Software Engineering
PhD Qualifying Examination
March 30, 2010

Answer six of the following nine questions including questions 1 and 2. Include citations for relevant literature where appropriate.

1. Your Research Interests – You must answer this question
   a) What area of research most interests you and why?
   b) Describe three important problems in the area along with the current state-of-the-art addressing those problems.
   c) What is the state of the practice for those problems—i.e., how do practitioners currently deal with the problems?
   d) List five people that are top researchers in your area and discuss, for each of them, what was his/her contribution to the field, why that contribution is important, and how it improved both the state of the art and the state of the practice.
   e) What problem or problems in the area do you currently plan to pursue, and what contributions do you plan to make?

2. Research Impact – You must answer this question
   It has been claimed that software engineering research has had relatively little impact on actual software development practice. Select three different areas of practice. The first should be an area for which you can show the direct impact of academic software engineering research. For this area, do the following.
   a) Describe the state of the practice in the area in terms of methods and technology (tools) currently used.
   b) Indicate the role that software engineering research has had on these methods and technologies citing at least two specific papers. Include a precise statement of the research contribution made in each paper.
   c) Argue why you think that research has been able to impact this area of practice.

   For the second area, select an area in which academic software engineering research has not had much practical impact. For this area, do the following.
   a) Describe the state of the practice in the area in terms of methods and technology (tools) currently used.
   b) Describe at least two major historical research results in this area, including citations of papers and precise statement of the contributions made.
   c) Explain why you think that research results have not had a practical impact on this area.

   For the third area, select an area for which you think current research will have a major impact in the future. For this area do the following.
   a) Describe the state of the practice in the area in terms of methods and technology (tools) currently actually used by developers today.
b) Describe at least two recent (last four years) research results that you believe will have future practical impact. Include citations and a precise statement of research contributions made for each.
c) Argue why you think this research will have an impact, including overcoming the kinds of barriers you discussed for the second area.

3. **General Software Engineering**
Barry Boehm has defined software engineering as "the application of science and mathematics by which the capabilities of computer equipment are made useful to man via computer programs, procedures, and associated documentation".
   a) Do you agree or disagree with Boehm? Defend your position.
   b) Give at least three examples of the use of science and mathematics in software engineering practice, carefully describing the scientific/mathematical principles employed.
   c) Give and justify at least three ways in which science and mathematics could be employed in current software development practice where it is not currently so employed.

F. L. Bauer has defined software engineering as "the part of computer science that is too difficult for the computer scientists."
   d) Do you agree or disagree with Bauer? Defend your position.
   e) How does Fred Brooks' *No Silver Bullet* argument relate to Bauer's position?

4. **Testing**
You are being interviewed for a position as a test manager, and your interviewer asks you the following questions. Provide concise but complete answers to the questions.
   a) What is a test plan? Provide a possible test plan for testing a login screen.
   b) Describe the main differences between testing a desktop application and testing a web application.
   c) What are the differences between a test scenario and a test case?
   d) What are the main benefits of test automation? What are its main drawbacks? In general, do you believe test automation can improve test effectiveness? Defend your position.
   e) Can you have a failure with high severity and low priority? What about the opposite (i.e., high priority and low severity)? Justify your answers and provide one example of each situation (if possible). How would you assign priority and severity for a failure?
   f) What is a non-reproducible failure? What would you do if you were asked to debug a non-reproducible failure?
   g) Provide an example of a program that contains a fault that may not be revealed by a statement-coverage-adequate test suite but would be revealed by a du-coverage-adequate test suite.
   h) Why is a path-coverage-adequate test suite in general not exhaustive?
   i) Why is it typically not possible to achieve 100% structural or data-flow coverage? How could this problem be alleviated?
Given are a program P, a modified program P', and a test suite T for P. A regression test selection technique selects T', a subset of T, to be rerun on P'. In this context, safety measures whether such a technique selects all tests that could reveal faults in P', whereas precision measures the degree to which the technique is able to omit unneeded tests. Consider the following three regression testing techniques:

- **T1**: select all tests in T that executed modified or deleted functions in P, or could reach new functions in P'.
- **T2**: select all tests in T that executed modified or deleted basic blocks in P, or could reach new blocks in P'.
- **T3**: select the minimal subset of tests in T that ensures coverage of all modified or deleted functions in P or new functions in P'.

Which of these techniques are safe for controlled regression testing, and which ones are not? Justify your answer. Which technique is more precise between T1 and T2? Prove it.

5. **Symbolic Evaluation**
   a) Using either static symbolic evaluation methods (Clarke and Richardson) or dynamic symbolic evaluation methods (Kousik, et al.) or both, present an algorithm for performing symbolic evaluation on programs with concurrency.
   b) Argue why you believe that your algorithm can be effective and efficient. Use Java threads as an example.

6. **Program Analysis**
   A dead definition is a definition of a variable that is never referenced (i.e., there is no definitional path from that definition to a use of the same variable).
   a) Formulate the problem of detecting dead definitions as a data-flow problem that can be solved using a generalized iterative data-flow analysis framework such as the one used in CS 6340. Assume that the data-flow analysis will be performed on programs in Static Single Assignment (SSA) form (i.e., that every variable in the program is assigned exactly once). Specifically:
      a.1) Specify whether it is a forward or backward data-flow problem and, based on that, which visiting order could improve the worst-case time complexity of the analysis.
      a.2) Define GEN(B) and KILL(B) for the problem, where B is a block in the program's CFG.
      a.3) Define the initial value of IN(B) and OUT(B).
      a.4) Define the equations to compute IN(B) and OUT(B).
      a.5) Define how the result is computed, that is, which set would contain the dead definitions at the end of the analysis.
   b) Is it possible to formulate this problem as an instance of the generalized iterative DF analysis framework, as you did above, for programs not in SSA form? Justify your answer.
   c) Exhibit a program for which your analysis would require four iterations to converge using the visiting order of Point a.1). (Convergence in four iterations means that the data-flow facts computed in the first three iterations are different,
but the data-flow facts computed in the fourth iteration are identical to those computed in the third iteration.)

7. **Refactoring and Testing**
   a) Define and briefly motivate program refactoring. Include a description of the steps a developer takes to apply a refactoring.
   How does refactoring affect the following. Explain your answers.
   b) Test coverage, and test-coverage criteria
   c) Regression test selection
   d) Test-suite criteria

8. **Software Process**
   Many small software companies are formed by extremely skilled and dedicated software developers, and one key determinant of success are exactly those skills. However, as the companies grow, they may not be able to continue to hire such high-quality people. Consequently, they must rely more on a well-defined development process.
   a) What are the costs and benefits of relying on highly skilled individuals?
   b) What are the costs and benefits of relying on a well-defined process?
   c) As a manager for such a growing company, describe a three-year plan for your development organization that addresses this situation. That is, which elements of process do you institute first? Which later? How do you judge the progress of your plan?
   d) As a software engineering researcher, describe a research program you could apply to situations like this to learn more about them.

9. **Software Architecture**
   a) Describe at least two ways that UML 1.4 supported the architectural part of software design.
   b) Describe at least two features that were added in UML 2 that directly supported architectural design.
   c) Compare and contrast the use of an architectural description language (ADL) like ACME with UML for describing software architectures. Give specific examples of UML inadequacies.
   d) Sketch a UML Profile (set of extensions to the UML metamodel) that you would use to make UML more ADL-like.