Fault-based Test Case Generation for Component Connectors

Sun Meng

CWI, Amsterdam, The Netherlands
http://www.cwi.nl/~sun

joint work with
Bernhard K. Aichernig (TU Graz)
Farhad Arbab, Lacramioara Astefanoaei, Frank S. de Boer, Jan Rutten (CWI)

CIC workshop, Braga, Portugal, May 7-8, 2009
Outline

Introduction

Connectors as Designs

Fault-based Test Case Generation

Conclusion
Problem

- The complexity and importance of coordination models in service-oriented applications necessarily lead to a higher relevance of testing issues for connectors during development of systems.
- Aim of testing: to show conformance or non-conformance of the final software system with some requirements or specifications.
- The behavior of connectors generally describes the manifold interactions among components / services rather than simple input-output behavior.
- we can use not simply sequences of input and output, but relations on different input / output sequences as test cases for connectors.
An Example

• For connectors, faults are caused by possible errors during the development process, such as wrongly used channels, missing or redundant subcircuits, or circuits with wrongly constructed topology, etc.

The intended connector: The faulty connector:

• In the faulty connector, the topology is changed and the sequence of values that appear through C consist of zero or more repetitions of the pairs of values written to B and A, in a different order, comparing with the intended one.
What and How?

• The idea: to automatically generate a test case that can detect such errors and exclude faulty implementations of connectors.

• The specification and implementation of connectors are given by pairs of pre- and post-condition.

• Every connector is given by a pair of predicates $P \vdash Q$ where
  • the assumption $P$ is what the designer can rely on when the communicating operation is initiated by inputs to the connectors, and
  • the commitment $Q$ must be true for the outputs when the communicating operation terminates.

• Test cases will detect certain faults in connectors by using refinement calculus.
Timed Data Sequences

- For an arbitrary connector $R$, the relevant observations come in pairs, with one observation on the source nodes of $R$, and one observation on the sink nodes of $R$. For every node $N$, the corresponding observation on $N$ is given by a timed data sequence.

- Let $TS = \{ a \in \mathbb{R}_+^* | \forall 0 \leq n < |a|.a(n) < a(n+1) \}$ and $DS = D^*$ be the set of time sequences and data sequences, the set of timed data sequences is defined by $TDS \subseteq DS \times TS$ that contains pairs $\langle \alpha, a \rangle$ consisting of $\alpha \in DS$ and $a \in TS$ with $|\alpha| = |a|$.

- Timed data sequences can be alternatively and equivalently defined as (a subset of) $(D \times \mathbb{R}_+)^*$ because of the existence of isomorphism

$$\langle \alpha, a \rangle \mapsto (\langle \alpha(0), a(0) \rangle, \langle \alpha(1), a(1) \rangle, \langle \alpha(2), a(2) \rangle, \cdots )$$
Design

- For connectors, a design $P \vdash Q$ has the following meaning:

  $$P \vdash Q = df (ok \land P \Rightarrow ok' \land Q)$$

  where $ok$ and $ok'$ are two variables being used to analyze explicitly the phenomena of communication initialization and termination.

- The variable $ok$ stands for a successful initialization and the start of a communication. When $ok$ is $\text{false}$, the communication has not started, so no observation can be made.

- The variable $ok'$ denotes the observation that the communication has terminated. The communication is divergent when $ok'$ is $\text{false}$. 
Connectors as Designs

- To specify input and output explicitly, for a connector $R$, we use $in_R$ and $out_R$ for the lists of timed data sequences on the input ends and output ends of $R$ respectively.
- Every connector $R$ can be represented as

$$R(in : in_R; out : out_R)$$
$$P(in_R) \vdash Q(in_R, out_R)$$

- $P(in_R)$ is the precondition that should be satisfied by inputs $in_R$ on the source nodes of $R$, and
- $Q(in_R, out_R)$ is the postcondition that should be satisfied by outputs $out_R$ on the sink nodes of $R$.
- A predicate $D$ is used to denote well-defined timed data sequence types. In other words, we define the behavior only for valid sequences expressed via a predicate $D$. 
Some Examples

