Bachelor/Master Thesis Topic
Grammar-Mining for Open Office Documents

Motivation and Background
Software testing is a crucial part of software development. It enables assurance, of correctness, completeness and reliability of software systems. Current state-of-the-art software testing techniques employ symbolic execution-based whitebox fuzzing techniques to generate automatically test inputs that expose software bugs. In practice, many programs require highly structured and complex files. The generation of such valid files is non-trivial and, hence, traditional whitebox fuzzing often lead to bad results spending too much time exploring parser errors, instead of the functional part of the program. Recent approaches like [1] and [2] try to tackle the problem by building a more accurate model of a valid program input, e.g. so-called grammar-based whitebox fuzzing requires the input grammar that is used to generate valid test inputs. The work in [3] shows how input grammars can be mined successfully by leveraging dynamic tainting.

Goals
The goal of this project is to mine an input grammar for Open Office documents, so that it can guide whitebox fuzzing testing.

Description of the Task
- Understand the current approaches in grammar mining and whitebox fuzzing
- Getting familiar grammar-mining tool support, eg. AUTOGRAM [3]
- Retrieve and annotate a grammar for Open Office documents
- Perform an experimental evaluation of the mined grammar by using whitebox fuzzing

Research Type
Theoretical Aspects:  
Industrial Relevance:  
Implementation  

Prerequisite
The student should be enrolled in the bachelor/master of computer science program, and has completed the required course modules to start a bachelor/master thesis.

Skills required
Programming skills in Java or C++, understanding of, or willingness to learn, the software engineering and software analysis foundations needed for the project.

References

Contacts
Lars Grunse, Humboldt-Universität zu Berlin, Institut für Informatik, Lehrstuhl Software Engineering, Unter den Linden 6, 10099 Berlin, Germany

Application
Please contact me during my office hours or send me an email with the title: “[ThesisProject]-GrammarMiningOpenOffice” to se-career@informatik.hu-berlin.de