Simplifying the topology

Domain testing overview

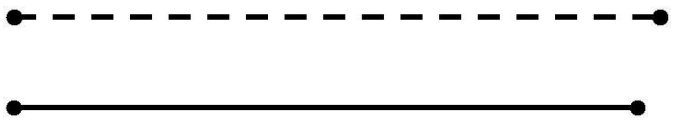
- Domains are defined by their boundaries; therefore, domain testing concentrates test points on or near boundaries.
- Classify what can go wrong with boundaries, then define a test strategy for each case. Pick enough points to test for all categorized kinds of boundary errors.
- Run the tests by post test analysis determine if any boundaries are faulty and if so, how.
- Run enough tests to verify every boundary of every domain.

Domain bugs and how to test for them

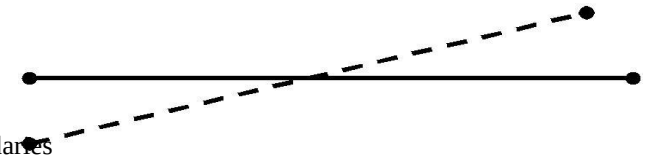
- An interior point is a point in the domain such that all points with in an arbitrarily small distance are also in the domain.
- A boundary point is one such that within an epsilon neighborhood there are points both in the domain and not in the domain.
- An extreme point is a point that does not lie between any two other arbitrary but distinct points of a domain.
- An on point is a point on the boundary.
- If the domain boundary is closed, an off point is a point near the boundary but in the adjacent domain.
- If the domain is open, an off point is a point near the boundary but in the domain being tested.

Generic Domain Bugs

- The generic domain Bugs are: shifted boundaries, tilted boundaries, open/closed errors, extra boundary and missing boundary.

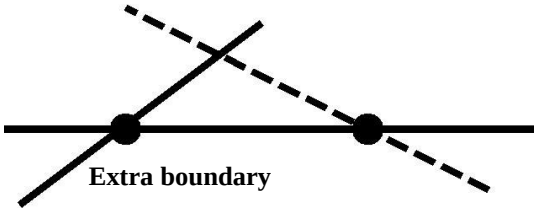


Shifted Boundaries

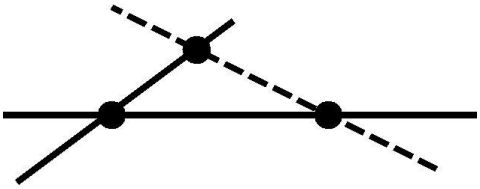


Tilted Boundaries

Open/closed error



Extra boundary



Missing Boundary

Testing one dimensional domains-open boundaries

Procedure for Testing

- The procedure is conceptually is straight forward.
- It can be done by hand for two dimensions and for a few domains and practically impossible for more than two variables.
- 1. identify input variables.
- 2. identify variable which appear in domain defining predicates, such as control flow predicates.
- 3. interpret all domain predicates in terms of input variables.
- 4. for p binary predicates, there are at most 2^p combinations of TRUE-FALSE values and therefore, at most 2^p domains. Find the set of all non null domains. the result is a boolean expression in the predicates consisting a set of AND terms joined by OR's. for example ABC+DEF+GHI….. Where the capital letters denote predicates. Each product term is a set of linear inequality that defines a domain or a part of a multiply connected domains.
- 5. Solve these inequalities to find all the extreme points of each domain using any of the linear programming methods.

UNIT V :

Paths, Path products and Regular expressions : Path products & path expression, reduction procedure, applications, regular expressions & flow anomaly detection.

PATHS, PATH PRODUCTS AND REGULAR EXPRESSIONS

This unit gives an in depth overview of Paths of various flow graphs, their interpretations and application.

PATH PRODUCTS AND PATH EXPRESSION:

- **MOTIVATION:**
  - Flow graphs are being an abstract representation of programs.
  - Any question about a program can be cast into an equivalent question about an appropriate flowgraph.
  - Most software development, testing and debugging tools use flow graphs analysis techniques.
- **PATH PRODUCTS:**
  - Normally flow graphs used to denote only control flow connectivity.
  - The simplest weight we can give to a link is a name.

- Using link names as weights, we then convert the graphical flow graph into an equivalent algebraic like expressions which denotes the set of all possible paths from entry to exit for the flow graph.
- Every link of a graph can be given a name.
- The link name will be denoted by lower case italic letters.
- In tracing a path or path segment through a flow graph, you traverse a succession of link names.
- The name of the path or path segment that corresponds to those links is expressed naturally by concatenating those link names.
- For example, if you traverse links a,b,c and d along some path, the name for that path segment is abcd. This path name is also called a **path product.** Figure 5.1 shows some examples:

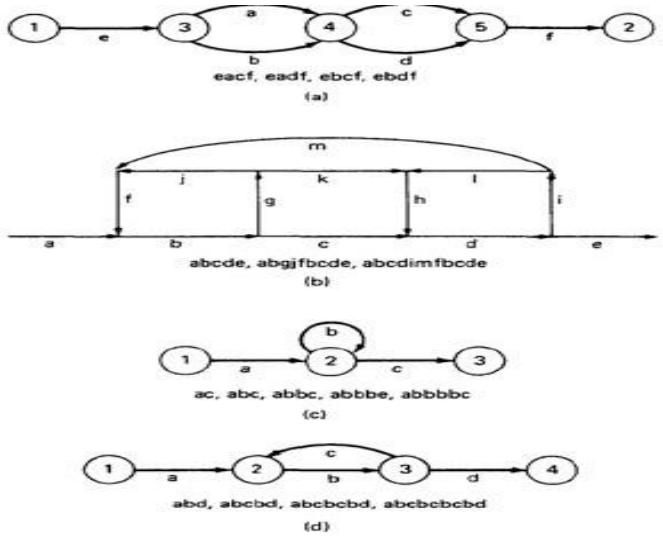


Figure 5.1: Examples of paths.

- **PATH EXPRESSION:**
  - Consider a pair of nodes in a graph and the set of paths between those node.
  - Denote that set of paths by Upper case letter such as X,Y. From Figure 5.1c, the members of the path set can be listed as follows:

    ac, abc, abbc, abbbc, abbbbc.....

  - Alternatively, the same set of paths can be denoted by :

    ac+abc+abbc+abbbc+abbbbc+.....

  - The + sign is understood to mean "or" between the two nodes of interest, paths ac, or abc, or abbc, and so on can be taken.
  - Any expression that consists of path names and "OR"s and which denotes a set of paths between two nodes is called a "**Path Expression.**".
- **PATH PRODUCTS:**
  - The name of a path that consists of two successive path segments is conveniently expressed by the concatenation or **Path Product** of the segment names.
  - For example, if X and Y are defined as X=abcde,Y=fghij,then the path corresponding to X followed by Y is denoted by

    XY=abcdefghij

- Similarly,
- YX=fghijabcde
- aX=aabcde
- Xa=abcdea
  XaX=abcdeaabcde

- If X and Y represent sets of paths or path expressions, their product represents the set of paths that can be obtained by following every element of X by any element of Y in all possible ways. For example,
- X = abc + def + ghi
- Y = uvw + z
  Then,
  XY = abcuvw + defuvw + ghiuvw + abcz + defz + ghiz

- If a link or segment name is repeated, that fact is denoted by an exponent. The exponent's value denotes the number of repetitions:
- $a^1 = a$; $a^2 = aa$; $a^3 = aaa$; a^n = aaaa . . . n times.

