Summary of Dynamic Testing

- Strengths:
  - Microscopic examination of execution details
  - Evaluation in actual runtime environment
  - Oldest approach, most familiar

- Weaknesses:
  - Cannot demonstrate absence of faults
  - Hard to generate test data
  - Hard to know when test sets are adequate
  - Testing aids (e.g. assertion checkers) alter execution

Some References for Testing

- Ghezzi, Jazayeri, Mandrioli, Fundamentals of Software Engineering, Chapter 6.3 (Testing)
  - Excellent annotated bibliography at the end of Ch. 6
- Many of those references are now rather old
- Much recent work in this area
- Proceedings of the International Symposium on Software Testing and Analysis (ISSTA)
- Proceedings of International Conference on Software Engineering (ICSE)
Summary of Problems in Doing Testing Effectively

- Hard to cover program execution space effectively
- Hard to select test data effectively
- Hard to tell if test results are right or wrong
  -- if program computes complex function(s)
  -- if the number of test cases gets large
- Best to detect errors early—before they become faults
- Testing comes at the end of the lifecycle, when time and budget tend to be short

What do you know when testing is done?
What relations have been checked?
How thoroughly?
To whose satisfaction?

Relations and Analysis

- A software product consists of
  - A collection of (types of) artifacts
  - Related to each other by myriad Relations
- The relations are essentially desiderata
  - At least initially
  - Before the product can be trusted, the relations need to be verified/confirmed
- That is the role of analysis
  - Does the software do what it is supposed to do?
  - What are its capabilities and its strengths?
  - What is the nature of the artifact(s) that have been built?
  - What can I count on?
  - What should I worry about?

Static Analysis

- Technique for demonstrating the absence of faults without the need for execution
- Specification of Intent: derived from requirements
- Specification of Behavior: derived from model(s)
- Comparison: Done analytically and mathematically
- Results: Theorems about the program (e.g., proofs that certain behaviors are impossible—or mandatory)

Dynamic Testing

- Specification of Intended Behavior
- Required Outputs
- Test Execution Results
- Comparator (Human or Machine)
- Failure Reporting
- Comparison of Behavior to Intent

Testing and Analysis Framework

- Specification of Intended Behavior
- Development Process
- Specification of Actual Behavior
- Evaluation Process
- Testing/Analysis Results
- Comparison of Behavior to Intent (relation checking)
**Static Analysis**

- Specification of Intended Behavior
- Derived From Source Text
- Specification of Actual Behavior
- Mathematical Reasoning
  - Theorems
  - Or Counterexamples
- Comparison of Behavior to Intent

**Dataflow Analysis**

- Specification of Intent: Sequence of events
- Specification of Behavior: Derived from flowgraph model
  - Nodes annotated with events of interest
  - All possible executions modeled as all sequences of events along all flowgraph paths
- Comparison: Analytic process
  - Are all possible event sequences the desired ones?
- Result: Theorems demonstrating absence of event sequence errors
- Examples:
  - No variable referenced before definition
  - No file read before it is opened
  - Elevator doesn’t move until doors are shut
  - Rocket won’t try to fire thrusters after fuel is exhausted

**Use of the Flow Graph (ImmFol Relation) to Determine Actual Behavior**

- Determine local information by looking at local behavior
  - Which artifacts are created, used,
  - Infer global information
- By looking at local information
- And tracing flow graph edges to find relevant information at other nodes

**Documenting Local Behavior**

- First, determine local information that is true at each node in the CFG
  - e.g., What variables are defined
  - What variables are referenced
  - Usually stored in sets
  - e.g., ref(n) is the set of variables referenced at node n
- This same type of information available for pre-code artifacts
  - Interface specifications for modules
  - Input and output specifications of requirements elements

**Natural Complementarity**

- Static Analysis excels at scanning for trouble
- Dynamic testing excels at focusing in sharply
- Both aim to identify violations of well-formedness properties
- Techniques should be mutually supportive, synergistic
- More on this later

**Standard Propagation Rules**

- Need to propagate information from IMMFLOL neighbors
  - Backwards or forwards
- Must take branching into account
- Must add in "local" information
- What is happening at this node
Inferring Global Behavior Example: Reaching Definitions

- Definition reaches a node if there is a def clear path from the definition to that node.
- Do this by tracing paths between nodes.
- Definition of x at node 1 reaches nodes 2, 3, 4, 5 but not 6.

