FACULTY OF SCIENCE AND BIO-ENGINEERING SCIENCES

Department of Computer Science

Software Languages Lab



Tier splitting using Static analysis

Thesis submitted in partial fulfilment of the requirements for the Master :

Master of Science in Applied Sciences and Engineering: Applied Computer Science, by

AI Xinran

Academic year: 2014-2015

Promoter: Prof. dr. Coen De Roover

Guidance by: Dr. Laure Philips

Dr. Jens Nicolay



©Vrije Universiteit Brussel, all rights reserved.

Abstract

Acknowledgments

I would never have been able to nish my dissertation without the guidance of several people. First of all I would like to thank professor Coen De Roover for promoting this thesis and giving me the opportunity to conduct my apprenticeship at the Software Languages Lab. Moreover, special thanks go to my advisors Laure Philips and Jens Nicolay for all the help they have given me. To start with, they came up with the subject of this dissertation, furthermore they always assisted when needed and without their proofreading and wise words, this document would not have been up to standards as it is now.

I would like to thank my parents a lots for supporting me study abroad. Without their continuous support, encouraging, unconditional love, I would not be here today.

Also thanks my friends in there sharing joy and care. Thanks Brussels is a quite, beautiful city and comfortable to live

**Content**

[Introduction 1](#_Toc23089)

[1.1 Problem statement 2](#_Toc22695)

[1.2 Contribution 3](#_Toc10822)

[1.3 Road map 3](#_Toc3352)

[Program Slicing 5](#_Toc24961)

[2.1 Static program slicing 5](#_Toc8665)

[2.1.1 Introduction 6](#_Toc19338)

[2.1.2 Data flow analysis 6](#_Toc25594)

[2.1.3 Static slicing on classic PDG. 9](#_Toc6418)

[2.2 Other approach of program slicing 12](#_Toc451)

[2.2.1 Method for static program slicing 12](#_Toc26160)

[2.2.2 Dynamic and conditional program slicing 18](#_Toc31259)

[2.3 Applications of program slicing 21](#_Toc25780)

[2.4 Conclusion 23](#_Toc13322)

[Abstract Interpreter 24](#_Toc21018)

[3.1 Syntax and semantics of programs 25](#_Toc26469)

[3.1.1 Syntax and semantics of programs 26](#_Toc24062)

[3.1.2 Static semantics of programs 27](#_Toc2037)

[3.2 Static analysis based on abstract interpretation 28](#_Toc3390)

[3.3 CESK-Machine 32](#_Toc11910)

[3.4 Meta-circular evaluator 36](#_Toc18228)

[3.5 Conclusion 36](#_Toc6086)

[Program Slicing on State Graph 38](#_Toc26025)

[4.1 Introduction of STIP.js 39](#_Toc20198)

[4.1.1 Jipda framework 39](#_Toc9582)

[4.1.2 Constructing PDG on top of state graph 43](#_Toc2635)

[4.2 Case study on state transition graph 44](#_Toc1220)

[4.2.1 Assumption 44](#_Toc13020)

[4.2.2 Conceive outline of interpreter 49](#_Toc17911)

[4.3 Implementation design of JipdaSlicer 50](#_Toc4784)

[4.4 Conclusion 55](#_Toc16343)

[Validation 56](#_Toc20001)

[5.1 Evaluation 56](#_Toc13044)

[5.2 Conclusion 66](#_Toc8855)

[Conclusion 68](#_Toc20939)

[6.1 Contribution 69](#_Toc133)

[6.2 Future work 70](#_Toc23366)

1

Introduction

The typical three-tier architecture of web applications includes client, server, and data tier. Figure 1 depicts a web application structure that conforms to this three-tier architecture introduced in [6].

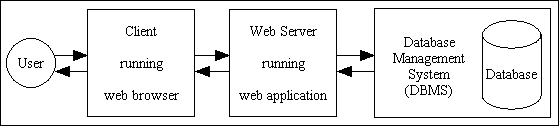


Figure 1 three-tier web application

Normally, the client tier represents the presentation layer such as a browser installed on user’s computer, the server tier is used to handle computations and events and the data tier acts as an interface that communicates to the database system. The traditional development method requires the programmer to use the appropriate technologies for each tier and a manual combination of each technology is required. In contrast with traditional methods, tier splitting of a web application is a rather new concept in the domain of web development. In distributed programming, tier splitting splits a single-tier application into its multi-tier variant [15]. This approach used on web applications tries to blur the distinction between the different tiers of web application such that the development is more straightforward.

Instead of investing in a novel programming language to realize tier splitting, existing technologies can be extended with distributed features,

such as JavaScript to cater for tier splitting seems more beneficial. To enable tierless programming for the JavaScript language, we use static analysis based on abstract interpretation. On top of that, we need to split the actual tiers of the program which can be achieved by program slicing.

**1.1 Problem statement**

Today people can’t live without internet. The applications built for the internet provides a lot of functionalities to convenience for people’s life. This kind of changes requires web applications to provide a huge amount of information and services, and also to be able to handle more interactions with users. In this case, developers have to create so-called rich internet applications, where the client contains program logic as well. Developing such applications in the traditional three-tier architecture increases the complexity of web development, because each tier comes with its own technology stack. This leads to complex glue between each technology of the different tiers. Normally, a combination of HTML, JavaScript and CSS is used for client tier, often together with Ajax and jQuery. The server tier itself is written in another programming language, often depending on which web framework is chosen by the programmer.

Tierless programming languages reduce this complexity by developing a web application in only one programming language. In these languages, a preprocessor or the runtime of the language can split the source code into client, server and sometimes a database tier. The communication between the different, distributed tiers is also handled by these languages. However, investment in novel programming languages is not an efficient way to relieve developers from handling the described complexity.

By allowing tierless programming in a general-purpose language the developer can profit from existing tool support for that language. JavaScript is an event-driven language that runs on the client tier and has been already extended to server tier, thanks to Node.js. By using program slicing mechanism, the tier-splitting tool called STIP.js1 is able to split a tierless JavaScript program into client and server tier. Program slicing[1, 2] is a mechanism to get a subset of the program on a graph presentation of program. Such a subset is a valid program itself and can be executed independently. This mechanism is widely used to analyse and understand

1<https://soft.vub.ac.be/~lphilips/jspdg/stip/stip-web/stip.html>

the behavior of code[11,12,13].

Currently, the process of generating a tier split by the STIP.js tool consists of two parts. First, a State Transition Graph (STG), also referred to as state graph in this paper, is generated based on abstract interpretation for JavaScript. Concretely, the Jipda2 framework is used for this analysis. This graph is a description of possible states the program can transition to. Then STIP.js constructs a Program Dependency Graph (PDG) out of the state graph. Such a graph contains data and control dependencies[7]. Data dependencies are used to represent the data flow relationships of a program, whereas, control dependency corresponds to the control flow relationships. A PDG is used as the input for program slicing algorithms. From the selected node in the PDG, a slicing algorithm performs a backward traversal on the basis of data and control dependencies, resulting in two subsets of the tierless program: the client and server program. To get the tier split code, the developer only needs to annotate the tierless JavaScript code with @client and @server.

**1.2 Contribution**

This dissertation introduces JipdaSlicer, a new program slicing algorithm based on the Jipda abstract interpreter framework. When slicing a program, this algorithm works on a state graph directly, instead of first constructing a Program Dependence Graph and using the classical slicing algorithms on this graph. JipdaSlicer accepts a variable pair containing a line number and a variable identifier from source JavaScript code, which are together the slicing criterion. This criterion is a representation of a point in the code where the users want to cut off code not involved in this point. Based on this criterion a backward slice is computed; this is the set of instructions in the program that influence the slicing criterion. The current algorithm supports a subset of JavaScript, which we discuss in more detail in Chapter 4. To validate the new solution, we test on different cases. These cases test if JipdaSlicer can compute a correct program slice or not from different point of view. As we show, the result of JipdaSlicer is satisfactory.

**1.3 Road map**

The rest of this document is structured as follows: in Chapter 2, we describe program slicing. We introduce the basic concept of static program slicing in this chapter first and then give a general introduction of slicing on PDG. In the last of this chapter we introduce the development and different use of program slicing technology. Chapter 3 introduces abstract interpretation. Chapter 4 is about program slicing on the state transition graph. We first introduce the structure of our tier splitting tool STIP.js. We first dissect the structure and working principle of STIP.js. Then give case analysis to explain how can we apply program slicing technique on STG. In the same chapter, we describe our implementation design of JipdaSlicer. Chapter 5 validates the new algorithm, by comparing the result to that of a current program slicing tool for JavaScript. This dissertation ends with a conclusion chapter, including work summary and future work.

2<https://github.com/jensnicolay/jipda/>

2

Program Slicing

In this chapter, we introduce program slicing. The original definition of program slicing is proposed by Weiser [1,2] mainly used for program debugging and maintenance. Weiser defined a slice of program *S* is a reduced program obtain from program *P* by removing statements based on data flow and control flow analysis. *S* can be executed independently and keeps the same behavior when it’s involved in *P*. Program slicing is the process to obtain the slice *S*.

So far, program slicing has been extended in variety method to adjust different use and properties in different applications. Tip [3] conducted a survey about all these types of program slicing technologies. Static slicing is distinct from dynamic slicing without input assumption. The survey summarizes static and dynamic slicing, starting from the basic algorithm, procedures, unstructured flow and composite data and gives an overview of slicing technologies applied in different application areas.

This thesis focuses on static program slicing. First in this chapter, we introduce the basic notions and character of static program slicing and Weiser’s data flow algorithm. Then we explain slicing on a program dependence graph (PDG). Lastly, we discuss the development of program slicing technology.

**2.1 Static program slicing**

2.1.1 Introduction

A program slice, conform to Weiser’s theory [1], has the following two properties:

1. A slice *S* of program *P*, is obtained from a specific slicing criteria denoted as a pair of value *<i, V>*, where i is the line number of statement in P, and V is set of variables defined or used at i.
2. A slice *S* can be obtained by deleting zero or more statements from program *P*. Meanwhile, *P* and *S* must behave the same with respect to *<i,V>*.

Figure 2.1(a) and 2.1(b) is an example to illustrate Weiser’s program slici-ng. This example is supporting Weiser’s theory from two aspects: slice *S* on slicing criteria *<10,{abc}>* is achieved by deleting several statements; looking at the 10th statement, executing the slice S and the program *P* both lead to the same result on variable ac.

1

2

3

4

5

6

7

8

9

10

var a=1;

var b=2;

var c=3;

var ab=0;

if (a>b){

ab=a-b;

}

else ab=a+b;

var ac = a + c;

var abc = ac + c;

var a=1;

var b=2;

var c=3;

~~var ab=0;~~

~~if (a>b){~~

~~ab=a-b;~~

~~}~~

~~else ab=a+b;~~

var ac = a + c;

var abc = ac + c;

Figure 2.1(a) a JavaScript program P; 2.1 (b) get slice S out of slicing criteria <10,{abc}>

2.1.2 Data flow analysis

Beside, Weiser also put forward some constructive opinions in [1]. Weiser’s first slicing theory establishes on graph representation of program. Each node in a graph represents for a statement in program. Weiser defined a flowgraph *G*=<*N*,*E*,*no*> where *N* is the set of nodes, *E* is the set of edges in set of *N×N* indicates the existing path from one node to another, an initial node *n0* as single entry where the program starts. A hammock graph structure *G=<N, E, n0, ne>* where *ne* is an exist nodewhere the program terminate is extensible definition of flowgraph. Another useful definition is *REF(n)* and *DEF(n)*. *REF(n*) indicates the set of variables whose value are used at statement n, *DEF(n)* indicates the set of variables whose value are changed at statement n. In [1], program slicing is done by flow datatype analysis [5].

Base on the original notions, Weiser proposed a data flow algorithm for program slicing in [2]. After discussing about how to slice a program by using reader’s intuitive understanding on a flowgraph, Weiser introduced a method to find slices by tracing backwards according to dataflow analysis.

1. Introduction of flowgraph

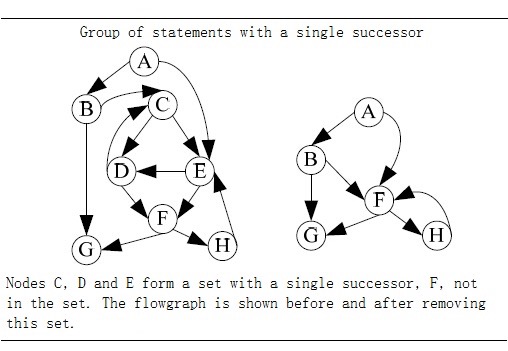


Figure 2.2 Deleting statement in a flowgraph[2].

Figure 2.2 is an example of slicing on flowgraph Weiser gave in [2]. From the figure we see that a flowgraph is an oriented graph with an initial node. The edge (*n*, *m*) in flowgraph indicates the execution progress can be from *n* to *m*, *n* is an immediate predecessor of *m*, and *m* is an immediate successor of *n*. A path of length *k* from *n* to *m* is all possible query on flow graph from node *n* to *m* with lengh *k*. A node *n* is dominated of node *m* when *n* is on every path from *n0* to *m*. If *m* is on every path from *n* to the terminate node *ne* on flowgraph, *m* is a inverse dominator of *n*. Deleting statements in a flowgraph to calculate slices must ensure that there is no increase of immediate successor of statement during deletion. With this concerning, Weiser defined statement deletion as: a set of nodes with single successor can be seen a deleted group, for all predecessors of a deleted group, set deleted group’s unique successor as their new successor. The left part of figure 2.2 shows the result of removing statement in deleted group.

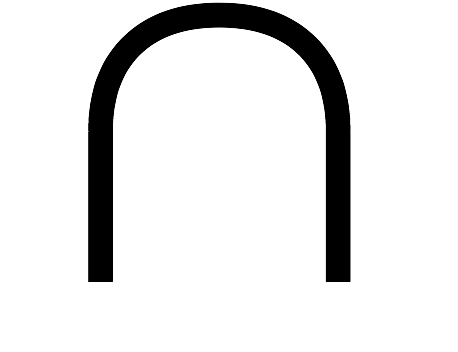
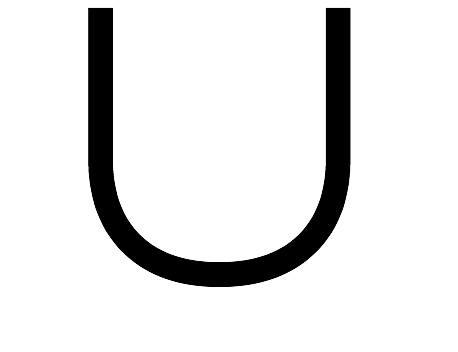
1. Dataflow algorithm

Weiser’s dataflow algorithm finds program slices by iteratively calculating the set of the related variables of each node in the flowgraph. The calculation steps are as follows:

Input: the flowgraph of program *P*, and slice criteria *C=<i, V>*

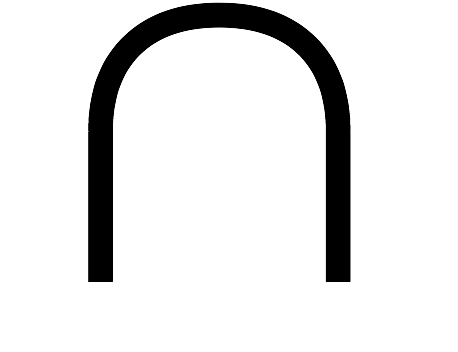
Output: slice *S* of program *P* on slice criteria *C*.

1. Calculation the directly relevant variables and directly relevant statement.
   1. For node *n* and *m* in flowgraph, if there exist path from *n* to *m*, then the set of directly relevant variables *R[0,C](n)* is denoted as: all variables v such that either:
      1. *n = i* and *v* is in *V*: *R[0,C](n)=V.*
      2. *n* is an immediate predecessor of a node *m*: *R[0,C](n) = {v | v∈ R[0,C](m), v∉DEF(n)}{v | v∈ REF(n), DEF(n)R[0,C](m) ≠ ϕ}*



* 1. directly relevant statements is denoted as:

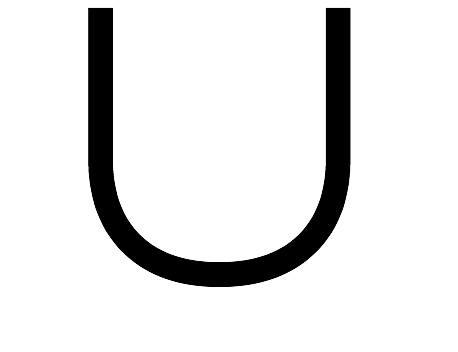
*S[0,C] = {N | DEF(n)R[0,C](n)≠ ϕ }*.



