Metrics: Dynamic Data Flow Analysis

Vincent Sanchez
Software Engineering
Sam Houston State University

October 20, 2015

Contents

1 General Considerations

Dynamic data flow analysis is a design methodology used to demonstrate the flow of data through a software system. A good software design, when analyzed with the dynamic data flow method will show a well defined input, or afferent, branch where data is collected and acted upon by related input processing functions (such as opening a file or reading the file contents to an array for use by another module). Data processed by the afferent branch should then be passed to a well defined central transform branch where the core operations of the software are performed on the data, and output data is generated and returned. Finally, this output data is sent a well defined output, or efferent, branch, where it is processed into its final output format, such as a report. Two data flow diagrams are used when plotting the path of data from input to output in a software system. The first is a flow diagram, which shows data coming into the system, how it moves through the system's various modules, and eventually is output in a useful way. The second diagram is a hierarchical diagram that shows the modules that make up the system, and how they are related. This diagram shows the flow of data, but also implies the control structure of each module as it relates to the others and to the system as a whole. Data couples and control couples are indicated here as well. Further, because this type of analysis focuses on the flow of data, with an implied control structure and without any concern for the contents of the actual function bubbles present in the diagram, it is not able to determine or show the cohesiveness of any particular module. This is detrimental because it potentially allows modules that are merely coincidentally cohesive to exist until the implementation phase, when the problems associated with that form of cohesion will likely begin to appear.

2 Analysis Metrics

2.1 Coupling

Dynamic data flow analysis focuses on two types of coupling: data and control.

2.1.1 Data Coupling

Data coupling occurs when one module depends on the output of another, and the data they exchange is composed of discreet, elementary units, providing exactly what each module needs and nothing more or less. This is the lowest form of coupling with the least dependency between coupled modules.

2.1.2 Control Coupling

Control coupling occurs when one module's output is used to influence or determine the execution logic of a subsequent module - for example, a boolean value or control flag is passed from one module into another that performs a different action depending on the value it receives. This type of coupling is detrimental to the software system because it implies that each of the coupled modules must know something about what is contained in the other, meaning one or both may not be a true black box, which indicates that one or both may not be separately implementable or maintainable.

2.2 Miller's Law

Miller's law tells us that, on average, the largest number of tasks that can be simultaneously held in memory is 7 ± 2. Thus, we should work for a design wherein each module complies with this law and its value. Using dynamic data flow analysis, we can see how the data flows into and out of each module, and with the hierarchical model, how control is implied to flow between modules as well. A good design will show, on average, each module having 4 to 5 tasks, and not more than 7 ± 2.

2.3 Graciunas's Law

Graciunas's law concerns itself with the number of control relations between modules, and predicts how this impacts the complexity of the software system. For a given set of modules, R, we have that $R = M(M^{2-1} + M - 1)$, where M is the number of subordinate modules. Because dynamic data flow analysis shows the flow of data and the implied flow of control throughout the system, we can use this law and its equation to minimize the number of inter-module relations to the ideal minimum, represented by the equation

$$I = \frac{N(N-1)}{2}$$

where I is the number of interactions between modules and N is the number of modules in the system.

2.4 Factoring

Factoring of a system occurs when the system is divided into an upper level of control modules, controlling a lower level of operations modules. Systems that are highly factored have few upper level modules and more lower level modules, and take on a hierarchical form.

2.5 Scope of Control

2.6 Black Boxes

2.7 Fan In/Out

3 Analysis

3.1 Coupling

3.2 Miller's Law

3.3 Graciunas's Law

3.4 Factoring

3.5 Scope of Control

3.6 Black Boxes

3.7 Fan In/Out

4 Conclusion