Originally used to determine data dependencies, for debugging, data flow testing, etc.

Another Example: Live Variables

- A variable, x, is live at node p if there exists a def-clear path for x from node p to a use of x.
- x is live at 2, 3, 4, 5 but not at node 6.

This is best done by tracing backwards.

Different Global Behavior Problems Require Different Propagation Rules

- Some use forward tracing.
- Some use backward tracing.
- Some require ANDing.
- Some require ORing.
- Some compute ALL results.
- Some compute SOME results.
- Some compute NONE results.
- And various combinations of the above.

“Might” or “Must” Reaching Behavior

Definitions that might reach a node

Definitions that must reach a node
Finite State Verification: Using Data Flow Analysis to Verify Properties

- Data flow analysis used to infer behavior
- Requires use of propagation among nodes
  - Multiple propagations may be needed
- Inferred behavior can then be compared to desired
  - Often comes from Requirements spec.
    - E.g. robustness specs.
- Works on any graph—not just graphs derived from code
  - E.g. DFGs in requirements or architectures

Examples of Properties

- Language specific
  - No undefined references
  - Can’t write to a file until it is opened
- Specific to a system
  - Can’t get money until pwd is OK
  - Can’t move elevator with doors open
  - Etc.

Some properties are much more complicated.
Longer sequences of events

Approach

- Define a desired (or undesired) property as a FSA
  - Based on identification of an alphabet of significant events
- Trace all paths through a graph that models the software
  - Where some nodes are annotated as the sites of events of interest
- Use the FSA as an acceptor
  - All paths from start to end through the model should reach an accepting state

Using FSAs to Define: The “Safe Elevator” Property

A Property, P: The elevator does not move while its doors are open.

Events: (open, close, move)

$L(P)$ is the set of all strings accepted by property P

All traces through the system should satisfy this property (drive FSA to accepting state)

Error: “trap” state
Using Quantified Regular Expressions

- Alphabet, quantification, regular expression

For the events {open, close, move} show that for all paths

((close ∨ move)*, (open +, close)*), open*

Event Stream Comes from Traces through an annotated CFG

- CFG G is (N, Pinit, Pfinal, E)
- Use annotation relation to associate events with nodes
  - \( \Sigma \) is the alphabet of G
  - Events must be indivisible wrt other events in the property
- \( L(G) \) is the language of G
  - The set of all strings in \( (\Sigma^*) \) that occur on paths from the initial node to the final node

Static Dataflow Analysis

Specification of Intended Behavior

Event Sequences

Possible Execution Sequences

Dataflow Propagation Algorithms

Proofs of the Presence or Absence of Faults

Comparison of Behavior to Intent

Quality Determination Architectural Framework

Specification of Actual Behavior

Development (Synthesis) Process

Evaluation (Analysis) Process

Testing/Analysis Results

Comparison of Behavior to Intent (relation checking)

Characteristics of Dynamic Analysis

- Dynamic analysis approaches are based on sampling the input space
  - Infer behavior or properties of a system from executing a sample of test cases
  - Functional (Black Box) and Structural (White Box) approaches
- Allows very detailed examination of one execution
  - But nothing about any other execution
  - Cannot prove the absence of faults, only the presence of failures

Informal Dynamic Testing

Specification of Intended Behavior

Test Cases

Evaluation Process

Execution Results

Comparison of Behavior to Intent

Failures Observed
Automated Dynamic Testing

- Specification of Intended Behavior
- Rigorous Specs. (Asserts)
- Evaluation Process
- Specification of Actual Behavior
- Execution Results
- Automated Comparator
- Failures Reported