- **Synchronous channel**: 
  \[
  \text{con} : \text{Sync}(\text{in} : (\langle \alpha, a \rangle); \text{out} : (\langle \beta, b \rangle))
  \]
  \[
  P : \mathcal{D}\langle \alpha, a \rangle
  \]
  \[
  Q : \mathcal{D}\langle \beta, b \rangle \land \beta = \alpha \land b = a
  \]

- **FIFO1 channel**: 
  \[
  \text{con} : \text{FIFO1}(\text{in} : (\langle \alpha, a \rangle); \text{out} : (\langle \beta, b \rangle))
  \]
  \[
  P : \mathcal{D}\langle \alpha, a \rangle
  \]
  \[
  Q : \mathcal{D}\langle \beta, b \rangle \land \beta = \alpha \land
  a < b \land (\text{tail}(b^R))^R < \text{tail}(a)
  \]

- **Synchronous drain**: 
  \[
  \text{con} : \text{SyncDrain}(\text{in} : (\langle \alpha, a \rangle, \langle \beta, b \rangle); \text{out} : ( ))
  \]
  \[
  P : \mathcal{D}\langle \alpha, a \rangle \land \mathcal{D}\langle \beta, b \rangle \land a = b
  \]
  \[
  Q : \text{true}
  \]
Connector Composition

- Three Composition Patterns:

  1. **Pattern 1**:
     - $R_i \{ \text{in} : \text{in}_{R_i}; \text{out} : \text{out}_{R_i} \}
     - $P_i(\text{in}_{R_i}) \vdash Q_i(\text{in}_{R_i}, \text{out}_{R_i})$

     where $i = 1, 2$ to denote the two connectors being composed.
Flow-through Composition

- For $R_1$ and $R_2$, suppose one sink node of $R_1$ and one source node of $R_2$ are joined together into a mixed node $B$. Such mixed nodes are hidden (encapsulated) by using existential quantification.

- If one sink node of $R_1$ and one source node of $R_2$ are joined together into a mixed node $B$, $\langle \beta_1, b_1 \rangle \in \text{out}_{R_1}$ and $\langle \beta_2, b_2 \rangle \in \text{in}_{R_2}$ are the timed data sequences on $B$ in $R_1$ and $R_2$ respectively, for two predicates $P$ and $Q$,

$$P(\langle \beta_1, b_1 \rangle); Q(\langle \beta_2, b_2 \rangle) = \exists \langle \beta, b \rangle. P(\langle \beta, b \rangle) \land Q(\langle \beta, b \rangle)$$
Let $\langle \beta_1, b_1 \rangle \in \text{out}_{R_1}$ and $\langle \beta_2, b_2 \rangle \in \text{in}_{R_2}$ be the timed data sequences on the node $B$ in $R_1$ and $R_2$, respectively. Then the resulting connector is

$$R(\text{in} : ( \bigcup_{i=1,2} \text{in}_{R_i}) \setminus \{\langle \beta_2, b_2 \rangle\}; \text{out} : ( \bigcup_{i=1,2} \text{out}_{R_i}) \setminus \{\langle \beta_1, b_1 \rangle\})$$

$P \vdash Q$

where

$$P = P_1 \land \neg (Q_1(\langle \beta_1, b_1 \rangle); \neg P_2(\langle \beta_2, b_2 \rangle))$$

$$Q = Q_1(\langle \beta_1, b_1 \rangle); Q_2(\langle \beta_2, b_2 \rangle)$$
Merging Sink Nodes

