故障检测模型,得出统计量值并与控制限比较;

6) 根据不同的检测策略,综合集成各子空间的检测结果,判断是否发生故障,并分离出故障源。

3 仿真

Tennessee Eastman (TE) 过程是由 Downs 和 Vogel 根据 Eastman 化学公司的实际化工联合反应过程建立的标准测试平台,是典型的复杂工业过程,被广泛地应用于过程控制研究。该过程包括反应器、冷凝器、循环压缩机、气液分离器和解吸塔 5 个操作单元,具有 12 个操作变量和 41 个测量变量。

仿真实验选取 41 个变量中的 10 个观测变量并从 1 到 10 编号:反应器压力,反应器温度,反应器液位,放空率,气液分离器温度,气液分离器压力,气液分离器液位,解吸塔压力,解吸塔温度,汽提塔液位。选取正常模式下的 4000 个采样数据,建立 10 个独立元子空间故障诊断模型。

仪表通常有 4 种故障,即偏差阶跃故障、冲击故障、漂移故障以及周期震荡故障。本文在 1000 个正常的测试数据中加入模拟的故障信号进行仿真实验。

在第 2 个仪表的第 300 个测试数据点开始加入模拟的偏差阶跃故障信号(故障 1),子空间故障模型检测结果如图 1~2 所示,分别为划分的子空间 1 和子空间 10 的检测结果,从图中可看出,子空间 1 在第 300 个数据点开始,连续检测出 I^2 统计量超过控制限,说明该子空间可明显检测出故障,而子空间 10 的三个统计量都未能检测出故障。

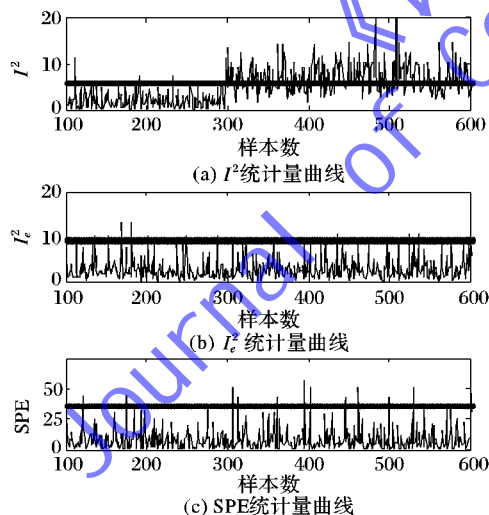


图 1 子空间 1 检测结果(故障 1)

图 3 为在整体变量空间内加入相同的模拟故障信号下,采用 ICA 故障检测方法的故障检测效果图。从图中可看出故障检测效果并不明显,说明在微小故障检测下,采用子空间的故障检测方法更有效。由于并非所有子空间的故障检测模型都能检测到故障,所以必须根据不同的实际需求制定相适应的集成策略中的 l 值。

在第 7 个仪表的第 250 个测试数据点开始加入模拟的微小漂移故障信号(故障 2),子空间故障模型检测结果如图 4~5 所示。

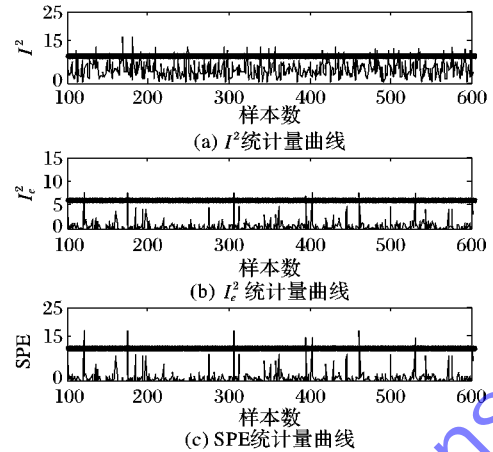


图 2 子空间 10 检测结果(故障 1)

