# 基于形式规范的协议一致性测试的可靠性分析

吕欣岩 赵保华 屈玉贵
(中国科学技术大学计算机系 合肥 230027)

**摘 要**：可以用于形式规范的描述语言很多，但存在一种统一的方法对形式规范进行一致性测试。该文采用统计的方法分析基于形式规范的协议一致性测试的可靠性，通过计算给出待测实体经过这种测试后所能达到的可靠性的置信区间，特别指出在“零错误”下影响可靠性的条件。

**关键词**：一致性测试；形式规范；测试假设；可靠性
**中图分类号**：TP393.04 **文献标识码**：A **文章编号**：1009-5896(2007)04-0781-04

# Analysis of Conformance Test Reliability Based on Formal Specification

Lü Xin-yan Zhao Bao-hua Qu Yu-gui
(*Computer Science and Technology Department*, *USTC*, *Hefei* 230027, *China*)

**Abstract**: Many description languages can be used by the formal specification, but there is a uniform method which can perform the conformance test under the formal specification. Adopting a statistical method, this paper analyses the conformance test reliability based on the formal specification, then it will give the confidence interval for the reliability and especially point out the conditions which can affect the reliability under ‘zero-error’.

**Key words**: Conformance test; Formal specification; Test hypotheses; Reliability

## 1 前言

目前运行在计算机和通信领域内的协议规范多为自然语言描述，由于自然语言本身所具有的二义性，使得人们对协议的理解可能会出现歧义。而采用形式化技术得到的形式规范可以消除二义，保证对协议理解的一致，同时形式规范为协议的一致性测试提供了有效的支持。

尽管协议描述的种类很多，但它们都可以抽象成相似的结构——数据类型描述以及作用在这些数据类型上的行为和性质描述。由这个特性可以从形式规范出发提出一种统一的协议一致性测试方法[1,2]，这种方法适用于：LOTOS，ESTEL，SDL以及新近提出的扩展RSL[3]等描述的协议。这种统一方法产生的测试集，可以通过对测试假设的限定，覆盖现有的多数基于模型或属性的测试方法所产生的测试集。

由于测试假设的存在，所以待测实体 IUT 即使通过了上述测试集，在现实中也不能保证协议实现与规范完全一致。因此 IUT 经过一致性测试之后所能达到的可靠程度是一个值得研究的问题。本文将以扩展 RSL 描述的协议规范为例，用统计的方法分析基于形式规范的协议一致性测试的可靠性，最终计算出待测实体经过这种一致性测试后所能达到的可靠性的置信区间。

本文的章节安排如下：第 2 节以扩展 RSL 为例简单介绍基于形式规范的协议一致性测试的一些概念；第 3 节给出基于形式规范的一致性测试的几个性质；第 4 节计算可靠性；第 5 节结束语。

## 2 基于形式规范的协议一致性测试简介

尽管协议描述的种类很多，但它们都可以抽象成相似的结构——数据类型描述以及作用在这些数据类型上的行为和性质描述。由这个结构，可以得到一种统一的基于形式规范的协议一致性测试方法，下面本文以扩展的 RSL 为例说明这种测试方法。

### 2.1 基于扩展 RSL 的描述

RSL 是一种广谱的形式化描述语言，它能统一地描述软件系统的抽象规范以及实现算法。文献[3]对 RSL 进行了扩展，并用扩展后的 RSL 描述协议规范。关于 RSL 的语法定义参见 RSL 规范，在此不做介绍。

限于篇幅，本文选取一个简单的例子，它由以下几条规则组成：(1)系统有一个缺省输出消息长度的最大值 Max；(2)系统从控制信道接收一个整数 newMax，如果 newMax 等于 0 则系统终止，否则用 newMax 代替 Max；(3)系统从输入信道读入两条消息，如果两条消息长度之和大于 Max，则从输出信道顺次输出这两条消息，否则，将两条消息组合成一条消息，然后从输出信道输出。这个系统的扩展 RSL 描述如下。