- In (2), let \( \langle \gamma_i, c_i \rangle \in out_{R_i}, \ i = 1, 2, \) be the timed data sequences on the node \( C \) in \( R_1 \) and \( R_2 \), respectively. Then the resulting connector is

\[
R(in : \bigcup_{i=1,2} in_{R_i}); out : (\bigcup_{i=1,2} (out_{R_i} \setminus \{\langle \gamma_i, c_i \rangle\})) \cup \{\langle \gamma, c \rangle\})
\]

\[P \vdash Q\]

where

\[
P = \bigwedge_{i=1,2} P_i(in_{R_i})
\]

\[Q : \exists \langle \gamma_1, c_1 \rangle, \langle \gamma_2, c_2 \rangle. (\bigwedge_{i=1,2} Q_i(in_{R_i}, out_{R_i})) \wedge M(\langle \gamma_1, c_1 \rangle, \langle \gamma_2, c_2 \rangle, \langle \gamma, c \rangle)\]
Merging Sink Nodes

In the definition, the ternary relation $M$ is defined as:

\[
M(\langle \gamma_1, c_1 \rangle, \langle \gamma_2, c_2 \rangle, \langle \gamma, c \rangle)
\]

\[
= \begin{cases} 
\langle \gamma, c \rangle = \langle \gamma_1, c_1 \rangle & \text{if } |\langle \gamma_2, c_2 \rangle| = 0 \\
\langle \gamma, c \rangle = \langle \gamma_2, c_2 \rangle & \text{if } |\langle \gamma_1, c_1 \rangle| = 0 \\
c_1(0) \neq c_2(0) \land \\
\gamma(0) = \gamma_1(0) \land c(0) = c_1(0) \land \\
M(\langle \gamma'_1, c'_1 \rangle, \langle \gamma_2, c_2 \rangle, \langle \gamma', c' \rangle) & \text{if } c_1(0) < c_2(0) \\
\gamma(0) = \gamma_2(0) \land c(0) = c_2(0) \land \\
M(\langle \gamma_1, c_1 \rangle, \langle \gamma'_2, c'_2 \rangle, \langle \gamma', c' \rangle) & \text{if } c_2(0) < c_1(0) \\
\text{otherwise}
\end{cases}
\]
Merging Source Nodes

In (3), Let $\langle \alpha_i, a_i \rangle \in \text{in}_{R_i}$, $i = 1, 2$, be the timed data sequences on the node $A$ in $R_1$ and $R_2$, respectively. Then the resulting connector is

$$R(\text{in} : ( \bigcup_{i=1,2} \text{in}_{R_i} \setminus \{ \langle \alpha_i, a_i \rangle \}) \cup \{ \langle \alpha, a \rangle \}; \text{out} : \bigcup_{i=1,2} \text{out}_{R_i})$$

$$P \vdash Q$$

where

$$P = \bigwedge_{i=1,2} P_i(\text{in}_{R_i})[\langle \alpha, a \rangle / \langle \alpha_i, a_i \rangle]$$

$$Q = \bigwedge_{i=1,2} Q_i(\text{in}_{R_i}, \text{out}_{R_i})[\langle \alpha, a \rangle / \langle \alpha_i, a_i \rangle]$$

For a predicate $P$, if $v$ is a variable in $P$, $P[u/v]$ is the predicate obtained by replacing all occurrence of $v$ in $P$ by $u$. 
Example

\textbf{Order}(in : (⟨α, a⟩, ⟨β, b⟩); out : ⟨γ, c⟩)

P \vdash Q

P = D⟨α, a⟩ \land D⟨β, b⟩ \land a = b

Q = D⟨γ, c⟩ \land ∃(γ_1, c_1), (γ_2, c_2).D⟨γ_1, c_1⟩ \land D⟨γ_2, c_2⟩ \land

\quad γ_1 = α \land a < c_1 \land ((c_1^R)')^R < a' \land γ_2 = β \land c_2 = b \land

