Testing: Coverage and Structural Coverage

Testing, Quality Assurance, and Maintenance Winter 2017

Prof. Arie Gurfinkel

based on slides by Prof. Marsha Chechik and Prof. Lin Tan



How would you test this program?

floor(x) is the largest integer not greater than x.

```
def Foo (x, y):
    """ requires: x and y are int
    ensures: returns floor(max(x,y)/min(x, y))"""
    if x > y:
        return x / y
    else
        return y / x
```



Testing

Static Testing [at compile time]

- Static Analysis
- Review
 - -Walk-through [informal]
 - -Code inspection [formal]

Dynamic Testing [at run time]

- Black-box testing
- White-box testing

Commonly, testing refers to dynamic testing.



Complete Testing?

Poorly defined terms: "complete testing", "exhaustive testing", "full coverage"

The number of potential inputs are infinite.

Impossible to completely test a nontrivial system

- Practical limitations: Complete testing is prohibitive in time and cost [e.g., 30 branches, 50 branches, ...]
- Theoretical limitations: e.g. Halting problem

Need testing criteria



CFG Exercise (1)

Draw a control flow graph with 7 nodes.

int binary_search(int a[], int low, int high, int target) { /* binary search for target in the sorted a[low, high] */

```
1 while (low <= high) {
```

```
2 int middle = low + (high - low)/2;
```

```
3 if (target < a[middle])
```

```
4 high = middle - 1;
```

```
5 else if (target > a[middle])
```

```
6 low = middle + 1;
```

```
else
```

7 return middle;

```
}
return -1; /* return -1 if target is not
```

```
found in a[low, high] */
```

8

CFG Exercise (1) Solution

Draw a control flow graph with 7 nodes.

int binary_search(int a[], int low, int high, int target) { /* binary search for target in the sorted a[low, high] */

1 while (low <= high) {

```
2 int middle = low + (high - low)/2;
```

- 3 if (target < a[middle])
- 4 high = middle 1;

```
5 else if (target > a[middle])
```

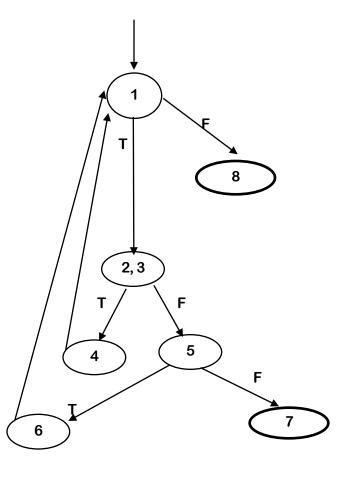
```
low = middle + 1;
```

else

6

```
7 return middle;
```

```
8 return -1; /* return -1 if target is not
found in a[low, high] */
```



Test Case

Test Case: [informally]

- What you feed to software; and
- What the software should output in response.

Test Set: A set of test cases

Test Case: test case values, expected results, prefix values, and postfix values necessary to evaluate software under test

Expected Results: The result that will be produced when executing the test if and only if the program satisfies its intended behaviour



Test Requirement & Coverage Criterion

Test Requirement: A test requirement is a specific element of a software artifact that a test case must satisfy or cover.

- Ice cream cone flavors: vanilla, chocolate, mint
- One test requirement: test one chocolate cone
- TR denotes a set of test requirements

A coverage criterion is a rule or collection of rules that impose test requirements on a test set.

- Coverage criterion is a recipe for generating TR in a systematic way.
- Flavor criterion [cover all flavors]
- TR = {flavor=chocolate, flavor=vanilla, flavor=mint}



Adequacy criteria

Adequacy criterion = set of test requirements

A test suite satisfies an adequacy criterion if

- all the tests succeed (pass)
- every test requirement in the criterion is satisfied by at least one of the test cases in the test suite.

Example:

the statement coverage adequacy criterion is satisfied by test suite S for program P if each executable statement in P is executed by at least one test case in S, and the outcome of each test execution was "pass"



Adequacy Criteria as Design Rules

Many design disciplines employ design rules

- e.g.: "traces (on a chip, on a circuit board) must be at least _____ wide and separated by at least ____"
- "Interstate highways must not have a grade greater than 6% without special review and approval"

Design rules do not guarantee good designs

• Good design depends on talented, creative, disciplined designers; design rules help them avoid or spot flaws

Test design is no different



Where do test requirements come from?

Functional (black box, specification-based): from software specifications

• Example: If spec requires robust recovery from power failure, test requirements should include simulated power failure

Structural (white or glass box): from code

• Example: Traverse each program loop one or more times.

Model-based: from model of system

- Models used in specification or design, or derived from code
- Example: Exercise all transitions in communication protocol model

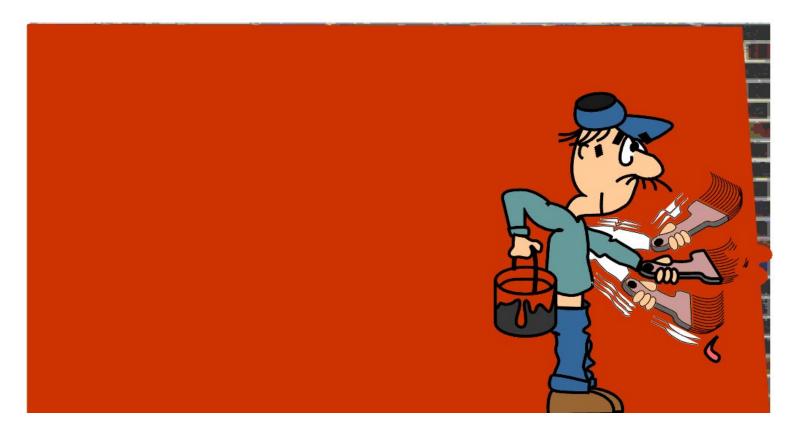
Fault-based: from hypothesized faults (common bugs)

 example: Check for buffer overflow handling (common vulnerability) by testing on very large inputs