1. Iteratively calculate indirectly relevant variables and indirectly relevant statements.
   1. indirectly relevant variables set is denoted as *R[k,C](n)(k0)*, when calculate *R[k,C](n)*, we have to take account into control dependency. INFL(b) is a set used to represent for the statements influence on statement b, then:

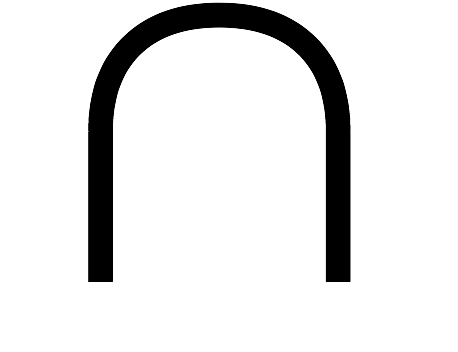
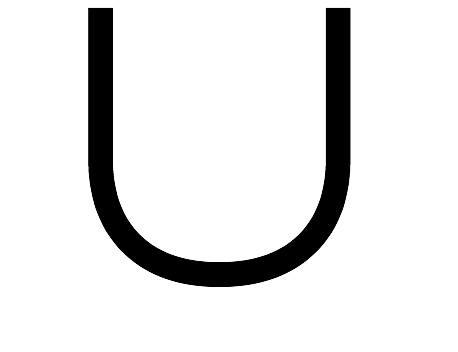


*R[k+1,C](n) = R[k,C](n){n|n∈R[k,C](n),b∈INFL(b)}*



* 1. Similarly for indirectly relevant statements S[k,C]:

*S[k+1,C] =* *{n|n∈R[k,C](n),b∈INFL(b)}{n|DEF(N)R[i+1,C](n)≠ ϕ}.*



1. Repeat step 2 until the size of set *S* doesn’t increase any more, and the statements in *S* consist of slice.

2.1.3 Static slicing on classic PDG.

Many program slicing approaches use a so-called program dependence graph (PDG)[4,7] as an intermediate representation of a program. Similar with a flowgraph, nodes in a PDG correspond to statements and control predicates of the program, and the edges correspond to data and control dependencies between nodes. Informally, when a variable is defined or appearing in one statement, another statement uses this variable without redefining it, then we say this dependency is data dependency. If one statement determines whether another statement will be executed, then the latter is control dependent on former. J.Ferrante et al. defined a PDG which can be use for static slicing of single-procedure programs[7]. According to their definition, the control dependency in PDG is constructed on top of a Control Flow Graph (CFG) associated with a Control Dependency Graph (CDG), and data dependency is identified through a Data Dependency Graph (DDG).

They follow the principle in [35,36] to define and construct CFG and post-dominator tree. A CFG describes the control flow of a program. A CFG is flowgraph augmented with a unique *Start* node represents for entry of the program and unique *Stop* node represents for exist of the program. Each node in CFG has at most two successors and they assume the attributes of outgoing edges get to the two successors is *“T”(true)* and *“F”(false)*. The *post-dominator* is defined as if every path from *n* to *Stop* contains *m*, then we can say *n* is post-dominated by *m*. If a node *X* is control dependent on *Y* in CFG, is must exist a path from *X* to *Y*, and *X* is not post-dominated by *Y*. In other word, if a statement *S’* is dependent on statement *S*, there will be two edges out of *S*, the *“T”* edge will lead to execute on *S’*, *“F”* edge will lead the program jump to another statement. *“F”* edges also ensure there always has an exist for loop in CFG. Figure 2.3 shows an example code with its corresponding CFG and *post-dominator* tree.

In the figure 2.3, we use the line number to mark each node in CFG. Whether line 5 and 6 will be executed is control dependent on the predicate in line 4. Another information in CFG to get control dependences is in term of *region node*. This term can be understand as an entry of a block of statements with the same control conditions. In the example above, the prediction in line 4 can be seen a *region node* of statement in line 5 and 6. In addition, each block has a exist node. The control dependences can be identified by examining the edges (*A, B*) label with *“T”* or *“F”* in post-dominator tree. As the definition of control dependency, the examining is aim to skip the post-dominator of *A* which actually is the parent node of *A*. Integrate with two type of region node, we can get the CDG that contains the control dependencies inside a program.

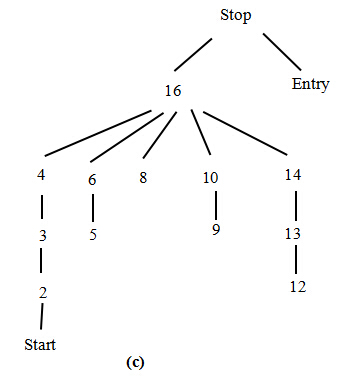
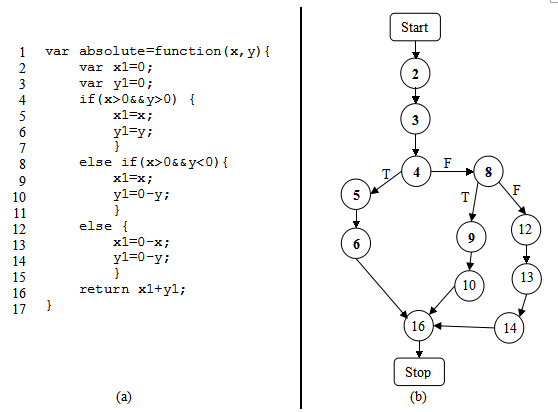


Figure 2.3 (a)An example code;(b) control flow graph of (a); (c) post-dominator tree.

Data dependent is defined as: there exists a path *k* from node *m* to node *n*, and a variable is defined at m and gets used at n without redefining it at any other node on path *k*, then we can say *n* is data dependent on *m*. DDG is a set of data dependencies between statements in a program. Their solution to build DDG is derived from [37,38] and perform data flow analysis[35] with additional assumption that every variable is initiated at entry. After constructing PDG by integrating the CDG and the DDG, a slice can be obtained by performing a backward traverse from an interested node in the graph, visiting all predecessors. In their later work[9], they state that slicing on a PDG is more accurate than the described earlier method.

Figure 2.4 shows corresponding PDG of figure 2.1(a). The gray node is the slice on slicing criterion *<10,{abc}>*. In the figure, solid line denotes for control dependences, dash line denotes for data dependences. Slicing starts by searching the node represents for line 10, and then backward traverse dependences edges from it and mark all the node you visited. In the first traversal, nodes represent for code “var b=2;” and “var ac=a+c;” will be visited. In step 2, backward traverse the dependences edges from the node you marked in first traversal. Iterative step 2 until all marked

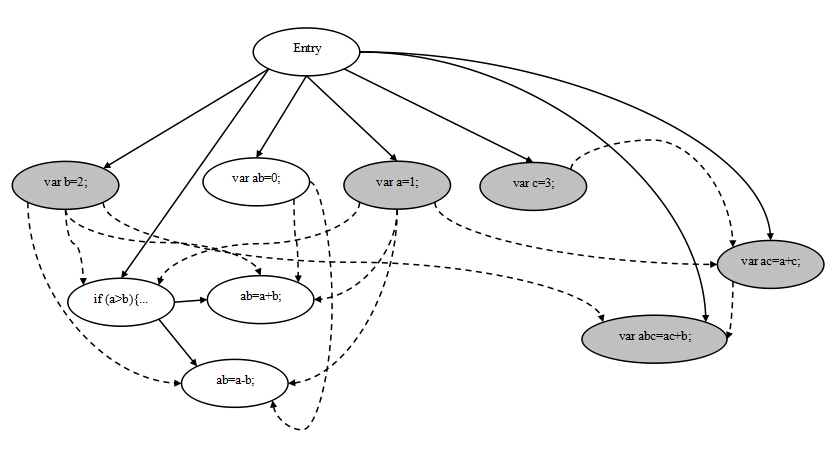


Figure 2.4 program dependency graph of program in figure2.1(a)

node on the path from slicing node to entry node. The second turn of traversal is through nodes represent for code “var a=1;” and “var c=3;”.

**2.2 Other approach of program slicing**

2.2.1 Method for static program slicing

The two approaches above describe the simple case to compute slices of structured, single-procedure programs with scalar variables. The slice we get from backward static slice, is the statements has influence on the the slicing criteria. In the similar way, forward static slice[17] determine slice by computing set of statements are influenced by the the slicing criteria, it requires forward tracing on dependences from slicing criterion. A statement is influenced by slicing criteria means the value computed in slicing criteria will effect on the statement or determine execution of it. In addition to this, program slicing also applies for inter-procedural program, unstructured control flow, composite variables and pointer, and concurrent program. The solutions to these problem will be discussed below.

***Procedures***

The main idea to solve inter-procedural static slicing is to construct a call-return structure to depict the interaction between procedures. Weiser’s approach [2,16] is able to approximate inter-procedural static slicing by taking account into *summary information* [8]: MOD(P) is a set of variables may be modified by procedure *P*, and variable may be used at P denoted as USE(P). The data flow algorithm is extended to identify a call effect with these *summary information*. Assuming a procedure *R* calls *P*, and *Q* is called by *P*, function Up(S) is defined as: a set of slice criteria (*nQ,VQ*) where *nQ*is the last statement of *Q* and *VQ* is all actual parameters from *P* pass to *Q*; function Down(S) is defined as: a set of (*NR,VR*) such that NR is statements in *R* that call *P*, and *VR* is the set of relevant variables at the first statement of *P* that substituted with formal parameters. The slice is got by iteratively generate Up(S) and Down(S) set until no new criteria are generated.

However this approach is inaccurate, because it fails to account for “calling context” problem as Horwitz, Reps, and Binkley pointed out in [18]. The shortcoming of Weiser’s approach is that it is infeasible in the case of entering a procedure *Q* fromprocedure *P* and exit *Q* to another, different procedure. To solve this, when a procedure *Q* is called by procedure *P*, all call sites that call procedure *Q* have to be considered, not only *P*. Horwitz et, al. had their PDG based on [7] with three other categories of vertices: a distinguished vertex called *entry vertex*, *initial definition of x* vertex represents initial state of variable *x* in program *P*, *final use of x* vertex represents final value of *x* compute by *P*. They define control dependency between two vertex as: if *v1* is an entry node, *v2* is control dependent on *v1* when *v2* is not nested in loop or conditional; if *v1* presents for predicate of while-loop, then *v2* is a component nested in loop body and edge *v1-> cv2* is labeled true; if *v2* represents for predicate of conditional statement, *v2* is the component in the conditional branch and edge *v1-> cv2* is labeled true or false depends on whether *v1* lead to execute it. The data-depencences edges occurs in their PDG contains two types: *flow dependences* and *def-order dependences*. Flow dependences is the path in standard CFG augmented with *loop carried* or *loop independent*. The edge *loop carried* flowdependence if the value change in vertex which inside a loop will effect on the predicate or another vertex inside the loop. If the execution of vertex have no effect on the loop, then edge is a loop independent flow dependence. Def-order dependences edges represent for reassignment a variable. Figure 2.4 is an example of the PDG define in [18]. Then they construct the *System Dependency Graph (SDG)* on top of their PDG. In the first step, they model procedure calls and parameter passing with five new kinds of vertices and three new kind of edges. A call site in a SDG is represented by a *call-site* vertex; on the calling side, *actual-in* and *actual-out* vertices are control dependent on the *call-site* vertex and copy the value of the actual parameter to temporary variables/ F*ormal-in* and *formal-out* vertices correspond to formal parameters of a procedure and are control dependent on this procedure’s entry vertex. A *call* edge start from *call-site* vertex to corresponding procedure entry vertex; a *parameter-in* edge is binding corresponding *actual-in* and *formal-in* vertex; similar, a *parameter-out* edge is binding corresponding *actual-out* and *formal-out* vertex. So the SDG includes two parts, a PDG for the main program, and a procedure dependency graph for each procedure. The second step to construct the SDG is to build transitive dependence edges. This dependent relationship due to the value changing of variables after the calling of a procedure. *Subordinate characteristic*

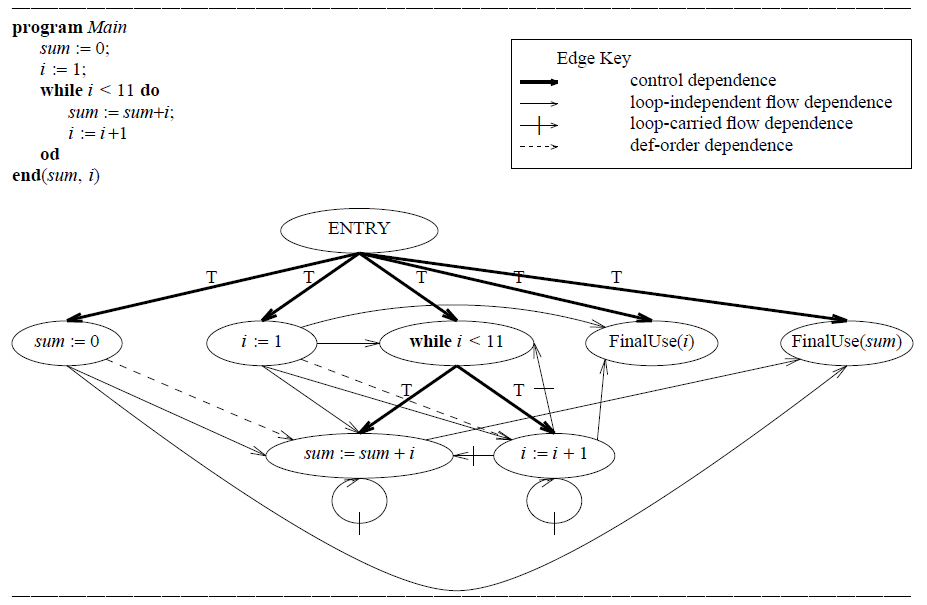
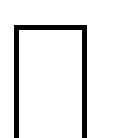
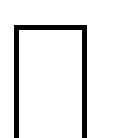
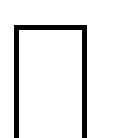
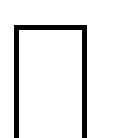
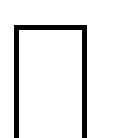


Figure 2.4 An example program, which sums the integers from 1 to 10 and leaves the result in the variable sum, and its program dependence graph. The boldface arrows represent control dependence edges, solid arrows represent loop independent flow dependence edges, solid arrows with a hash mark represent loop-carried flow dependence edges, and dashed arrows represent def-order dependence edges[18].

*graphs* are used to compute transitive dependencies. This graph depicts the transitive flows between the procedure’s input and output, and also, the call and return relations between the call site and procedure. To perform an inter-procedural slicing, their algorithm includes two phases that can be derived by starting vertex *s*: first, identify the vertices which can reach *s* without descending into procedure calls, it’s either in *P* itself or in the procedure that calls on *P*; second, identify the vertices that can reach s from procedures called by *P* or from procedures called by procedures that call *P*. In addition, they also use a data flow analysis when building the procedure dependency graph to identify modified variables and referenced variables. This results in a more accurate slice.

Later, Lakhotia [22] presents an alteration algorithm for inter-procedural static slicing on a SDG. Lakhotia’s approach labels vertices in the SDG with a three-valued vertex tag: “” marks node has not been visited; “Τ” marks the visited node; “*β* ” indicate a visited node and some of its reachable node should be visited. These tag permits to traverse each edge at most once to get slices. The node in slice labeled with “Τ” at beginning and all other node is “”, and backward slicing algorithm traverse on the node and change the tag value based on operation they configured in algorithm where:

 *β = β* = *β, x*  Τ = Τ *x*, and *x* *x* = *x*.



All the nodes labeled *β* and Τ in the slice when the algorithm stops. However, it may require to change the value change twice when slice program. In a subsequent study, the SDG is also extended into object-oriented programs [20] to identify reference between classes. Donglin and Mary discuss a SDG [19] based on existing approach[20, 21] that can distinguish data members for different instances of the same class.

Another approach is also suitable for inter-procedural static slicing proposed by Bergerretti and Carre [17], they define a information-flow relations to identify the information transition statements that can be used to compute slices. In their approach set of variable denoted by V and E denote the set of predict expressions in program. For each statement S, they define three relations between V and E: (e,v)λs if for the value of v on entry to S potentially affects the value computed for e; (e,v)μs if the value computed for *e* potentially affects the value of *v* on exit from *S*; *(e,v)ρs* if the value of *v* on entry to *S* may affect the value of *v*’ on exit from *S*. They also introduced a notion partial statement denoted as which indicates those value may be used in obtaining the value of *v* on exist from *S*. Slices can be got by replacing the those statements doesn’t contain some member of by an empty statement.