1. MESSAGE = class
2. type Message = char*
3. value

2005-09-02 收到，2006-04-03 改回
国家自然科学基金重大研究计划项目(90104010)，国家自然科学基金(60241004)和国家 973 计划(2003CB314801)资助课题

```
4.    ": -> Message
5.    Pack : Message, Message -> Message
6.    Size : Message -> Nat
7.  axiom forall m1, m2: Message.
8.    Pack(m1, m2) ≡ m1^m2,
9.    Size(") ≡ 0;
10.   Size(m1) ≡ Succ(Size(tail(m1)));
11. end
12. PROCESS = extend Message with class
13. type Process
14.  channel
15.     ctrl:Nat,
16.     inGate,outGate:Message
17.  value process:Nat -> in control,inGate out outGate Message
18.  axiom forall Max:Nat.
19.    process(Max) ≡
20.  let newMax=?ctrl in newMax>0; process (newMax); end
21.          []let newMax =?ctrl in newMax = 0;exit;end
22.          []let x=?inGate in
23.           let y=?inGate in Size(x)+Size(y)>Max ;
24.              outGate!x !y; process(Max); end
25.           []let y=?inGate in Size(x)+Size(y)≤Max;
26.              outGate!Pack(x,y);process(Max);end
27.           end
28. end
```

### 2.2 基于形式规范的一致性测试[1,2]

因为协议描述可以抽象成数据类型描述以及作用在这些数据类型上的行为和性质描述，所以基于形式规范的一致性测试可以分两个部分，数据类型测试和行为测试，文献[2]统一了这两部分测试，下面从抽象数据类型开始介绍基于形式规范的一致性测试。

**定义 1** 一个代数规范由两部分组成：基调 $\Sigma=(S, F)$ 和公理 Ax。其中 $S$ 是类别(sort)的有限集合，$F$ 是关于 $S$ 的操作(函数)的有限集合，Ax 是一些公理的有限集合，这些公理主要是等式公理，用以描述函数间应该满足的性质。

下面定义基于代数规范的穷尽测试集。

**定义 2** 给定一个规范SP，穷尽测试集$\text{Exhaust}_{\text{SP}}$是满足SP的公理的所有基项实例集。

对于形式规范中的行为或过程描述，可以引入进程迹的概念来生成测试用例。

**定义 3** 一个进程的迹是这个进程的可执行事件序列的集合。

如果将进程描述也视为一条公理，则可以将穷尽测试集的概念扩展，以包含进程的所有迹。比如上文中扩展 RSL 描述进程 process(Max)的一条迹是：let newMax=?ctrl in newMax=0;exit;end。

由于通常$|\text{Exhaust}_{\text{SP}}|=\infty$，所以测试是基于特定假设$H$的，通过测试假设的约束，可以将$\text{Exhaust}_{\text{SP}}$精化成一个有限实用的测试集$T$，最常用的两个假设是均一假设和正则假设。

**定义 4** 正则假设：如果一个由所有的基项的复杂程度都小于或者等于给定值的测试集通过测试，则$\text{Exhaust}_{\text{SP}}$也通过测试。本文中基项的复杂程度由基项中出现的函数个数表征，如果一个基项中允许最多出现$k$个相同的函数，则称为$k$级正则假设。

**定义 5** 均一假设：假定输入空间能够被划分成一些子空间，如果一个包含每个子空间中单一元素的测试集通过测试，则$\text{Exhaust}_{\text{SP}}$也通过测试。

例如上文中扩展 RSL 描述中的例子仅对 process 函数应用 2 级正则假设，并对子空间(newMax1＞0)∧(Size($x$1)+Size($y$1)＞newMax1)∧(newMax2＞0)∧(Size($x$2)+Size($y$2)＜newMax2)应用均一假设实例化变量得到的一条实际测试用例是

Ctrl!5,inGate!“abc”,inGate!“xyz”, outGate?“abc”, outGate?“xyz”, Ctrl!3，inGate!“y”, inGate!“t”, outGate?“yt”;

在假设$H$下，如果测试集$T$都通过测试，则可以判定$P$与SP在假设$H$的条件下一致。测试假设实质上是测试开销和测试质量的妥协，由于不能完全覆盖$\text{Exhaust}_{\text{SP}}$，因此一致性测试的可靠程度就不能达到 100%。

## 3 基于形式规范的协议一致性测试的几个性质

本节将给出几个基于形式规范的协议一致性测试的性质。

**性质 1** 令$k$级正则假设下生成的抽象测试集是$T_k$，其中$k$是自然数，则有$T_k \subseteq T_{k+1}$。

这个性质很容易从 $\Sigma$ 代数的性质出发对测试集进行结构归纳证明，本文略。

**性质 2** 对于一个给定的形式规范，$k$ 级正则假设下的测试集$T_k$中的测试用例个数$|T_k|$是一个关于$k$的增函数(事实上这个函数往往近似为一个指数函数)，或者是存在自然数 $\Delta$，使得当$k>\Delta$时$|T_k|$为一常数。

