CS189A - Capstone

Chandra Krintz
Department of Computer Science
UC Santa Barbara
http://www.cs.ucsb.edu/~ckrintz/
Verification, Validation, Testing

- **Verification**: Demonstration of consistency (contradictions), completeness (omissions), and correctness (errors) of the software artifacts at each stage of and between each stage of the software life-cycle.
  - Different types: manual inspection, testing, formal methods
  - Verification answers the question: Am I building the product right?

- **Validation**: The process of evaluating software at the end of the software development (each release) to ensure compliance with respect to the customer needs and requirements.
  - Validation can be accomplished by verifying the artifacts produced at each stage of the software development life cycle
  - Validation answers the question: Am I building the right product?
Verification, Validation, Testing

- **Testing**: Examination of the behavior of a program by executing the program on sample data sets.
  - Testing is a verification technique used at the **implementation stage**.
Manual Verification

• When we have an executable program we can use testing methods for verification

• Can we find a way to check requirements specifications, design specifications, and source code?

• Manual verification techniques help in these situations:
  – *Walkthroughs, Inspections, Reviews, Audits*
    • Each is a meeting with phases: planning, meeting, and post-mortem
    – Sometimes all of these techniques together are called *Reviews*

• These tasks are done typically in meetings, manually
• They are useful when no automated technique is available
Reviews

• **Review**: A process or meeting during which a work product, or a set of work products, is presented to project personnel, managers, users, customers, or other interested parties for comment or approval. Types include code review, design review, and requirements review.

• **Characteristics for a review:**
  – Review should generate a written report on status of the product reviewed—a report that is available to everyone involved in the project, including management;
  – Review requires active and open participation of everyone in the review group;
  – Review requires full responsibility of all participants for the quality of the review—that is, for the quality of the information in the written report.
Walk-Through

- **Walk-through**: A manual static analysis technique in which a designer or programmer leads members of the development team and other interested parties through a segment of documentation or code, and the participants ask questions and make comments about possible errors, violation of development standards, and other problems.

- Walk-throughs are a form of manual simulation

- Two variations
  - led by a reader or presenter who could be the person responsible for the product
  - led by a moderator independent of the person responsible for the product

- In a code walkthrough, you go over the code statement-by-statement explaining what each statement does
Walk-Through

- **Objectives**
  - detect, identify, and describe software element defects
  - examine alternatives and stylistic issues
  - provide a mechanism that enables the authors to collect valuable feedback on their work, yet allows them to retain the decision-making authority for any changes
  
  - Git comments on pull requests
Inspection

- **Inspection**: A static analysis technique that relies on visual examination of development products to detect errors, violations of developing standards, and other problems. Types include code inspection; design inspections.

- Inspections are used to manually check for common errors

- A method of rapidly evaluating material by confining attention to a few selected aspects, one at a time.

- In an inspection, the inspector uses a rigid set of guidelines or a checklist to assess the degree of compliance with the checklist or guidelines

- In a code inspection you have a checklist that looks for errors such as uninitialized variables, division by zero etc. and check each item in the checklist one by one
Inspections

A typical inspection checklist for code inspections may include:

- wrongful use of data: uninitialized variables, array index out of bounds, dangling pointers
- faults in declarations: use of undeclared variables, declaration of the same name in nested blocks
- faults in computations: division by zero, overflow, wrong use of variables of different types in one and the same expression, faults caused by an erroneous conception of operator priorities
- faults in relational expressions: using an incorrect operator, an erroneous conception of priorities of Boolean operators
- faults in control flow: infinite loops, a loop that gets executed n+1 or n-1 times rather than n
- faults in interfaces: incorrect number of parameters, parameters of the wrong types, inconsistent use of global variables
Audits

- **Audit**: An independent examination of a work product or a set of work products to assess compliance with specifications, standards, contractual agreements, or other criteria.

- Similar to inspections but
  - More interactive than inspections
  - Less structured than inspections

- You can consider your project demo an audit
Manual Verification

Software Requirements Analysis

- Requirements Walkthroughs
- Requirements Specification Inspection
- Software Requirements Review

Preliminary Design

- Design Walkthroughs
- Preliminary Design Review

Detailed Design

- Design Specification Inspection
- Critical Design Review

Coding and Unit Test

- Code Walkthroughs
- Critical Design Review

System Test and Integration

- User Manual Inspection
- Code Inspections
- Functional and Physical Configuration Audits
Software Testing

• Correctness
  – software should match its specifications
  – software should meet its functional requirements

• Testing is necessary because we cannot guarantee correctness in the software development process

• Testing: techniques of checking software correctness by executing the software on some data sets
Software Testing