M(⟨γ_1, c_1⟩, ⟨γ_2, c_2⟩, ⟨γ, c⟩)
**Example**

\[ \text{EXRouter}(\text{	extit{in}} : (⟨α, a⟩); \text{	extit{out}} : (⟨β, b⟩, ⟨γ, c⟩)) \]

\[ P \vdash Q \]

\[ P = D⟨α, a⟩ \]

\[ Q = D⟨β, b⟩ \land D⟨γ, c⟩ \land L(⟨α, a⟩, ⟨β, b⟩) \land L(⟨α, a⟩, ⟨γ, c⟩) \land M(⟨β, b⟩, ⟨γ, c⟩, ⟨α, a⟩) \]

where

\[ L(⟨α, a⟩, ⟨β, b⟩) \equiv ⟨β, b⟩ = ( ) \lor \]

\[ \begin{cases} a(0) \leq b(0) \land \{ α(0) = β(0) \land L(⟨α', a'⟩, ⟨β', b'⟩) \quad \text{if} \; a(0) = b(0) \} \\ L(⟨α', a'⟩, ⟨β, b⟩) \quad \text{if} \; a(0) < b(0) \} \]
Refinement

- Implication of predicates establishes a refinement order over connectors. Thus, more concrete implementations imply more abstract specifications. For two connectors

\[
R_i(in : in_{R_i}; out : out_{R_i})
\]

\[
P_i(in_{R_i}) \vdash Q_i(in_{R_i}, out_{R_i})
\]

where \(i = 1, 2\), if \(in_{R_1} = in_{R_2}\) and \(out_{R_1} = out_{R_2}\), then

\[
R_1 \sqsubseteq R_2 \equiv (P_1 \Rightarrow P_2) \land (P_1 \land Q_2 \Rightarrow Q_1)
\]

- In other words, preconditions on inputs of connectors are weakened under refinement, and postconditions on outputs of connectors are strengthened.
Faulty Connectors

Definition
Given an intended connector specification

\[ R(in : in_R; out : out_R) \]
\[ P(in_R) \vdash Q(in_R, out_R) \]

and a connector implementation

\[ R'(in : in_R; out : out_R) \]
\[ P'(in_R) \vdash Q'(in_R, out_R) \]

with the same input and output nodes as in \( R \), which may contain an error. \( R' \) is called a faulty connector if and only if \( R \not\subseteq R' \).
Test Cases

- For connectors, we consider test cases as specifications that define the expected list of timed data sequences on the output nodes for a given list of timed data sequences on the input nodes.

Definition (deterministic test case)
For a connector $\mathbf{R}(\text{in} : \text{in}_R; \text{out} : \text{out}_R)$, let $i$ be the input vector and $o$ be the output vector, both lists of timed data sequences with the same lengths as $\text{in}_R$ and $\text{out}_R$ respectively. A deterministic test case for $\mathbf{R}$ is defined as

$$td(\text{in}_R, \text{out}_R) = \text{in}_R = i \vdash \text{out}_R = o$$
Test Cases

• Sometimes the behavior of a connector can be non-deterministic. In this case, we can generalize the notion of test case as follows:

Definition (test case)
For a connector \( R(\text{in} : \text{in}_R; \text{out} : \text{out}_R) \), let \( i \) be the input vector and \( O \) be a possibly infinite set containing the expected output vector(s). Both \( i \) and any \( o \in O \) are lists of timed data sequences with the same lengths as \( \text{in}_R \) and \( \text{out}_R \) respectively. A test case for \( R \) is defined as

\[
t(\text{in}_R, \text{out}_R) = \text{in}_R = i \vdash \text{out}_R \in O
\]
Test Cases, Specifications and Implementations

- Test cases, as well as connector specifications and implementations, can be specified by designs.
- An implementation that is correct with respect to its specification should refine its test cases.
- Test cases are abstractions of an implementation if and only if the implementation passes the test cases.
- Taking specifications into consideration, test cases should also be abstractions of a specification if they are properly derived from the specification.
Correctness of Test Cases

Definition
For a connector specification $S$, its implementation $R$ and a test case $t$, which satisfy

$$t \sqsubseteq S \sqsubseteq R$$

- $t$ is called a correct test case with respect to $S$.
- $R$ passes the test case $t$ and conforms to the specification $S$. 
**Fault-Adequate Test Cases**

