# 1 Introduction

## 1.1 Introduction

Software developers, quality assurance engineers, and security analysts have been entrusted with safeguarding the sensitive, personal, and corporate data flowing through computers and networks. These guardians of our data have a variety of tools available for use as a part of the “best practices” for software security development. Some of the available techniques include static analysis, dynamic analysis, and formal code verification. Each tool comes with its own set of strengths and weaknesses, but using them in tandem should provide a strong defense against release of vulnerabilities through which hackers can penetrate. The fact remains that even though these tools are being used, the rate at which vulnerabilities are being discovered and exploited is escalating rapidly which implies that something more needs to be done to minimize the attack surface of our computer-driven world.

Static analysis incorporates several techniques for examining programs that are not currently running. Static code analysis (SCA) is an automated process which examines source code for bugs and vulnerabilities based on sets of pre-defined rules. "Static program analyses are used by many developers to test their programs because they are effective in finding some trivial bugs that can be caught by the rules that define security violations with very small resource." (1) Static code analysis can provide comprehensive code coverage but at the cost of high false positive and false negative rates. SCA is a linear technique that is only as good as the rules girding it, but is foundational for other types of analyzers.

Fuzz testing, or *fuzzing*,is the primary dynamic analysis tool of choice. Fuzz testing is the process of monitoring a running application while providing invalid or malformed input data. As increasing amounts of bad data is inserted, the application is audited for a failure – a crash, hang, or unexpected behavior. Fuzzing can achieve only 10-50 percent code coverage, but every failure detected is a valid bug to be fixed. The appeal of fuzz testing lay in its depth, the ability to test the robustness of an application.

The newest dynamic analysis tool being researched and used is concolic execution which combines the code coverage of symbolic execution with dynamic test generation. Concolic execution suffers from the same resource exhaustion constraints of symbolic execution, but provides a theoretical 100% code coverage. Most concolic executors are configured to limit code coverage to stay within acceptable resource limits.

In this thesis, I intend to identify the sources of resource exhaustion in both symbolic and concolic execution, and optimize the code to reduce either runtime or memory usage. The contribution of this thesis will be to improve the efficiency of symbolic and concolic execution program so that the imposed coverage limits can be raised or possibly eliminated and therefore increase code coverage and the number of bugs and vulnerabilities identified.

## 1.2 Software Bugs and Vulnerabilities

In simplest terms, a bug is a flaw in a program’s logic. A bug can cause the program to not function in the way it was designed or it can, given the right set of circumstances, create a way for a hacker to gain unauthorized access to an application, computer, or network. This condition is not trivial; in fact, “it is estimated that there are as many as 20 flaws per thousand lines of software code.” (1) Left to their own devices in the wild, these resource bugs can do tremendous economic damage to individual and corporate bank accounts.

A vulnerability is neither a subset of software bug nor a flavor of it. A vulnerability is its own species of bug which allows any unauthorized access to or control over an application, the computer on which it is running, or the network to which the computer is attached. Give an example or description here.

Despite all the efforts that have been made to finding and eradicating vulnerabilities, the number of vulnerabilities found in the wild and the corresponding number of software breaches reported is prolific. The traditional bug-hunting tools appear to be insufficient to the task of finding the vulnerabilities. [need quote here about how software will never be 100% bug free]

details about symbolic and concolic execution will be presented later in chapter 4.—considerable research has been done on symbolic execution and I want to capitalize on the work already in the literature to expand upon it.

## 1.3 Thesis Organization

This thesis is organized into five chapters. Chapter 1 is an introduction and overview of several bug-hunting techniques, definitions of applicable terms, problems to be solved, and the goals of my work. Chapter 2 will discuss in more depth the current tools being used to find bugs. The strengths and weaknesses of these tools will be compared. Chapter 3 is a review of relevant literature on the topics of static code analysis, fuzz testing, symbolic execution and dynamic test generation. Chapter 4 will contain details of my work, the results of my tests, and analysis of those results. Chapter 5 will contain a summary of my results and suggested future work.