• Goal of testing
  – finding faults in the software
  – demonstrating that there are no faults in the software

• It is not possible to prove that there are no faults in the software using testing

• Testing should help locate errors, not just detect their presence
  – a “yes/no” answer to the question “is the program correct?” is not very helpful

• Testing should be repeatable
  – could be difficult for distributed or concurrent software
  – effect of the environment, uninitialized variables
Testing Software is Hard

• If you are testing a bridge’s ability to sustain weight, and you test it with 1000 tons you can infer that it will sustain weight ≤ 1000 tons

• This kind of reasoning does not work for software systems
  – software systems are not linear nor continuous

• Exhaustively testing all possible input output combinations is too expensive
  – the number of test cases increase exponentially with the number of input/output variables
Some Definitions

• Let $P$ be a program and let $D$ denote its input domain

• A **test case** $d$ is an element of input domain $d \in D$
  – a test case gives a valuation for all the input variables of the program

• A **test set** $T$ is a finite set of test cases, i.e., a subset of $D$, $T \subseteq D$

• The basic difficulty in testing is finding a test set that will uncover the faults in the program

• Exhaustive testing corresponds to setting $T = D$
Exhaustive Testing is Hard

- Number of possible test cases (assuming 32 bit integers)
  - \(2^{32} \times 2^{32} = 2^{64}\)

- Do bigger test sets help?
  - Test set
    \[\{(x=3,y=2), (x=2,y=3)\}\]
    will detect the error
  - Test set
    \[\{(x=3,y=2),(x=4,y=3),(x=5,y=1)\}\]
    will not detect the error although it has more test cases

- It is not the number of test cases
- But, if \(T_1 \subseteq T_2\), then \(T_1\) will detect every fault detected by \(T_2\)

```c
int max(int x, int y)
{
    if (x > y)
        return x;
    else
        return x;
}
```
Exhaustive Testing

- Assume that the input for the $\text{max}$ procedure was an integer array of size $n$
  - Number of test cases: $2^{32 \times n}$

- Assume that the size of the input array is not bounded
  - Number of test cases: $\infty$

- The point is exhaustive testing is pretty hopeless
Random Testing

- Use a random number generator to generate test cases
- Derive estimates for the reliability of the software using some probabilistic analysis
- Coverage is a problem
  - More on coverage in a bit
Generating Test Cases Randomly

bool isEqual(int x, int y) {
    if (x == y)
        z := false;
    else
        z := false;
    return z;
}

• If we pick test cases randomly it is unlikely that we will pick a case where \(x\) and \(y\) have the same value.
• If \(x\) and \(y\) can take \(2^{32}\) different values, there are \(2^{64}\) possible test cases. In \(2^{32}\) of them \(x\) and \(y\) are equal:
  – probability of picking a case where \(x\) is equal to \(y\) is \(2^{-32}\).
• It is not a good idea to pick the test cases randomly (with uniform distribution) in this case.
Testing

• Testing can be categorized in different ways:
  – Functional vs. Structural testing
    • Functional testing: Generating test cases based on the functionality of the software
    • Structural testing: Generating test cases based on the structure of the program
  – Black box vs. White box testing
    • Black box testing is the same as functional testing. Program is treated as a black box, its internal structure is hidden from the testing process.
    • White box testing is same as structural testing. In white box testing internal structure of the program is taken into account
  – Module vs. Integration testing
    • Module testing: Testing the modules of a program in isolation
    • Integration testing: Testing an integrated set of modules
Testing Boundary Conditions

• For each range \([R_1, R_2]\) listed in either the input or output specifications, choose five cases:
  – Values less than \(R_1\)
  – Values equal to \(R_1\)
  – Values greater than \(R_1\) but less than \(R_2\)
  – Values equal to \(R_2\)
  – Values greater than \(R_2\)

• For unordered sets select two values
  – 1) in, 2) not in

• For equality select 2 values
  – 1) equal, 2) not equal

• For sets, lists select two cases
  – 1) empty, 2) not empty
Coverage Metrics

• Coverage metrics
  – *Statement coverage*: all statements in the programs should be executed at least once
  – *Branch coverage*: all branches in the program should be executed at least once
  – *Path coverage*: all execution paths in the program should be executed at least once

• The best case would be to execute all paths through the code, but there are some problems with this:
  – the number of paths increases fast with the number of branches in the program
  – the number of executions of a loop may depend on the input variables and hence may not be possible to determine
  – most of the paths can be *infeasible*
Statement Coverage