***Unstructured control flow***

The approach of program slicing introduced above compute the projection for structured control flow. In [23], Ball and Horwitz point out the previous approach[1, 2, 7] is failure to deal with unstructured control flow, such as a unconditional jump statement **break**. Their solution to this problem is construct a PDG on top of an augmented CFG in which a unconditional jump (i.e, **break**, **goto**, **halt**) is represented as a *pseudo-predicate* vertex. The value of *pseudo-predicate* vertex is always set true, and its true-successor is the target of jump; its false-successor indicate the jump statement is continuation. Similar with Ottenstein’s approach, the slicing is graph reachability problem.

Choi and Ferrante [25] represented two methods to deal with **goto** statement. As same as [23], their first method is to build augmented program dependency graph base on augmented CFG. Choi and Ferrante extend augmented CFG with *fall-through* edge. Informally, a *fall-through* edge between two statement Si and Sj indicates a relationship: if Si is replaced by **null** statement, the program control flow coming into Si would be forwarded to Sj. The notion *fall-through* edge is renamed by *lexical successor* in [28]. A *fall-through* edge is used to capture the relevant control dependent when the **goto** statement is deleted. Their testing example shows the augmented slice may contain statements that don’t affect the statement for which the slice is computed. After analysis the disadvantage of this method, Choi and Ferrante proposed the second method that computes smaller slice as the executable slice by remove redundant cascaded **goto** statement. In this method, they define a new kind of slice named *CFG slice*. *CFG slice* is extract from CFG that includes nodes from which there is a path to the statement to compute slice. The executable slice is out of classical PDG and *CFG slice* by replaced with each statement’s immediate post-dominator.

Another PDG-based static slicing algorithm for unstructured flow is proposed by Agrawal [28]. Agrawal demonstrated the conventional slicing algorithm (refers to slicing algorithm for classic PDG) is easy to be extended with conditional jump statement in the form of **if** P **then goto** L: if the predicate P is in included in the slice because there is other statement control dependent on it, then the associated jump statement L must be included in the slice. In the case of unconditional jump statement,

Agrawal found the way to determine a statement in a slice by: if and only if the the nearest post-dominator and nearest lexcial successor of statement J is different, then J must be included in the slice. A statement S’ is post-dominator of another statement S, when every path from S to exist contains S’[35]. Considering a sequence of statements S1, S2, S3 in original program, if nearest post-dominator of S2 is different with nearest lexcial successor, it means exclude S3, S2 cause control jump to elsewhere. In this case S2 is unconditional jump statement in original program, omit S2 from the slice will cause the control always transfer from S1 to S3. Agrawal distinguished structured jump statement if its target statement is also its lexical successor. The property of structured jump ensure a single traversal of the post-dominator tree is sufficient to obtain correct slice. Second, if a predicate statement is in slice by conventional slicing algorithm, the jump statement directly dependent on it will be also included in the slice.

***Composite variables and pointer***

In order to construct flow dependences for pointers, aliasing which refers to the phenomenon of two or more variables point to the same memory location has to be determined. The James R. and David proposed [29] algorithm use symbolic execution to construct aliasing identical list. They define two type of address: address for static object and address for dynamic object. Address for dynamic object enable to keep indirect assignment through dynamic address can be treated as assignment to array. Their algorithm consists of two phases. In the fisrt step, extract Pointer State Subgraph (PSS) by deleting the nodes in CFG without pointer value, and also deleting the branch node under the node being deleted, then repeat propagate address from address creation node to all successor on PSS. Second, the point state information we get in first step is used to determine which statements should be included in a slice[34].

**Concurrency**

Multiple flows, synchronization, communication, non-deterministic selection cause identify primary program dependency more difficult in concurrent program. Cheng introduced an graph-theoretical approach[31] for concurrent program slicing based on his previous work. Nondeterministic parallel control-flow net introduced in [31] is used to represent control flows, and can be bundled with information concerning definitions and the uses of variables and communication channels. In addition to data and control dependency, Cheng defines selection dependency, synchronization dependency, communication dependency to construct *Program Dependence Net* (PDN). Selection dependency is a kind of control dependency involves non-deterministic selection statement; synchronization dependency reflect the synchronous execution between two statement; communication dependency corresponds to modification of a variable at a point directly has influence on the value of a variable at another point via interprocess communication. Program slicing is a depth-first or breadth-first graph traversal algorithm on PDN. In final, Cheng also discuss how to alter his algorithm into dynamic slicing and forward slicing.

### 2.2.2 Dynamic and conditional program slicing

So far, static program slicing we introduced is independent of input value. On the contrast, dynamic and conditional program slicing have to take account into input at slicing criterion. The difference between dynamic and conditional slicing is that the former only considers one specific input when compute slices, but the latter computes slices with a range of possible input.

Dynamic program slicing is first proposed by Korel and Laski [21,26]. Their approach depends on data flow analysis. They redefine flowgraph of program *P* is a tuple *<N, A, en, ex>* where *N* is a set of nodes represent for statement, *A* is a set of binary relationships between nodes in *N*, *en* and *ex* are used to label the unique entry and exist node of program. The notion of *arc(n,m)*  *A* identifiesthe control transition from node *n* to *m*. They distinguished the path whose execution invoked by input feasible path. After the execution of feasible path derived by a specific input, the tracing history through the path is referred as program trajectory *T*. The slicing criterion is defined as a triple *(x, Iq,V)* where x is the input, *V* is subset of program variables, and *Iq* is the *qth* element of the program trajectory. The dynamic flow they introduce includes three type: Data-Data relations(DD), Test-Control relations(TC) and Identity Relation(IR). DD models the relationship of definition and usage of same variable, and TC is used to determine which statements will be executed upon the predicate in **while** or **if-then-else**. IR refers to duplicate elements in *T*. Furthermore, they subdivide the use of a variable with its last definition as definition-use(DU). The slice is computed by iterative finding directly and indirectly relevant statement. The role of three type of relation ensure traverse on dynamic flow. They formalize the process as follow[21]:



S0 = A0 = LD(q,V) + LT(Iq),

Si+1 = Si + Ai+1,

where

Ai+1={ Xp | Xp  Si, (Xp, Yt)(DDTCIR) for some YtSi,p < q}

In the formalization, Xp is referred to statement X at execution position p; LD(q, V) is the set of last definitions of variables in *V* at execution position *q*, and LT(Iq) is the set of test instructions which have control influence on the execution of Iq. The example of this approach is shown in figure 2.5, also the figure gives the comparison of static program slicing.

In the example, dynamic slice on slicing criterion <1, 99, z> is computed. In the process of computation, A0 is {76} and produced in subsequently A1={33, 22, 65}, A2={11,38, 44} A3={87,37}. From figure 2.5(d) and (e) we can see the different result from dynamic slice and static slice. In static program slicing without assumption of input value, the result slice always contains statement 6, on contrast, dynamic slice get rid of statement 6 because this statement won’t be executed when variable *n* equals to 1. In the sake of take account into a specific input, dynamic generate a smaller slice of program. However, Korel and Laski don’t give an proof about sufficient of their approach. The trajectory of when input value *n* equals 2 is shown in Figure 2.6 (a). In this time, the dynamic concept follow for this trajectory is :{ (65,76), (510,711) }DU, (76, 711)IR. Statement 6 will be included in slice, however, this statement is not necessary to slice cause by reassignment of variable *z* on statement 5.

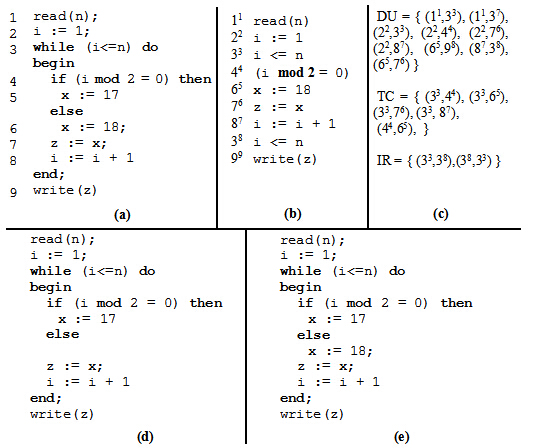


Figure 2.5 (a) Example code; (b) Trajectory of (a) for input n =1; (c) Dynamic flow concept for this trajectory; (d) Dynamic slice for slicing criterion <1, 99, z>; (e) Static slice for slicing criterion <9,z>.

Based on Korel and Laski’s method, Agrawal and Horgan discussed dynamic program slicing on top of Dynamic Dependency Graph(DDG)[32] where DDG is an extension of PDG that represents for execution history with a specific input. The slice is get from traversal on all vertex that can be reached from criterion. Figure 2.6 gives the trajectory of for input *n* = 2,corresponding PDG and DDG of the program of in figure 2.5 (a). Gray nodes in figure 2.6 (c) is the slice result on < 2, 914, z>.

ConSIT is a conditional program slicing tool introduce by Danicic et al.[24]. The slicing criterion is a tuple <*V, n, π*>, where *V* is a set of variables, n is the position and *π* is some condition. Their algorithm consists of three phase. The first phase is symbolically execute to propagate state information to each statement in program. The second phase is a theorem proving phase to determine which statement should be eliminate from slice under the input condition. The last phase is carry out traditional static program slicing on conditional program extract from second phase. As they state in conclusion, their approach on conditional is limited on single-procedural program.



Figure 2.6 (a) Trajectory of the program of figure 2.5 (a) for input n =2; (b) PDG of the program of figure 2.5 (a); (c) DDG of the program of figure 2.5 (a) [3].

**2.3 Applications of program slicing**

The original purpose of program slicing is for debugging[1, 12]. In the case of an erroneous value out of source program, backward slicing enable tracing backward from erroneous position to get all statements have effect on this position. The statement cause erroneous value or some unexpected

statement or omissions is likely occurs in slice. A forward slice can be also used in debugging from two purpose. First, it can query the statements influenced by slicing criterion; second, what kind of effect on program’s behavior when make changes on a specific point in program. When slice with inserting execution, the behavior carried out by statements in dynamic slice are all dependent on insertion. Dynamic program slicing enable programmers to check if the program exhibit abnormal behavior in slice.

In later, the use of program slicing is also extended to other applications. In software testing, program slices reduce the range of testing. When applying program slicing on data flow testing, each def-use pair out of dependency graph in slice would have influence on output value of program[39]. By comparing the output get from slice with expecting value, programmers can check if the program return correct value of output variable. Program slicing is also useful to regression testing[40, 41]. Both backward and forward slices can be used to determine test portion of program depends on the change. Also, the portion can be used to compare the equivalent behaviors between components of new version and old version of program[25].

Gallagher and Lyle propose applying program slicing on software maintenance[43]. A decompose slice includes all computations of a variable and also keeps the same behavior as program. Thus, the decompose slices can be seen as independent components of program and are highly reusable. These components is useful to analysis whether a change in one component will influence on the behavior of other component. In similar idea, Wang et al. [10] discuss functional independent analysis of web security in the use of program slicing. Merge all slices of in each component, if intersection of two different components is empty, then the error or insecurity result from one component will not effect on another one; otherwise, we can’t be sure. Program slicing is also useful in programming understanding and re-engineering by identifying possible coherent components[11, 44].

Weiser[45] also introduced how program slicing can be used in parallelization. His spliced algorithm requires the set of slices should

cover the behavior of original program, the slices can be out of any executable static slice algorithm. Several slices are executed in parallel, the I/O behavior of slices can be spliced together in consistent with original program is preserved.

Beyond traditional program slicing which operate on source code of program, Zhao applies slicing technique on software architecture[14] in assistant of architectural understanding, maintenance and reuse.

**2.4 Conclusion**

In this chapter, we first introduced Weiser’s static program slicing theory: the basic concept and dataflow algorithm. The slices is obtained by deleting statement from original program in such a way that the original behavior is preserved. Intuitionally, static slicing is backward analysis on control flows depends on a interested point in original program to get slices. Then, we introduced the concept of using PDG to describe data dependences and control dependences between statements of a program and give an example to illustrate static slicing on basic PDG by backward traversal. In second section, we summarize several different program slicing approaches in two categories: static program slicing, dynamic and conditional program slicing. At the beginning of this section, we also give a short introduction of forward static slice which only different with backward static slice from the direction of tracing dependences. In last of the chapter, we introduce the applications of program slicing can be use in debugging, testing, software maintenance, program analysis and understanding, parallelizition, and software architecture. In the next chapter we will discuss abstract interpreter.

3

Abstract Interpreter

We can say all the world is an abstract interpretation. Abstraction refers to a process that extract essence from complicated things; or a process from concreteness to conception. Such as by giving the words, apples, bananas, pears, grapes, peaches, we can know their common feature is fruit. Drawn they are fruit is a process of abstraction. Another example of abstraction such as a student is identified by enroll number in their school and is able to sign in a lecture with their enroll number. On top of this, we can build a abstract interpretation that if the student’s enroll number doesn’t occur in a sign-in list of a lecture, then we can conclude that this student doesn’t participate this lecture. In this case, the abstract interpretation is a prediction of attendance based on abstraction relationship between student and enroll number.

In the domain of computer science, abstraction is crucial important and widespread used. In Unix, all devices is abstracted to concept of file, the interaction between each programs is in abstraction of pipe. Android abstracts out a mobile application as four components: Activity, Intent, Service, Provider. A computer program can be seen as a computation in abstraction concept which abstracted from universal object. An interpreter is also a [computer program](https://en.wikipedia.org/wiki/Computer_program" \o "Computer program) that directly [executes](https://en.wikipedia.org/wiki/Execution_(computers)" \o "Execution (computers))[49] and performs the behavior of code, normally written in instruction or [scripting language](https://en.wikipedia.org/wiki/Scripting_language" \o "Scripting language), for the use of parse code, translate source code into [intermediate representa-tion](https://en.wikipedia.org/wiki/Intermediate_representation" \o "Intermediate representation), and pre-compile code.

The possible behavior of computation in program is referred as semantics. Compare with the lecture enrollment example we just mentioned above, an abstract interpreter of program can be informally defined as an approximation of the program’s semantics according to its language property. This method is mainly used to program optimization, transformation. Beside, by designing new feature of abstract interpreter, we can therefore extend a programming language to deal with new computation pattern. The deign of STIP.js is based on the idea abstract interpreter by static analysis. More concrete, we develop an abstract interpreter JavaScript based on a CESK machine for Jipda framework. The role of this abstract interpreter is generate an intermediate structure of program to build state graph. Also, the translation from state graph to PDG is achieved by meta-circular evaluator.

The formal specification of abstract interpretation is first proposed by Cousot P. and Cousot R. in 1977 [46]. In the first section of this chapter, we introduce the Cousots model of abstract interpretation: syntax and semantics of programs. Then we will introduce the notion of static analysis which the most underlies of abstraction interpretation. In the last, we’ll introduce implementation of two kind of interpreter. One is a machine-based interpreter named CESK machine[53], another one is meta-circular evaluator[54].

**3.1 Syntax and semantics of programs**

Syntax[50] describes the rules of a programming language by symbolic, and semantics focus on relations between group of syntax to describe computation behavior. At beginning of [46], Cousots give an intuitive example to explain what is abstract interpretation. The example is, arithmetic computation like -1515\*7 can been abstracted as a sign computation, where (-) \* (+)(-). Even this abstract interpretation lead to imprecise, the rule of calculation on sign is the semantics used to predicate the result of expression is a negative number is enough to reflect of knowledge about sign operation of program. Then they elaborate their mathematical syntax and semantics model of program to build abstract interpreter. Their method concerned with underlying structure is a reflection of actual execution of a program. We give an overview of their method in this section.

3.1.1 Syntax and semantics of programs

Similar with the graph representation we introduced in chapter 2. Cousots[46] use set of “Nodes” and *directed graph*, they call it finite flowchart, to define syntax of a program. For each nod in set, it would have successor and predecessor. Notation *Arcs* presents for node *n* is in predecessor set of node *m* as well as *m* is in successor set of *n*. The node set contains five subsets, the definition for each of node *n* as below:

*Entry node*: the node has no predecessor *(n-pred(n) = )* and one successor *(n-succ(n) = 1)*.

*Assignment node*: the node has one predecessor *(n-pred(n) = 1)* and one successor *(n-succ(n) = 1)*. This node represents for the expression *expr* in right hand-side assign its value to identifier *id* in left hand-side.

*Test node*: the node has one predecessor *(n-pred(n) = 1)* and two successor*(n-succ(n) = 2)*. One is *true* successor *(n-succ-t(n))* another is *false* successor *(n-succ-f(n))*.