Code Coverage

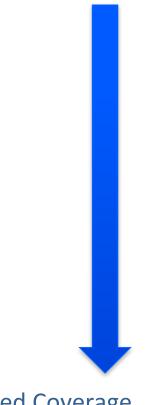
Introduced by Miller and Maloney in 1963





Coverage Criteria

Basic Coverage



Advanced Coverage



- Statement
- Function/Method coverage
- Branch coverage
- Decision coverage
- Condition coverage
- Condition/decision coverage
- Modified condition/decision coverage
- ·Path coverage
- ·Loop coverage

• • • •

Mutation adequacy



Line Coverage

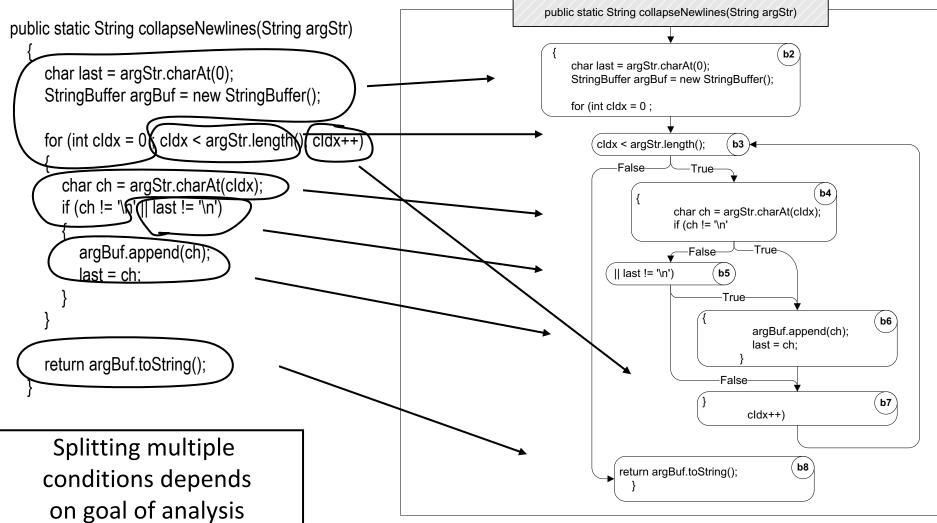
Percentage of source code lines executed by test cases.

- For developer easiest to work with
- Precise percentage depends on layout?
 - -int x = 10; if (z++ < x) y = x+z;
- Requires mapping back from binary?

In practice, coverage not based on lines, but on control flow graph



Deriving a Control Flow Graph





Adequacy criterion: each statement (or node in the CFG) must be executed at least once

```
void foo (int z) {
    int x = 10;
    if (z++ < x) {
        x=+ z;
    }
}</pre>
```

Coverage:

executed statements



Adequacy criterion: each statement (or node in the CFG) must be executed at least once

```
void foo (int z) {
    int x = 10;
    if (z++ < x) {
        x=+ z;
    }
}</pre>
```

```
@Test
void testFoo() {
  foo(10);
}
```

Coverage:

executed statements



Adequacy criterion: each statement (or node in the CFG) must be executed at least once

```
void foo (int z) {
    int x = 10;
    if (z++ < x) {
        x=+ z;
    }
}</pre>
```

```
@Test
void testFoo() {
  foo(10);
}
```

Coverage:

executed statements



Adequacy criterion: each statement (or node in the CFG) must be executed at least once

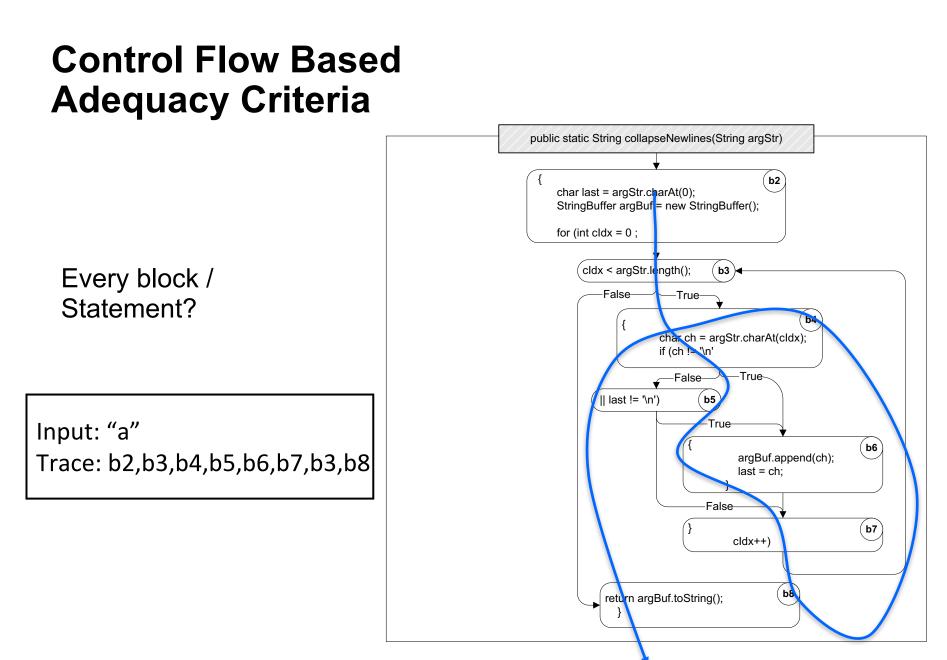
```
void foo (int z) {
    int x = 10;
    if (z++ < x) {
        x=+ z;
    }
}</pre>
```

```
@Test
void testFoo() {
  foo(5);
}
// 100% Statement coverage
```

Coverage Level:

executed statements







Branch / Edge Coverage

Every branch going out of node executed at least once

- Decision-, all-edges-, coverage
- Coverage: percentage of edges hit.