  Similarly, if

  X = abcde

  then

  X^1 = abcde
  X^2 = abcdeabcde = $(abcde)^2$
  X^3 = abcdeabcdeabcde = $(abcde)^2$abcde
  = abcde$(abcde)^2$ = $(abcde)^3$

- The path product is not commutative (that is XY!=YX).
- The path product is Associative.

  RULE 1: A(BC)=(AB)C=ABC

  where A,B,C are path names, set of path names or path expressions.

- The zeroth power of a link name, path product, or path expression is also needed for completeness. It is denoted by the numeral "1" and denotes the "path" whose length is zero - that is, the path that doesn't have any links.
- $a^0 = 1$
- $X^0 = 1$

- **PATH SUMS:**
  - The "+" sign was used to denote the fact that path names were part of the same set of paths.
  - The "PATH SUM" denotes paths in parallel between nodes.
  - Links a and b in Figure 5.1a are parallel paths and are denoted by a + b. Similarly, links c and d are parallel paths between the next two nodes and are denoted by c + d.
  - The set of all paths between nodes 1 and 2 can be thought of as a set of parallel paths and denoted by eacf+eadf+ebcf+ebdf.
  - If X and Y are sets of paths that lie between the same pair of nodes, then X+Y denotes the UNION of those set of paths. For example, in Figure 5.2:

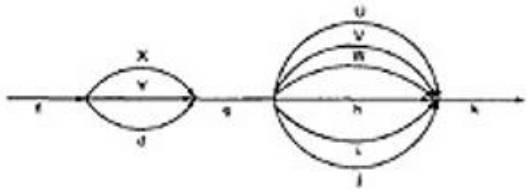


Figure 5.2: Examples of path sums.

The first set of parallel paths is denoted by X + Y + d and the second set by U + V + W + h + i + j. The set of all paths in this flowgraph is f(X + Y + d)g(U + V + W + h + i + j)k

- The path is a set union operation, it is clearly Commutative and Associative.
- RULE 2: X+Y=Y+X
- RULE 3: (X+Y)+Z=X+(Y+Z)=X+Y+Z

- **DISTRIBUTIVE LAWS:**
  - The product and sum operations are distributive, and the ordinary rules of multiplication apply; that is

    RULE 4: A(B+C)=AB+AC and (B+C)D=BD+CD

  - Applying these rules to the below Figure 5.1a yields
  - e(a+b)(c+d)f=e(ac+ad+bc+bd)f = eacf+eadf+ebcf+ebdf
- **ABSORPTION RULE:**
  - If X and Y denote the same set of paths, then the union of these sets is unchanged; consequently,

    RULE 5: X+X=X (Absorption Rule)

  - If a set consists of paths names and a member of that set is added to it, the "new" name, which is already in that set of names, contributes nothing and can be ignored.
  - For example,
  - if X=a+aa+abc+abcd+def then
    X+a = X+aa = X+abc = X+abcd = X+def = X

    It follows that any arbitrary sum of identical path expressions reduces to the same path expression.

- **LOOPS:**
  - Loops can be understood as an infinite set of parallel paths. Say that the loop consists of a single link b. then the set of all paths through that loop point is b0+b1+b2+b3+b4+b5+.....

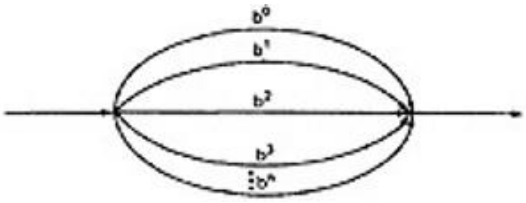


Figure 5.3: Examples of path loops.

- This potentially infinite sum is denoted by b* for an individual link and by X* when X is a path expression.

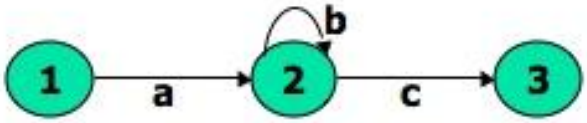


Figure 5.4: Another example of path loops.

- The path expression for the above figure is denoted by the notation:

  ab*c=ac+abc+abbc+abbbc+.....

- Evidently,

  aa*=a*a=a+ and XX*=X*X=X+

- It is more convenient to denote the fact that a loop cannot be taken more than a certain, say n, number of times.
- A bar is used under the exponent to denote the fact as follows:

  $X^{\underline{n}} = X^0+X^1+X^2+X^3+X^4+X^5+.....+X^n$

- **RULES 6 - 16:**
  - The following rules can be derived from the previous rules:
  - RULE 6: $X^{\underline{n}} + X^{\underline{m}} = X^{\underline{n}}$ if n>m
    RULE 6: $X^{\underline{n}} + X^{\underline{m}} = X^{\underline{m}}$ if m>n
    RULE 7: $X^nX^m = X^{\underline{n+m}}$
    RULE 8: $X^nX^* = X^*X^n = X^*$
    RULE 9: $X^nX^+ = X^+X^n = X^+$
    RULE 10: $X^*X^+ = X^+X^* = X^+$
    RULE 11: 1 + 1 = 1
    RULE 12: 1X = X1 = X
    Following or preceding a set of paths by a path of zero length does not change the set.
    RULE 13: $1^n = 1^{\underline{n}} = 1^* = 1^+ = 1$
    No matter how often you traverse a path of zero length,It is a path of zero length.
    RULE 14: $1^+ +1 = 1^* = 1$
    The null set of paths is denoted by the numeral 0. it obeys the following rules:
    RULE 15: X+0=0+X=X
    RULE 16: 0X=X0=0
    If you block the paths of a graph for or aft by a graph that has no paths , there wont be any paths.

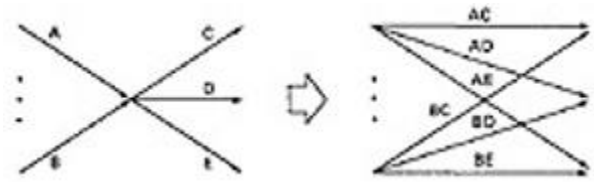
REDUCTION PROCEDURE:

- **REDUCTION PROCEDURE ALGORITHM:**

- This section presents a reduction procedure for converting a flowgraph whose links are labeled with names into a path expression that denotes the set of all entry/exit paths in that flowgraph. The procedure is a node-by-node removal algorithm.
- The steps in Reduction Algorithm are as follows:
    1. Combine all serial links by multiplying their path expressions.
    2. Combine all parallel links by adding their path expressions.
    3. Remove all self-loops (from any node to itself) by replacing them with a link of the form X*, where X is the path expression of the link in that loop.