这个性质同样可以由 $\Sigma$代数的性质证明，事实上如果把$T_k$看成一棵树，则$T_{k+1}$或者是在$T_k$的基础上扩展一层得到的树，或者是由于协议规范不能扩展此时$T_{k+1}=T_k$，对比树的性质，可以得到性质 2。例如上文中扩展RSL描述中的例子，如果只限定process函数的出现次数，并且删除相互覆盖和重复的测试用例，则可以得到$|T_k|=(3^{k+1}-1)/2$。由于$|T_k|$为常数的情况比较简单，所以本文只分析$|T_k|$是$k$的增函数的情况。

**性质 3** 设测试用例中的每个函数的平均出错的概率为

$p(p\in(0,1])$，则一条测试用例在第 $r$ 步出错的概率是 $(1-p)^{r-1}p$，整条测试用例的出错概率是 $\sum_{r=1}^{R}(1-p)^{r-1}p=1-(1-p)^R$，其中 $R$ 为测试用例的长度，而 $\lim_{R\to\infty}(1-(1-p)^R)=1$，这说明只要函数的出错概率不为 0，则一定能通过一条足够长的测试用例检测到这个错误。

## 4　基于形式规范的协议一致性测试的可靠性分析

本节将使用统计的方法计算基于形式规范的协议一致性测试的可靠性。首先给出一些假设：设 $S=\{s_1,\cdots,s_m\}$ 是类别的集合，$F=\{f_1,\cdots,f_n\}$ 是函数的集合，$\mathrm{Ax}=\{a_1,\cdots,a_l\}$ 是公理的集合，其中 $m$，$n$，$l>0$。

为考虑更一般的情况，扩展$k$级正则假设，定义一个times函数，times：$F\to$Nat，其中Nat是自然数集，表示在正则假设下每条测试用例中出现函数$f_i$的次数最多为times($f_i$)。很容易看出经过这样扩展的正则假设保持上述 3 个性质不变，证明略。我们用$n$元组($N1,\cdots,Nn$)表示times函数。

由性质 1 和性质 3 知($N1,\cdots,Nn$)的各个元素越大，则测试的越彻底，但由于性质 2，使得对($N1,\cdots,Nn$)的选取被限定。

下面分两种情况计算在正则假设下基于形式规范的协议一致性测试的可靠性。

### 4.1　情况 1

首先分析当 $N1,\cdots,Nn$ 不是很大时的情况。

由于采用均一假设实例化抽象测试集的输入变量并不能覆盖所有情况，所以需要通过分解(unfold)过程对输入变量进行子域划分[1]，这个分解过程的本质是对变量输入空间进行等价类划分。例如可以将上文中扩展RSL描述的第 25 行中自然数上的“≤”函数分解成两个函数：“<”和“=”。经过分解过程后抽象测试集$T$中有多少条测试用例则整个测试空间$M$就被划分为多少个等价类，设$T$中有$l$条测试用例，则$M$被划分为$l$个等级类，即 $M=\bigcup_{i=1}^{l}M_i$ 且 $M_i\cap M_j=\phi$，其中 $i,j\in[1,l]$且$i\neq j$，$M_i$与$T_i$一一对应，且 $T_i\in T$。根据均一假设只需从$M_i$中随机选取 1 个成员(可能是一个元组)对$T_i$进行实例化，但在实际测试中往往是随机选择$\mathrm{EM}_i$个元素对输入变量进行实例化，此时对于 $\forall\ x\in M_i$，$x$导致测试用例出现错误的概率相同，设为$p_i(i\in[1,l\ ]$且 $0\leqslant p_i\leqslant 1)$，则这个模型与文献[4, 5]中的模型等价，所以当IUT存在与协议规范不一致的地方时，其出错概率$p$及其估计 $\hat{p}$ 可以表示为

$$p=\sum_{i=1}^{l}\left(p_i\times\frac{|M_i|}{|M|}\right),\quad \hat{p}=\sum_{i=1}^{l}\left(\bar{Y}_i\times\frac{|M_i|}{|M|}\right)\tag{1}$$