Each branch predicate must be both true and false

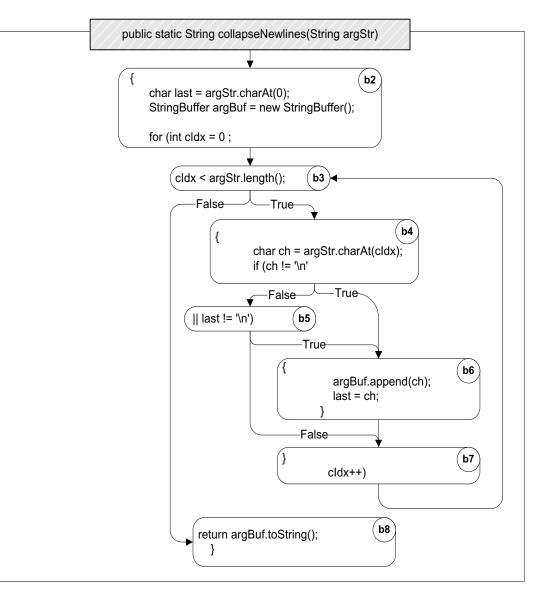


Branch Coverage

One longer input: "a\n\n"

Alternatively:

Block ("a") and "\n" and "\n\n"





Infeasible Test Requirements

if (false)
 unreachableCall();

Real code from the Linux kernel:

while (0)
 {local_irq_disable();}

Statement coverage criterion cannot be satisfied for many programs.



Coverage Level

Given a set of test requirements **TR** and a test set **T**, the *coverage level* is the ratio of the number of test requirements satisfied by **T** to the size of **TR**.

TR = {flavor=chocolate, flavor=vanilla, flavor=mint} Test set 1 T1 = {3 chocolate cones, 1 vanilla cone} Coverage Level = 2/3 = 66.7%

Coverage levels helps evaluate the goodness of a test set, especially in the presence of infeasible test requirements.



Subsumption

Criteria Subsumption: A test criterion C1 *subsumes* C2 if and only if **every** set of test cases that satisfies criterion C1 also satisfies C2

Must be true for every set of test cases



Subsumption is a rough guide for comparing criteria, although it's hard to use in practice.



More powerful coverage criterion helps find more bugs!

int d[2];

Path [N1, N2, N3, N4, N5]: satisfies node coverage but not edge coverage. The corresponding test case passes. No bug found.

Path [N1, N3, N4, N5]: buffer overflow bug!

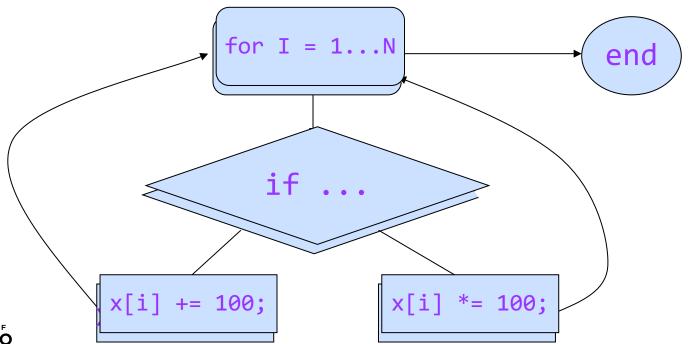


Path Coverage

Adequacy criterion: each path must be executed at least once Coverage:

executed paths

paths

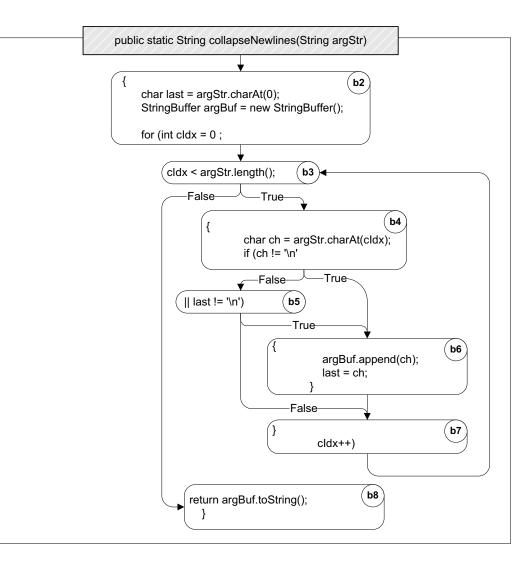




Path-based criteria?

All paths?

Which paths?





if(cond1) f1(); else f2(); if(cond2) f3(); else f4();

How many test cases to achieve branch coverage?



if(cond1) f1(); else f2(); if(cond2) f3(); else f4();

How many test cases to achieve branch coverage?

Two, for example:

cond1: true, cond2: true
 cond1: false, cond2: false



if(cond1) f1(); else f2(); if(cond2) f3(); else f4();

How about path coverage?



```
if( cond1 )
   f1();
else
   f2();
if( cond2 )
   f3();
else
   f4();
```

How about path coverage?

Four:

- 1. cond1: true, cond2: true
- 2. cond1: false, cond2: true
- 3. cond1: true, cond2: false
- 4. cond1: false, cond2: false



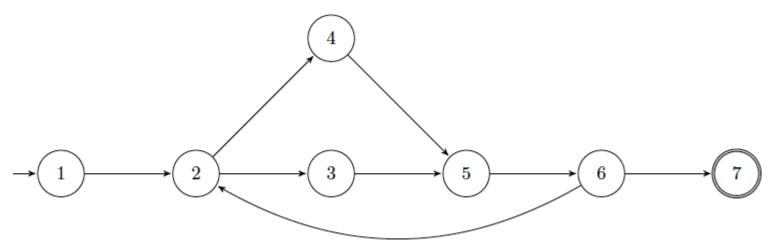
```
if( cond1 )
   f1();
else
   f2();
if( cond2 )
   f3();
else
   f4();
if( cond3 )
   f5();
else
   f6();
if( cond4 )
   f7();
else
   f8();
if( cond5 )
   f9();
else
   f10();
if( cond6 )
   f11();
else
   f12();
if( cond7 )
   f13();
else
   f14();
```

How many test cases for path coverage?