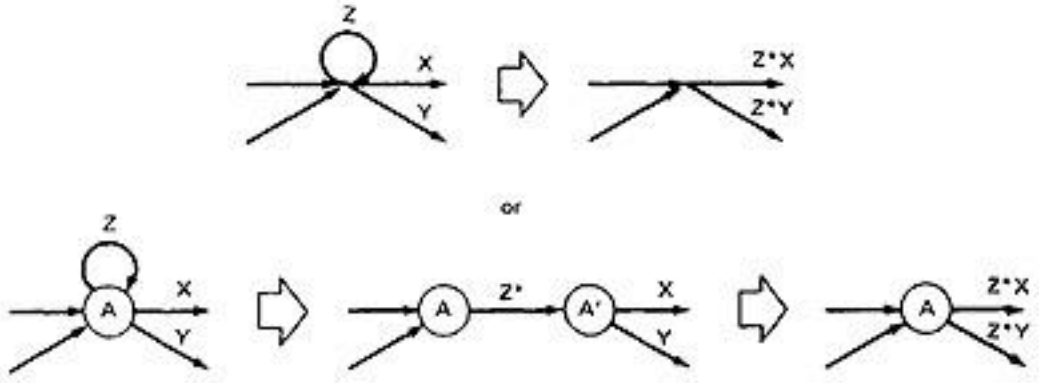
    STEPS 4 - 8 ARE IN THE ALGORIHTM'S LOOP:
    4. Select any node for removal other than the initial or final node. Replace it with a set of equivalent links whose path expressions correspond to all the ways you can form a product of the set of inlinks with the set of outlinks of that node.
    5. Combine any remaining serial links by multiplying their path expressions.
    6. Combine all parallel links by adding their path expressions.
    7. Remove all self-loops as in step 3.
    8. Does the graph consist of a single link between the entry node and the exit node? If yes, then the path expression for that link is a path expression for the original flowgraph; otherwise, return to step 4.
- A flowgraph can have many equivalent path expressions between a given pair of nodes; that is, there are many different ways to generate the set of all paths between two nodes without affecting the content of that set.
- The appearance of the path expression depends, in general, on the order in which nodes are removed.

- **CROSS-TERM STEP (STEP 4):**
    - The cross - term step is the fundamental step of the reduction algorithm.
    - It removes a node, thereby reducing the number of nodes by one.
    - Successive applications of this step eventually get you down to one entry and one exit node. The following diagram shows the situation at an arbitrary node that has been selected for removal:



    - From the above diagram, one can infer:
    - (a + b)(c + d + e) = ac + ad + + ae + bc + bd + be

- **LOOP REMOVAL OPERATIONS:**
    - There are two ways of looking at the loop-removal operation:



- In the first way, we remove the self-loop and then multiply all outgoing links by Z*.
- In the second way, we split the node into two equivalent nodes, call them A and A' and put in a link between them whose path expression is Z*. Then we remove node A' using steps 4 and 5 to yield outgoing links whose path expressions are Z*X and Z*Y.

- **A REDUCTION PROCEDURE - EXAMPLE:**
  - Let us see by applying this algorithm to the following graph where we remove several nodes in order; that is

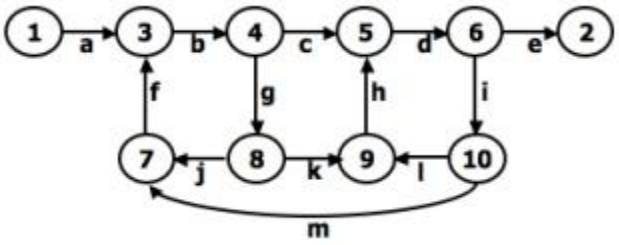
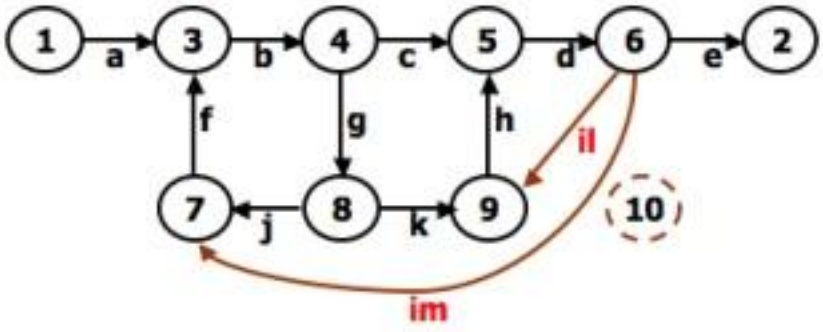
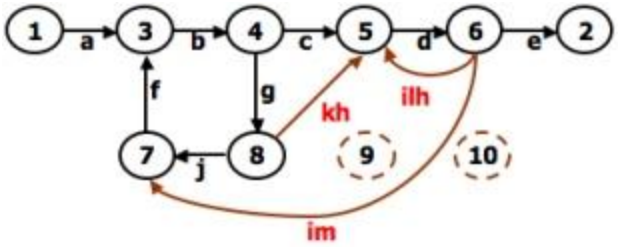


Figure 5.5: Example Flowgraph for demonstrating reduction procedure.

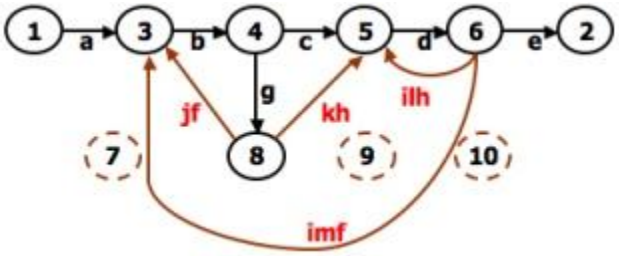
- Remove node 10 by applying step 4 and combine by step 5 to yield



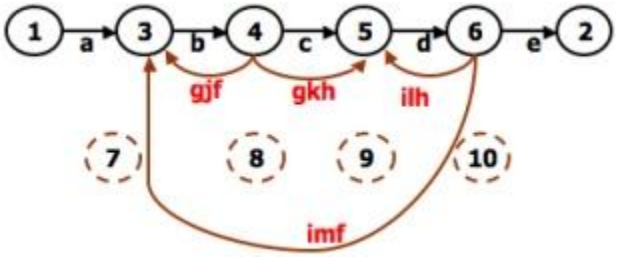
- Remove node 9 by applying step4 and 5 to yield



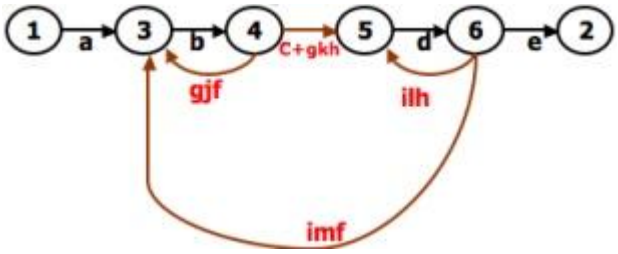
- Remove node 7 by steps 4 and 5, as follows:



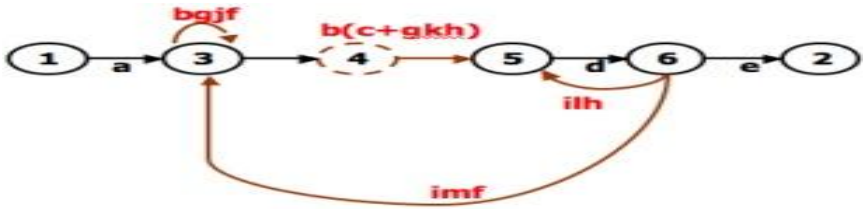
- Remove node 8 by steps 4 and 5, to obtain:



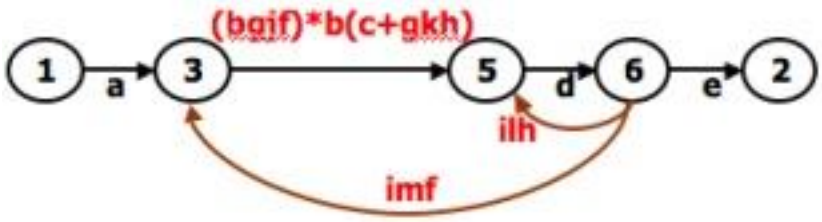
- **PARALLEL TERM (STEP 6):**
  Removal of node 8 above led to a pair of parallel links between nodes 4 and 5. combine them to create a path expression for an equivalent link whose path expression is c+gkh; that is



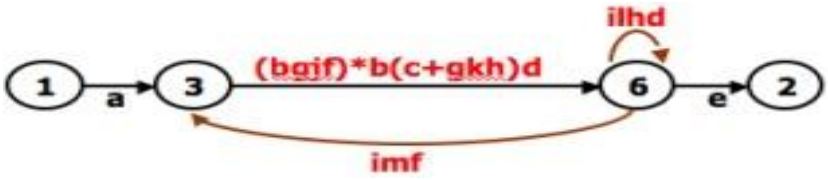
- **LOOP TERM (STEP 7):**
  Removing node 4 leads to a loop term. The graph has now been replaced with the following equivalent simpler graph:



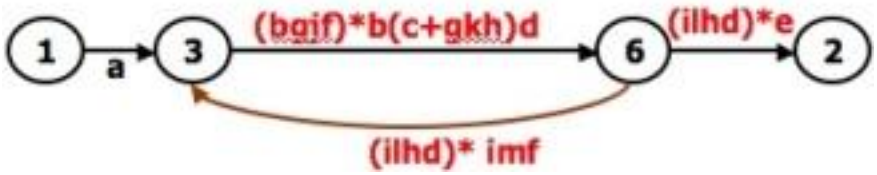
- Continue the process by applying the loop-removal step as follows:



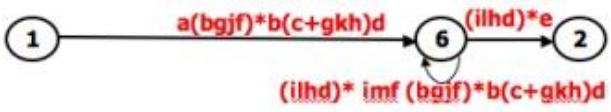
- Removing node 5 produces:



- Remove the loop at node 6 to yield:



- Remove node 3 to yield:



- Removing the loop and then node 6 result in the following expression:
- a(bgjf)*b(c+gkh)d((ilhd)*imf(bjgf)*b(c+gkh)d)*(ilhd)*e
- You can practice by applying the algorithm on the following flowgraphs and generate their respective path expressions:

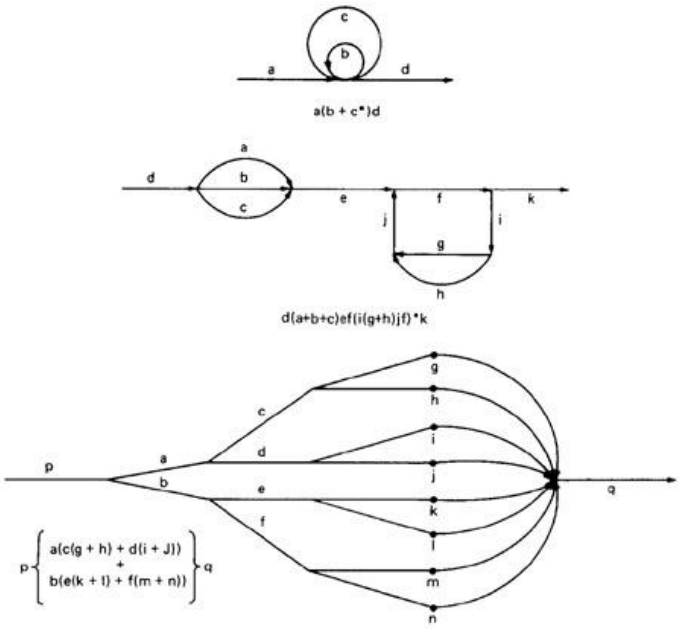


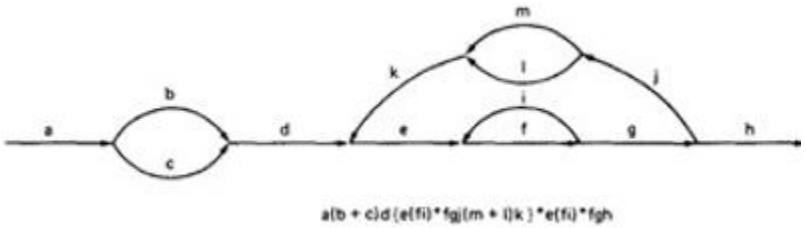
Figure 5.6: Some graphs and their path expressions.

APPLICATIONS:

- **APPLICATIONS:**
  - The purpose of the node removal algorithm is to present one very generalized concept- the path expression and way of getting it.
  - Every application follows this common pattern:
    1. Convert the program or graph into a path expression.
    2. Identify a property of interest and derive an appropriate set of "arithmetic" rules that characterizes the property.
    3. Replace the link names by the link weights for the property of interest. The path expression has now been converted to an expression in some algebra, such as ordinary algebra, regular expressions, or boolean algebra. This algebraic expression summarizes the property of interest over the set of all paths.
    4. Simplify or evaluate the resulting "algebraic" expression to answer the question you asked.
- **HOW MANY PATHS IN A FLOWGRAPH ?**
  - The question is not simple. Here are some ways you could ask it:
    1. What is the maximum number of different paths possible?
    2. What is the fewest number of paths possible?
    3. How many different paths are there really?
    4. What is the average number of paths?
  - Determining the actual number of different paths is an inherently difficult problem because there could be unachievable paths resulting from correlated and dependent predicates.
  - If we know both of these numbers (maximum and minimum number of possible paths) we have a good idea of how complete our testing is.
  - Asking for "the average number of paths" is meaningless.
- **MAXIMUM PATH COUNT ARITHMETIC:**
  - Label each link with a link weight that corresponds to the number of paths that link represents.
  - Also mark each loop with the maximum number of times that loop can be taken. If the answer is infinite, you might as well stop the analysis because it is clear that the maximum number of paths will be infinite.
  - There are three cases of interest: parallel links, serial links, and loops.

Case	Path expression	Weight expression
Parallels	A+B	W_A+W_B
Series	AB	W_AW_B
Loop	A^n	$\sum_{j=0}^{n} W_A^j$

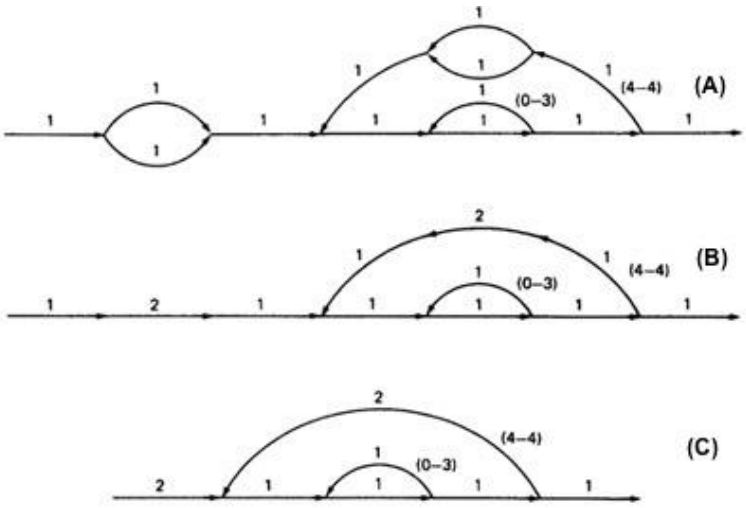
- This arithmetic is an ordinary algebra. The weight is the number of paths in each set.
- **EXAMPLE:**
  - The following is a reasonably well-structured program.



