SETL NEWSLETTER 193

ON A STATIC SCHEME TO FIND PROCEDURE VARIABLES

MICHA SHARIR
MAY 18 1977

WE PRESENT HERE AN APPROACH TO THE TREATMENT OF PROCEDURE VARIABLES IN THE SETL OPTIMIZER. THE ASSUMED INPUT TO THE OPTIMIZER IS A MODULE CONTAINING SEVERAL PROCEDURES. THE NAMES OF THESE PROCEDURES ARE GLOBAL WITHIN THE MODULE AND THEREFORE NO OTHER VARIABLE IN THE MODULE CAN MAKE USE OF THESE NAMES. EXTERNAL VARIABLES MUST BE DEFINED AS SUCH, BUT THE USER NEED NOT SPECIFY WHETHER EXTERNAL OBJECTS ARE PROCEDURE NAMES, MAPS OR WHATEVER.

AS FAR AS SUBROUTINES ARE CONCERNED, THE PROBLEMS ARE NOT SERIOUS, SINCE A SUBROUTINE CALL IS SYNTACTICALLY RECOGNIZABLE BY THE COMPILER, ITS FORMAT BEING PROCNAME (EXP1,EXP2, ... ,EXPN). IF PROCNAME IS ONE OF THE PROCEDURE NAMES WITHIN THE MODULE, THEN THIS CALL INSTRUCTION ESTABLISHES A UNIQUE WELL DEFINED LINK IN THE CALL GRAPH. OTHERWISE, PROCNAME MAY BE A PROCEDURE VARIABLE AND THE ABOVE INSTRUCTION ESTABLISHES LINKS TO ALL ≠COMPATIBLE≠ SUBROUTINES (IF WE LACK ADDITIONAL INFORMATION, A COMPATIBLE SUBROUTINE MUST BE DEFINED AS AN INTERNAL ONE HAVING THE SAME NUMBER OF PARAMETERS AS IN THE CALLING INSTRUCTION, OR ANY EXTERNAL PROCEDURE). SINCE IT MAKES NO DIFFERENCE TO THE OPTIMIZER WHICH EXTERNAL PROCEDURE IS CALLED, IT IS BEST TO TREAT ALL SUCH CALLS AS CALLS TO A SINGLE EXTERNAL ≠BLACK BOX≠ SUPER PROCEDURE WHICH MAY USE AND MODIFY ALL EXTERNAL AND GLOBAL PUBLIC VARIABLES, AND MAY CALL ANY PUBLIC PROCEDURE BELONGING TO THE MODULE BEING OPTIMIZED. NOTE THAT WHEN PROCEDURE VARIABLES OCCUR THE CALL GRAPH MAP CAN BE MULTI-VALUED.

MORE SERIOUS PROBLEMS ARISE WHEN WE CONSIDER FUNCTION CALLS. ALTHOUGH THE BASIC APPROACH THAT WE USE IS PRECISELY THE SAME AS FOR SUBROUTINES, A NEW DIFFICULTY ARISES FROM ONE OF THE UNDERLYING PRINCIPLES OF THE SETL LANGUAGE; NAMELY - THAT THERE SHOULD BE NO DISTINCTION BETWEEN THE FUNCTIONAL APPLICATION OF A PROCEDURE OR A MAP, BOTH HAVING THE SAME SYNTAX   Y = F(X). IF F IS AN INTERNAL PROCEDURE NAME, OR AN EXTERNAL NAME, THEN WE CAN PROCEED IN THE SAME WAY AS ABOVE, AND INCUR THE SAME PROBABLE AMOUNT OF OVER - ESTIMATION. HOWEVER, IF F IS AN INTERNAL VARIABLE AND NOT A PROCEDURE NAME, WE WOULD HAVE TO ASSUME, IN THE ABSENCE OF

OTHER INFORMATION, THAT F MAY BE A PROCEDURE VARIABLE AND LINK THE ABOVE ≠CALLING≠ INSTRUCTION TO ALL COMPATIBLE FUNCTION PROCEDURES. THUS, IN AN UNDECLARED SETL PROGRAM, ONE MIGHT FEAR THAT ALL MAP RETRIEVALS WILL BE INTERPRETED AS VARIABLE FUNCTION CALLS.

OBVIOUSLY SO CRUDE A MANNER OF PROCEEDING IS TOO COSTLY, FOR MAP RETRIEVALS ARE QUITE COMMON IN SETL. A COMPLETE VALUE FLOW ANALYSIS COULD IN MOST CASES DISTINGUISH RATHER ACCURATELY BETWEEN MAP RETRIEVALS AND CALLS, BUT TO DO A FULL VALUE FLOW ANALYSIS FOR THIS PURPOSE WE WOULD HAVE TO ITERATE OVER MOST OF THE OPTIMIZER. WE PRESENT NOW A COARSER, BUT EFFICIENT METHOD TO OVERCOME THIS PROBLEM.

LET US INTRODUCE THE NOTION OF ≠PROCEDURE CONTAMINATION≠. A VARIABLE WILL BE CALLED DIRECTLY CONTAMINATED IF IT CAN BE A PROCEDURE VARIABLE. A ≠PRIMITIVE CONTAMINATION STATUS≠ (PCS) IS A STRING OF THE SYMBOLS ≠S≠ (≠SET OF≠) AND ≠T≠ (≠TUPLE OF≠), WITH THE FOLLOWING MEANING.

A) ALL DIRECTLY CONTAMINATED VARIABLES HAVE THE EMPTY STRING NULC AS A PCS.

