**MCFromC Project Overview**

# Glossary

* AST – Abstract Syntax Tree – a tree data structure that contains the given source code as a syntax tree.
* CFG – Control Flow Graph
* MC – Monotonicity Constraints
* ACFG – Annotated Control Flow Graph is CFG that contains MCs on its edges.

# General Architecture Overview

The project architecture is based on stages where each stage takes an input, processes it, and outputs it as input to the next stage. **Figure 1** below shows a high level overview of the data flow between the system stages.

Figure 1 - System Stages

Output

CFG

CFG

AST

C Source

Syntax Parser

CFG Creation

CFG Simplifier

MC Creation

## Syntax Parser

This stage is responsible for converting a text file containing a subset of the  
**C programming language**[[1]](#footnote-1) into an AST that which will be processed into a CFG in the next stage. This stage is implemented using the following 3rd party tools:

* **Flex** – This tool uses **lex** file and produces source file that holds function that takes input stream and returns tokens according to the **lex** file.
* **Bison** – This tool is an implementation for **YACC**. The tool receives a **YACC** input file (language grammar in a format similar to BNF, where each rule contains C code snippet) and generates a C source file which is the code of a parser for the given language.

Since the above tools produce a source files in the C language (and not C++), while the system main algorithms are coded in C++, this stage is implemented as a separate DLL with an interface which takes a file name and returns a pointer to the root node of the AST.

## CFG Creation

This stage is responsible for generating a CFG based on the AST input. The output is a CFG and all the program state variable names. The graph data structure implementation is based on a 3rd party library – the **boost graph library** (**BGL**). The library is a C++ library, implemented in templates and contains only header files. This stage is implemented in a separate DLL named CFGGen, which is written in C++. The algorithm for CFG generation based on the AST of the input program will be covered in a later section.

## CFG Simplifier

The CFG simplifier is a small stage, implemented in the same DLL as the previous stage (CFG Creation) and is responsible to removal of redundant flow points from the CFG. One example for such redundancy is appearance of consecutive expression flow points (basic block). Those flow points can be reduced into a single flow point since there are no branches within the block. The current stage aggregates those flow points into a single flow point that which represents and holds the information of all other expression flow points.

## MC Creation

This MC Creation layer is responsible for trying to evaluate changes in the values of the program state variables in each abstract transition of the program (represented as an edge in the CFG). The input for this stage is a CFG (simplified from the previous stage) and the names of all state variables (found by traversing the AST, during the CFG generation). Abstract Interpretation is used in order to track changes in the values of the program state variables. This stage will be covered in detail in **MC generation** section.

The following chart describes the normal application flow between the different code projects.

Figure 2 - Code Project Flow

# User Manual

## Installation

The installation of the project is pretty simple all you need to do is to perform double click on **MCFromCSetup** and proceed with the installation. Note that you have to be administrator in order to perform installation.

### Inspecting the Install Directory

Under the installation directory you will find the following items:

* Main project's binary outputs that includes the following: SyntaxParserDLL.dll,CFGGen.dll, ParserDotNetBridge.dll, CFGViewer.exe
* A Zip file, containing the project's source along with documentation.

## Usage

When launching the **CFGViewer.exe** application the main application window will be opened:

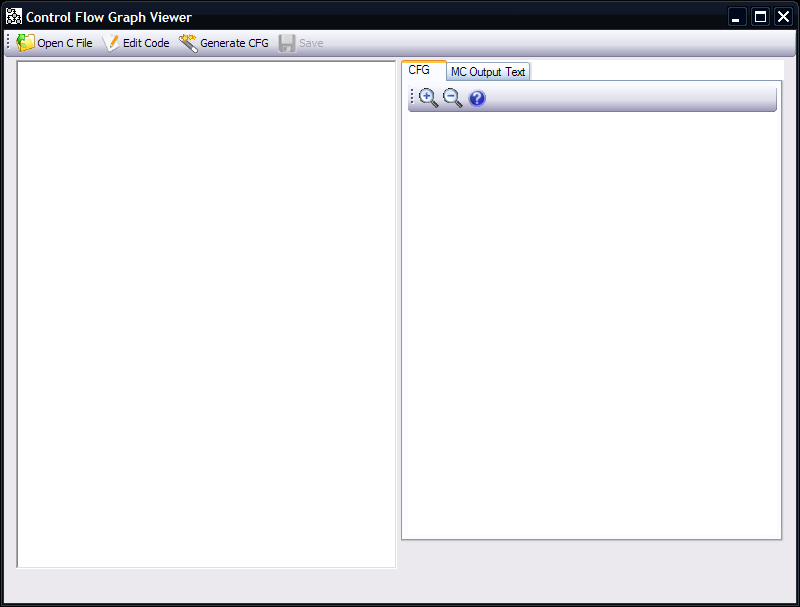


Figure 3 - Main Application Window

### Generating CFG from C code

In order to generate a CFG from a "C" source you will need to provide the "C" source by either loading the code from a source file:

1. Click on the "**Open C File**" button on the toolbar.
2. Select the "C" source file.

Or by manually editing the code:

1. Click on the "**Edit Code**" button on the toolbar.
2. Edit the code in the left pane.

Then just click on the "**Generate CFG**" button on the toolbar.