Comparison of Behavior to Intent

Move from Dynamic Analysis to Static Analysis

- Dynamic analysis approaches are based on sampling the input space
  - Infer behavior or properties of a system from executing a sample of test cases
  - Black Box versus White Box approaches
- Static analysis approaches tend to be based on a “global” assessment of the behavior
  - Based on an understanding of the program (artifact)
  - Again, usually must approximate the semantics to keep the problem tractable
  - Can prove the absence of faults

Static Analysis

- Specification of Intended Behavior
- Assertions, Theorem Statements
- Evaluation Process
- Specification of Actual Behavior
- Annotated Flow Graph
- Rigorous Formal Reasoning, Inference
- Proofs of theorems About absence or presence of specific kinds of faults

Comparison of Behavior to Intent

Static Analysis Approaches

- Data Flow Analysis
- Concurrency Analysis
- Symbolic Evaluation
- Formal Verification

Static Analysis Approaches

- Data Flow Analysis
- Concurrency Analysis
- Symbolic Evaluation
- Formal Verification

Architecture of a Static Finite State Verification System

- Event alphabet
- Reasoning Engine
- Consistent
- Inconsistent + counter example
- System Translator
- Annotated CFG
- FSA
- Property Translator
- Property
### Simple Example: Actual Behavior

1: if (stopped at call_floor) then
2: open;
end if;
...
3: if (stopped at non_call_floor) then
4: close;
end if;
5: move;
...

### Intended Behavior

1: close
2: open
3: move
4: close
5: move

### Verification of consistency between system and requirement

- **System**: Event alphabet
  - Ada, Java, C++, Jovial
- **Property**: Often a robustness requirement
- **Translator**: Consistent
- **Annotated CFG**: Inconsistent + counter example

### Determining consistency between functionality and robustness requirement

- **Worklist**: X, X, X, X

#### Violation Path

1: if
2: open
3: if
4: close
5: move
Appears to be Unexecutable: A “False Positive”

But is it really necessary to modify the program?

When maybe the problem is with the precision of the analysis
Produced this Result

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1: if</td>
<td>(1)</td>
</tr>
<tr>
<td>2: open</td>
<td>(2)</td>
</tr>
<tr>
<td>3: if</td>
<td>{1,2}</td>
</tr>
<tr>
<td>4: close</td>
<td>(1)</td>
</tr>
<tr>
<td>5: move</td>
<td>(1,3)</td>
</tr>
</tbody>
</table>

With this Spurious Report of a Fault

Spurious because the fault is on a path that is not executable

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1: if</td>
<td>(1)</td>
</tr>
<tr>
<td>2: open</td>
<td>(2)</td>
</tr>
<tr>
<td>3: if</td>
<td>{1,2}</td>
</tr>
<tr>
<td>4: close</td>
<td>(1)</td>
</tr>
<tr>
<td>5: move</td>
<td>(1,3)</td>
</tr>
</tbody>
</table>

How to address these kinds of problems attributable to lack of precision?

Improving Precision

Constraints

- Are represented as FSAs
- Describe conditions necessary for feasible execution
- Have a special violation state that is entered when an infeasible path is detected
  - Violation is a trap state; once it is entered, never leave that state

Can add constraints incrementally as needed
How do constraints affect the data flow equations

- IN and OUT are now sets of tuples of FSA states
- Merge is still union
- Transfer function now has to look at each FSA state in the input tuple when computing the output tuple
- Property states not propagated when any constraint automaton is in the violation state
- Result looks at paths that are feasible with respect to the constraints
  - The property state is the same as before
  - Every constraint must be in an accepting state
Concurrent and Distributed Systems

- Extremely difficult to develop and test such systems
- Non-determinism means that
  - the same inputs might produce different outputs on different executions
  - When reasoning about a system there are numerous different alternatives to consider
    - Usually more than a programmer can reasonably consider
- In addition to the problems that can arise with sequential programs, have problems that are unique to concurrent systems
  - Data access problems
  - Synchronization problems

Power and Value of Static Analysis is Best Seen When Applied to Concurrent Systems

- Concurrent software is much harder, more complicated
- Requires very powerful analysis
- Finite state verification shines in this domain

Concurrent systems are complex

- Non-determinism means that
  - the same inputs might produce different outputs on different executions
  - When reasoning about a system there are numerous alternatives to consider
    - Usually more than a human can reasonably consider
- In addition to the problems that can arise with sequential programs, have problems that are unique to concurrent systems
  - Data access problems
  - Synchronization problems
Data Access Anomalies

- Typically you want mutual exclusion
  - shared resource (e.g., data) that should only have a single access at a time
  - e.g., don’t want two travel agents assigning the last seat on a plane
- Typically you don’t want race conditions
  - order of execution affects results
  - undesirable nondeterminism

Typically you want two travel agents assigning the last seat on a plane.