2ⁿ test cases, where n is the number of conditions



Test Path



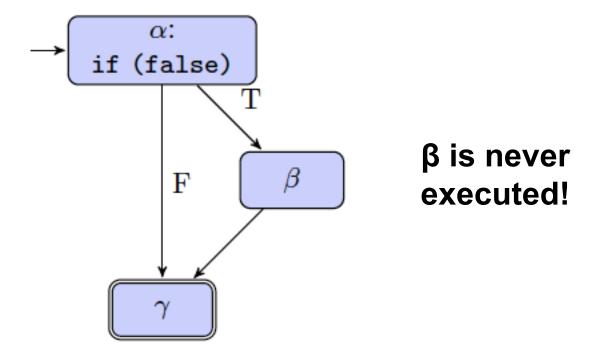
A test path is a path p [possibly of length 0] that starts at some node in N₀ and ends at some node in N_f.

Test path examples:

- [1, 2, 3, 5, 6, 7]
- [1, 2, 3, 5, 6, 2, 3, 5, 6, 7]



Paths and Semantics



Some paths in a control flow graph may not correspond to program *semantics*.

In path coverage, we generally only talk about the *syntax* of a graph -- its nodes and edges -- and not its *semantics*.



Syntactical and Semantic Reachability

A node n is *syntactically* reachable from m if there exists a path from m to n.

A node n is *semantically* reachable if one of the paths from m to n can be reached on some input.

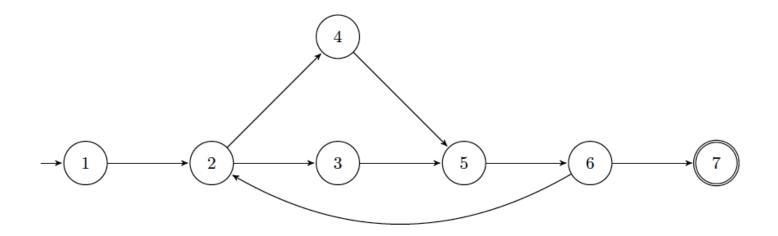
Standard graph algorithms when applied to Control Flow Graph can only compute syntactic reachability.

Semantic reachability is undecidable.



Reachability

Let $reach_G(X)$ denote the sub-graph of G that is (syntactically) reachable from X, where X is either a node, an edge, a set of nodes, or a set of edges.

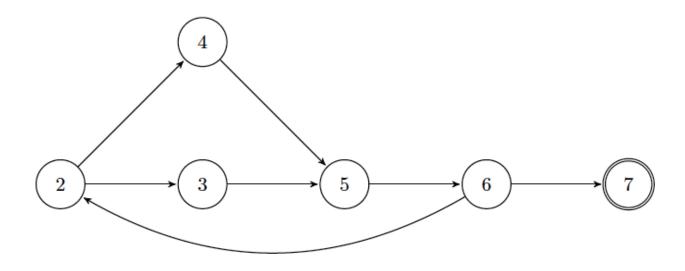


In this example, reach_G(1) is the whole graph G.



Syntactical Reachability

• reach_{G#}(2) is the subgraph that is syntactically reachable from node 2.



• reach_{G#}(7) is:





Connect Test Cases and Test Paths

Connect test cases and test paths with a mapping $path_{G}$ from test cases to test paths

- e.g., $path_{G}[t]$ is the set of test paths corresponding to test case t.
- Usually just write *path*, as G is obvious from the context
- Lift the definition of path to test set T by defining path(T)

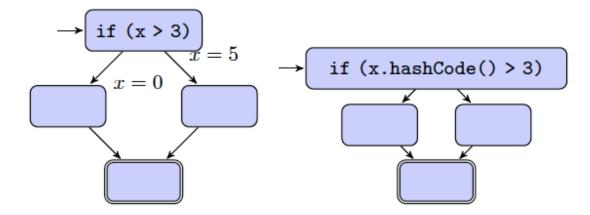
$$\operatorname{path}(T) = \{\operatorname{path}(t) | t \in T\}.$$

• Each test case gives at least one test path. If the software is deterministic, then each test case gives exactly one test path; otherwise, multiple test cases may arise from one test path.



Deterministic and Nondeterministic CFG

Here's an example of deterministic and nondeterministic control-flow graphs:

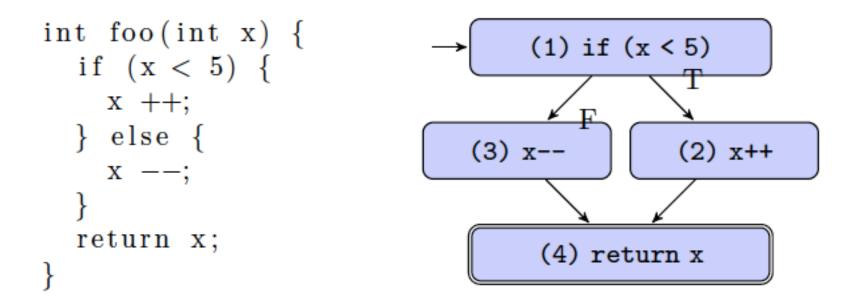


Causes of nondeterminism include dependence on inputs; on the thread scheduler; and on memory addresses, for instance as seen in calls to the default Java hashCode() implementation.

Nondeterminism makes it hard to check test case output, since more than one output might be a valid result of a single test input.