### Inspecting Results

After the CFG is generated you can view the results in the right pane:

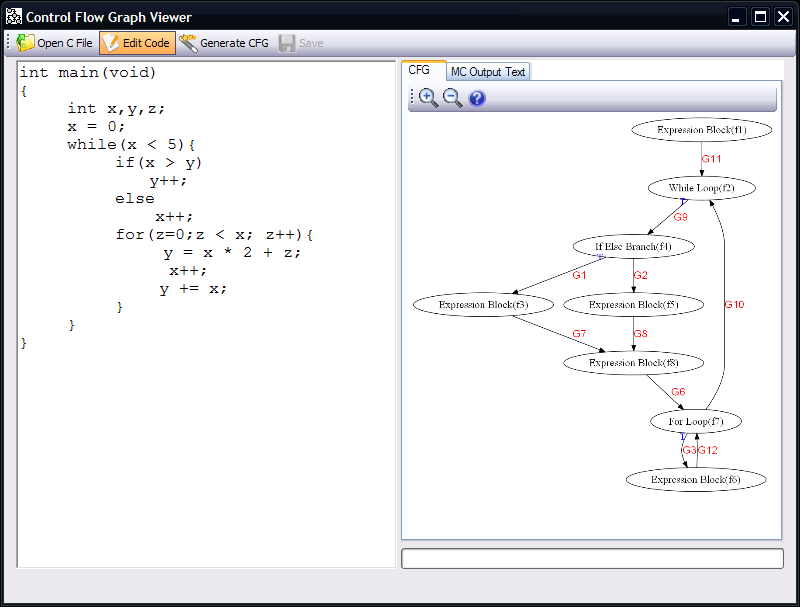


Figure 4 - Result Window

As you can see in Figure 4 above, the input "C" source code is on the left pane and the result CFG is on the right pane. When hovering with the mouse cursor over a flow point (vertex in the CFG), the code represented by the flow point is displayed in a tool-tip and in the status-bar below the CFG. In order to inspect a transition you can double click on the MC graph red label, located on the edge representing the requested transition. This action will pop a window where you can inspect how specific transition affects the program state variables:

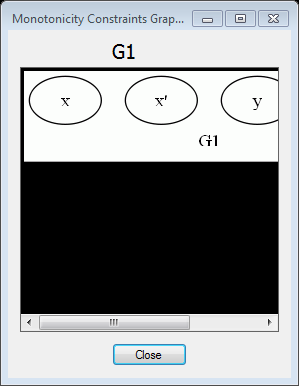


Figure 5 - MC Inspection window

In case you would like you can inspect the result in "C. S. Lee MCS Format" (you can find the format in Appendix A: C. S. Lee MCS Format) by clicking the "**MC Output Text**" tab that is located below the right pane toolbar. You can also save this text to file by clicking on the "**Save**" button that is located at the main application window's toolbar.

# Algorithms

## Converting AST to CFG

This page describes general outlines of the algorithm used for converting the Abstract Syntax Tree(AST), as received by the Syntax Parser, into a Control Flow Graph(CFG).

### Algorithm Details

*Input*: **AST** - an Abstract Syntax Tree.

*Output*: **CFG** - a Control Flow Graph, containing Flow Points and directed edges between them.

1. For each handled node in the AST (mainly expressions and control flow constructs), create a Flow Point in the CFG.
2. Merge consecutive commands Flow Points into a compound block - a block of commands which are consecutive in the program code. This block does not appear in the CFG.
3. Merge consecutive expression Flow Points into a single Flow Point in the CFG(representing an expression block)
4. Recursively traverse the AST:
   1. Connect each node Flow Point to its sub-nodes Flow Points.
   2. If current node is a loop node, connect all end-nodes Flow Points to its Flow Point.
   3. If current node's Flow Point is part of a compound block, connect its end points (might be the node's Flow Point itself) to the next Flow Point inside the compound block.
   4. Return current end points to the calling level.
5. If there are consecutive expression blocks, merge them into a single block.

### Remarks

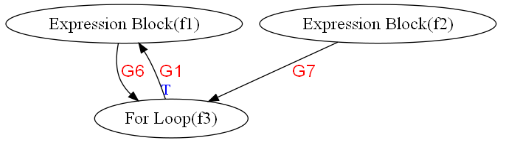
Each branch command (if / if else / while / for / do while) in the AST is handled differently by the AST to CFG conversion algorithm, according to the specific command semantics. For instance, a For Loop of the following form:

for (i = 0; i < 100; ++i)

{

x = x + i;

}

Is interpreted into the following graph:

Where flow point **f2** represents the for loop initialization (i = 0;), and flow point **f1** represents the commands:   
 1. x = x + i;  
 2. ++i;

## MC generation

### Description of how expressions are evaluated

### Algorithm

# Supported grammar (in BNF format)

# Appendix A: C. S. Lee MCS Format

**Syntax**

*MCS* ::= *MC* { *MC* } every rule on separate line

*MC* ::= *head* **':-'** *body* **';'** *tail*

*head, tail* ::= *FPid* **'('** *Vid* { , *Vid* } **')'**

*body* ::= **'['** *inequality* { , *inequality* } **']'**

*inequality* ::= *Vid* [ **'>'** | **'>='** | **'='** | **'<'** | **'<='** ] *Vid*

**Lexical Tokens:**

*FPid* flow-point identifier

*Vid* variable identifier

**Rules:**

Every *Vid* that appears in a rule's body has to appear either among the variables of the head or those of the tail.

# Benchmark programs and result graphs and MCs.

1. See **Supported grammar (in BNF format)** section for the exact language definition [↑](#footnote-ref-1)