Each link represents a single link and consequently is given a weight of "1" to start. Lets say the outer loop will be taken exactly four times and inner Loop Can be taken zero or three times Its path expression, with a little work, is:

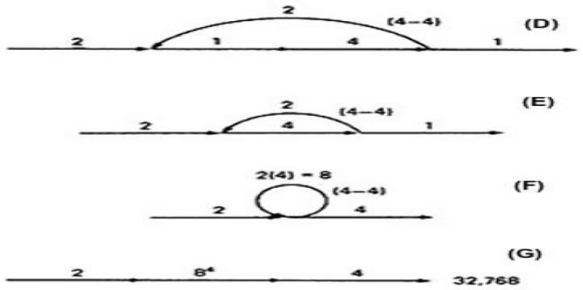
Path expression: a(b+c)d{e(fi)*fgj(m+l)k}*e(fi)*fgh

- **A:** The flow graph should be annotated by replacing the link name with the maximum of paths through that link (1) and also note the number of times for looping.
- **B:** Combine the first pair of parallel loops outside the loop and also the pair in the outer loop.
- **C:** Multiply the things out and remove nodes to clear the clutter.



- **For the Inner Loop:**
  D:Calculate the total weight of inner loop, which can execute a min. of 0 times and max. of 3 times. So, it inner loop can be evaluated as follows: $1^{\underline{3}} = 1^0 + 1^1 + 1^2 + 1^3 = 1 + 1 + 1 + 1 = 4$

- **E:** Multiply the link weights inside the loop: 1 X 4 = 4
- **F:** Evaluate the loop by multiplying the link wieghts: 2 X 4 = 8.
- **G:** Simpifying the loop further results in the total maximum number of paths in the flowgraph:
  2 X 8^4 X 2 = 32,768.



Alternatively, you could have substituted a "1" for each link in the path expression and then simplified, as follows:

a(b+c)d{e(fi)*fgj(m+l)k}*e(fi)*fgh

= $1(1 + 1)1(1(1 \times 1)^{\underline{3}}1 \times 1 \times 1(1 + 1)1)^{\underline{4}}1(1 \times 1)^{\underline{3}}1 \times 1 \times 1$

= $2(1^{\underline{3}}1 \times (2))^4 1^{\underline{3}}$

= $2(4 \times 2)^4 \times 4$

= 2 x 8^4 x 4 = 32,768

This is the same result we got graphically.

Actually, the outer loop should be taken exactly four times. That doesn't mean it will be taken zero or four times. Consequently, there is a superfluous "4" on the outlink in the last step. Therefore the maximum number of different paths is 8192 rather than 32,768.

STRUCTURED FLOWGRAPH:

Structured code can be defined in several different ways that do not involve ad-hoc rules such as not using GOTOs.

A structured flowgraph is one that can be reduced to a single link by successive application of the transformations of Figure 5.7.

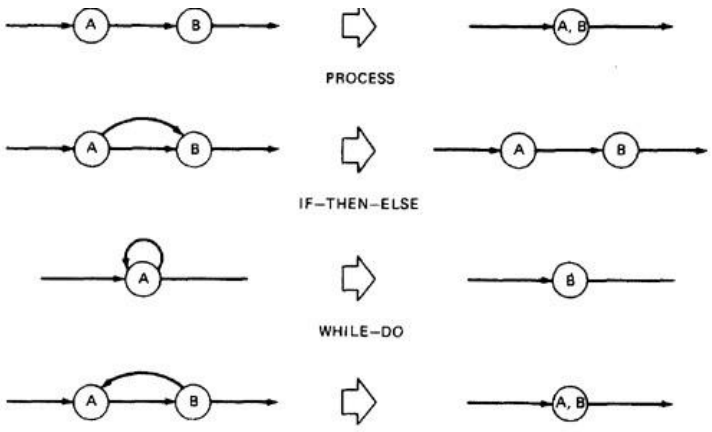


Figure 5.7: Structured Flowgraph Transformations.

The node-by-node reduction procedure can also be used as a test for structured code.

Flow graphs that DO NOT contain one or more of the graphs shown below (Figure 5.8) as subgraphs are structured.

0. Jumping into loops
1. Jumping out of loops
2. Branching into decisions
3. Branching out of decisions

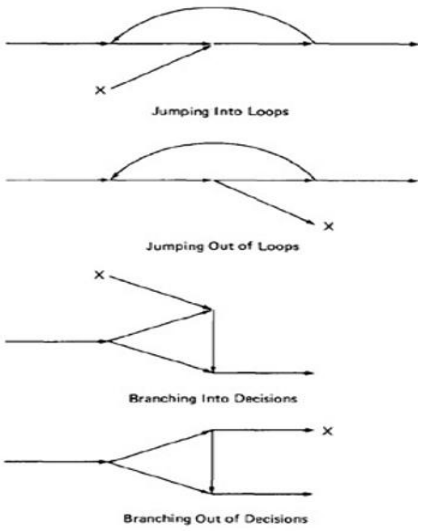


Figure 5.8: Un-structured sub-graphs.

LOWER PATH COUNT ARITHMETIC:

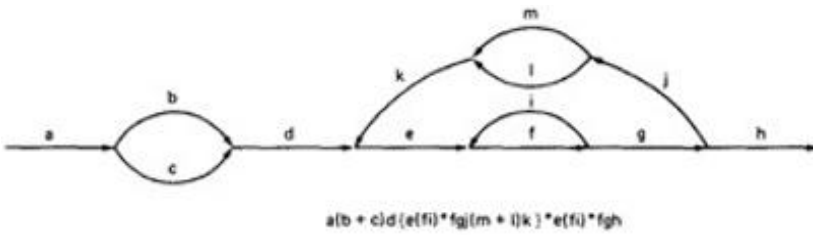
A lower bound on the number of paths in a routine can be approximated for structured flow graphs.
The arithmetic is as follows:

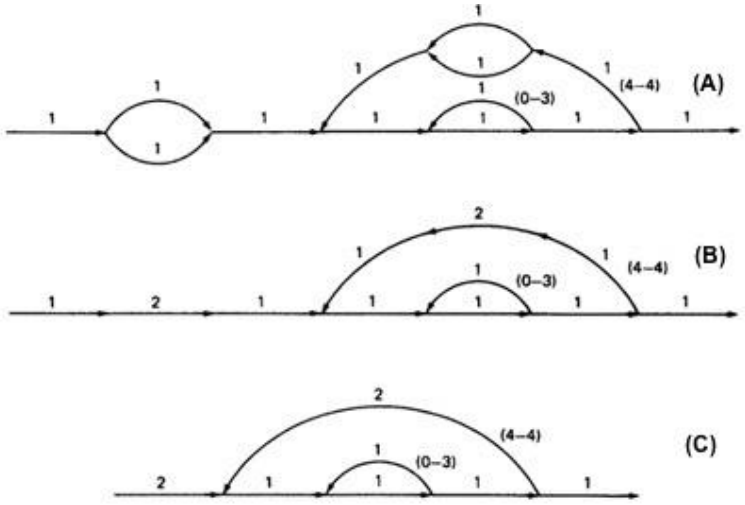
Case	Path expression	Weight expression
Parallels	A+B	W_A+W_B
Series	AB	$\max(W_A,W_B)$
Loop	A^n	1, W_1

The values of the weights are the number of members in a set of paths.