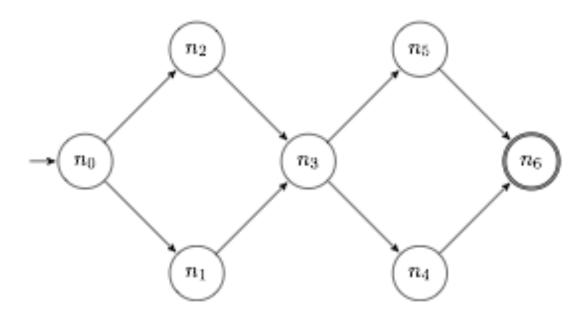
Connecting Test Cases, Test Paths, and CFG



- Test case: x = 5; test path: [(1), (3), (4)].
- Test case: x = 2; test path: [(1), (2), (4)].



Node Coverage



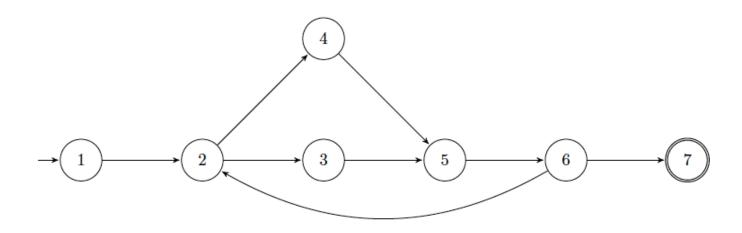
Node coverage: For each node $n \in reachG[N_0]$, TR contains a requirement to visit node n.

Node Coverage [NC]: TR contains each **reachable** node in G. TR = {n0,n1,n2,n3,n4,n5,n6}

a.k.a. statement coverage



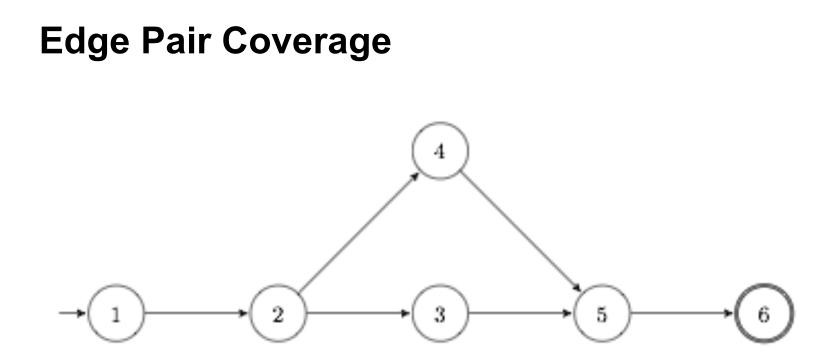
Edge Coverage (a.k.a. Branch Coverage)



Edge Coverage [EC]: TR contains each **reachable** path of length up to 1, inclusive, in G.

 $TR = \{[1,2], [2,4], [2,3], [3,5], [4,5], [5,6] [6,7], [6,2]\}$





Edge-Pair Coverage [EPC]: TR contains each **reachable** path of length up to 2, inclusive, in G. TR= {[1,2,3], [1,2,4], [2,3,5], [2,4,5], [3,5,6], [4,5,6]}



Simple Path

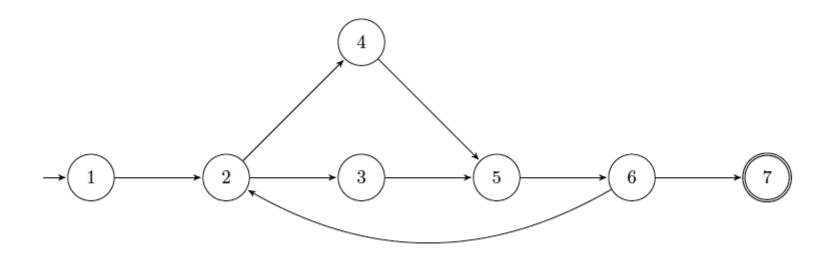
A path is simple if no node appears more than once in the path, except that the first and last nodes may be the same.

Some properties of simple paths:

- no internal loops;
- can bound their length;
- can create any path by composing simple paths; and
- many simple paths exist [too many!]



Simple Path Examples



Simple path examples:

- [1, 2, 3, 5, 6, 7]
- [1, 2, 4]
- [2,3,5,6,2]

Not simple Path: [1,2,3,5,6,2,4]



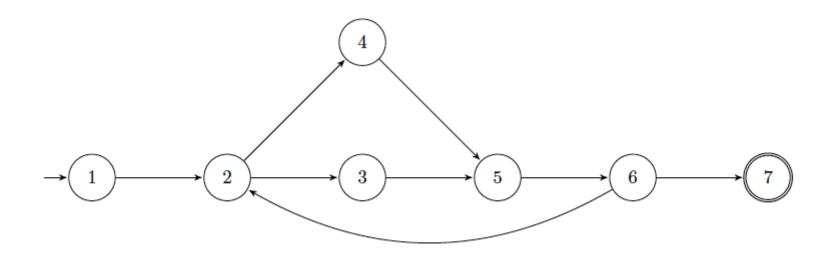
Prime Path

Because there are so many simple paths, let's instead consider prime paths, which are simple paths of maximal length.

A path is prime if it is simple and does not appear as a proper subpath of any other simple path.



Prime Path Examples



Prime path examples:

- [1, 2, 3, 5, 6, 7]
- [1, 2, 4, 5, 6, 7]
- [6, 2, 4, 5, 6]

Not a prime path: [3, 5, 6, 7]



Prime Path Coverage

Prime Path Coverage [PPC]: TR contains each prime path in G.

There is a problem with using PPC as a coverage criterion: a prime path may be infeasible but contains feasible simple paths.

• How to address this issue?