- Choose a test set $T$ such that by executing program $P$ for each test case in $T$, each basic statement of $P$ is executed at least once.
- Executing a statement once and observing that it behaves correctly is not a guarantee for correctness, but it is an heuristic.
  - this goes for all testing efforts since in general checking correctness is undecidable.

```cpp
bool isEqual(int x, int y)
{
    if (x == y)
        z := false;
    else
        z := false;
    return z;
}

int max(int x, int y)
{
    if (x > y)
        return x;
    else
        return x;
}
```
areTheyPositive(int x, int y) {
    if (x >= 0)
        print("x is positive");
    else
        print("x is negative");
    if (y >= 0)
        print("y is positive");
    else
        print("y is negative");
}

Following test set will give us statement coverage:
T₁ = {(x=12,y=5), (x=-1,y=35),
(x=115,y=-13),(x=-91,y=-2)}

There are smaller test cases which will give us statement coverage too:
T₂ = {(x=12,y=-5), (x=-1,y=35)}

There is a difference between these two test sets though
Control Flow Graphs (CFGs)

- Nodes in the control flow graph are basic blocks
  - A **basic block** is a sequence of statements always entered at the beginning of the block and exited at the end
- Edges in the control flow graph represent the control flow

```java
if (x < y) {
    x = 5 * y;
    x = x + 3;
} else
    y = 5;
```

- Each block has a sequence of statements
- No jump from or to the middle of the block
- Once a block starts executing, it will execute till the end
assignAbsolute(int x) {
    if (x < 0)
        x := -x;
    z := x;
}

Identify a test that gives statement coverage but not branch coverage
assignAbsolute(int x) 
{ 
    if (x < 0) 
        x := -x; 
    z := x; 
}

Consider this program segment, the test set T = {x=-1} will give statement coverage, however not branch coverage

Test set {x=-1} does not execute this edge, hence, it does not give branch coverage
Branch Coverage

- Construct the control flow graph

- Select a test set $T$ such that by executing program $P$ for each test case $d$ in $T$, each edge of $P$’s control flow graph is traversed at least once.

```plaintext
x := -x

Test set \{x=-1\} does not execute this edge, hence, it does not give branch coverage.

Test set \{x=-1, x=2\} gives both statement and branch coverage.
```
Path Coverage

- Select a test set $T$ such that by executing program $P$ for each test case $d$ in $T$, all paths leading from the initial to the final node of $P$'s control flow graph are traversed
Path Coverage

areTheyPositive(int x, int y) {
    if (x >= 0)
        print("x is positive");
    else
        print("x is negative");
    if (y >= 0)
        print("y is positive");
    else
        print("y is negative");
}

Test set:
T₂ = {(x=12, y=-5), (x=-1, y=35)}
gives both branch and statement coverage but it does not give path coverage

Set of all execution paths: {(B0,B1,B3,B4,B6), (B0,B1,B3,B5,B6), (B0,B2,B3,B4,B6), (B0,B2,B3,B5,B6)}
Test set T₂ executes only paths: (B0,B1,B3,B5,B6) and (B0,B2,B3,B4,B6)
Path Coverage

areTheyPositive(int x, int y) {
    if (x >= 0)
        print("x is positive");
    else
        print("x is negative");
    if (y >= 0)
        print("y is positive");
    else
        print("y is negative");
}

Test set:
T_1 = {(x=12,y=5), (x=-1,y=35),
       (x=115,y=-13), (x=-91,y=-2)}
Gives branch, statement and path coverage
Path Coverage

- Number of paths is exponential in the number of conditional branches
  - Testing cost may be expensive
- Note that every path in the control flow graphs may not be executable
  - It is possible that there are paths which will never be executed due to dependencies between branch conditions
- In the presence of cycles in the control flow graph (for example loops) we need to clarify what we mean by path coverage
  - Given a cycle in the control flow graph we can go over the cycle arbitrary number of times, which will create an infinite set of paths
  - Redefine path coverage as: each cycle must be executed 0, 1, ..., k times where k is a constant (k could be 1 or 2)
Condition Coverage

• In branch coverage we make sure that we execute every branch at least once
  – For conditional branches, this means that, we execute the TRUE branch at least once and the FALSE branch at least once

• Conditions for conditional branches can be compound boolean expressions
  – A compound boolean expression consists of a combination of boolean terms combined with logical connectives AND, OR, and NOT

• Condition coverage:
  – Select a test set $T$ such that by executing program $P$ for each test case $d$ in $T$, (1) each edge of $P$’s control flow graph is traversed at least once and (2) each Boolean term that appears in a branch condition takes the value TRUE at least once and the value FALSE at least once

• Condition coverage is a refinement of branch coverage (part (1) is same as the branch coverage)
something(int x)
{
    if (x < 0 || y < x)
    {
        y := -y;
        x := -x;
    }
    z := x;
}

T = \{(x=-1, y=1), (x=1, y=1)\} will achieve statement, branch and path coverage, however T will not achieve condition coverage because the boolean term \((y < x)\) never evaluates to true. This test set satisfies part (1) but does not satisfy part (2).