其中 $\bar{Y}_i=\dfrac{\text{与}T_i\text{相关的测试用例执行失败个数}}{EM_i}$，在大样本条件下$p$的区间估计为$[\hat{p}-z_{\alpha/2}\cdot\hat{\sigma}$，$\hat{p}+z_{\alpha/2}\cdot\hat{\sigma}]$，其中 $z_{\alpha/2}$ 为标准正态分布的 $1-\alpha/2$ 分位数，$\hat{\sigma}$ 为 $\hat{p}$ 的方差；当IUT经过一致性测试但没有发现错误时，即测试满足“零错误”条件，此时式(1)中的 $\bar{Y}_i=0$，所以 $\hat{p}=0$，这表明IUT的可靠性达到 100%，但由于测试的不完全性，这个结论并不正确。考虑从$M_i$中选取元素时的随机性，则IUT的可靠性问题转化为在 $P(\bar{Y}_i=0$，$i=1,2,\cdots,w)\geqslant 1-\alpha$ 的条件下求解 $\sum_{i=1}^{l}\left(p_i\times\frac{|M_i|}{|M|}\right)$ 的最大值$p_0$。因为 $P(\bar{Y}_i=0,i=1,2,\cdots,l)=\prod_{i=1}^{l}(1-p_i)^{EM_i}$，所以利用 Kuhn-Tucker 条件即求得

$$p_0=\sum_{i=1}^{l}\frac{|M_i|}{|M|}\left(1-\frac{EM_i}{|M_i|}|M|\cdot\exp\left(\frac{\log(1-\alpha)-\sum_{i=1}^{l}EM_i\log(|M|(EM_i/|M_i|))}{\sum_{i=1}^{l}EM_i}\right)\right)\tag{2}$$

因此在times函数确定的正则假设下IUT的可靠性处于区间[1−$p_0$，1)的概率为 100×(1−$\alpha$)%。式(2)经化简可以表示为

$$p_0=1-(1-\alpha)^{(\sum_{i=1}^{l}EM_i)^{-1}}\cdot\exp\left(-\frac{\sum_{i=1}^{l}\left(EM_i\cdot\log\left(\frac{|M|}{|M_i|}\cdot\frac{EM_i}{\sum_{i=1}^{l}EM_i}\right)\right)}{\sum_{i=1}^{l}EM_i}\right)\tag{3}$$

由式(3)知，当测试用例总数一定时，如果对于 $\forall\ i\in[1,\ w]$ 有等式 $\dfrac{EM_i}{\sum_{i=1}^{l}EM_i}=\dfrac{|M_i|}{|M|}$ 成立，则此时$p_0$达到最小值 $1-(1-\alpha)^{(\sum_{i=1}^{l}EM_i)^{-1}}$，因此上面的这个等式在“零错误”条件下可以从增加可靠性的角度指导测试用例的选取。

需要注意的是式(1)和式(3)中的$|M_i|/|M|$可以用一个权值$w_i$替换，$w_i$可以根据测试的实际情况进行调整。

### 4.2　情况 2

本节分析当 $N1,\cdots,Nn$ 很大时的情况。

由性质 1 和性质 3 知 $N1,\cdots,Nn$ 越大，测试的覆盖范围越广，但由于性质 2，使得当 $N1,\cdots,Nn$ 很大时即使只采用正则假设，抽象测试用例的个数也是非常庞大的。为了做到测试质量与测试开销的平衡，可以分两种情况处理：(1) 如果测试集能按某种方式划分成有限个子类(比如按照功能划分等)，则从每个子类中随机抽取出若干测试用例进行测试，直到测试预算耗尽；(2) 如果测试集不能划分，则将整个测试集视为一个整体，并从中随机抽取测试用例进行测试，直到测试预算耗尽。对情况(1)，假设每个子类中的测试用例出错概率相同或者设定一个平均出错概率，则这个模型与上一节相似，这里不再讨论，下面我们讨论情况(2)。

设测试集 $T$ 中每个测试用例的平均出错概率为 $p$，并设 $\hat{p}$ 是 $p$ 的估计，则在测试时第一个错误出现在第 $i$ 条测试用例的概率是 $p\times(1-p)^{i-1}$，整个实际测试集的出错概率是

$\sum_{i=1}^{N} p(1-p)^{i-1}$，其中 $N$ 是从 $T$ 中实际选出的测试用例个数。因为当 $0<p<1$ 时有 $\sum_{i=1}^{N} p(1-p)^{i-1} = 1-(1-p)^N$ 成立，由此可知只要测试用例的出错概率不为零，则当$N$足够大时错误必然会发生。令 $X_i$ 表示第 $i$ 条测试用例的执行情况，$X_i = \begin{cases} 0, \text{pass} \\ 1, \text{fail} \end{cases}$，其中 pass/fail 表示公式的执行结果，则由极大似然估计或矩估计都可以得到 $\hat{p} = \frac{1}{N} \cdot \sum_{i=1}^{N} X_i$。又因为 $\hat{p}$ 的均值 $E(\hat{p}) = \frac{1}{N} \sum_{i=1}^{N} E(X_i) = p$，所以 $\hat{p}$ 是 $p$ 的无偏估计。由中心极限定理知在大样本条件下 $\frac{1}{\sqrt{N \cdot p \cdot (1-p)}} \cdot \left( \sum_{i=1}^{N} X_i - N \cdot p \right)$ 服从标准正态分布，所以

$$P\left( -\mu_{\alpha/2} \le \frac{1}{\sqrt{N \cdot p \cdot (1-p)}} \left( \sum_{i=1}^{N} X_i - N \cdot p \right) \le \mu_{\alpha/2} \right) = 1-\alpha \tag{4}$$