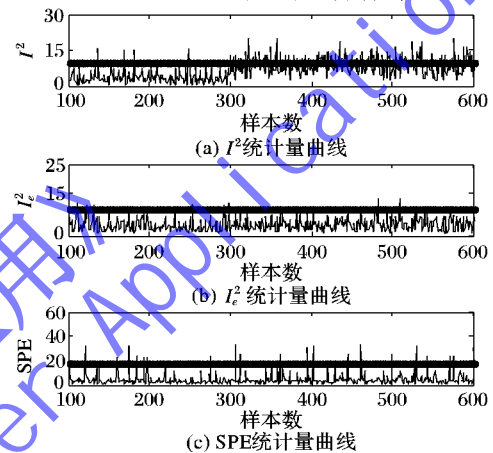


图 3 仪表 2 第 300 点模拟偏差阶跃 ICA 故障检测

在图 4 中,子空间 3 最先在第 320 个数据点左右开始, I_e^2 统计量连续多点超过控制限,检测到故障。

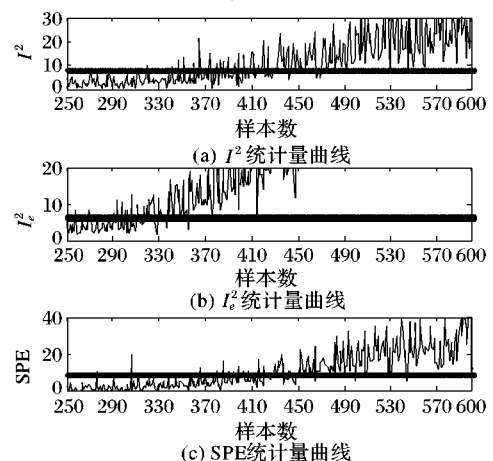


图 4 子空间 3 检测结果(故障 2)

图 5 中,子空间 9 在第 340 个数据点左右开始, I^2 统计量最先连续多点超过控制限,检测到故障。

通过对比可发现, I_e^2 统计量可对独立元数目选取不适造成的影响进行补偿,得到更好的检测效果,说明了 I_e^2 统计量建立的必要性。

图 6 为加入相同的模拟故障信号,采用 ICA 故障检测方法在整体变量空间下的检测结果。 I^2 统计量最先在 360 个数据点之后连续多点超过控制限检测到故障异常发生。

通过效果对比可发现采用子空间的故障检测模型在微小偏移故障的检测中可较早检测出故障,对漂移故障诊断具有较好的实时性。

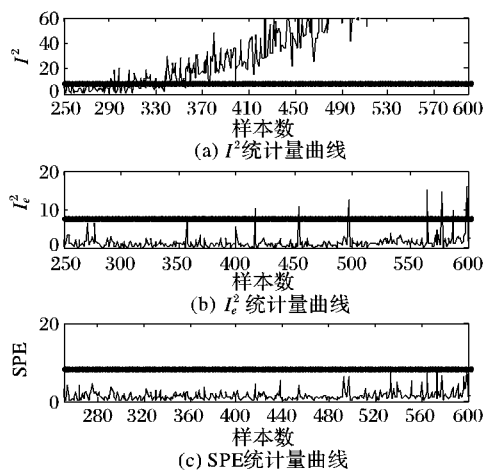


图5 子空间9 检测结果(故障2)

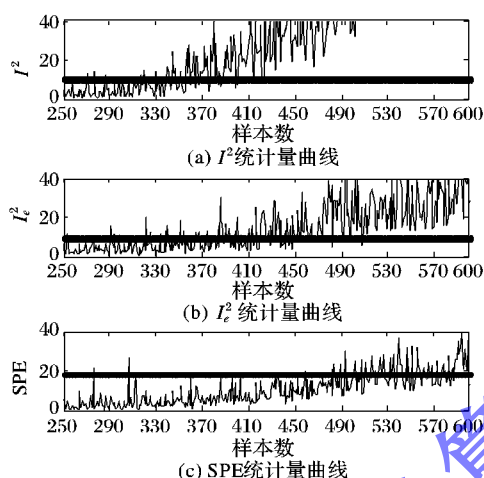


图6 仪表7 第250 点模拟微小漂移 ICA 故障检测

4 结语

为诊断流程工业中多仪表的微小故障,本文通过独立元的贡献度将整体空间划分为多个独立元子空间,在各子空间内原有的故障诊断模型基础上增加了用于补偿的统计量。利用 TE 过程数据进行仿真实验研究,与整体故障诊断方法在相同故障信号下的对比,说明子空间算法和增加的补偿统计量提高了微小故障诊断的检测效果,增强了微小漂移故障检测的实时性。提供的可根据不同故障诊断需求而变动的故障诊断策略,使该方法更具有实用性和灵活性。

参考文献:

- [1] SIMOGLU A, GEORGIECA P, MARTIN E B, *et al.* On-line monitoring of a sugar crystallization process [J]. *Computers and Chemical Engineering*, 2005, 29(6): 1411 - 1422.
- [2] RUMANA S, SIRISH L S, UTTANDARAMAN S. A PCA based fault detection scheme for an industrial high pressure polyethylene reactor [J]. *Macromolecular Reaction Engineering*, 2008, 2(1): 12 - 30.
- [3] LIU X Q, XIE L, KRUGER U, *et al.* Statistical-based monitoring of multivariate non-Gaussian systems [J]. *AIChE Journal*, 2008, 54(9): 2379 - 2391.
- [4] 范玉刚, 李平, 宋执环. 基于特征样本的 KPCA 在故障诊断中的应用[J]. *控制与决策*, 2005, 20(12): 1415 - 1418.
- [5] 邓晓刚, 田学民. 基于免疫核主元分析的故障诊断方法[J]. *清华大学学报: 自然科学版*, 2008, 48(10): 1794 - 1798.
- [6] LEE J-M, YOO C K, LEE I-B. Statistical process monitoring with independent component analysis [J]. *Journal of Process Control*, 2004, 14(5): 467 - 485.
- [7] GE Z Q, SONG Z H. Process monitoring based on Independent Component Analysis-Principal Component Analysis (ICA-PCA) and similarity factors [J]. *Industrial and Engineering Chemistry Research*, 2007, 46(7): 2054 - 2063.
- [8] 王培良, 葛志强, 宋执环. 基于迭代多模型 ICA-SVDD 的间歇过程故障在线监测[J]. *仪器仪表学报*, 2009, 30(7): 1347 - 1352.
- [9] 薄翠梅, 乔旭, 张广明, 等. 基于 ICA-SVM 的复杂化工过程集成故障诊断方法[J]. *化工学报*, 2009, 60(9): 2259 - 2264.
- [10] 薄翠梅, 柏杨进, 杨海荣, 等. 多切面分类改进独立成分与支持向量机集成故障诊断方法[J]. *控制理论与应用*, 2012, 29(2): 229 - 234.
- [11] HYVARINEN A, OJA E. Independent component analysis: algorithms and applications [J]. *Neural Networks*, 2000, 13(4): 411 - 430.
- [12] HYVARINEN A. Fast and robust fixed-point algorithms for independent component analysis [J]. *IEEE Transactions on Neural Networks*, 1999, 10(3): 626 - 634.
- [13] 张沐光, 宋执环. 一种基于独立元贡献度的子空间故障检测方法[J]. *控制理论与应用*, 2010, 27(3): 296 - 302.
- [14] 张沐光, 宋执环. 独立元子空间算法及其在故障检测上的应用[J]. *化工学报*, 2010, 61(2): 425 - 431.
- [15] CHEN Q, WYNNE R J, GOULDING P, *et al.* The application of principal component analysis and kernel density estimation to enhance process monitoring [J]. *Control Engineering Practice*, 2000, 8(5): 531 - 543.

(上接第2050页)

- [5] 肖枫涛, 王维, 刘波, 等. 一种基于进程流量行为的蠕虫检测系统[J]. *计算机工程与科学*, 2011, 33(4): 19 - 24.
- [6] ZHANG B, YANG J H, WU J P, *et al.* Diagnosing traffic anomalies using a two-phase model [J]. *Journal of Computer Science and Technology*, 2012, 27(2): 313 - 327.
- [7] Microsoft MSDN Library. Windows Sockets2 (Windows) [EB/OL]. [2012 - 10 - 26]. [http://msdn.microsoft.com/en-us/library/windows/desktop/ms740673\(v=vs.85\).aspx](http://msdn.microsoft.com/en-us/library/windows/desktop/ms740673(v=vs.85).aspx).
- [8] Microsoft MSDN Library. About WinINet [EB/OL]. [2012 - 10 - 26]. [http://msdn.microsoft.com/en-us/library/aa383630\(v=vs.85\).aspx](http://msdn.microsoft.com/en-us/library/aa383630(v=vs.85).aspx).
- [9] Microsoft MSDN Library. About WinHTTP [EB/OL]. [2012 - 10 - 26]. [http://msdn.microsoft.com/en-us/library/aa383630\(v=vs.85\).aspx](http://msdn.microsoft.com/en-us/library/aa383630(v=vs.85).aspx).
- [10] NICHOLAS N. Dynamic binary analysis and instrumentation or building tools is easy [D]. Cambridge: University of Cambridge, 2004.
- [11] SION B. Pin - a dynamic binary instrumentation tool [EB/OL]. [2012 - 06 - 13]. <http://www.pintool.org/>.
- [12] 王乾, 舒辉, 李洋, 等. 基于 DynamoRIO 的恶意代码行为分析[J]. *计算机工程*, 2011, 37(18): 139 - 141.
- [13] 刘豫, 王明华, 苏璞睿, 等. 基于动态污点分析的恶意代码通信协议逆向分析方法[J]. *电子学报*, 2012, 40(4): 661 - 668.
- [14] 赵天福, 周丹平, 王康, 等. 一种基于网络行为分析的反弹式木马检测方法[C]//第26次全国计算机安全学术交流会论文集. 北京: 中国学术期刊电子出版社, 2011: 80 - 83.
- [15] 段刚. 加密与解密[M]. 3版. 北京: 电子工业出版社, 2010: 121 - 126.