*Junction node*: the node has more than one predecessor *(n-pred(n) > 1)* and one successor *(n-succ(n) = 1)*.

*Exist node*: the node has one predecessor *(n-pred(n) = 1)* and no successor (*n-succ(n) = *).

Based on node sets, their *directed graph* is denoted as *<Node, Arcs>* consist of flowing functions:

*origin, end*: a Arcs, a = <origin(a), end(a)>

*a-succ*: <n, m> where m n-succ(n)

*a-pred*: <m,n> where m n-pred(n)

*a-succ-t*: <n, n-succ-t(n)>

*a-succ-f*:<n, n-succ-f(n)>

Their basic semantics model is based on complete *lattice* which proposed in [51]. A complete *lattice S0* is obtained from a domain *S* who has a unique supreme and a unique infimum for all elements *x* inside it. It is denoted as:

*S0* = {S , Τs} where *x**S*, S < *x* < Τs.

In their model, the complete *lattice* of value is the lattice of *Boolean* addition with other primitive type. For example, the supreme of *Boolean* is *true* and infimum is *false*, supreme of natural number is infinite and infimum is zero. They define the notion of *environment* is binding of identifiers to their values. The set of “State” is a joint of *Arcs0* and *environment*. In this case, a state is define as <*cs(s), env(s)*> where *env(s)* is related information in current *environment* and *cs(s)* indicates possible transition. A conditional function *cond(b, e1, e2)* is denotes for if *b* then *e1*, else *e2*. Review the *directed graph* we introduced above, the possible transition of state of their basic model consist of:

Assignments → <a-succ-(n), evaluate expression with binding identifier>

Tests → cond(predicate, <a-succ-t(n), env(s)> , <a-pred-f(n), env(s)>)

Junctions → <a-succ(n), env(s)>

Exists → s

3.1.2 Static semantics of programs

Above theory models program in the sequence of state during execution. On top of this, Cousots define two notations, *context* and *context-vector*, which also respect to the conclusion in [52] about static properties of a program that each program point is concerned a set of associated states. Primarily, a sequence states has a initial state. *Context* is defined as the environments involved in a sequence state transition from initial state associated to point *q*. *Context-vector* is a context of each points in program to summary possible execution of program. Their definition of abstract interpreter for static semantics of program is a tuple:

I = < Contexts, , , Env, , n-context>

where *Env* is supreme and  is infimum of *Contexts*. denotes relation between *a*, *b* in *Contexts* where, *a, b  Contexts*, we have *a**b ab=b*.

**3.2 Static analysis based on abstract interpretation**

The model we introduced above, is analysis of all possible execution on a point of program. On step further, there is a method to implement abstract interpretation is use abstract semantics simulates the computation traces of concrete semantics. Concrete semantics means statements of a program. These statements in a particular order that carry out an execution with un-evaluated identifier. It’s also referred as symbolic execution of program. Based on “symbolic execution”, static analysis is the most common way to make most precise abstract interpretation.

First of all, we have to introduce the notion Galois connection. Galois connection is build on top of value abstraction. The graphic illustration of value abstraction is in figure 3.1.

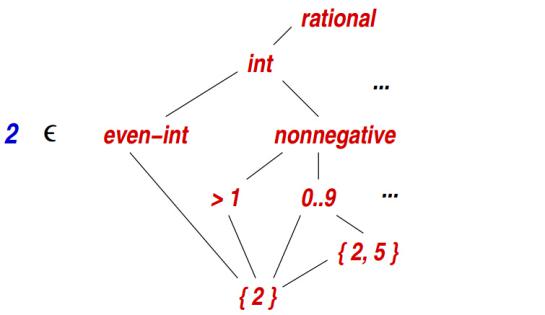


Figure 3.1 Value abstractions

The right hand list all the properties abstract of concrete value number 2. We can see the properties connected with upward lines became general and loss precision. This attribute of abstract is denoted by . For every value abstraction, we also have a supreme Τ, and infimum **. The infimum of abstraction is *{2}*, and supreme is *rational* in above example.



A function  is the mapping from each abstract value to the set of concrete values it represents. A function  maps each set of concrete values to the abstract value that best describe it. Following figures show the example of function  and .  is a set of mapping relations between concrete values and abstract value in a specific domain.

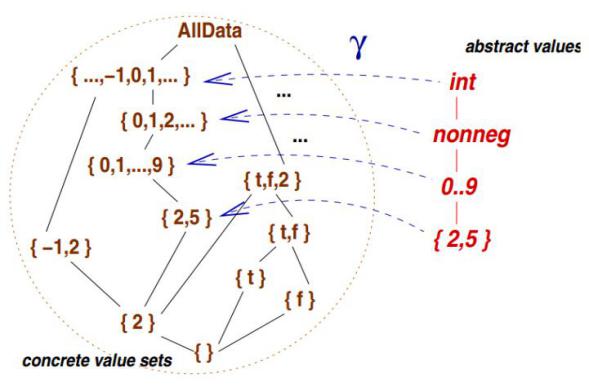


Figure 3.2 (a) Function 

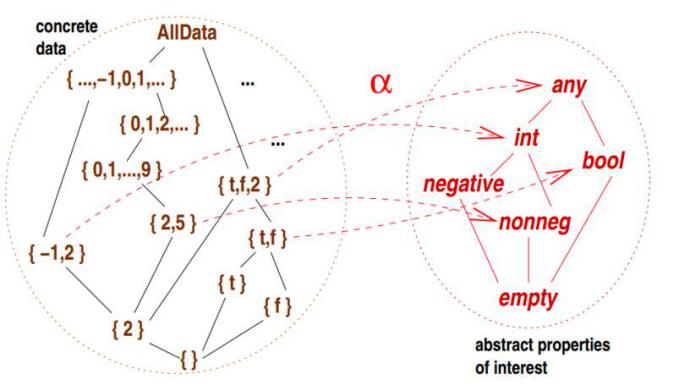


Figure 3.2 (b) Function 

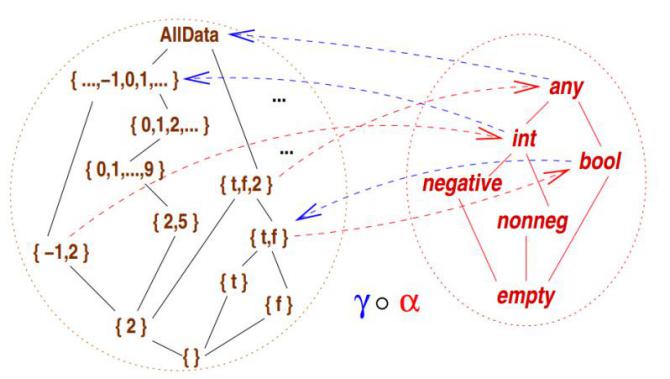


Figure 3.2 (c) 

From figure 3.2(b) we can know, use a abstract value to describe a set of concrete value will loss of accuracy. However ** dedicates to make the description as precise as it possibly can be. The Galois connection formalizes the situation, and the graphical explanatory of it is in figure 3.3.

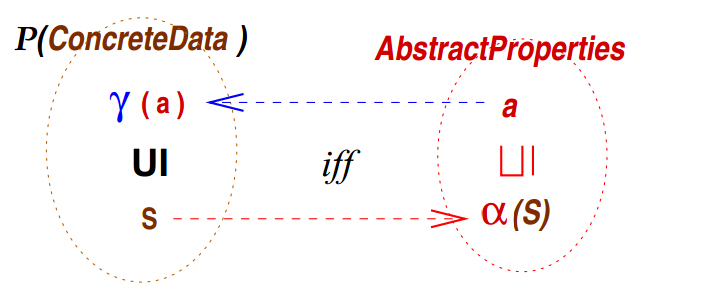


Figure 3.3 Formalization of Galois connection

The definition is, for all SP(ConcreteData), aAbstractProperties,

S(a) iff (s) a



When and  monotone, this is equivalent to

S (S) and  (a) a .



By Galois connection we can interpret concrete semantics with abstract value. Look at the example below:

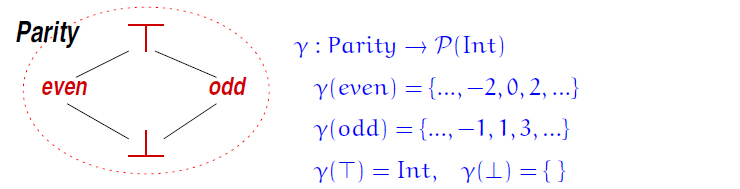


Figure 3.4 Example of abstract interpretation based on Galois connection

This is an abstract interpreter of judgment the parity of an integer. So the abstraction value set contains two value to indicate the [category](C:/Users/aixinran/AppData/Local/Yodao/DeskDict/frame/20150804010944/javascript:void(0);) of interger, one is *even*, another is *odd*. The supreme of this value abstraction is whole set of integer, and infimum is an empty set. Function  is mapping the [category](C:/Users/aixinran/AppData/Local/Yodao/DeskDict/frame/20150804010944/javascript:void(0);) value in the set of abstraction to its concrete value set. Function  in this example is defined as P(Int), to predicates an concrete number weather is an odd or an even basic the rule of {(v)|vS}, where (2n) = even and (2n+1) = odd.

Based on the abstract interpretation of semantics in above example, we can predict the computation traces of concrete semantics. Such as, if we have a piece of code like below:

S0

S1

S2

S3

while( x div 2 == 0) {

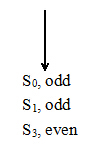
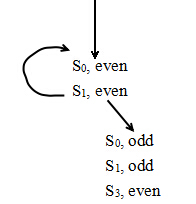
x = x/2;

}

x=x+1;

return x;

In the code we can know, if the value of *x* in *S0* is even, the program will go into the body of loop, otherwise it will jump to *S2*. Assuming the operation of program is denoted by a pair of a program point and the abstract value of variable, we can get two trace trees from basic on underlying abstract interpretation semantics as showing in figure 3.5:



(a) (b)

Figure 3.5 (a)Trace trees if input value is even; (b)Trace tree if input value is odd.

To ensure every transition in concrete semantics is simulated by one in the abstract semantics, we have to follow this principle: if a concrete variable *c* at *Si* can transit to *c’* at *Sj*, then we have to ensure there is a corresponding transition in abstract interpreter from a within *Si* to a’ within *Sj,* where *Sj*, c’ {*Sj*, c’| c’  (a’)}.

So far, we can know a static analysis of a program is a sound, finite, and approximate calculation of the program’s executions. Additional, static analysis is an important method to deal with halting problem. This kind of problem is generated by undecidable termination by a given input. In the parity example, because of loss precise on abstraction, we can’t say we can get termination for all even value, even if only 0 will cause non-termination. In other word, static analysis attempts to predicate program behavior as precise as it can be.

Beside above solution of abstract interpretation of static program analysis, in the following section of this chapter, we’ll introduce two types of abstract interpreter which can be used as a framework to do static analysis.

**3.3 CESK-Machine**

Mtthias Felleisen[55] introduced a *fg* machine interpreter which include four component: a (C)ontrol component, a (E)nvironment, a (S)tore and a (K)ontinuation. The control component an be seen as instructions based on the programming language being reduced. In functional programming language, this component can be expressions. Environment and store can be a map or a first-class function. Environment is meaning the machine will assign an address in storage to each variables. The store is used to map from address to value. The continuation component K is the stack of program, it represents for instruction’s control context.

Their purpose to develop such kind of machine interpreter is translate a source language, which defines source component, to a target language which augment source language with *frame expression* to specify the permission of component Their model based on -calculus language and allowing explicit name -calculus. The intersection of appropriate permissions *R* in caller and callee -expression used to ensure the translation is reliable. They also define an *Eval* function which is a combination of each permission of component in source to determine the meaning of source program. In *fg* machine, the role of *Eval* function is abstract context transition, denoted as fg. The following figure illustrate the structure of *fg* machine. In the first part, is their definition of principle of evaluation sub-expression and extending corresponding continuation. Symbol *M* denotes for target language. The basic call-by-value -calculus is attached with *test expression* to check the branch of enabled permission, *grant expression* enables privilege, and *fail expression* direct halt to security failure.

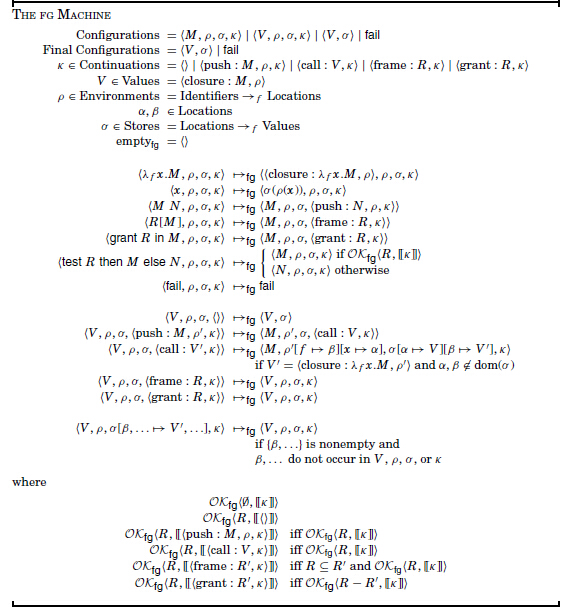


Figure 3.6 Definition of *fg* machine[55].

In figure, the second group in definition of *fg* machine is a complementary of party one to determine the transition as shown in figure 3.7. The evaluated program, *R[M],* is target language adds a framing expression to source language. *OK* is predicate applied on current context and desired permissions. The last part of their machine is garbage collection in storage. The *Eval* function takes charge of validating target in a valid machine configuration, and also replaces variables by their bound value in environment.

[53] gives an overview of implementation of a simple CESK machine for A-Normalized lambda calculus. They review the machine-based interpret-er of program is model source program by group statements with its execution within an environment. The environment is a immutable hasp maps that mapping between variables to values. In general, a machine-based interpreter should contains, a set of program used to be abstracted, a set of machine states, an *injection* function used to initial the state of program and a *step* function used to transit state. They assuming the terminate of state transition is deterministic, and for undecidable termination we have to set a function to generate multiple potential successor states. There are three kinds of expression used to evaluated: atomic expression (*aexp*), complex expression (*cexp*) and expression (*exp)*. Atomic expression is the expression which always has terminate; a complex can contains error, side effect, or defer execution such as conditional statement; expression can be atomic, complex or let-bound.

The component of their CEK-machine is followed Felleisen’s design. Control string is expression. The environment maps variable to its address and a store maps address to values. The trace of variable query in CESK-machine is first through some environment and then through the store to get value. Their machine includes five type of value: void, booleans, integers, closures and first-class continuation. The set of value is defined as:

*val* Value :: = **void** | **z** | **#t** | **#f** | **clo**(lam,) | **cont**(k)

where **z** is integer, **#t** and **#f** are boolean.

By the assumption that a continuation execution is always triggered by the result it awaiting for, continuation can deal with sub-expression in *let-bound* expression by keeping resume execution. For instance, given a *let-bound* expression (*let* ([*v exp*]) *body*), the machine will first evaluated *exp* and bind the computation result to *v*, then holding the resume to *body*. A special initial continuation triggered by result of program is called halt. Finally, their continuation includes address information, that is the scope of variable.

They elaborate their implementation from two-stage. The first is evaluate on *aexp* with auxiliary semantic function: A: AExp  Env  Store  Value. The evaluate result of integers and booleans is themselves: A (z, *ρ*, ) = **z** ; A (**#t**, *ρ*, ) = #t ; A (**#f**, *ρ*, ) = #f. Lambda becomes to closures: A (lam, *ρ*, ) = col(lam, *ρ*). By use function O(*prim*) to define mapping from primitive operation to its corresponding operation, primitive expression are evaluated recursively:

A ((*prim aexp1, ... aexpn* ), *ρ*, ) = O(*prim*).

In second stage, they use *step* function to evaluate each expression type. When evaluation of procedure, the return statement of each procedure is an atomic expression and return the result to the continuation is waiting for its result. The procedure call is evaluated the expression of invocation first and the expressions in arguments, and then evaluate the body of procedure with current arguments. The first class continuation as call-with-current-continuation (call/cc) in functional program, is evaluated the current continuation as first-class procedure, and returns the value to the continuation of the call/cc application. The next step of conditional expression is evaluation by predicate expression. Overwrite is done by searching the address and modify the value of it. Recursive is valuated by self-reference which is the extending of environment. The *ste*p function is defined as:

*step*((*letrec*(*[v1 aexp1]…[vn aexpn]*) *body*)*, ρ, σ, κ*) *=* (*body,ρ′, σ′, κ*),

where: *a1,…,an* are fresh addresses in *σ*

*ρ*′=*ρ*[*viai*]

*valuei*=A(*aexpi*, *ρ*′, *σ*)

*σ*′=*σ*[*aivaluei*].