其中 $\mu_{\alpha/2}$ 是标准正态分布的上 $\alpha/2$ 分位点，解式(4)，因此 $p$ 的区间估计是[$A$，$B$]，其中

$$A, B = \frac{N}{N+\mu_{\alpha/2}^2} \left( \hat{p} + \frac{\mu_{\alpha/2}^2}{2N} \pm \mu_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{N} + \frac{\mu_{\alpha/2}^2}{4N^2}} \right),$$

$A$ 取负号，$B$ 取正号 (5)

区间的置信系数为 $1-\alpha$。若所有样本都通过测试，即发生“零错误”，此时 $\frac{1}{N} \cdot \sum_{i=1}^{N} X_i = 0$，则可靠性问题转化为求 $p$ 的置信上界，因此式(4)被改写成

$$P\left( \frac{1}{\sqrt{N \cdot p \cdot (1-p)}} \left( \sum_{i=1}^{N} X_i - N \cdot p \right) \le \mu_{\alpha} \right) = 1-\alpha \tag{6}$$

解式(6)，得到 $p$ 的置信上界为 $\frac{\mu_\alpha^2}{N+\mu_\alpha^2}$，即在样本都通过测试的情况下，$T$ 中测试用例的平均出错概率的置信区间为 $\left[0, \frac{\mu_\alpha^2}{N+\mu_\alpha^2}\right]$，置信系数为 $1-\alpha$，所以在 times 函数确定的正则假设下 IUT 的可靠性处于区间 $\left[1-\frac{\mu_\alpha^2}{N+\mu_\alpha^2}, 1\right]$ 的概率为 $100\times(1-\alpha)\%$。另外从这个结果可以得到：在“零错误”条件下，IUT 的可靠性的置信区间只与测试用例的总数 $N$ 和系数 $\alpha$ 相关。

## 5 结束语

基于形式规范的测试，可以将现有的各种协议一致性测试方法生成的测试用例，统一成穷尽测试集加上一些特定的测试假设。本文用统计的方法分析了基于形式规范的协议一致性测试在均一假设下的可靠性，并计算出这个可靠性所能达到的置信区间。最后特别指出：当函数调用次数不是很大时，“零错误”条件下的可靠性在 $\frac{EM_i}{\sum_{i=1}^{l} EM_i} = \frac{|M_i|}{|M|}$ 时达到最大，其置信区间是[$1-p_0$，1)，置信系数是 $1-\alpha$，其中 $p_0 = 1-(1-\alpha)^{(\sum_{i=1}^{l} EM_i)^{-1}}$；当函数调用次数很大时，“零错误”条件下的可靠性只与测试用例的个数 $N$ 和系数$\alpha$相关。

## 参 考 文 献


[1] Bernot G, Gaudel MC, and Marre B. Software testing based on formal specifications: a theory and a tool. *Software Engineering Journal*, 1991, 6(6): 387–405.

[2] Marie-Claude Gaudel and Perry R James. Testing algebraic data types and processes. *Formal Aspects of Computing*, 1998, 10(5-6): 436–451.

[3] 赵静, 屈玉贵, 赵保华. 一种基于RSL的协议形式化描述技术的研究. 计算机科学, 2003, 30(1): 97–99.

[4] Hagwood C, Kacker R, Yen J, Banks D, Rosenthall L, Gallagher L, and Black P. Reliability of conformance tests. Computer Software and Applications Conference, Vinenna, Aug. 1998.

[5] Hagwood C and Lynne Rosenthal. Reliability of conformance tests. *IEEE Trans. on Reliability*, 2001, 50(2): 204–208.



吕欣岩： 男，1976年生，博士生，研究方向为通信软件测试理论与方法.

赵保华： 男，1947年生，教授，博士生导师，研究方向为协议理论与工程.

屈玉贵： 女，1945年生，教授，博士生导师，研究方向为计算机体系结构、协议理论与工程.