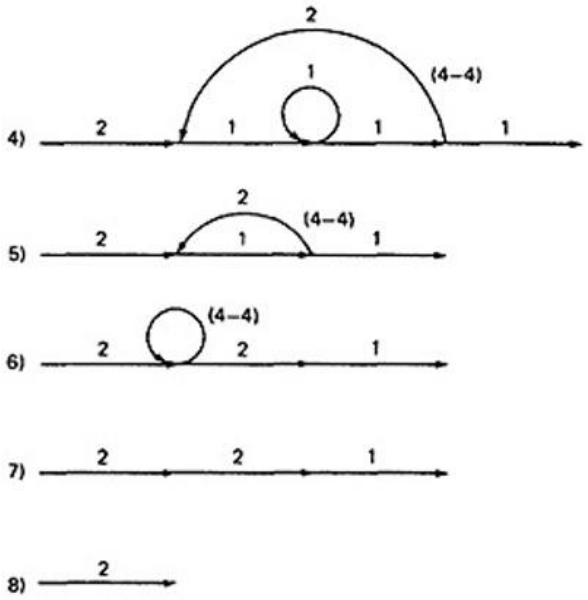
EXAMPLE:

- Applying the arithmetic to the earlier example gives us the identical steps unitl step 3 (C) as below:





- From Step 4, the it would be different from the previous example:



- If you observe the original graph, it takes at least two paths to cover and that it can be done in two paths.
- If you have fewer paths in your test plan than this minimum you probably haven't covered. It's another check.

CALCULATING THE PROBABILITY:

Path selection should be biased toward the low - rather than the high-probability paths.

This raises an interesting question: ***What is the probability of being at a certain point in a routine?*** This question can be answered under suitable assumptions, primarily that all probabilities involved are independent, which is to say that all decisions are independent and uncorrelated.

We use the same algorithm as before : node-by-node removal of uninteresting nodes.

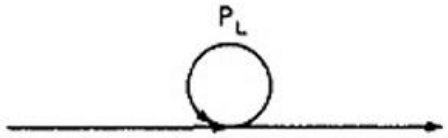
Weights, Notations and Arithmetic:

- Probabilities can come into the act only at decisions (including decisions associated with loops).
- Annotate each outlink with a weight equal to the probability of going in that direction.

- Evidently, the sum of the outlink probabilities must equal 1
- For a simple loop, if the loop will be taken a mean of N times, the looping probability is N/(N + 1) and the probability of not looping is 1/(N + 1).
- A link that is not part of a decision node has a probability of 1.
- The arithmetic rules are those of ordinary arithmetic.

Case	Path expression	Weight expression
Parallel	A+B	P_A+P_B
Series	AB	P_AP_B
Loop	A*	$P_A / (1-P_L)$

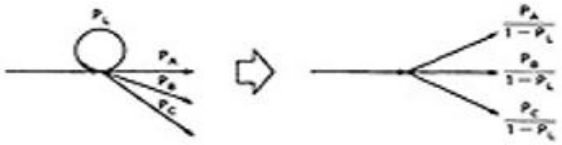
- In this table, in case of a loop, P_A is the probability of the link leaving the loop and P_L is the probability of looping.
- The rules are those of ordinary probability theory.
    1. If you can do something either from column A with a probability of P_A or from column B with a probability P_B, then the probability that you do either is $P_A + P_B$.
    2. For the series case, if you must do both things, and their probabilities are independent (as assumed), then the probability that you do both is the product of their probabilities.
- For example, a loop node has a looping probability of P_L and a probability of not looping of P_A, which is obviously equal to I - PL.



$$P_A = 1 - P_L$$

$$P_{NEW} = \frac{P_A}{1 - P_L} = \frac{1 - P_L}{1 - P_L} = 1$$

- Following the above rule, all we've done is replace the outgoing probability with 1 - so why the complicated rule? After a few steps in which you've removed nodes, combined parallel terms, removed loops and the like, you might find something like this:



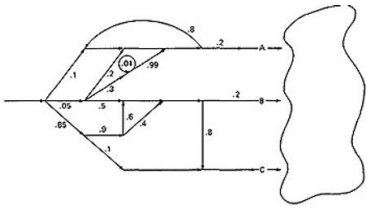
because $P_L + P_A + P_B + P_C = 1$, $1 - P_L = P_A + P_B + P_C$, and

$$\frac{P_A}{1 - P_L} + \frac{P_B}{1 - P_L} + \frac{P_C}{1 - P_L} = \frac{P_A + P_B + P_C}{1 - P_L} = 1$$

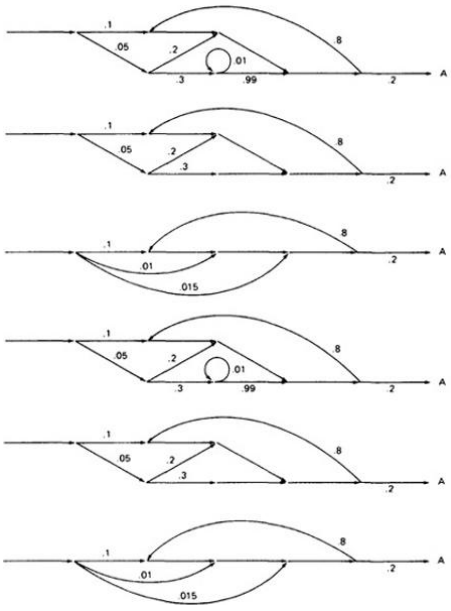
which is what we've postulated for any decision. In other words, division by 1 - P_L renormalizes the outlink probabilities so that their sum equals unity after the loop is removed.

EXAMPLE:

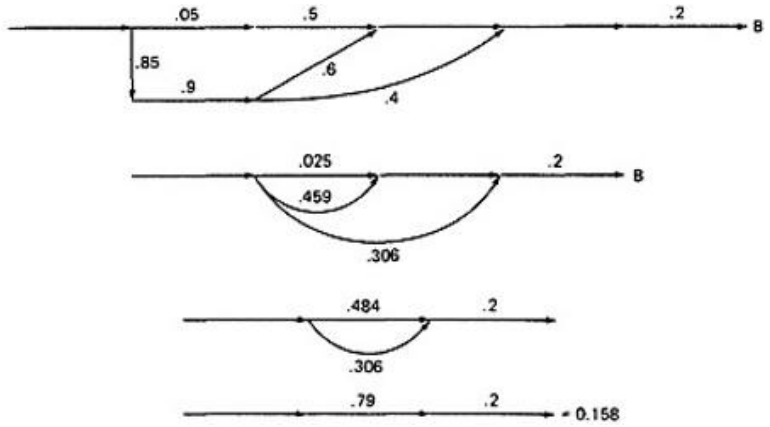
- Here is a complicated bit of logic. We want to know the probability associated with cases A, B, and C.