The two auxiliary function, one is *applyproc* is used to apply a procedure to value, another one is *applykont* is used to apply a continuation to a return value. The definition of two function are:

*applyproc*( **clo** (( (v1…vn) body), *ρ*) , *σ*, *κ*) =

(body, *ρ*′, *σ*′, *κ*), where

a1,…,an are fresh addresses in *σ*

*ρ*′=*ρ*[*viai*]

*σ*′=*σ*[*aivaluei*].

*applykont*( **letk** (*v*, *e*, *ρ*, *κ*), *value*, *σ*) = (*e*, *ρ*[*va*], *σ*[ *avalue*], *κ*),

where *a**dom*(σ) is a fresh address.

**3.4 Meta-circular evaluator**

**3.5 Conclusion**

In this chapter, we start with real-life example to explain the notion of abstract interpretation, and briefly introduce in computer science, by computation on the abstracted concept, the abstract interpreter of program is also a program in the use of process code for analyst and reuse. Then we state the static semantic model to build abstract interpretation of program proposed by Cousot. Their model contains two part. They first abstract the syntax and semantic of a program by finite flowchart, and integrate semantic model with lattice to approximate the behavior. Then summary the possible behavior of program by sequence states during execution. In the second part of this chapter, we introduce static analysis which is most widely used pattern in setting up abstract interpretation. In the same section, we mainly introduced Galois connection which is a common method to build connection between concrete data and abstraction data. In the last, we introduce implementation design of two types of abstract interpreter which can be used as a framework to do static analysis. One is a machine based interpreter proposed by Felleisen. Another one is meta-circular evaluator.

4

Program Slicing on State Graph

By building a PDG on top of the Jipda framework, the current version of STIP.js is able to calculate slices on the PDG. The slices can be used to generate distributed code to complete the tier splitting. Our program slices is based on Weiser’s static slicing theory. This chapter will introduce the new algorithm, JipdaSlicer, to get JavaScript slices from a state graph directly. JipdaSlicer is a new program slicing tool installed in original version of STIP.js. It can slice JavaScript code on state graph without building PDG. In the first section of this chapter, we explain how the original STIP.js tool completes the tier splitting via two phase. The first phase is generate state graph by Jipda framework, the second phase is backward slicing on PDG constructed out of state graph to generate distributed code. After review the structure of current tool, we give some case analysis on the state graph in the Jipda framework to illustrate how can we find data dependencies and control dependencies on state graph. By the observation of state graph, we set three assumptions as essential conceive guideline of implementation design. In the last of this chapter, we propose our implementation design of JipdaSlicer which permits to extract slices from a state graph without building any intermediate representation.

**4.1 Introduction of STIP.js**

In the original version of STIP.js, tier splitting consists of two phase. The first phase is converting source code into state graph by Jipda framework installed it. In second phase, it constructs PDG on top of state graph, then completes tier splitting by static program slicing. In following we will give more details about how STIP.js works.

4.1.1 Jipda framework

The analysis done by the Jipda framework on the source code results in a state graph in run-time. STG represents a program through a sequence str-ucture graph to indicate finite representation of behavior of the program. The first step is parse the code, achieved by the Esprima.js tool[47]. Esprima.js is a parser of JavaScript used to convert source code into abstract syntax tree[48]. Such a syntax tree is a tree structure representati-on of program code which is context-free. Syntax tree describes composi-tion and structure of grammatical feature of programming language and implicit unnecessary information such as parentheses. Beside, Esprima.js also results in *tokens* of source code. *Tokens* is lexical information of prog-ram code. In general, the structure of *token* is a pair of “type” and “value” to identify the lexical attribute of each element. Here, we give an example to explain the structure of sensible syntax tree generated by Esprima.js. If the input code is:

a = 3+4;

Intuitionally, this is an assignment of arithmetic expression to operand *a*. Figure 4.1 shows parse analysis result of this code out of Esprima.js. In (a) is the list of *tokens* of this assignment. On left of expression, variable a is lexing as an *Identifier* with value *a*; on the right of expression, 3 and 4 is a *Numeric* type with integrate value. All the operators is classified as *Punctuator*. Figure 4.1(b) gives syntax information of the code in JavaScript.



1. (b)

Figure 4.1 (a) tokens of assignment a=3+4;; (b) syntax tree structure of assignment a=3+4;.

After get the parsed code, our tool invokes the abstract interpreter which implement by static analysis based on CESK-machine mechanism. Look back to the definition we introduced in section 3.3, our CESK-machine contains four components. The control string in this interpreter is expressions. When the interpreter is launching, CESK-machine will install a global pointers and references. This component is used to map a variable to its address. The launching also leads to initiate storage, this component maps address to its value. The continuation component suffix with *Kont* is the program stack to track computation and operation information of program. This component is also used to associated the scope information of variables and a continuation of current continuation. The resume of an execution in this interpreter is kept by *body stack (BodyKont)*. The functions suffix with *State* is associated to calculate state transition. Then apply static analysis on CESK-machine to get *halt*, *push* and *pop* edges with address information in state graph to reflect behavior of program.

1

2

var a=1;

var b=a+1;

Figure 4.2 Example code

Figure 4.3 shows the analysis result derived by above code. From the figure we can know, the node in state graph is either a element on syntax tree or a node contains one or more statements. The digit in parentheses on edges is the edges’ number. Each edge also contains stack information of the abstract interpreter. This consists of a *tag* of a node. This *tag* can be seen as an *id* of each *block*. The *block* in state graph is the node between corresponding positive and negative edges. This node may be a function declaration or an operand in code. A number of *blocks* compose more complex semantics, such as variable declaration, function call. For examp-le, statement in first line is identified with two *blocks*: node between edge +1 and -1 identify the value of variable, then edge +2 and -2 identify variable declaration statement. For some edges, it also contains reference and scope information such as the 8th edge. 8th edge tells us, this node refers to an identifier whose value is *a*, and the scope of *a* is point to global variable. The edges labeled with positive digit denote for the operation of *push stack*; on contrast, negative digit denotes for operation of *pop stack*. In our CESK-machine, *push stack* indicates starting evaluate on a identifier or a statement; *pop stack* indicates the evaluation is terminate. Each corresponding positive and negative digit pair is used to identify semantics of the program.

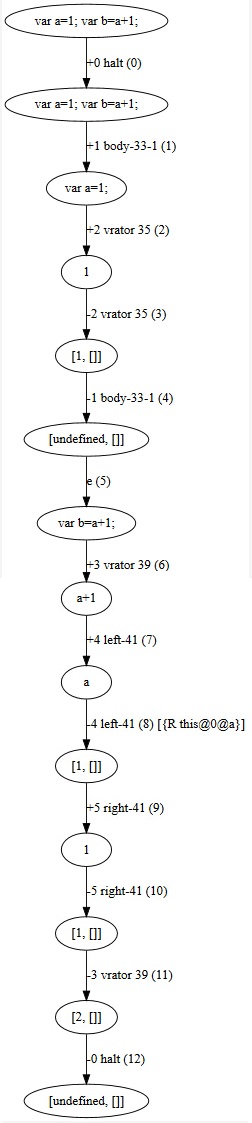


Figure 4.3 Corresponding state graph of code in figure 4.2.

4.1.2 Constructing PDG on top of state graph

The PDG in our tier splitting tool is based on classic PDG augmented with summary edges. The summary edges are added between input actual parameter to output actual parameter if a path between the corresponding formal input to formal output exists. The constructing process is impleme-nted by the notion of *meta-circular evaluator* we introduced in section 3.4. By using *evaluato*r to estimate the type of nodes in state graph, the abstract interpreter will result in corresponding adding operation of node as well as the connection with its predecessors and successors in PDG. The type of node in PDG is identified by corresponding positive and negative digit pair of edge in state graph.

For example, in figure 4.3, the edge labeled *+2* and *-2* also stamped with string *vrator*, this string indicates the node between these edges is in *variable declaration stack (VariableDeclarationKont)*. The corresponding edges in *body stack* identify the first variable declaration statement in code and outgoing edge to second line. The *evaluator* in STIP.js handles this information get from state graph and makes a statement node of variable declaration in PDG. The data dependence in figure 4.3, is identified with scope information in 8th edge. Following two figures show the PDG out of STIP.js. To get slices, the program slicing algorithm is backward traversal as we discussed in section 2.2.

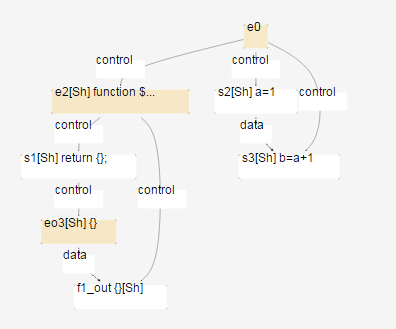


Figure 4.4 Corresponding PDG of code in figure 4.2.

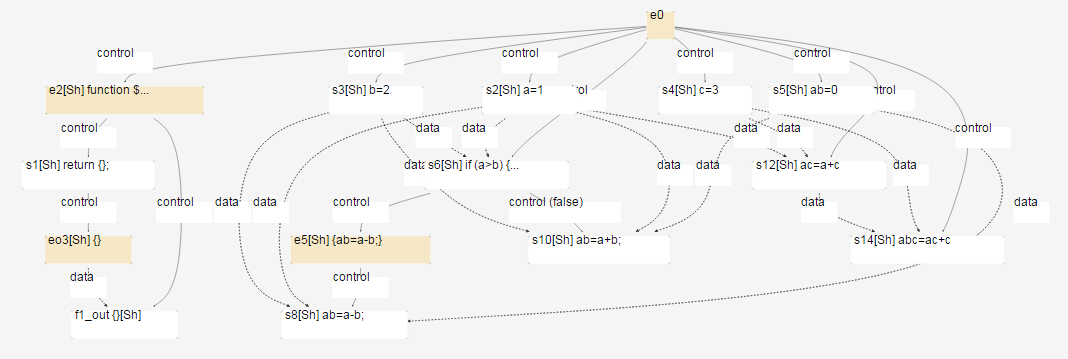


Figure 4.5 Corresponding PDG of code in figure 2.1(a).

**4.2 Case study on state transition graph**

Beside semantic links between nodes, the predecessors and successors of each nodes is linked depending on data and control dependences when we construct the dependence graph. So when we find the slicing node corresponding to the slicing criterion, we can simply get all nodes related to it by performing a back traversal from this node to entry. Since the state graph is also a graph representation of the program code, slicing via Jipda also needs to locate nodes relate to slicing criterion by traversal on graph.

4.2.1 Assumption

By observing figure 4.3 and 4.4, we can set the state graph apart from PDG by two points. The first one is, a node in a PDG represents one corresponding statement, however, a node in a STG represent identifiers in state graph. The second one, the edges in a PDG identify the type of relationship between each node. Looking at figure 4.3, we can know, edges store semantics information linked with interpreter in state graph. A statement is identified a sequence of edges in state graph. Based on the characteristic of state graph we observed in figure 4.3 and 4.4, we make an assumption that by point out an interested line to slice on program, if this line does not contain an identifier then this line isn’t data dependent on other code; otherwise, locating the value of identifier from top to this line. In this case, the first step to drive slicing algorithm is calculate the range of interested line to get all corresponding nodes in state graph. Then, from the start of the range, we examine each node to confirm data dependences. In order to illustrate our assumption, we make an observation of instance in below.

1

2

3

4

var a=1;

var b=2;

var c=a+1;

var d=b-1;

The corresponding state graph of this experimental code get from Jipda framework is shown in figure 4.6. We can know variable *a* is directly influence on variable *c*; and variable *b* is directly influence on variable *d*. This code can be divided into two slices. The first slice can be get by slicing on criterion *< 3, {c} >* is:

1

2

var a=1;

var c=a+1;

The second slice result from slicing criterion *< 4, {d} >* is:

1

2

var b=2;

var d=b-1;

Checking the first slice on state graph by regarding our assumption, we locate the range of 3th statement in source code is between edge 9 to 16. While going through this range, we receive identifier reference at edge number 12. Then looking up this identifier from the top to 9, the variable declaration matching with this identifier is between edge 1to edge 4. Therefore, we can add these nodes to slice. And return to range of 3th statement to go on querying. When reaching the end of the range, we accomplish program slicing. We can get the second slice by same method of query on state graph.

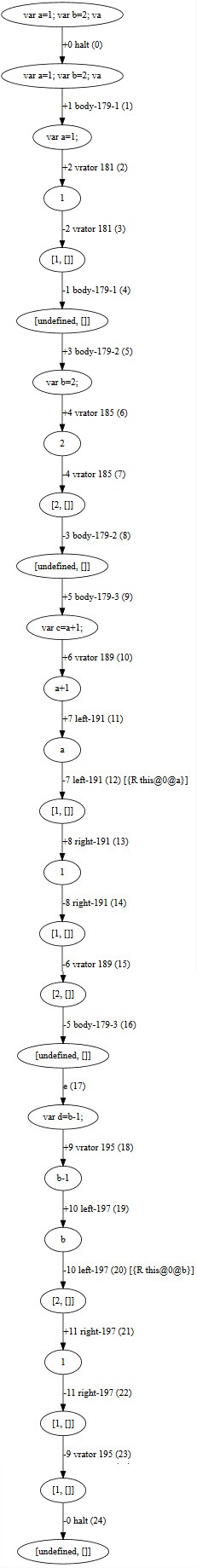
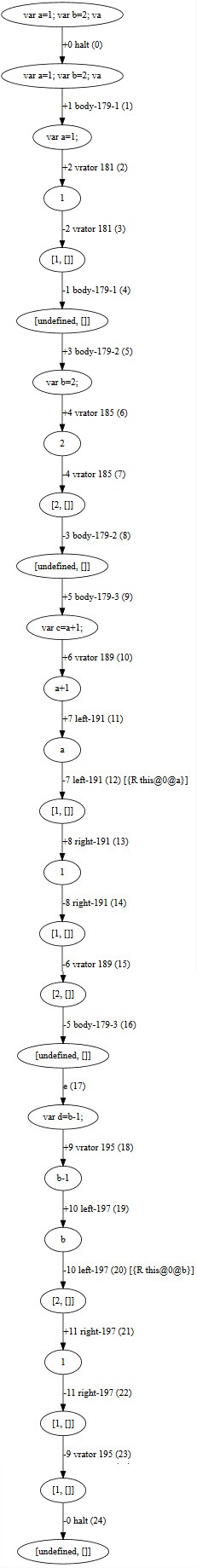
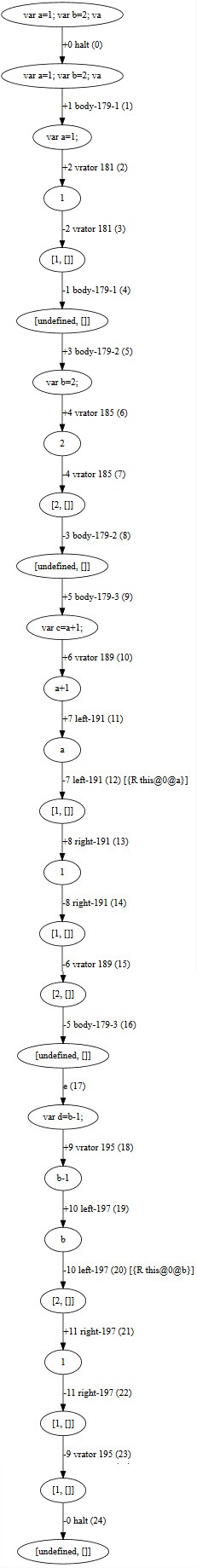


Figure 4.6 State graph of experimental code

We use the same method of observation to find possible solution to get control dependency. The experimental code is in below:

1

2

3

4

5

6

7

var a=1;

var b=3;

var c=2;

var add=function(x,y){

return x+y;

}

var sum=add(a,b);

In this experimental code, slicing on line 7th will invoke function add with actual parameter pass in. By query on state graph in figure 4.7, the range of statement 7th is between edge 18 to 35. As the identifier information labeled on edge 20. We can know when interpreter building state graph, function reference is recognized in the same way with variable reference, by compare identifier with function name, we can determine which function is called. Also the actual parameters a and b, are identified with their declaration statement. On the state graph, we can see the graph listing all states of how a program walk through into function *add* after a statement pass actual parameter in that function. The final state of function returns computation result is assigned to the variable sum. The details of the state transition is not needed for the slice. We only need to involve the part of the function declaration in sliced code. To go on querying to the end edge of this statement, we receive identifier a and b subsequently, to each identifier, query from top again. The function declaration and the variable binding with actual parameter is directly influenced on slicing criterion.

