Control Flow Graph Visualization in Compiled Software Engineering

Andrey Mikhailov[*], **Aleksey Hmelnov**[*],
Evgeny Cherkashin[*][**], Igor Bychkov[*]

{mikhailov,alex,eugeneai,bychkov}@icc.ru

[*]Matrosov Institute for System Dynamics and Control Theory of Siberian Branch of Russian Academy of Sciences;

[**]Irkutsk National Research Technical University,
Irkutsk, Russian Federation

ISDCT SB RAS, INRTU
31 May 2016
Opatija, Croatia

Applications

Control flow graph (CFG) amalysis is a common stage in syntactic approaches of data mining and pattern recognition, e.g., in source and binary code analysis in software and instrumental tool quality assessment:

- compiled binary code productivity;
- quality of compiler and system libraries;
- features of hardware platforms;
- reconstruction of legacy source code;
- malware and virus code analysis.

Control flow graph

Definition

An directed graph $G(V, E)$ is a ***control flow graph*** *if the following holds:*

1. *graph G does not contain multiple edges;*
2. *node $start \in V$ is the only entrance to the graph;*
3. *node $end \in V$ is the exit from the graph;*
4. *each node $v \in V$ is accessible from $start$;*
5. *node end is accessible from each node $v \in V$.*

Definition

A node x is a ***dominator*** *of y ($x \quad dom \quad y$) in a directed graph, if any path from $start$ to y includes x.*

Definition

A node x is an ***immediate dominator*** *of y ($x \quad idom \quad y$), if $x \quad dom \quad y$ and there are no such p that $x \quad dom \quad p$ and $p \quad dom \quad y$.*

Hierarchic layout engine

Software: uDraw (daVinci), VCG, Graphlet, GraVis, Graph Drawing Server, graphViz, VisualGraph.

1. **Distribution of graph nodes between layers.** Each node is assigned a rank. All directed edges can connect nodes from a lower rank to a higher one. Rank distribution of the nodes is performed, *e.g.*, on the base of path length calculation in depth-first graph traversal procedure.
2. **Defining order on the nodes in a layer.** The nodes of a layer are ordered according to principle of minimization of intersections of edges, e.g., by means of Method of median.
3. **Figuring out of the node coordinates in a layer.** Each node of each layer is assigned a coordinate so as the graph will correspond to predefined aesthetic criteria.
4. **Edge drawing.** The edges are drawn according to rules of visualization, for example, as arrows.

Quality criteria of graph visualization

A display of the nodes and the edges of a graph on a surface (or in a 3d-space) is referred to as a *graph layout*.

- **Visual arrangement** is the main set of rules that a graph representation must obey to be acceptable as a desired result, *e.g.*, to visualize programs as a flowchart, the rules of flowchart layout is used.
- **Aesthetics** is a subset of the criteria that defines attributes of the constructed image, **improving visual quality**.
- **Restrictions** are a subset of the criteria that define layout rules for specific elements and subgraphs of the constructed image, *e.g.*, place root at the center of image, place nodes outside of a region.

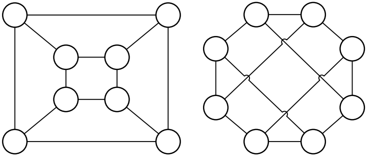


Рис.: Various layouts of the same graph

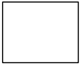
Visual arrangements for flowcharts

a) **Operator shape** represents a node where the control flow passed only one direction.

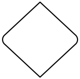
b) **Branching shape** corresponds to conditional operators in high-level programming languages; in a control flow graph it is a node, where flow control splits up.

c) **Cycle edge shapes** denote two graph nodes, one is for beginning of the cycle and one for its end, the cycle body is located between these shapes.

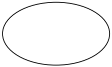
d) **Starting and terminal shapes** mark the entrance and the exit from a function or a program.