**Definition**
Let $t(in_R, out_R)$ be a test case (which can be either deterministic or non-deterministic), $R$ an expected connector, and $R'$ its faulty implementation. Then $t$ is a fault-adequate test case if and only if

$$t \sqsubseteq R \land t \not\sqsubseteq R'$$

A fault-adequate test case detects a fault in $R'$. Alternatively we can say that the test case distinguishes $R$ and $R'$. All test cases that detect a certain fault form a fault-adequate equivalence class.
Test Case Generation

- A test case for finding errors in a faulty connector has to,
  - first, be a correct test case of the intended connector;
  - second, it must not be an abstraction of the faulty connector.
Algorithm

Consider an intended connector

\[ R(in : in_R; out : out_R) \]
\[ P(in_R) \vdash Q(in_R, out_R) \]

and its faulty implementation

\[ R'(in : in_R; out : out_R) \]
\[ P'(in_R) \vdash Q'(in_R, out_R) \]

as inputs. A test case \( t \) is generated by the following steps:
Test Case Generation Algorithm

1. A test case $t$ is searched by
   1. find a pair $\langle \hat{i}, \hat{o} \rangle$ as a solution of

   $$P(\hat{i}) \land Q'(\hat{i}, \hat{o}) \land \neg Q(\hat{i}, \hat{o})$$

   2. if it exists, then the test case $t(i, O)$ is generated by finding
      the maximal set $O$ of output vectors, such that for all $o \in O$,
      $$i = \hat{i} \land P(i) \land Q(i, o)$$
      is satisfied.

2. If the previous step does not succeed, then look for a test
   case $t(i, O)$ with $O$ the maximal set of output vectors, such
   that for all $o \in O$,
   $$\neg P'(i) \land P(i) \land Q(i, o)$$
   is satisfied.
Correctness

**Theorem (Correctness)**

*Given an intended connector*

\[ R(in : in_R; out : out_R) \]
\[ P(in_R) \vdash Q(in_R, out_R) \]

*its faulty implementation*

\[ R'(in : in_R; out : out_R) \]
\[ P'(in_R) \vdash Q'(in_R, out_R) \]

*and a test case \( t(i, O) \) generated by the algorithm,*

\[ t(i, O) \sqsubseteq R \]
Correctness

Proof: The proof is divided into two cases, corresponding to steps 1 and 2 of the algorithm, respectively.

1. Since $O$ is the maximal set such that for all $o \in O$, $P(i) \land Q(i, o)$ is satisfied, we have

$$t(i, O) \subseteq R$$

$$\equiv \{ \text{Definition of refinement and non-deterministic test cases } \}$$

$$(in_R = i \Rightarrow P(in_R)) \land (Q(in_R, out_R) \land in_R = i \Rightarrow out_R \in O)$$

$$\equiv \{ O \text{ is the maximal set such that } P(i) \land Q(i, o) \text{ is satisfied for all } o \in O \}$$

true $\land$ true

$\equiv$true
2. Since $O$ is the maximal set of output vectors $o$ each of which satisfies $\neg P'(i) \land P(i) \land Q(i, o)$, it follows that $P(i) \land Q(i, o)$ is also satisfied for all $o \in O$ and $O$ is still maximal. By the same style of reasoning as in the first case, we can derive that

$$t(i, O) \sqsubseteq R \equiv \text{true}$$
Fault Coverage

Theorem (Fault Coverage)

Given an intended connector

\[ R(\text{in} : \text{in}_R; \text{out} : \text{out}_R) \]
\[ P(\text{in}_R) \models Q(\text{in}_R, \text{out}_R) \]

its faulty implementation

\[ R'(\text{in} : \text{in}_R; \text{out} : \text{out}_R) \]
\[ P'(\text{in}_R) \models Q'(\text{in}_R, \text{out}_R) \]

and a test case \( t(i, O) \) generated by the algorithm,

\[ t(i, O) \not\subseteq R' \]
Fault Coverage

Proof: This proof is also divided into the two cases of the algorithm.