Over the sample code, we assuming variable *sum* would have effect on some statements in follow-up code, such as an assignment come after the code, like:

var d = sum;

This time, we have to take into account statements that are indirectly influenced by the slicing criterion. While querying on corresponding node to this statement on state graph, we will get an identifier point to declaration of sum. After find out the start edge of sum declaration statement, we further query the corresponding end edge of it. During the query in this turn, we will first receive function identifier to search corresponding function declaration. Then we need to search data dependency on identifier a and b. When we meet the end edge of statement in second query, the process to find relative code which effect on *sum* in sliced line is finished. We can draw conclusion based on above, the process of dependences collection is iteratively execution to find indirectly dependences.

For conditional statement, Jipda is able to analyse state transitions on **if-else** statements. By checking on corresponding state graph of **if-else**

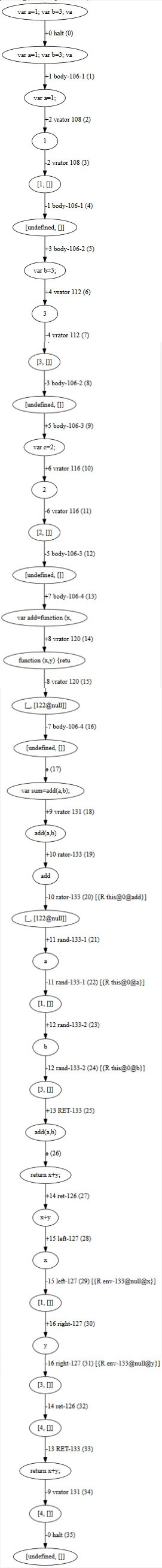
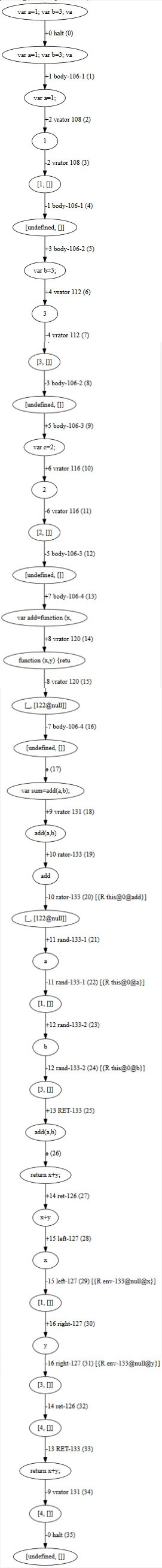
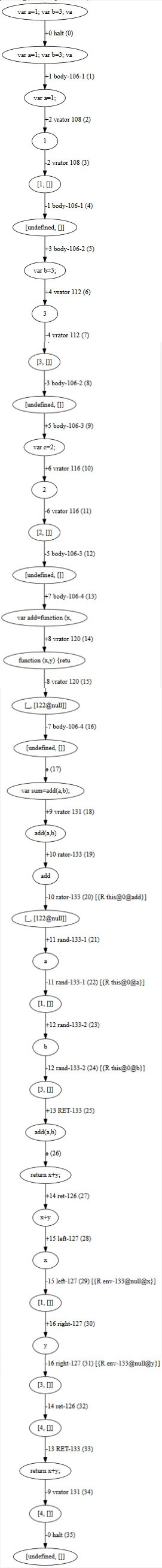


Figure 4.7 Corresponding state graph of experimental code.

statement, we can know each predict expression has a boolean state relies on its computation result. The followed state transits into its clause only when the state’s value of predict is true. Our hypothesis to handle conditional statement is: if a statement in the clause of false predicate, then we can say in static analysis, this code is never used. Therefore, we can replace the code in that clause with null value. If line of slicing criterion is involved in clause of **if-else** statement, we say the statement in predicate is directly influence on slicing criterion; otherwise, is indirectly influence.

### 4.2.2 Conceive outline of interpreter

Respect to Weiser’s theory, our JipdaSlicer requires user input in the form of a line number and a variable name which denotes the slicing criterion or the start condition to drive the slicing algorithm. Based on our assumptions, we need a driving function to detect the range of sliced code. Then put nodes in this range into evaluator in order. The evaluator is used to detect the identifier and launch the search function. The variable is always declared by reserved word *var*, matching a variable identifier on state graph is always searching its *var* declaration statement by compare the variable name. For the declaration of functions, JavaScript allows definition in two ways, one is starting with *var*, another is function declaration with function’s name. So, there should be two methods of searching a function. The searching method, in last, identify dependent by comparing function name with its corresponding deceleration. An iterator to get indirect dependences node of slicing criterion. Finally, there is a function to convert the set of sliced nodes into the target, generated code. Since the state graph results in boolean value of conditional statement, we can determine which statement in clause is executed by the boolean value. Figure 4.8 is a sketch of our conceived outline to design the algorithm. The process modeled in figure will be recursive execution until the all the node in selected line is visited.

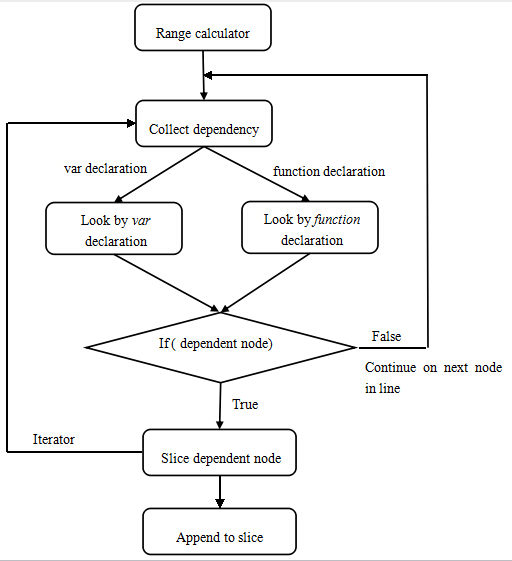


Figure 4.8 Basic idea of design

**4.3 Implementation design of JipdaSlicer**

Following we elaborate on the implementation of JipdaSlicer. JipdaSlicer is an extension of the Stip.js tool. So the implementation is still based on the framework of original Stip.js. At the beginning of the program slicing, JipdaSlicer collects the nodes and edges in the state graph generated by the original version. As we mentioned above, the first step of slicing is to find the range of statement based on the input line number. The implementation of calculation is quite simple, we use *split()* method of JavaScript to detect line breaks, and use *length* method of array to sum total length before the selected line. This value is the start point, and along with the length of selected line, we get the end point of statement. The computation process is in below code:

var lines = (src + "\n").split('\n');

var line\_range = [0, 0];

for(var i = 0; i < line\_num - 1; i++){

line\_range[0] = line\_range[0] + lines[i].length + 1; }

line\_range[1] = line\_range[0] + lines[line\_num - 1].length - 1;

Listing 4.1 Range calculator

Then we pass the edges and nodes in this line to the *collect\_dep* function. This function is responsible for collection data or control dependences via searching corresponding identifier. By comparing the value of *tag* on edges we can get corresponding end edge of each *block*. Then the evaluation of current identifier is end. The code start *collect\_dep* function as below:

***for*** each node ***in*** select line\_range[0,1]

{

***collect\_dep***(stg\_node, program, edges,stg\_nodes,slicing\_nodes)

}

The definition of *collect\_dep* function is in listing 4.2, the meaning of each formal parameter in list 4.2 is: *stg\_node* is every single node in selected line, *program* is an array used to keep sliced nodes, *edges* is all the edge on state graph, and the last two are represent for all nodes in state graph. In order to avoid repeat query nodes in selected line, we set *skip* as a span value, to skip the nodes in the same *block* with the identifier which the *collect\_dep* function has finished to evaluate. This value can improve efficiency when the selected line includes a function call. As we checked in last section on state graph, the state only go into function’s body when that function is called by actual parameter pass in. So all state inside function invocation occurs in selected line. When the algorithm start to evaluate first identifier in the line, the iterator will go through all nodes encapsule in function call to reach corresponding end edge. So when the algorithm return back to fetch next node in selected line, we can step over all nodes in function call *block*.

function collect\_dep(stg\_node, program, edges,stg\_nodes,slicing\_nodes){

***for***(***var*** i = 0; i < edges.length; i++){

***for***(***var*** j = i + 1; j < edges.length; j++){

***dep\_var*** is the variable name contains in node;

***base*** is scope

**dep\_node** is the dependent node we are searching for

Step 1:

1. First look up **dep\_node** by **var** declaration
2. If **dep\_node** is not found in 1, then look up **dep\_node** by **function** declaration

Step 2:

***If*** (**dep\_node** is found after execute 1 and 2 in step 1){

Put **dep\_node** with necessary information to iterator in order to find other code dependent on **dep\_node**.

}

Step 3: Looking for corresponding end edge.

Set **skip** = index\_of\_end\_edge - index\_of\_start\_edge

}

}

return **skip**;

}

List 4.2 Design of function used to collection dependency.

As we discussed above, we need two functions to identify data and control dependencies by comparing identifiers with its scope. The function is used to look up by var declaration is in listing 4.3. The principle to find dependent variable or function declaration is in the same way, just compare the variable name or function name with every node. In our assumption, we were planning to query from top to bottom, however, we found the same effect if we query from the current node to top node.

function lookup\_dep\_node\_by\_var(stg\_nodes, start\_index, var\_name, base){

***for***(***var*** i = start\_index; i > 0; i--){

***if***(stg\_nodes[i].node){

***var*** parsenode = stg\_nodes[i].node;

***if***(parsenode.declarations){

***for***(***var*** j = 0; j < parsenode.declarations.length; j++){

***if***(parsenode.declarations[j].id.name == var\_name){

***if***(stg\_nodes[i].benva && stg\_nodes[i].benva.base == base)

***return*** stg\_nodes[i];

}

}

}

}

}

***return false***;

}

Listing 4.3 Look up dependence node by *var* declaration

Look up by *function* declaration is in same principle with *var* declaration, the only different operation is, we first need to check if the type of node is a *“FunctionDeclaration”* node. the function implementation is:

function lookup\_dep\_node\_by\_func\_name(stg\_nodes, start\_index, var\_name, base){

***for***(***var*** i = 0; i < stg\_nodes.length; i++){

***if***(stg\_nodes[i].node){

***var*** parsenode = stg\_nodes[i].node;

***if***(parsenode.type=="FunctionDeclaration" && parsenode.id && parsenode.id.name==var\_name){

***if***(stg\_nodes[i].benva && stg\_nodes[i].benva.base == base){

***return*** stg\_nodes[i];

}

}

}

}

***return false***;

}

Listing 4.4 Look up dependence node by *function* declaration

The iterator function used to find indirectly dependences is in listing 4.5.

function slice\_dep\_node(dep\_node, program, edges, stg\_nodes, slicing\_nodes){

Step 1:

call **collect\_dep** function to find dependent nodes

collect\_dep(dep\_node, program, edges, stg\_nodes, slicing\_nodes);

Step 2: append the current node to sliced array

}

Listing 4.5 Function is responsible for finding iterative dependences.

Above algorithm is a *general pattern* to find slices. This algorithm is able to append statements have effect on selected code into slice without reflecting the structure of **if-else** statement. In order to keep deterministic process of predicate, we need a function to read stack information of **if-else** statement in CESK-machine. Respect our hypothesis in last section about **if-else** statement, we first set a function to process the boolean result of predicate for each clause. The node label with same *tag*, is in same predicate expression. By identifying the boolean result of this expression, we can decide which branch of **if-else** statement should be replaced by null. The function implementation of pre-process **if-else** statement is in listing 4.6.

***var*** tag = stg\_nodes[if\_node\_index].node.tag;

***if***(stg\_nodes[i].frame && stg\_nodes[i].frame.node

&& stg\_nodes[i].frame.node.type == "IfStatement" && stg\_nodes[i].frame.node.tag == tag){

***if***(stg\_nodes[i].value.prim.cvalue == true){

*//if the current predicate is true*

*//cut off its alternate clause by fill in null;*

stg\_nodes[i].node.alternate = null;

}***else***{

stg\_nodes[i].node.consequent.type = "BlockStatement";

stg\_nodes[i].node.consequent.body = [];

}

}

Listing 4.6 Pre-process of **if-else** statement.

The query of dependences in **if-else** statement consist of two parts, one is searching dependences of expression in predicates, another one is searching dependences of each clause after *true* predicates. The way of query is the same to general pattern.

## 4.4 Conclusion

In this chapter we first illustrated how we implemented tier splitting tool STIP.js on top of the Jipda framework. We pointed out two abstract interpreter style embedded in our tier splitting tool and sketched how theywork. A CESK-machine based interpreter in Jipda framework to get state graph from source code. A self-interpreter in STIP.js used to construct a PDG based on Jipda. In the second part, we put forward with case studies on state graph to analysis how to query correctness dependences on state graph to get slice derived by slicing criterion. We suppose the query on state graph is always from the top to bottom at first. We put forward three assumptions in this section:

1. Directly influence statement and variable is found by matching identifiers and scope information on the edges.
2. Indirectly influence statement and variable is found by iteratively executing to find dependences on directly influence statement and variable we get in each stage of query.
3. We can replace the code in clause result from false predicates with *null* value.

According the conclusion we observed from case study, we elaborate our implementation of JipdaSlicer in last part of this chapter. We sketch the design of algorithm and list some key function. In next chapter, we will evaluate JipdaSlicer using some sample code and compare the slicing result with which we get from STIP.js.

5

Validation

In previous chapter, we introduced our tier splitting tool STIP.js and a new slicing algorithm JipdaSlicer. In this chapter we validate JipdaSliser, by comparing the slices result from STIP.js and it. First we test both of them on the code in figure 2.1(a). From the testing result we can see, two sliced code result from STIP.js and JipdaSliser is conform to Weiser’s slice theory. We further carry out three test case to evaluate accuracy of two tools. From the slicing result we can see the interpreter installed in two tools has different feature. When slicing on condition statement, JipdaSlicer returns more pleasant result code. Finally, we give the conclusion about our test result.

## 5.1 Evaluation

1. Evaluation of slicing on example code in figure 2.1(a).

From the figure 5.1 and figure 5.2, we can see the pleasing result we get from STIP.js and JipdaSlicer. Both of them return correctness slice based on Weiser’s slicing theory: by selected an interested point, the program slice can be get from deleting the statements have no effect on the selected point. The computation result of sliced code keep consistent with the original code. Hence, in two figures we can see, both two slicing tools delete line no.2 and the statements between line 5 to 8 in code in the left side. The computation result of variable *ac* and *abc* in all code segment is 4 and 7.

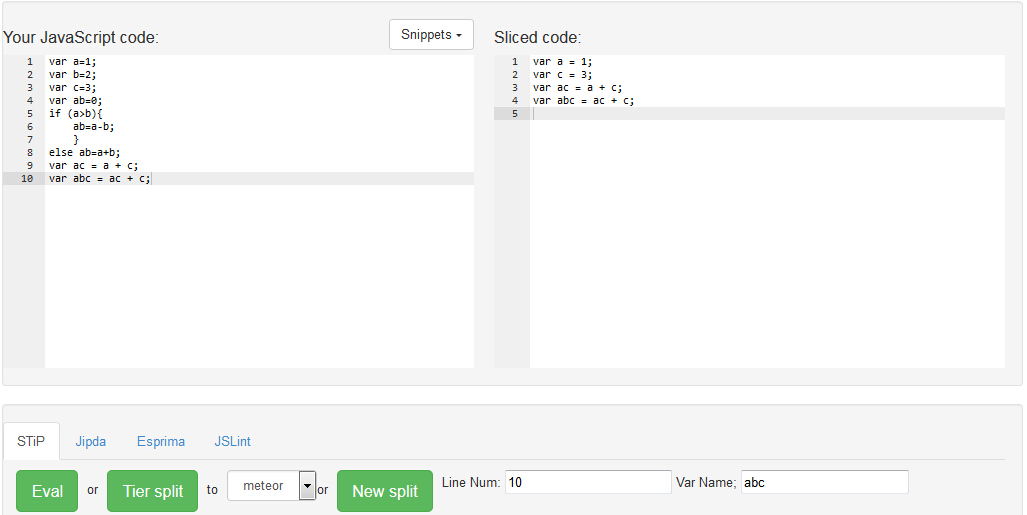


Figure 5.1 Slicing result get from Jipdaslicer on slice criterion <10, {abc}>.



Figure 5.2 Slicing result get from STIP.js on slice criterion <10, {abc}>.