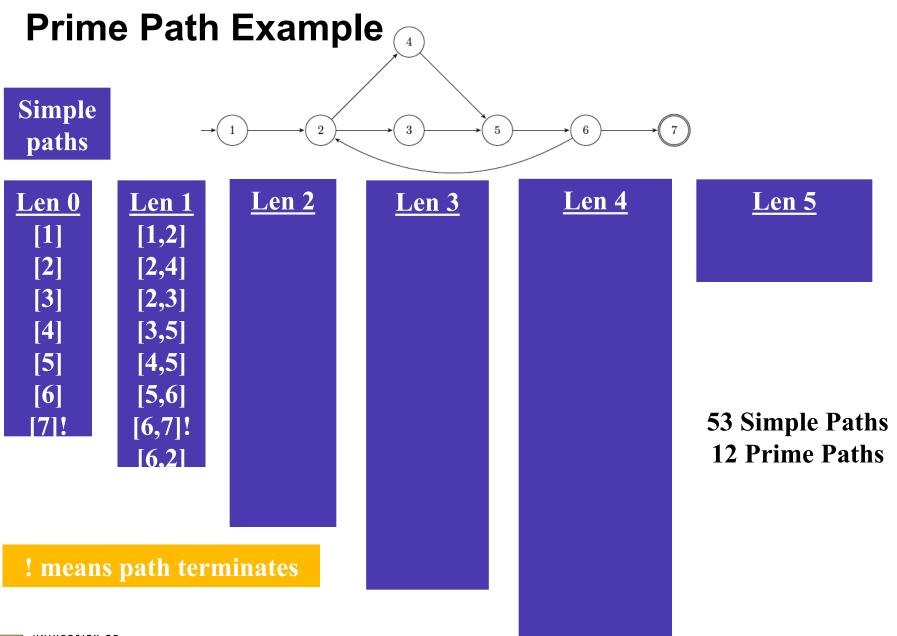


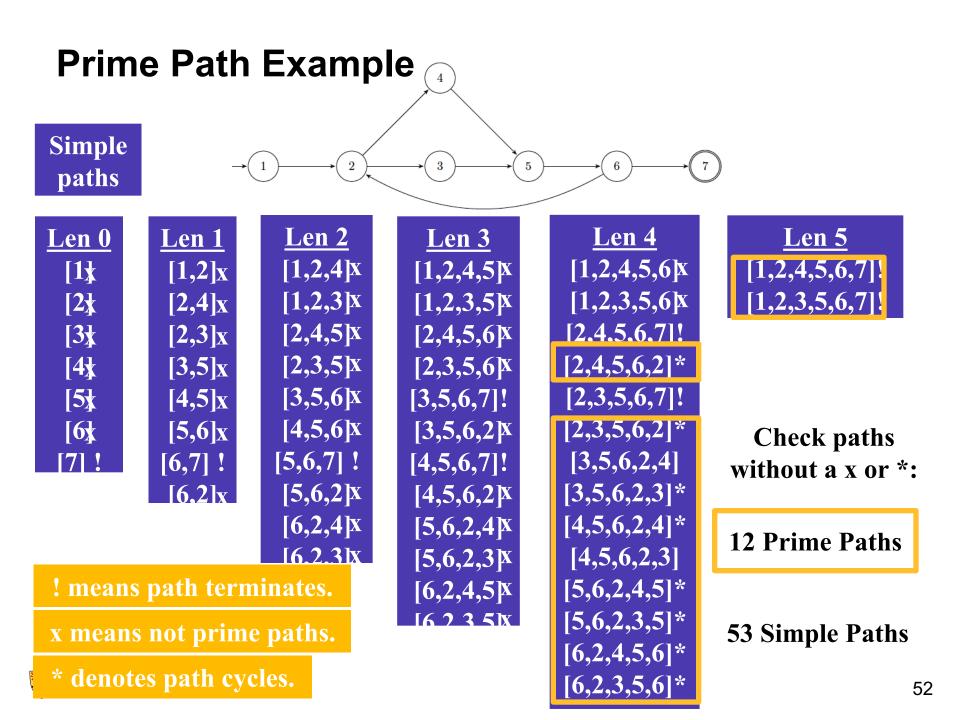
More Path Coverage Criterions

Complete Path Coverage [CPC]: TR contains all paths in G.

Sof paths.

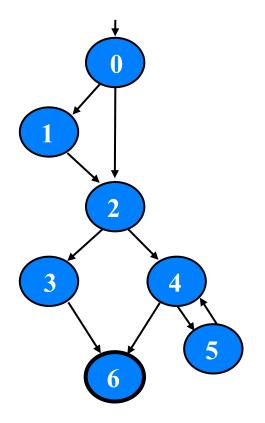






Prime Path Example (2)

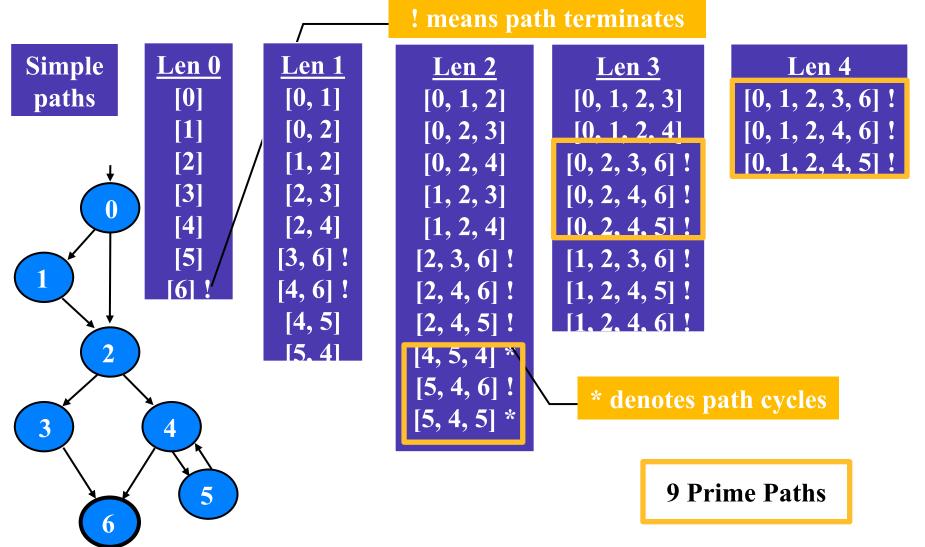
This graph has 38 simple paths Only 9 *prime paths*