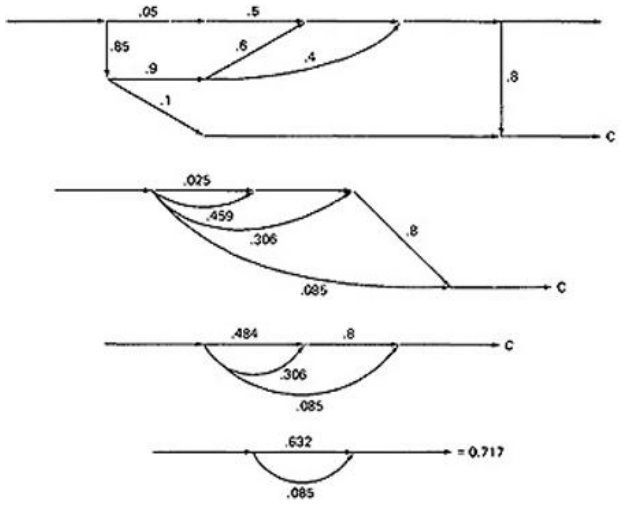
- Let us do this in three parts, starting with case A. Note that the sum of the probabilities at each decision node is equal to 1. Start by throwing away anything that isn't on the way to case A, and then apply the reduction procedure. To avoid clutter, we usually leave out probabilities equal to 1.
CASE A:



- Case B is simpler:



- Case C is similar and should yield a probability of 1 - 0.125 - 0.158 = 0.717:



- This checks. It's a good idea when doing this sort of thing to calculate all the probabilities and to verify that the sum of the routine's exit probabilities does equal 1.
- If it doesn't, then you've made calculation error or, more likely, you've left out some branching probability.
- How about path probabilities? That's easy. Just trace the path of interest and multiply the probabilities as you go.
- Alternatively, write down the path name and do the indicated arithmetic operation.
- Say that a path consisted of links a, b, c, d, e, and the associated probabilities were .2, .5, 1., .01, and I respectively. Path *abcbcbcdeabddea* would have a probability of 5×10^{-10}.
- Long paths are usually improbable.

MEAN PROCESSING TIME OF A ROUTINE:

Given the execution time of all statements or instructions for every link in a flowgraph and the probability for each direction for all decisions are to find the mean processing time for the routine as a whole.

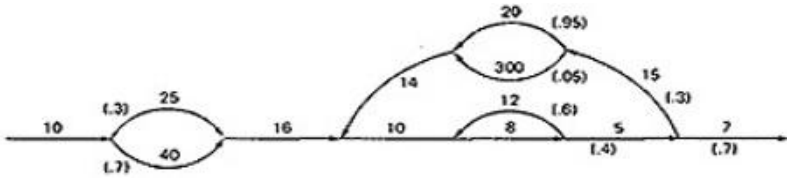
The model has two weights associated with every link: the processing time for that link, denoted by **T**, and the probability of that link **P**.

The arithmetic rules for calculating the mean time:

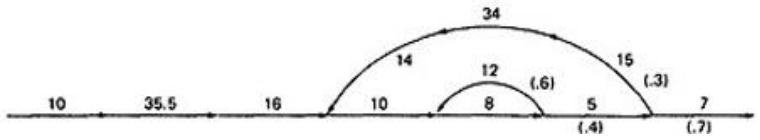
Case	Path expression	Weight expression
Parallel	A+B	$T_{A+B}=(P_AT_A+P_BT_B)/(P_A+P_B)$ $P_{A+B}= P_A+P_B$
Series	AB	$T_{AB}=T_A+T_B$ $P_{AB} =P_AP_B$
Loop	A^n	$T_A= T_A+ T_LP_L/(1-P_L)$ $P_A =P_A/(1-P_L)$

EXAMPLE:

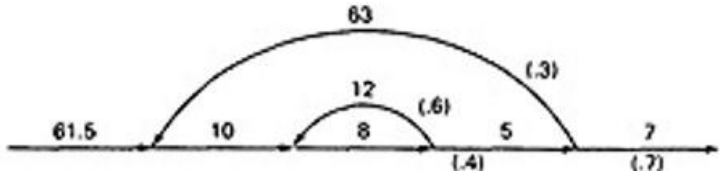
0. Start with the original flow graph annotated with probabilities and processing time.



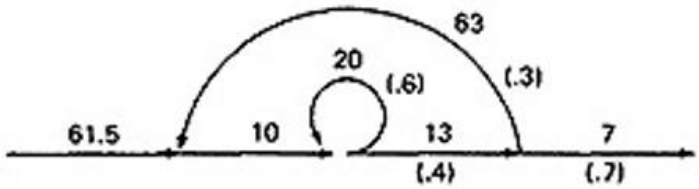
1. Combine the parallel links of the outer loop. The result is just the mean of the processing times for the links because there aren't any other links leaving the first node. Also combine the pair of links at the beginning of the flowgraph..



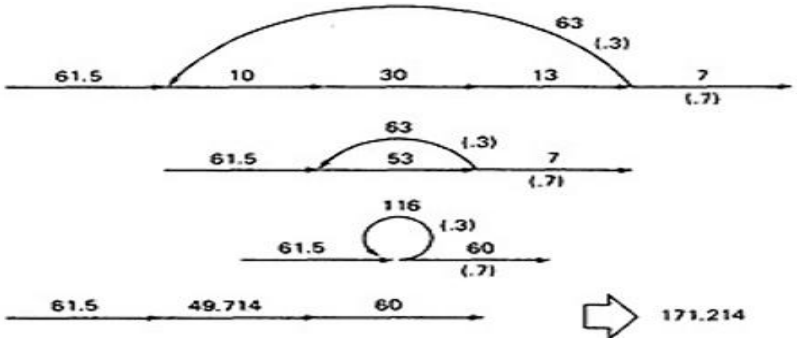
2. Combine as many serial links as you can.



3. Use the cross-term step to eliminate a node and to create the inner self - loop.



4. Finally, you can get the mean processing time, by using the arithmetic rules as follows:



PUSH/POP, GET/RETURN:

This model can be used to answer several different questions that can turn up in debugging.

It can also help decide which test cases to design.

The question is:

Given a pair of complementary operations such as PUSH (the stack) and POP (the stack), considering the set of all possible paths through the routine, what is the net effect of the routine? PUSH or POP? How many times? Under what conditions?

Here are some other examples of complementary operations to which this model applies:

GET/RETURN a resource block.

OPEN/CLOSE a file.

START/STOP a device or process.

EXAMPLE 1 (PUSH / POP):

- Here is the Push/Pop Arithmetic:

Case	Path expression	Weight expression
Parallels	A+B	W_A+W_B
Series	AB	W_AW_B
Loop	A^*	W_A^*

- The numeral 1 is used to indicate that nothing of interest (neither PUSH nor POP) occurs on a given link.
- "H" denotes PUSH and "P" denotes POP. The operations are commutative, associative, and distributive.

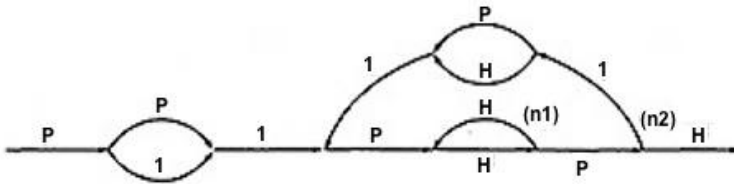
PUSH/POP MULTIPLICATION TABLE

x	H PUSH	P POP	1 NONE
H	H^2	1	H
P	1	P^2	P
1	H	P	1

PUSH/POP ADDITION TABLE

+	H PUSH	P POP	1 NONE
H	H	P + H	H + 1
P	P + H	P	P + 1
1	H + 1	P + 1	1

- Consider the following flowgraph:



$P(P + 1)1\{P(HH)^{n1}HP1(P + H)1\}^{n2}P(HH)^{n1}HPH$

- Simplifying by using the arithmetic tables,
- $=(P^2 + P)\{P(HH)^{n1}(P + H)\}^{n1}(HH)^{n1}$
- $=(P^2 + P)\{H^{2n1}(P^2 + 1)\}^{n2}H^{2n1}$
- Below Table 5.9 shows several combinations of values for the two looping terms - M1 is the number of times the inner loop will be taken and M2 the number of times the outer loop will be taken.