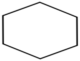
a)
operator



b)
branching



c)
cycle edges



d)
start/termination

Two terminal (TT) region

Definition

A subgraph having one entry and one exit node is a TT region.
A node pair $\langle a, b\rangle$ of a graph G is a TT region if

1) a *idom* b;
2) b *postidom* a;
3) *any graph cycle containing* a *also contains* b *and vice versa.*

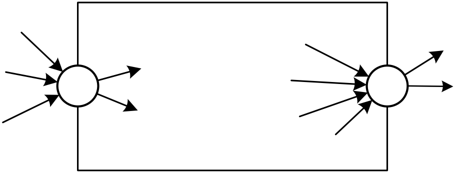


Рис.: A two terminal region

Recognizable regions

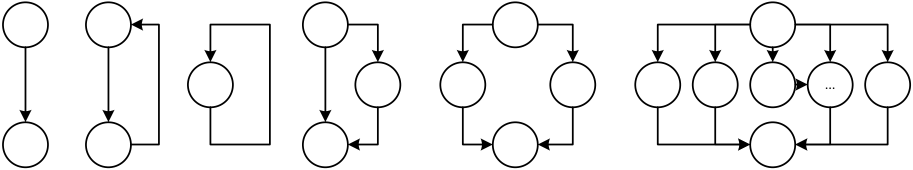


Рис.: Patterns of regions

Algorithm of control flow graph structuring

```
**Input parameters:** G, D, P
**Result:** An abstract node containing a hierarchy of folded subgraphs
for each $v \in D$ in a backward breadth-first order execute
    for each $p \in Children(v)$ execute
        if $p \quad pidom \quad v$ then
            $S \leftarrow Children(v) \setminus p$
            if $Classify\_Region(S) \neq$ undeterminated then
                $Apply\_Template(S)$
            end of condition
            else
                $Hierarchical\_Layout(S \cup p)$
                $Recognize\_Undeterminanted\_Region(S)$
            end of condition
            $Modify(G, D, P)$
        end of condition
    end of loop
end of loop
```

Layout procedure

The layout process is a top-down recursive procedure of region recognition and visualization. For the top region the initial coordinates are specified.

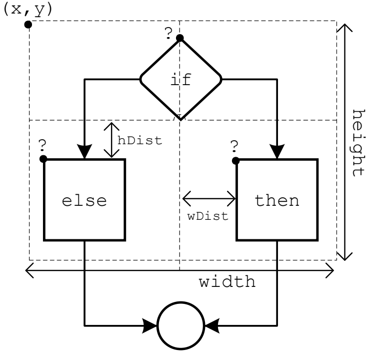
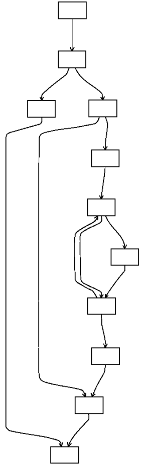
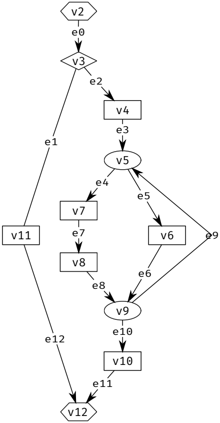


Рис.: pattern of if-then-else operator

A layout example of a control flow graph



a) Hierarchical layout



b) Structural layout

Testing results

- 197.parser[1]
- 252.eon[2]

About 70% of graphs are structured completely **without undeterminanted** regions. Around 96% of recognized regions are structured.

Main advantages of the approach:

- Visual arrangement rule set change by means of new templates.
- Similar visualization for the same operators of different programming languages.
- The possibility to emphasize graph regions according to a recognized semantics.

[1] Syntactic parsing for natural language

[2] Ray tracing

Control Flow Graph Visualization in Compiled Software Engineering

Andrey Mikhailov[*], **Aleksey Hmelnov**[*],
Evgeny Cherkashin[* **], Igor Bychkov[*]
{mikhailov,alex,eugeneai,bychkov}@icc.ru

[*]Matrosov Institute for System Dynamics and Control Theory of Siberian Branch of Russian Academy of Sciences;

[**]Irkutsk National Research Technical University,
Irkutsk, Russian Federation

ISDCT SB RAS, INRTU
31 May 2016
Opatija, Croatia