RE: Lecture 9; theory notes required to understand “CFG, Graph Cyclomatic Complexity, basis paths coverage method”

Appendix 8.10.3 McCabe's testing strategy

I. Background

In a **strongly connected graph G**, for any nodes x, y there is a path from x to y and vice versa. Each path can be represented as an n-tuple where n is the number of nodes. For example, in the Graph G of Figure 8.30 each path can be represented as a 6-tuple (vector with 6 components):

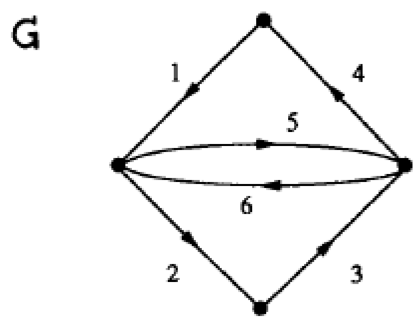

Figure 8.30: A strongly connected graph

$< 1, 2, 3, 4 > = (1\ 1\ 1\ 1\ 0\ 0)$
$< 1, 5, 6, 2, 3, 4, 1 > = (2\ 1\ 1\ 1\ 1\ 1)$
$< 1, 5, 4 > = (1\ 0\ 0\ 0\ 1\ 1)$ etc.

That is, the ith position in the vector is the number of occurrences of edge i.

A **circuit** is a path that begins and ends at the same node, e.g. $< 1, 2, 3, 6, 5, 4 >$.

A **cycle** is a circuit with no node (other than the starting node) included more than once, e.g. $< 1, 2, 3, 4 >$ and $< 5, 6 >$.

A path p is said to be a **linear combination** of paths $p_1,\ldots,p_n$ if there are integers $a_1,\ldots,a_n$ such that $p = \sum a_i p_i$ in the vector representation, e.g. path $< 1, 2, 3, 4, 1, 5, 6 >$ is a linear combination of paths $< 1, 2, 3, 4 >$ and $<1, 5, 6 >$ since

$$(2\ 1\ 1\ 1\ 1\ 1) = (1\ 1\ 1\ 1\ 0\ 0) + (1\ 0\ 0\ 0\ 1\ 1)$$

As another example, let:

$$a = < 1, 2, 3, 4 > = (1\ 1\ 1\ 1\ 0\ 0)$$
$$b = < 5, 6 > = (0\ 0\ 0\ 0\ 1\ 1)$$
$$c = < 1, 5, 4 > = (1\ 0\ 0\ 1\ 1\ 0)$$
$$d = < 2, 3, 6 > = (0\ 1\ 1\ 0\ 0\ 1)$$

Then $$a + b - c = d \qquad (*)$$

A set of paths is **linearly independent** if no path in the set is a linear combination of any other paths in the set. Thus $\{ a, b, c \}$ is linearly independent, but $\{ a, b, c, d \}$ is not by virtue of (*).

A **basis set** of cycles is a maximal linearly independent set of cycles. In a graph of e edges and n nodes, the basis has $e - n + 1$ cycles. Although the size of the basis is invariant, its content is not. For example, for the graph G above:

$$\{ a, b, c \}, \{ a, b, d \}, \{ b, c, d \}, \{ a, c, d \}$$

are different basis sets of cycles. Every path is a linear combination of basis cycles.

II. Strategy

Any flowgraph can be transformed into a strongly connected graph by adding an edge from stop node to start node. Figure 8.31 shows how we transform a flowgraph to obtain the same graph as that of Figure 8.30.

McCabe's test strategy is based on choosing a basis set of cycles in the resulting graph G.

As we have seen above, the number of these is $v(G)$ (McCabe's "cyclomatic complexity") satisfying

$$v(G) = e - n + 1$$

Since the original flowgraph has had one edge added, the formula for computing $v(G)$ for an arbitrary flowgraph G with e edges and n nodes is:

$$v(G) = e - \overset{2}{n} + 2$$

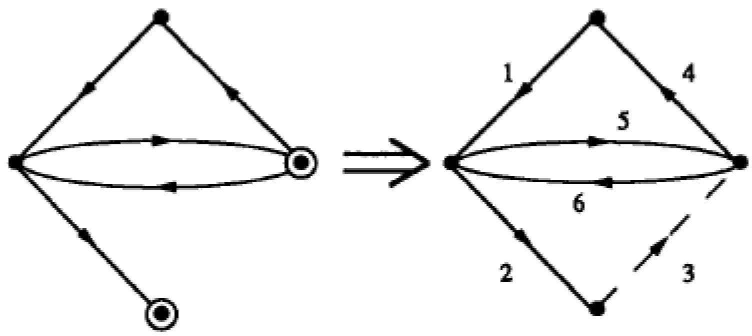

Figure 8.31: Transforming a flowgraph into a strongly connected graph

The idea is that these cycles are "representative" of all the paths since every path is a linear combination of basis cycles. In the example, one possible basis set of cycles is:

$$\{ < 4, 1, 2, 3 >, < 6, 5, >, <6, 2, 3> \}$$

Unfortunately these do not correspond to paths through the flowgraph (because of the artificial introduction of edge 3) which is what we need for a structural testing strategy. We thus have to derive the "smallest" associated paths. These are:

$$< 4, 1, 2 >, < 6, 5, 6, 2 >, < 6, 2 >$$

Note that a different basis set of cycles such as

$$\{ < 6, 5 >, < 6, 2, 3 >, < 4, 1, 5 > \}$$

leads to a different set of testing paths:

$$< 6, 5, 6, 2 >, < 6, 2 >, < 4, 1, 5, 6, 2 >$$

We can also show that if all predicate nodes have out-degree 2 then $v(G) = d + 1$ where d is the number of predicate nodes in G. For if there are p procedure nodes, then $n = p + d + 1$ since the nodes of G are: the procedure nodes, the predicate nodes, plus a single stop node. Now $e = p + 2d$ since each procedure node contributes 1 to the total number of edges and each predicate node contributes 2. Thus $e - n + 2 = (p + 2d) - (p + d + 1) = d - 1$.