1. When $O$ is the maximal set such that for all $o \in O$, $i = \widehat{i} \land P(i) \land Q(i, o)$ is satisfied. The test case is generated only if there exists $(\widehat{i}, \widehat{o})$, such that $P(\widehat{i}) \land Q'(\widehat{i}, \widehat{o}) \land \neg Q(\widehat{i}, \widehat{o})$ is satisfied. Thus we have $i = \widehat{i}$, $\widehat{o} \notin O$ and

$$t(i, O) \not\sqsubset R'$$

$\equiv \{\text{Definition of refinement and non-deterministic test cases}\}$

$$\neg (\text{in}_R = i \Rightarrow P'(\text{in}_R)) \lor \neg (\text{in}_R = i \land Q'(\text{in}_R, \text{out}_R) \Rightarrow \text{out}_R \in O)$$

$\equiv (\text{in}_R = i \land \neg P'(\text{in}_R)) \lor$

$$\exists \text{in}_R, \text{out}_R. (\text{in}_R = i \land Q'(\text{in}_R, \text{out}_R) \land \text{out}_R \notin O)$$

$\equiv \{\text{let } \text{in}_R = \widehat{i}, \text{out}_R = \widehat{o}\}$

$$(\text{in}_R = i \land \neg P'(\text{in}_R)) \lor \text{true}$$

$\equiv \text{true}$
Fault Coverage

2. If \( O \) is the maximal set of output vectors \( o \) each of which satisfies \( \neg P'(i) \land P(i) \land Q(i, o) \), then

\[
  t(i, O) \not\subseteq R' \\
  \equiv \{ \text{Definition of refinement and non-deterministic test cases} \} \\
  \neg (\text{in}_R = i \Rightarrow P'(\text{in}_R)) \lor \\
  \neg (\text{in}_R = i \land Q'(\text{in}_R, \text{out}_R) \Rightarrow \text{out}_R \in O) \\
  \equiv \exists \text{in}_R. (\text{in}_R = i \land \neg P'(\text{in}_R)) \lor \\
  \neg (\text{in}_R = i \land Q'(\text{in}_R, \text{out}_R) \Rightarrow \text{out}_R \in O) \\
  \equiv \{ \text{let } \text{in}_R = i \} \\
  \text{true} \lor \neg (\text{in}_R = i \land Q'(\text{in}_R, \text{out}_R) \Rightarrow \text{out}_R \in O) \\
  \equiv \text{true}
\]
The intended connector:  

The faulty connector:

- Generated test case: \(((\langle \alpha, a \rangle, \langle \beta, b \rangle), \langle \gamma, c \rangle))\)
- \(\langle \alpha, a \rangle\) and \(\langle \beta, b \rangle\) are two timed data sequences with length 1, while \(\langle \gamma, c \rangle\) with length 2, and the constraint \(\alpha \neq \beta\) should be satisfied.
Conclusion

- A semantic model of Reo connectors, which makes it possible to specify both finite and infinite behavior.
- Faults are modeled on the same level as the specification and implementation of connectors by using the notion of design.
- Test cases are generated based on a general theory of refinement for pre- and post-condition specifications, and has been instantiated in a prototype developed in Maude.

- more complex case studies,
- testing data selection based on constraint solving and its integration into the generated symbolic test cases, and
- specification-based testing of black-box connectors.
Conclusion

- A semantic model of Reo connectors, which makes it possible to specify both finite and infinite behavior.
- Faults are modeled on the same level as the specification and implementation of connectors by using the notion of design.
- Test cases are generated based on a general theory of refinement for pre- and post-condition specifications, and has been instantiated in a prototype developed in Maude.

- more complex case studies,
- testing data selection based on constraint solving and its integration into the generated symbolic test cases, and
- specification-based testing of black-box connectors.
Thank you!

Questions?