2. Test case 1: slicing on function invoking.

This example aims to evaluate the correctness of function invoking in different way. Therefore, we set for sub case to define functions by using reserved word *var* as well as function declarations. Figure 5.3(a) is used to test on invoking function expression by *var* declaration. As we use Weiser’s theory analysis on the test code, the slicing result is shown in figure(b).

1

2

3

4

5

6

7

8

9

10

11

12

var a=1;

var b=2;

var sum=function(x,y){

var xx=x\*x+a;

var yy=y\*y+b;

var xy=xx+yy;

return xy;

}

function foo(z){

return z+1;

}

var s=sum(a,b);

(a)

var a=1;

var b=2;

var sum=function(x,y){

var xx=x\*x+a;

var yy=y\*y+b;

var xy=xx+yy;

return xy;

}

~~function foo(z){~~

~~return z+1;~~

~~}~~

var s=sum(a,b);

(b)

Figure 5.3 (a) test code; (b) slicing with criteria<12,{s}>;

The following pictures is two algorithms resulting in sliced code of above example. In the right part of figures, we can see the correctness of two algorithm.



Figure 5.4(a) Slicing result on <12,{s}> of code in figure 5.3 (a) by JipdaSlicer.



Figure 5.4(b) Slicing result on <12,{s}> of code in figure 5.3 (a) by STIP.js.

The test code we set in figure 5.5 is used to test on invoking function declaration. Compare with the test code in figure 5.3, we only change the last line of code, and in this time the sliced code should get rig of function declaration of *sum* function. Figure 5.6 shows the slice results. As the correctness of result we can say the two slicing tools can recognize *function* declaration.

1

2

3

4

5

6

7

8

9

10

11

12

var a=1;

var b=2;

var sum=function(x,y){

var xx=x\*x+a;

var yy=y\*y+b;

var xy=xx+yy;

return xy;

}

function foo(z){

return z+1;

}

var f=foo(a,b);

(a)

var a=1;

var b=2;

~~var sum=function(x,y){~~

~~var xx=x\*x+a;~~

~~var yy=y\*y+b;~~

~~var xy=xx+yy;~~

~~return xy;~~

~~}~~

function foo(z){

return z+1;

}

var f=foo(a,);

(b)

Figure 5.5 (a) test code; (b) slicing with criteria<12,{f}>;

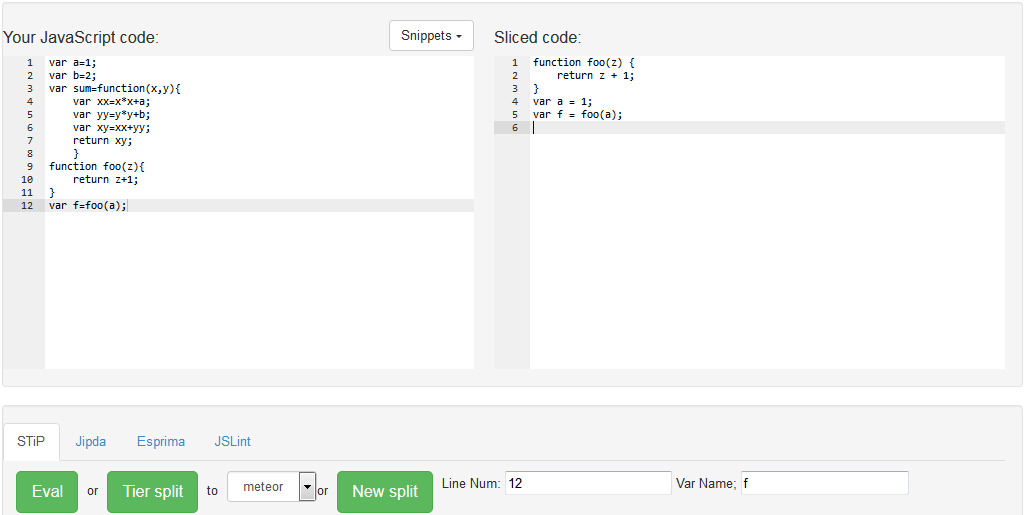


Figure 5.6 (a) Slicing result on <12,{f}> of code in figure 5.5 (a) by JipdaSlicer.

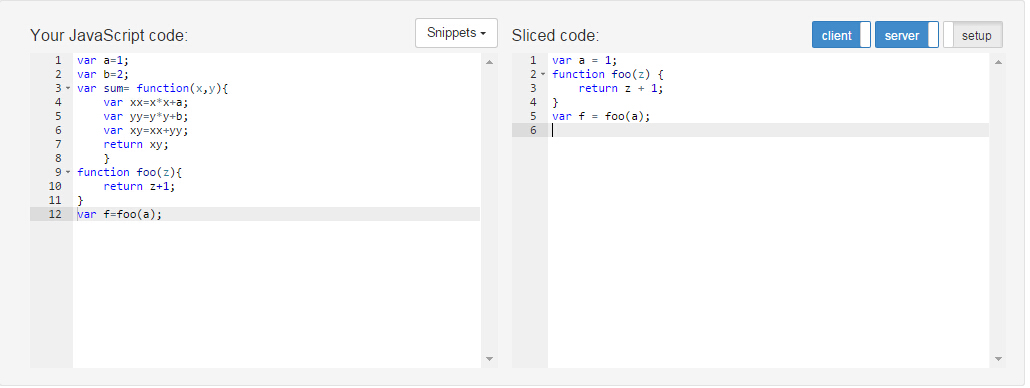


Figure 5.6 (b) Slicing result on <12,{f}> of code in figure 5.5 (a) by STIP.js.

The next test case shows slicing with inter procedural call.

1

2

3

4

5

6

7

8

9

10

11

12

var a=1;

var b=2;

var sum=function(x,y){

var xx=x\*x+a;

var yy=y\*y+b;

var xy=xx+yy;

return xy;

}

function foo(z){

return z+1;

}

var c=sum(foo(b),a);

Figure 5.7 Test code used to evaluate inter procedural call.

The following pictures show the slicing result on slicing criterion *<12,{c}>*, since in line 12, the expression includes two function invocation, the result code should keep all code in slicing result.

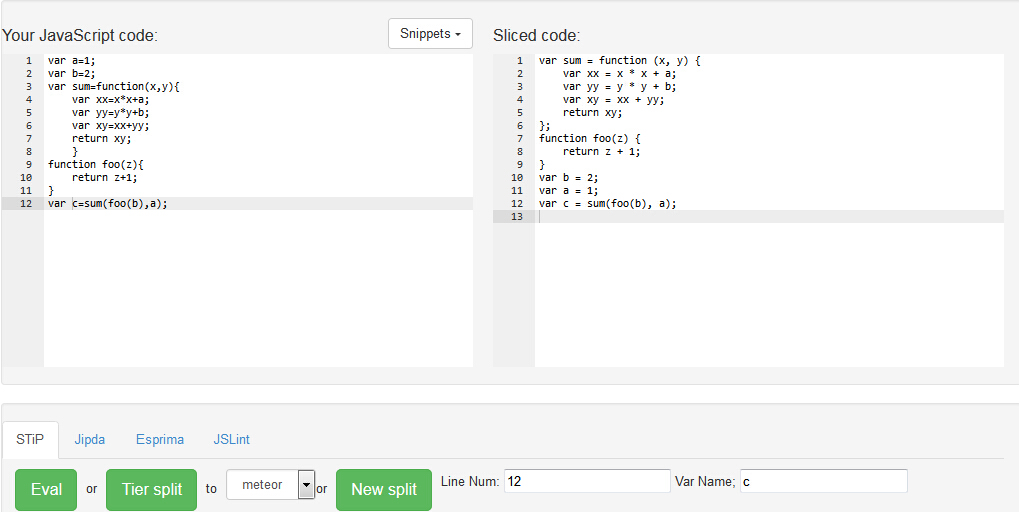


Figure 5.8 (a) Slicing result on <12,{c}> of code in figure 5.7 by JipdaSlicer.



Figure 5.8 (b) Slicing result on <12,{c}> of code in figure 5.7 by STIP.js.

Looking at the results we get from two algorithm in above test cases, we can know both of them return us the correct code of program slicing. The different thing is the order of line in sliced code. The different way of query of node result in this kind of different ordering. In JipdaSlicer, after abstract interpreter invoked by each node in selected line, the algorithm evaluate each node in that line in sequence. If the node is bundled with identifier, the node on which it dependent found is always found by traversal on state graph to matching the variable with scope. So the order of code in JipdaSlicer depends on the position of identifier in selected line.

Another case of function invoking is by nesting. We set this sub case in order to checking if the two algorithm binding with right scope when analysis the source code. The test code is in figure 5.9. In the test code we define a function *foo* not only in global environment, but also inside the *sum* function. If we slice the code on line 13, only the nested function should involved in sliced code.

1

2

3

4

5

6

7

8

9

10

11

12

13

var a=1;  
var b=3;  
var foo=function(z){  
    return z+1;  
}  
var sum=function(x,y){  
    function foo(x){  
        var xx=x+1;  
        return xx;  
        }  
    return foo(x)+y;  
    }  
var c=sum(b,1);

Figure 5.9 Test code.

The following figures show us we can get correct sliced code by two algorithm slicing on criterion <13,{c}>. The different is JipdaSlicer keep the function nested but STIP.js decomposes the two function in same level. This is because we Stip.js build PDG, it create an entry node for each function and after traverse all related node, the algorithm prints the function based on each entry node. In JipdaSlicer, after the algorithm find corresponding function declaration, it just fetch the whole body of that function. In the line used to start slice, is doesn’t contain an identifier point to *foo* function define in the same level with *sum*, so JipdaSlicer didn’t check on this function declaration.



Figure 5.10(a) Result of slicing on <11,{d}> of the code in figure 5.9 by JipdaSlicer.



Figure 5.10(b) Result of slicing on <11,{d}> of the code in figure 5.9 by STIP.js.

3. Test case 2: slicing on slicing on conditional statement.

We make this example in order to check grammatical structure of conditional statement in the result code. The test code is in figure 5.11(a). As shown in figure 5.11, if we want to get slice on slicing criteria *<16,{max}>* the result would like in figure 5.11(b). The result code will get rid of the false branch of **if-else** statement based on computation result in conditional expression. Based on our design of JipdaSlicer, in the pre-process phase of conditional statement, it will first cuts off branch which isn’t accessed when slicing on program. In this example, since the statement in 16th is in **if-else** statement, it should return an empty block of each false branch. The running the result of JipdaSlicer is shown in figure 5.12.

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

var a = 1;

var b = 2;

var c = 3;

var d = 4;

var f =0;

var g=1.2;

var max=0;

if(a<b) {

if(b<c){

if(c<d){

if(d<f)

max=f;

else if(d<g)

max= g;

else

max=d;

}

}

}

(a)

var a = 1;

var b = 2;

var c = 3;

var d = 4;

var f =0;

var g=1.2;

var max=0;

if(a<b) {

if(b<c){

if(c<d){

if(d<f)

~~max=f;~~

else if(d<g)

~~max= g;~~

else

max=d;

}

}

}

(b)

Figure 5.11 (a) sample code; (b) slicing with criteria<16,{max}>

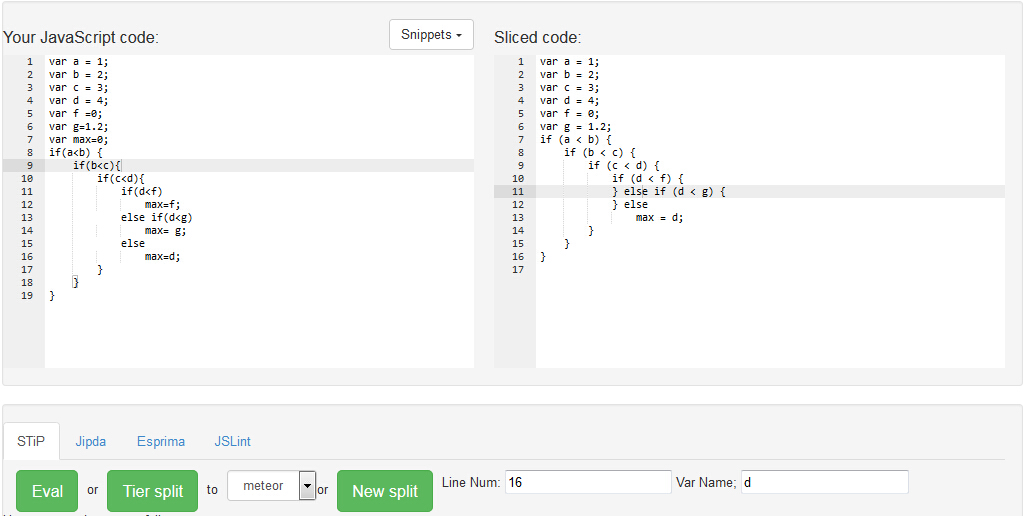


Figure 5.12 Slicing result of figure 5.11(b) by JipdaSlicer

The following picture shows the result we get from STIP.js. From the figure we can see STIP.js remains all source code. By check the sliced PDG used to generate this result code, the STIP.js marks correct node in PDG, as the nodes in reach. So STIP.js also traces on right data and control dependences. However, when generate target code from PDG, STIP.js keeps all code in order to sustain **if-else** structure. The different between the two solutions is JipdaSlice use *null* value fill in the clause when it isn’t accessed based on prediction.



Figure 5.13 (a) Slicing result of figure 5.11 (b) by STIP.js

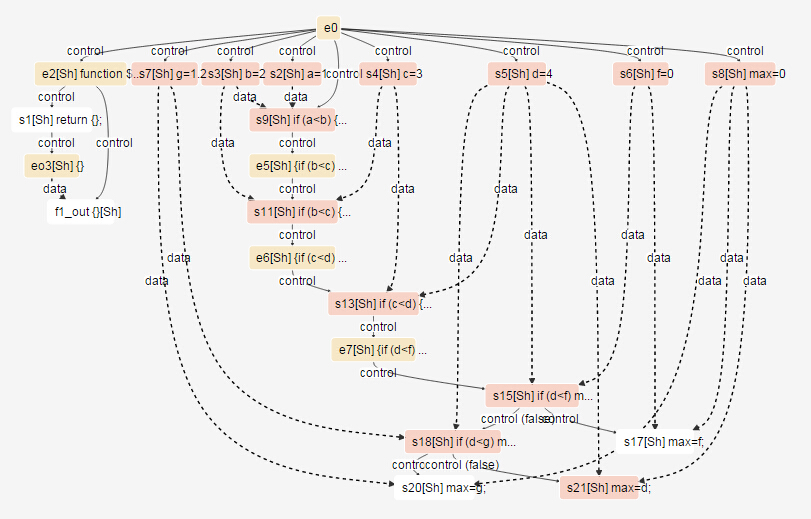


Figure 5.13 (b) PDG generate by STIP.js of the sample code in figure 5.11(a).

4. Test case 3: slicing on function invoking with conditional statement.

This example shows another situation where the process of slicing involves conditional statement. The difference with test case 2 is the slicing criterion is not in any clause of **if-else** statement. We set this example because JipdaSlicer has different mechanism on this situation.

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

var a=1;

var b=3;

var c=2;

var sum=function(x,y){

return x+y;

}

var foo=function(z){

return z+1;

}

if(a<b){

if(b<c)

var d=sum(0,1);

else {

var d=sum(a,b);

}

}

var s=d;

(a)

var a=1;

var b=3;

var c=2;

var sum=function(x,y){

return x+y;

}

~~var foo=function(z){~~

~~return z+1;~~

~~}~~

if(a<b){

~~if(b<c)~~

~~var d=sum(0,1);~~

else {

var d=sum(a,b);

}

}

var s=d;

(b)

Figure 5.14 (a) Sample code; (b)slice on criteria<17,{s}>.

As analysed in figure 5.14(b), line 17th will rely on variable *d* whose value is decided by predicate result of two nested **if-else** statement. The result of code in figure 5.14(b) we get from JipdaSlicer is in figure 5.15 and STIP.js is in figure 5.14(a).

As we can see from the result, JipdaSlicer gets rid of all clause of **if-else** statement and only leaves the statement on which the identifier in 17th is dependent. It’s all because of we configure the interpreter of JipdaSlicer only extract relevant statements if the identifiers contained in slice criterion isn’t involved in **if-else** statement. Such design, we assuming each predicate result of clause in **if-else** statement is not relevant to target code, only the clause is real executed by semantics should be concerned.

We can also see same problem with test case 2 in the result from STIP.js in figure 5.16(a). By check the corresponding PDG of this testing example, we can find the node represents for statement in line 17 isn’t binding with correctness data dependency. This possible due to STIP.js failure to consider of the scope outside **if-else** statement.