Prime Paths	
[0, 1, 2, 3, 6]	
[0, 1, 2, 4, 5]	
[0, 1, 2, 4, 6]	
[0, 2, 3, 6]	
[0, 2, 4, 5]	
[0, 2, 4, 6]	
[5,4,6]	
[4,5,4]	
[5, 4, 5]	



Prime Path Example (2)





Examples of NC, EC, EPC, CPC

<u>Node Coverage</u> TR = { 0, 1, 2, 3, 4, 5, 6 } Test Paths: [0, 1, 2, 3, 6] [0, 1, 2, 4, 5, 4, 6]

Edge Coverage TR = { [0,1], [0,2], [1,2], [2,3], [2,4], [3,6], [4,5], [4,6], [5,4] } Test Paths: [0, 1, 2, 3, 6] [0, 2, 4, 5, 4, 6]

 $\frac{Edge-Pair Coverage}{TR = \{ [0,1,2], [0,2,3], [0,2,4], [1,2,3], [1,2,4], [2,3,6], \\ [2,4,5], [2,4,6], [4,5,4], [5,4,5], [5,4,6] \} \}}{Test Paths: [0, 1, 2, 3, 6] [0, 2, 3, 6] [0, 2, 4, 5, 4, 5, 4, 6]}$

<u>Complete Path Coverage</u> Test Paths: [0, 1, 2, 3, 6] [0, 1, 2, 4, 6] [0, 1, 2, 4, 5, 4, 6] [0, 1, 2, 4, 5, 4, 5, 4, 6] [0, 1, 2, 4, 5, 4, 5, 4, 5, 4, 6] ...



2

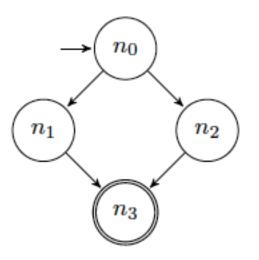
6

4

5

3

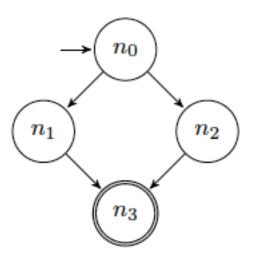
Prime Path Coverage vs. Complete Path Coverage



- Prime paths:
- $\operatorname{path}(t_1) =$
- $\operatorname{path}(t_2) =$
- $T_1 = \{t_1, t_2\}$ satisfies both PPC and CPC.



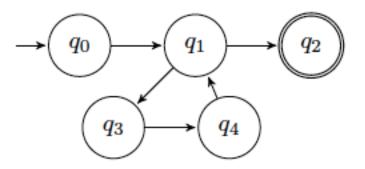
Prime Path Coverage vs. Complete Path Coverage



- Prime paths: [n0, n1, n3], [n0, n2, n3]
- $path(t_1) = [n0, n1, n3]$
- $path(t_2) = [n0, n2, n3]$
- $T_1 = \{t_1, t_2\}$ satisfies both PPC and CPC.



Prime Path Coverage vs. Complete Path Coverage (2)

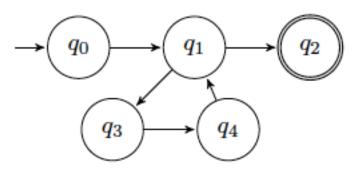


- Prime paths:
- $path(t_3) =$
- $\operatorname{path}(t_4) =$
- $T_1 = \{t_3, t_4\}$ satisfies

PPC but not CPC.



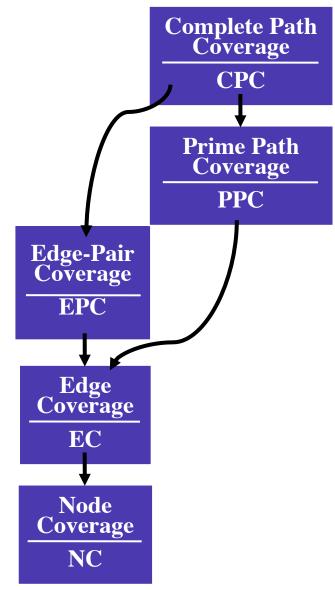
Prime Path Coverage vs. Complete Path Coverage (2)



- Prime paths: [q0, q1, q2], [q0, q1, q3, q4], [q3, q4, q1, q2], [q1, q3, q4, q1], [q3, q4, q1, q3], [q4, q1, q3, q4]
- $path(t_3) = [q0, q1, q2]$
- $path(t_4) = [q0, q1, q3, q4, q1, q3, q4, q1, q2]$
- $T_1 = \{t_3, t_4\}$ satisfies PPC but not CPC.



Graph Coverage Criteria Subsumption

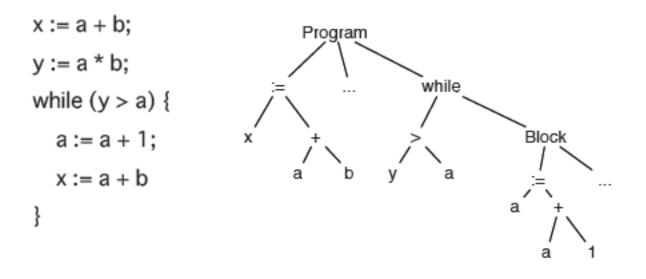




How do we measure coverage?

