## Overview of the company-specific problem

A cellular base station connects cell phones to a voice and data network via LTE (4G) or 3G standards. The software that runs on these base stations contains not only complex signaling logic with stringent real-time constraints, but also must be highly reliable, providing safety critical services, such as 911 calling. At Ericsson Ottawa, testing base station software is time consuming and involves expensive specialized hardware. For example, testers may need to simulate cellular devices, the backhaul network and high capacity scenarios for example, a base station that is overwhelmed by requests from cell users at a music concert. In order to maximize the return on investment of their testing efforts, we plan to use historical test runs and other historical artifacts to create statistical models (1) to prioritize tests and (2) to help developers categorize and locate the cause of a failure. The outcome of this work will advance the state-of-the-art in test prioritization and localization and provide Ericsson with a more efficient test infrastructure.

## Background, Research Approach, and Data

The proposed work intersects two areas of software engineering: test prioritization and empirical software engineering. There has been much excellent work using past test runs and static code analysis information such as paths and control flow to investigate the order in which tests should be run in the future \cite{Porter, Hadi, SunKim}. We plan to build on this work and to use it as a baseline to compare new techniques.

Empirical software engineering researchers use historical data to create models to help developers work more productively. Relatively little work has combined test archives with information from other artifacts, such as version history and bug reports. One of the first studies to combine these artifacts was done by a colleague at Concordia, Dr. Shihab, during his PhD. In this work, he created a statistical model from the version history to help developers decide where to write tests for a legacy system at Blackberry \cite{Emad}.

The major factor limiting research into test prioritization is information about past runs and links to other development artifacts. Ericsson collects massive histories of logs, test runs, and development artifacts, in part to provide for reproducibility of software and test lineups and auditing. Much of the current analysis is done manually where developers must sift through this information to locate the root cause of a failure.

My research team has extensive training in mining and linking software development artifacts. We have worked extensively with semi-structured datasets and been able to provide useful statistical models. For example, we have mined millions of code reviews, bug reports, and tests results, and StackOverflow discussions. Rigby currently teaches these mining, linking, and analytics techniques in a graduate course. Rigby has mined test data for Google Chrome in the past. While the Ericsson data will be more complete, the techniques remain similar. The novel contribution lies in advancing our knowledge of test prioritization and localization and in improving the efficiency of quality assurance at Ericsson Ottawa.

## Detailed Proposal and Methodology

The proposal consists of three milestones, which will each run for approximately 2 months. First, we will investigate, clean, and link the available data. Second we will create a test prioritization model (RQ1). Third, we will create a fault localization model (RQ2).

**Investigate Process and Data:** A masters student will be onsite at Ericsson Ottawa. His goal will be to work with Griffiths’s team to understand the quality assurance processes and to clean and link data. Ericsson’s in-kind contribution will be substantial as they will work with the masters student and help in interpreting the findings. I will visit Ericsson twice a month to help the student. A PhD student, Rahman, who researches release engineering, will also be investigating test prioritization and the impact on release effectiveness and help generalize the master’s student findings to other software projects.

**RQ1. Test prioritization model:** Not all tests are equally important, but all tests must be run. With Ericsson’s hard quality requirements, all tests must be run before a change can be released. However, some tests can span days, while others take seconds. Further, some tests have not failed in months, while some are new or fail regularly. Since a failure will stop the build process and require an engineer to intervene, a clear research direction will be to prioritize based on past failures and test running time \cite{Porter}. Adding to these measures, we will develop change risk and social measures, such as risk analysis based on histories of defects across software areas and development teams. The research outcome of this stage will be a better understanding of the measures that provide good test prioritization.

**RQ2. Fault categorization and localization model:** Did that test fail because of a problem in the development environment or product? This problem is unusually difficult at Ericsson because the environment is exceedingly complex. For example, test equipment is so expensive that developers in Ottawa can request their tests be run on specialized equipment that is only available in Sweden, and vice versa.

Once a test fails, a developer must determine the root cause of the failure and whether it is an environment or product problem. Unfortunately, much of the other information in the report is added sporadically. We will investigate the reliability of existing fields in the reports and supplement this with freeform data extracted using existing NLP techniques. Our goal is to train a classifier based on past manual categorizations to help developers differentiate these failure types. We plan to refine this coarsegrained categorization to a finegrained localization at the file or method level by linking issue reports to other software artifacts.

## Benefits to Canada

In 2009, Ericsson acquired the majority of Nortel's North American wireless access business, including its market share and became the world leader in this market. This human expertise remains in Canada where Ericsson employs about 1000 R&D staff in Ottawa. Our contribution will be to make these individuals more productive by improving the efficiency of the test process. We hope that this contribution leads to a long term relationship with Ericsson Ottawa where we can attack their practical problems and abstract them to the general advancement of knowledge in software engineering.