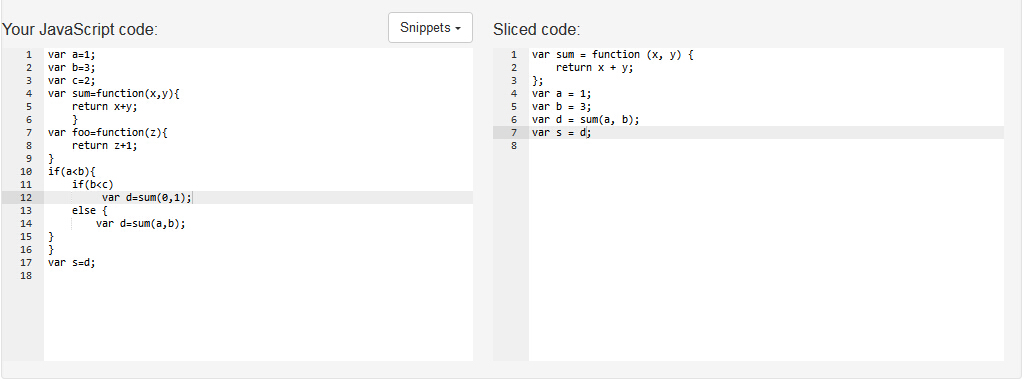


Figure 5.15 Slicing result of figure 5.14 (b) by JipdaSlicer.



Figure 5.16 (a) Slicing result of figure 5.15 (b) by STIP.js.

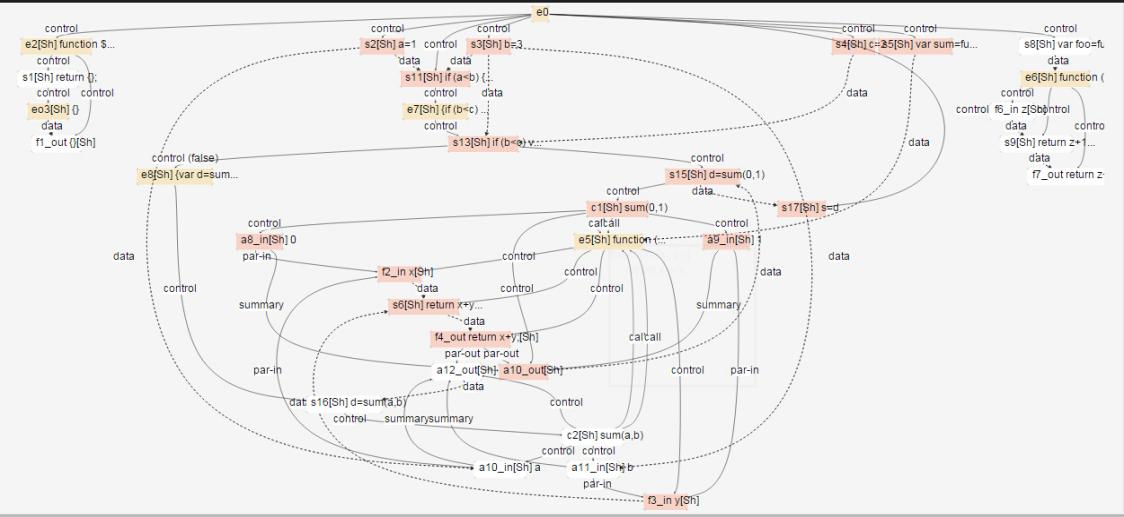


Figure 5.16 (b) Corresponding PDG of code in figure 5.14 (a) by STIP.js.

## 5.2 Conclusion

Fisrt of all, both of JipdaSlicer and STIP.js respect to Weiser’s static slicing theory. From the testing result, we can see the different properties

of abstract interpreter in JipdaSlicer and STIP.js. As we discussed in the first test case, the order of sliced code in JipdaSlicer is dependent on the identifiers’ position in slicing criterion, whereas, STIP.js can keep sequence of statement as same as source code. In the last two cases, we discuss about different mechanism to deal with clause in **if-else** statement in JipdaSlicer based on if slicing criterion invloved in **if-else** statement or not. If the slicing criterion invloved in clause of predicate, we regard for grammatical structure of **if-else** statement is relevant for sliced code. In this case, we configure JipdaSlicer fill *null* value in the clause which isn’t executed based on predicate. The syntax structure of **if-else** statement still be kept in result code so that we can know tracing of clause based on condition expression. On contrast, we only need to fetch executed clause in **if-else** statement if slicing criterion isn’t in any clauses. By comparing with the result code, we can say JipdaSlicer is better in analysis of **if-else** statement. Both of JipdaSlicer and STIP.js are powerful enough to deal with function invoking in any method.

6

Conclusion

As we state our study background at beginning, writing a web application is not easy since it requires developers to have knowledge about each tier and it’s technologies. This complexity can be reduced by tierless programming languages, which enable developers to use a single language, then distribute the tierless code into different logic tier, such as client and server tier in web development. Instead of investment in a novel programming language, we are trying to extend an existing general-purpose language, JavaScript, to have tierless programming feature. Aim at enable Javascript to achieve tier splitting in the case of adorn a minimum of code with @client and @server, we already have STIP.js which built on top of Jipda framework to generate distributed code by program slicing on PDG.

The development of our tier splitting tool is mainly based on two technologies: program slicing and abstract interpreter. The Jipda frame-work results in state graph by static analysis of source code depends on a machine-based abstract interpreter. On top of state graph, STIP.js constru-ct PDG by reconstitution each node and edge through evaluator. Then split the tierless code by static program slicing to generate sliced code in server and client tier.

In this thesis, we develop a new algorithm to slice on state graph instead of PDG, we call it JipdaSlicer. The slice result of JipdaSlicer can be act as prerequisites to accomplish tier splitting. After illustrate our implemen-tation design based on our assumption in section 4.2.1, we set up the eval-uation based on three case study in order to check if JipdaSlicer could compute correct program slices and keep the syntax structure of if-else statement. Finally, we validate JipdaSliser by comparing the slicing result with STIP.js.

## 6.1 Contribution

The focus of this dissertation was applying program slicing technology on a state graph directly. Therefore we make an implementation of JipdaSlic-er which contributes to slice on a state graph directly without constructing a Program Dependency Graph anymore. We give an overview of contribu-tion in this section.

**Improvement**

JipdaSlicer breaks with the traditional slicing algorithm that perform a tra-versal on a Program Dependency Graph. When Stip.js first constructs a Program Dependency Graph, it have to take consider of the type of node, connection with successor and predecessor based on data and control dep-endeces or syntax structure. Instead of building a complex tree structure via analysis of each node to present data and control dependencies, Jipda-Slicer develops a general pattern to deal with dependency matching. This is beneficial from the feature of state graph, all dependent relationship can be identified with identifier information marked on edges. Queries can be executed from top to bottom, as well as, from the slicing node to top. This general pattern isolate dependencies query from analysis syntax structure. To keep the syntax of conditional statements in sliced code, we developed a component in the interpreter to deal with predicates and clauses in condi-tional statements by analysis stack information in the CESK-machine.

**Correctness**

As we can see from evaluation in chapter 5, the slice we got from JipdaSlicer respect to Weiser’s theory. By comparing with STIP.js, JipdaSlicer returns more pleasing result in the case of dealing with **if-else** statements.

## 6.2 Future work

In this section we will discuss possible improvements and extensions for JipdaSlicer.

**Interpreter design**

Our current version of JipdaSlicer deals with conditional statements very well. Next step we will extend JipdaSlicer to process more syntax structure, such as loop statement like **while**, **for**. It is not difficult to accomplish such an extension. The general pattern of searching data and control dependenc-es also fits for loop statement. What we need to do is to design a compone-nt in the interpreter to handle the syntax structure of loop statement, like what we did for **if-else** statement.

**Distributed code**

To achieve the goal of tier splitting, we also need to extend JipdaSlicer to deal with distributed node.

**Algorithm performance**

As we discussed above, slicing on state graph is simplified the process of seeking data and control dependeces. However, query on liner structure cost much time of query than tree structure. So in further work, we need to develop a dynamic programming algorithm to reduce time complexity of our algorithm.

Reference

[1] M. Weiser. Program slices: formal, psychological, and practical investigations of an automatic program abstraction method. PhD thesis, University of Michigan, Ann Arbor, 1979.

[2] M. Weiser. Program slicing. *IEEE Transactions on Software Engineering*, 10(4):352–357, 1984.

[3] Tip F. A Survey of Program Slicing Techniques.[J]. Journal of Programming Languages, 1994, 3(93):121--189.

[4] D.J. Kuck, R.H. Kuhn, D.A. Padua, B. Leasure, and M. Wolfe. Dependence graphs and compiler optimizations. In Conference Record of the Eighth ACMSymposium on Principles of Programming Languages, pages 207–218, 1981.

[5] "Experience with a data flow datatype” J. Comput. Languages, to be published, 1983.

[6] http://www.tonymarston.net/php-mysql/3-tier-architecture.html

[7] J. Ferrante, K.J. Ottenstein, and J.D. Warren. The program dependence graph and its use in optimization. ACM Transactions on Programming Languages and Systems, 9(3):319–349, 1987.

[8] J.M. Barth. A practical interprocedural data flow analysis algorithm. ommunications of the ACM, 21(9):724–736, 1978.

[9] Ottenstein K, L. Ottenstein: "The program dependence graph in a software development environment[J]. Acm Sifsoft, 1984.

[10] Wang L M, Xian Y, Zhang L, et al. JSSlicer: A Static Program Slicing Tool for JavaScript[J]. Applied Mechanics & Materials, 2013, 241-244:2690-2695.

[11] Rodrigues N F, Barbosa L S. Component Identification Through Program Slicing[J]. Electronic Notes in Theoretical Computer Science, 2006:291–304.

[12] Weiser M. Programmers Use Slices When Debugging.[J]. Communications of the Acm, 1982, 25(7):446-452.

[13] Mohapatra D P, Mall R, Kumar R. An Overview of Slicing Techniques for Object-Oriented Programs[C]// INFORMATICA. 2006.

[14] Zhao J. Applying Slicing Technique to Software Architectures[C]// In Proc. of 4th IEEE International Conferencei on Engineering of Complex Computer Systems. 1998:87--98.

[15] Meijer, Henricus Johannes Maria, Manolescu, Dragos. Tier splitting for occasionally connected distributed applications: WO, WO2011100171 A2[P]. 2011.

[16] M. Weiser. Private communication, 1994.

[17] J.-F. Bergeretti and B.A Carre. [Information-flow and data-flow analysis of while-programs](http://dl.acm.org/citation.cfm?id=2366" \t "http://xueshu.baidu.com/_blank). ACM Transactions on Programming Language and Systems, 7(1):37-61, 1985.

[18] S. Horwitz, T. Reps, and D. Binkley. Interprocedural slicing using dependence graphs. ACM Transactions on Programming Languages and Systems, 12(1):26–61, 1990.

[19] Liang D, Harrold M J. Slicing objects using system dependence graphs[C]// Software Maintenance, 1998. Proceedings., International Conference on. IEEE, 1998:358-367.

[20] Larsen L, Harrold M J. Slicing object-oriented software[C]// icse. IEEE Computer Society, 1996:495.

[21] Korel B, Laski J. Dynamic program slicing[J]. Information Processing Letters, 1988, 29(88):155–163.

[22] Lakhotia A. Improved Interprocedural Slicing Algorithm[J]. Report Cacs Tr, 1992.

[23] Ball T, Horwitz S. Slicing Programs with Arbitrary Control Flow[C]// 1ST CONFERENCE ON AUTOMATED ALGORITHMIC DEBUGGING. 1993:206--222.

[24] G. Canfora, et al. Conditioned program slicing. Information and Software Technology, 40(11-12):595~607 1988.

[25] Choi J D, Ferrante J. Static Slicing in the Presence of GOTO Statements[J]. Acm Transactions on Programming Languages & Systems, 1994, 16(4):1097--1113.

[26] Korel B, Laski J. Dynamic program slicing[C]. Information Processing Letters, 1988, 29(88):155–163.

[27] T. Reps and W. Yang. The semantics of program slicing and program integration. In Proceedings of the Colloquium on Current Issues in Programming Languages, volume 352 of Lecture Notes in Computer Science, pages 60–74. Springer Verlag, 1989.

[28] Agrawal H. On Slicing Programs with Jump Statements[J]. Acm Sigplan Notices, 1994, 29(6):302-312.

[29] Landi W, Ryder B G. A safe approximate algorithm for interprocedural pointer aliasing[J]. Acm Sigplan Notices, 2004, 39(7):473-489.

[30] R. Gupta, M.J. Harrold, and M.L. Soffa. An approach to regression testing using slicing. In Proceedings of the Conference on Software Maintenance, pages 299–308, 1992.

[31] Cheng J. Slicing Concurrent Programs - A Graph-Theoretical Approach[C]// Proceedings of the First International Workshop on Automated and Algorithmic Debugging. Springer-Verlag, 1993:223--240.

[32] Agrawal H, Horgan J R. Dynamic Program Slicing[J]. Sigplan Notices, 1990, 25(6):246-256.

[33] Weihl W E. Interprocedural Data Flow Analysis in the Presence of Pointers, Procedure Variables and Label Variables.[J]. POPL '80 Proceedings of the 7th ACM SIGPLAN-SIGACT symposium on Principles of programming languages, 1980.

[34] J.R. Lyle. Evaluating Variations on Program Slicing for Debugging. PhD thesis, University of Maryland, 1984.

[35] R. Compilers : principles, techniques, and tools[J]. Compilers Principles Techniques & Tools, 1988.

[36] ALLEN, F. E. Control flow analysis. In Proceedings of a Symposium on Compiler Optimization. SZGPLAN Not. 5, 7 (July 1970), 1-19.

[37] Ottenstein K J. Data-Flow Graphs as an Intermediate Program Form[J]. Purdue University, 1978.

[38] Ottenstein K J. And Ottenstein L M.[An intermediate program form based on a cyclic data-dependence graph](http://citeseerx.ist.psu.edu/showciting?cid=2115552" \t "http://xueshu.baidu.com/_blank). CS-TR 81-1,Dept. Of Computer Science, Michgan Technological Univ., Houghton, MI, Oct. 1981; July 1982, errata.

[39] E. Duesterwald, R. Gupta, and M.L. Soffa. Rigorous data flow testing through output influences. In proceedings of the Second Irvine Software Symposium ISS’92, pages 131–145, California, 1992.

[40] R. Gupta, M.J. Harrold, and M.L. Soffa. An approach to regression testing using slicing. In Proceedings of the Conference on Software Maintenance, pages 299–308, 1992.

[41] Binkley D. Using Semantic Differencing to Reduce the Cost of Regression Testing[C]// In Proceedings of the Conference on Software Maintenance1992:41--50.

[42] Horwitz S. Identifying the Semantic and Textual Differences Between Two Versions of a Program[J]. Proceedings of the Acm Sigplan Conference on Programming Language Design & Implementation, 1990:234--245.

[43] Gallagher K B, Lyle J R. Using Program Slicing in Software Maintenance[C]// IEEE Transactions on Software Engineering1991:751--761.

[44] [On Choosing Program Refactoring and Slicing Re-engineering Practice Towards Software Quality](http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.217.8119" \t "http://xueshu.baidu.com/_blank).

[45] Weiser M. Reconstructing Sequential Behavior from Parallel Behavior Projections.[J]. Information Processing Letters, 1983, 17:129-135.

[46] Cousot P, Cousot R. Abstract interpretation: a unified lattice model for static analysis of programs by construction or approximation of fixpoints[C]// Proceedings of the 4th ACM SIGACT-SIGPLAN symposium on Principles of programming languagesACM, 1977:238-252.

[47] <http://esprima.org/index.html.>

[48] https://github.com/estree/estree/blob/master/spec.md

[49] <https://en.wikipedia.org/wiki/Interpreter_(computing)#Self-interpreter>

[50] <https://en.wikipedia.org/wiki/Syntax_(programming_languages)>

[51] Scott D. The lattice of flow diagrams[M]// Symposium on Semantics of Algorithmic LanguagesSpringer Berlin Heidelberg, 1971:311-366.

[52] Floyd R W. Assigning meanings to programs[J]. Studies in Cognitive Systems, 1993, 14:19-32.

[53] <http://matt.might.net/articles/cesk-machines/>

[54] https://en.wikipedia.org/wiki/Meta-circular\_evaluator.

[55] Clements J, Felleisen M. A tail-recursive machine with stack inspection[J]. Acm Transactions on Programming Languages & Systems, 2004, 26(6):1029-1052.