T = \{(x=-1, y=1), (x=1, y=0)\} will not achieve condition coverage either. This test set satisfies part (2) but does not satisfy part (1). It does not achieve branch coverage since both test cases take the true branch, and, hence, it does not achieve condition coverage by definition.

T = \{(x=-1, y=-2), (x=1, y=1)\} achieves condition coverage.
Types of Testing

- **Unit (Module) testing**
  - testing of a single module in an isolated environment

- **Integration testing**
  - testing parts of the system by combining the modules

- **System testing**
  - testing of the system as a whole after the integration phase

- **Acceptance testing**
  - testing the system as a whole to find out if it satisfies the requirements specifications
Unit Testing

• Involves testing a single isolated module

• Note that unit testing allows us to isolate the errors to a single module
  – we know that if we find an error during unit testing it is in the module we are testing

• Modules in a program are not isolated, they interact with each other. Possible interactions:
  – calling procedures in other modules
  – receiving procedure calls from other modules
  – sharing variables

• For unit testing we need to isolate the module we want to test, we do this using two things
  – drivers and stubs
Drivers and Stubs

- **Driver:** A program that calls the interface procedures of the module being tested and reports the results
  - A driver simulates a module that calls the module currently being tested

- **Stub:** A program that has the same interface as a module that is being used by the module being tested, but is simpler.
  - A stub simulates a module called by the module currently being tested
Drivers and Stubs

Driver

Module Under Test

Stub

- Driver and Stub should have the same interface as the modules they replace
- Driver and Stub should be simpler than the modules they replace
Integration Testing

• Integration testing: Integrated collection of modules tested as a group or partial system

• Integration plan specifies the order in which to combine modules into partial systems

• Different approaches to integration testing
  – Bottom-up
  – Top-down
  – Big-bang
  – Sandwich
We assume that the use hierarchy is a directed acyclic graph.

If there are cycles merge them to a single module.
Bottom-Up Integration

• Only terminal modules (i.e., the modules that do not call other modules) are tested in isolation

• Modules at lower levels are tested using the previously tested higher level modules

• Non-terminal modules are not tested in isolation

• Requires a module driver for each module to feed the test case input to the interface of the module being tested
  – However, stubs are not needed since we are starting with the terminal modules and use already tested modules when testing modules in the lower levels
Bottom-up Integration
Top-down Integration

• Only modules tested in isolation are the modules which are at the highest level

• After a module is tested, the modules directly called by that module are merged with the already tested module and the combination is tested

• Requires stub modules to simulate the functions of the missing modules that may be called
  – However, drivers are not needed since we are starting with the module which is not used by any other module and use already tested modules when testing modules in the higher levels
Top-down Integration
Other Approaches to Integration

- **Sandwich Integration**
  - Compromise between bottom-up and top-down testing
  - Simultaneously begin bottom-up and top-down testing and meet at a predetermined point in the middle

- **Big Bang Integration**
  - Every module is unit tested in isolation
  - After all of the modules are tested they are all integrated together at once and tested
  - No driver or stub is needed
  - However, in this approach, it may be hard to isolate the bugs!
System Testing, Acceptance Testing

- System and Acceptance testing follows the integration phase
  - testing the system as a whole

- Test cases can be constructed based on the requirements specifications
  - main purpose is to assure that the system meets its requirements

- Manual testing
  - Somebody uses the software on a bunch of scenarios and records the results
  - Use cases and use case scenarios in the requirements specification would be very helpful here
  - manual testing is sometimes unavoidable: usability testing
System Testing, Acceptance Testing

• Alpha testing is performed within the development organization

• Beta testing is performed by a select group of friendly customers

• Stress testing
  – push system to extreme situations and see if it fails
  – large number of data, high input rate, low input rate, etc.
Regression Testing

• You should preserve all the test cases for a program

• During the maintenance phase, when a change is made to the program, the test cases that have been saved are used to do regression testing
  – figuring out if a change made to the program introduced any faults

• Regression testing is crucial during maintenance
  – It is a good idea to automate regression testing so that all test cases are run after each modification to the software

• When you find a bug in your program you should write a test case that exhibits the bug
  – Then using regression testing you can make sure that the old bugs do not reappear
Test Plan

• Testing is a complicated task
  – it is a good idea to have a test plan

• A test plan should specify
  – Unit tests
  – Integration plan
  – System tests
  – Regression tests