First:

- 1. Parse the source code to build an Abstract Syntax Tree (AST)
- 2. Analyze the AST to build a Control Flow Graph (CFG)
- 3. Count points of interest
 - (total # of statements, branches, etc.)
- 4. Instrument the AST using the CFG
 - add tracing statements in the code





How do we measure coverage?

Then:

- 1. Transform AST back to instrumented code
- 2. Recompile and run the test suite on the recompiled code
- 3. Collect tracing data
 - (line 1 executed, line 3 executed, etc.)
- 4. Calculate coverage:
 - # traced points / total # points



Coverage May Affect Test Outcomes

Heisenberg effect

• the act of observing a system inevitably alters its state.

Coverage analysis changes the code by adding tracing statements

Instrumentation can change program behaviour



Enabled In-code Assertions Mess Up Branch Coverage Reporting

assert P Turned into:

if assertion-enabled then
 if P then skip()
 else abort()
else skip()

Thus 4 branches!

Reported as such

Assertions shouldn't fail

Resulting branch coverage reports:

- Not useful with assertion checking enabled
- Without it, they miss invariants



Coverage: Useful or Harmful?

Measuring coverage (% of satisfied test obligations) can be a useful indicator ...

- Of progress toward a thorough test suite, of trouble spots requiring more attention
- ... or a dangerous seduction
- Coverage is only a proxy for thoroughness or adequacy
- It's easy to improve coverage without improving a test suite (much easier than designing good test cases)

The only measure that really matters is effectiveness



EXERCISES



Exercise 1: Bridge Coverage

Bridge Coverage (BC): If removing an edge adds unreachable nodes to the graph, then this edge is a bridge. The set of test requirements for BC contains all bridges.

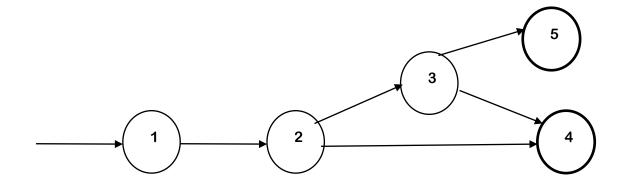
Assume that a graph contains at least two nodes, and all nodes in a graph are reachable from the initial nodes.

- (a) Does BC subsume Node Coverage (NC). If yes, justify your answer. If no, give a counterexample.
- (b) Does NC subsume BC? If yes, justify your answer. If no, give a counterexample



Bridge Coverage: Part (a)

Bridge Coverage does not subsume Node Coverage.



TRBC = {[1,2], [2,3], [3,5]}

 $\mathsf{TRNC} = \{[1, 2, 3, 4, 5\}$

Test path [1,2,3,5] satisfies BC, but not NC because node 4 is not visited.



Bridge Coverage: Part (b)

NC subsumes BC

Key points for the proof:

- For any bridge [a, b], any test case that visits b must also visit the edge [a, b] (can be proved by contradiction).
- Any test set that satisfies NC must visit node b (TR of NC contains all nodes, including node b). Therefore, for any bridge [a, b], the test set will visit it. Therefore, NC subsumes BC.



Exercise 2 (1/2)

Answer questions [a]-[g] for the graph defined by the following sets:

- N ={1, 2, 3, 4, 5, 6, 7}
- N0 = {1}
- Nf = {7}
- E = {[1, 2], [1, 7], [2, 3], [2, 4], [3, 2], [4, 5], [4, 6], [5, 6], [6, 1]}

Also consider the following test paths:

- t0 = [1, 2, 4, 5, 6, 1, 7]
- t1 = [1, 2, 3, 2, 4, 6, 1, 7]



Exercise 2 (2/2)

[a] Draw the graph.

- [b] List the test requirements for EPC. [Hint: You should get 12 requirements of length 2].
- [c] Does the given set of test paths satisfy EPC? If not, identify what is missing.
- [d] List the test requirements for NC, EC and PPC on the graph.
- [e] List a test path that achieve NC but not EC on the graph.
- [f] List a test path that achieve EC but not PPC on the graph.



Exercise 2: Partial Solutions (1/2)

[a] Draw the graph.

[b] List the test requirements for EPC. [Hint: You should get 12 requirements of length 2].

• The edge pairs are: {[1, 2, 3], [1, 2, 4], [2, 3, 2], [2, 4, 5], [2, 4, 6], [3, 2, 3], [3, 2, 4], [4, 5, 6], [4, 6, 1], [5, 6, 1], [6, 1, 2], [6, 1, 7] }

[c] Does the given set of test paths satisfy EPC? If not, identify what is missing.

• No. Neither t0 nor t1 visits the following edge-pairs: {[3, 2, 3], [6, 1, 2]}



Exercise 2: Partial Solutions (2/2)

[d] TR for NC, EC, and PPC.

- TRNC =
- TREC =
- TRPPC = {[3, 2, 4, 6, 1, 7], [3, 2, 4, 5, 6, 1, 7], [4, 6, 1, 2, 3], [4, 5, 6, 1, 2, 3], [3, 2, 3], [2, 3, 2], [1, 2, 4, 5, 6, 1], [1, 2, 4, 6, 1], [2, 4, 6, 1, 2], [2, 4, 5, 6, 1, 2], [4, 6, 1, 2, 4], [4, 5, 6, 1, 2, 4], [5, 6, 1, 2, 4, 5], [6, 1, 2, 4, 6], [6, 1, 2, 4, 5, 6]}

[e] A test path that achieve NC but not EC.

• [1, 2, 3, 2, 4, 5, 6, 1, 7] does not cover edge [4, 6].

[f] A test path that achieve EC but not PPC.

• [1, 2, 3, 2, 4, 5, 6, 1, 2, 4, 6, 1, 7] does not cover prime paths such as [3,2,3].