M_1	M_2	PUSH/POP
0	0	$P + P^2$
0	1	$P + P^2 + P^3 + P^4$
0	2	$\sum_{1}^{6} P^i$
0	3	$\sum_{1}^{8} P^i$
1	0	$1 + H$
1	1	$\sum_{0}^{3} H^i$
1	2	$\sum_{0}^{5} H^i$
1	3	$\sum_{0}^{7} H^i$
2	0	$H^2 + H^3$
2	1	$\sum_{4}^{7} H^i$
2	2	$\sum_{6}^{11} H^i$
2	3	$\sum_{8}^{15} H^i$

Figure 5.9: Result of the PUSH / POP Graph Analysis.

- These expressions state that the stack will be popped only if the inner loop is not taken.
- The stack will be left alone only if the inner loop is iterated once, but it may also be pushed.
- For all other values of the inner loop, the stack will only be pushed.

EXAMPLE 2 (GET / RETURN):

- Exactly the same arithmetic tables used for previous example are used for GET / RETURN a buffer block or resource, or, in fact, for any pair of complementary operations in which the total number of operations in either direction is cumulative.
- The arithmetic tables for GET/RETURN are:

Multiplication Table

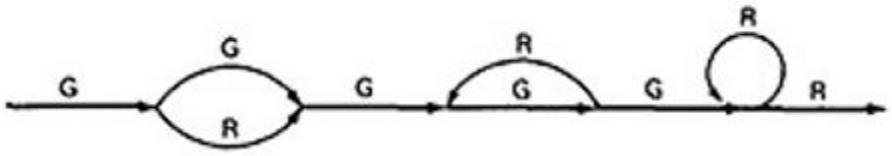
×	G	R	1
G	G^2	1	G
R	1	R^2	R
1	G	R	1

Addition Table

+	G	R	1
G	G	G + R	G + 1
R	G + R	R	R + 1
1	G + 1	R + 1	1

"G" denotes GET and "R" denotes RETURN.

- Consider the following flowgraph:



- G(G + R)G(GR)*GGR*R
  = G(G + R)G^3R*R
  = (G + R)G^3R*
  = (G^4 + G^2)R*
- This expression specifies the conditions under which the resources will be balanced on leaving the routine.
- If the upper branch is taken at the first decision, the second loop must be taken four times.
- If the lower branch is taken at the first decision, the second loop must be taken twice.
- For any other values, the routine will not balance. Therefore, the first loop does not have to be instrumented to verify this behavior because its impact should be nil.

LIMITATIONS AND SOLUTIONS:

The main limitation to these applications is the problem of unachievable paths.

The node-by-node reduction procedure, and most graph-theory-based algorithms work well when all paths are possible, but may provide misleading results when some paths are unachievable.

The approach to handling unachievable paths (for any application) is to partition the graph into subgraphs so that all paths in each of the subgraphs are achievable.

The resulting subgraphs may overlap, because one path may be common to several different subgraphs.

Each predicate's truth-functional value potentially splits the graph into two subgraphs. For n predicates, there could be as many as 2^n subgraphs.

REGULAR EXPRESSIONS AND FLOW ANOMALY DETECTION:

- **THE PROBLEM:**
  - The generic flow-anomaly detection problem (note: not just data-flow anomalies, but any flow anomaly) is that of looking for a specific sequence of options considering all possible paths through a routine.
  - Let the operations be SET and RESET, denoted by s and r respectively, and we want to know if there is a SET followed immediately a SET or a RESET followed immediately by a RESET (an *ss* or an *rr* sequence).
  - Some more application examples:

1. A file can be opened (o), closed (c), read (r), or written (w). If the file is read or written to after it's been closed, the sequence is nonsensical. Therefore, *cr* and *cw* are anomalous. Similarly, if the file is read before it's been written, just after opening, we may have a bug. Therefore, *or* is also anomalous. Furthermore, *oo* and *cc*, though not actual bugs, are a waste of time and therefore should also be examined.
2. A tape transport can do a rewind (d), fast-forward (f), read (r), write (w), stop (p), and skip (k). There are rules concerning the use of the transport; for example, you cannot go from rewind to fast-forward without an intervening stop or from rewind or fast-forward to read or write without an intervening stop. The following sequences are anomalous: *df*, *dr*, *dw*, *fd*, and *fr*. Does the flowgraph lead to anomalous sequences on any path? If so, what sequences and under what circumstances?
3. The data-flow anomalies discussed in Unit 4 requires us to detect the *dd*, *dk*, *kk*, and *ku* sequences. Are there paths with anomalous data flows?

- **THE METHOD:**
  - Annotate each link in the graph with the appropriate operator or the null operator 1.
  - Simplify things to the extent possible, using the fact that a + a = a and 12 = 1.
  - You now have a regular expression that denotes all the possible sequences of operators in that graph. You can now examine that regular expression for the sequences of interest.
  - **EXAMPLE:** Let A, B, C, be nonempty sets of character sequences whose smallest string is at least one character long. Let T be a two-character string of characters. Then if T is a substring of (i.e., if T appears within) AB^nC, then T will appear in AB^2C. (**HUANG's Theorem**)
  - As an example, let
    A = *pp*
    B = *srr*
    C = *rp*
    T = *ss*

    The theorem states that *ss* will appear in $pp(srr)^nrp$ if it appears in $pp(srr)^2rp$.
  - However, let
    A = *p* + *pp* + *ps*
    B = *psr* + *ps*(*r* + *ps*)
    C = *rp*
    T = P^4
    Is it obvious that there is a p^4 sequence in AB^nC? The theorem states that we have only to look at
    $(p + pp + ps)[psr + ps(r + ps)]^2rp$
    Multiplying out the expression and simplifying shows that there is no p^4 sequence.
  - Incidentally, the above observation is an informal proof of the wisdom of looping twice discussed in Unit 2. Because data-flow anomalies are represented by two-character sequences, it follows the above theorem that looping twice is what you need to do to find such anomalies.
- **LIMITATIONS:**
  - Huang's theorem can be easily generalized to cover sequences of greater length than two characters. Beyond three characters, though, things get complex and this method has probably reached its utilitarian limit for manual application.
  - There are some nice theorems for finding sequences that occur at the beginnings and ends of strings but no nice algorithms for finding strings buried in an expression.
  - Static flow analysis methods can't determine whether a path is or is not achievable. Unless the flow analysis includes symbolic execution or similar techniques, the impact of unachievable paths will not be included in the analysis.
  - The flow-anomaly application, for example, doesn't tell us that there will be a flow anomaly - it tells us that if the path is achievable, then there will be a flow anomaly. Such analytical problems go away, of course, if you take the trouble to design routines for which all paths are achievable.