Reachability-based Model Checking: some history

- Originally proposed for hardware
- Early 80’s: E. Clarke and Emerson;
  - L’emann and Sifakis
- Late 80’s: Improved algorithms and property notations (E. Clarke, Emerson, Sistla)
- 90’s: Symbolic Model Checking (SMV) and other optimizations
  - Burch, E. Clarke, Dill, Long, and McMillan
- Current: Hybrid approaches that combine model checking with
  - Theorem proving techniques
  - Symbolic execution
  - Optimization techniques (e.g., points to analysis)

Interleaved Model of Execution; Examples

- task A
  - x := x + 1;
  - write x;
- task B
  - x := x - 1;
  - write x;

Reachability Graph

- Typically, each edge represents progress in a single task
  - Multiple concurrent events may be possible, but allowing only single events captures all states and simplifies the graph structure (interleaved execution model)
  - Only have multiple tasks progress when required by the semantics of the programming construct
    - E.g., rendezvous
  - Only contains states that are potentially reachable from the start state
Reachability Graph Example

Reachability graph

(b1,b2)
(q1,q2)
(e1,e2)
(b1,q)
(b2,q)

clarity

task control flow graphs
T1 T2

begin
T2.Q
end

begin
Accept Q
end

T1 T2

Reachability Graph Example (clarified)

Reachability graph

(b1,b2)
(q1,q2)
(e1,e2)
(b1,q)
(b2,q)

(b_q means that a task is blocked at q
r(q,q') means that the rendezvous
between q and q' occurs)

Use a worklist to build the reachability graph

Worklist: <b1,b2> <b_q,b2> <b1,b_q> <r(q,q')> <e1,q> <q,e2> <e1,e2>

Task control flow graphs

T1 T2

begin
T2.Q
end

begin
Accept Q
end

T1 T2

Useful in determining possible/impossible synchronizations

Infeasible synchronizations

T1 begin
T2 begin
T1 begin
T2 begin

1 2 5 4 6 8 7 3

1,5
b2, 5
1,b7
3,6
R(2,6)
3,8
4,7
4,8

But this is "precise up to symbolic execution"

Two possible synchronizations

This is a bit harder

begin
T2.Q
end

begin
T2.Q
end

begin
T2.Q
end

begin
Accept Q
end

begin
Accept Q
end

synchronizations

1.8
1.6
1.2
1.3
1.5
3.9
3.4
5.9
2.8
4.3
2.5
2.3
3.6
4.8
5.4
9.8
3
5
Evaluation of Static Analysis

- **Strengths:**
  - Can demonstrate the absence of faults
  - Proofs can be automatically generated and proven
  - Algorithms are fast (low-order polynomial)
  - No need to generate test data
  - You know when you are done

- **Weaknesses:**
  - Behavior specification is a model with inaccuracies
    - Not all paths are executable
  - Only certain classes of faults analyzable
    - Mostly sequence specific
    - Weak on functionality

Symbolic Execution

- **Specification of Intent:** Formulae, functions
- **Specification of Behavior:** Functions derived from annotated flowgraph, symbol table
  - Annotate nodes with function(s) computed there
  - Specify path to be studied
  - Compute function(s) computed as composition(s) of functions at path nodes, constraints of path edges
    - Comparison: Solving simultaneous constraints; symbolic algebra
- **Results:** Demonstrations that given paths computed the right function(s)

Conclusions

- Finite state verification approaches are improving
- Being used in industry for hardware systems
- With the increasing interest in software security and quality, becoming more widely used for software systems