B) A VARIABLE WHICH MIGHT BE A SET CONTAINING A MEMBER WITH A PCS W, WILL HAVE SW AS A PCS.

C) A VARIABLE WHICH MIGHT BE A TUPLE CONTAINING AN ELEMENT WITH A PCS W, WILL HAVE TW AS A PCS.

THE ≠GENERAL CONTAMINATION STATUS≠ (GCS) OF A VARIABLE IS THE SET OF ALL PCS≠S OF THIS VARIABLE. VARIABLES FOR WHICH THIS SET IS NON-EMPTY ARE CALLED ≠CONTAMINATED≠. THUS, THE CONTAMINATION STATUS TELLS US IN WHAT WAYS THE VARIABLE CAN BE RELATED TO POTENTIAL PROCEDURE VARIABLES. NOTE THAT THE PROCEDURE CONSTANTS WITHIN THE MODULE ARE NOT CONSIDERED CONTAMINATED. THEY MIGHT, HOWEVER, BECOME SO DURING THE ANALYSIS (SEE REMARK 3 BELOW).

THE CONSTRUCTION OF THE GCS≠S IS COMPLETELY STATIC. WE START BY SETTING THE GCS OF ALL VARIABLES TO WHICH PROCEDURE CONSTANTS ARE ASSIGNED, TO ≤ NULC ≥ , AND BY SETTING THE GCS OF ALL UNDECLARED EXTERNALS TO THE UNIVERSAL SET OF ALL PCS≠S (UP TO A CERTAIN FIXED LENGTH, SEE REMARK 1 BELOW). BECLARED EXTERNALS GET A GCS WHICH DEPENDS ON THEIR DEFINITION. E.G. IF AN EXTERNAL IS DEFINED AS A SET, ITS GCS WILL CONTAIN ALL STRINGS WHICH STARTS WITH ≠S≠.

IF AFTER THIS STEP NO VARIABLE IS CONTAMINATED (WHICH, HOPEFULLY, WILL BE THE COMMON CASE) THEN THERE CAN NOT EXIST PROCEDURE VARIABLES. IT IS THIS CASE FOR WHICH THE OPTIMIZER WILL MOST EASILY BE ABLE TO PRODUCE SUBSTANTIALLY EFFICIENT CODE.

OTHERWISE, WE PROCEED BY THE USUAL ≠WORKPILE METHOD≠ TO UPDATE THE GCS OF VARIABLES WHICH APPEAR AS OVARIABLES IN INSTRUCTIONS WITH A CONTAMINATED IVARIABLE. THE FOLLOWING ARE SOME TYPICAL EXAMPLES OF SUCH UPDATES. (Y IS ASSUMED TO BE ALREADY CONTAMINATED.)

A) S WITH Y;

```
    GCS(S) = GCS(S) + ≤ ≠S≠↑↑W : W ↱ GCS(Y) ≥
```

B) F(X) = Y;

```
    GCS(F) = GCS(F) + ≤ ≠T≠↑↑W : W ↱ GCS(Y) ≥
                    + ≤ ≠ST≠↑↑W : W ↱ GCS(Y) ≥
```

THE FIRST SET CORRESPONDS TO F BEING A TUPLE, AND SECOND ONE TO F BEING A MAP.

C) P FROM Y;

```
    GCS(P) = GCS(P) + ≤ W(2:) : W ↱ GCS(Y) ↑ W(1) = ≠S≠ ≥
```

D) P = Y(K);

```
    GCS(P) = GCS(P) + ≤ W(2:) : W ↱ GCS(Y) ↑ W(1) = ≠T≠ ≥
                    + ≤ W(3:) : W ↱ GCS(Y) ↑ W(1:2) = ≠ST≠ ≥
```

IF Y IS DIRECTLY CONTAMINATED, THERE IS FURTHER UPDATE TO BE DONE. SEE REMARKS 3 AND 4 BELOW.

AFTER COMPLETING THIS ANALYSIS, WE CAN DETERMINE WHICH FUNCTIONAL APPLICATIONS CAN STILL BE VARIABLE PROCEDURE CALLS. MORE PRECISELY - IF Y = F(X) IS SUCH AN APPLICATION, THEN IF F IS DIRECTLY CONTAMINATED, THEN THIS MAY BE A VARIABLE PROCEDURE CALL, OTHERWISE IT HAS TO BE A MAP OR A TUPLE RETRIEVAL. THUS, WE CAN ELIMINATE MOST FALSE CALLS FROM THE CODE, AND ESTABLISH THE CALL GRAPH RATHER ACCURATELY (ESPECIALLY IF THERE ARE NO EXTERNALS). ONE SHOULD ALSO NOTE THAT IN THE REMAINING CASES OF A FUNCTIONAL APPLICATION WHICH MIGHT BE A MAP RETRIEVAL OR VARIABLE PROCEDURE CALL, THE CORRESPONDING Q1 CODE SHOULD INCLUDE THE CALL AND THE RETRIEVAL ON ALTERNATIVE PATHS.

LET US NOW OBSERVE SEVERAL SPECIAL PROBLEMS WHICH ARISE IN THE CONTAMINATION ANALYSIS.

1) THE LENGTH OF A PCS MUST BE LIMITED BY SOME M ≥ 0. IF WE ENCOUNTER A CONTAMINATED VARIABLE WITH TOO COMPOUND A STRUCTURE TO BE DESCRIBED BY M SYMBOLS, WE ASSIGN TO IT THE PCS OBTAINED BY APPENDING A SPECIAL SYMBOL ≠*≠ TO THE LEFT OF THE M RIGHTMOST SYMBOLS OF THE ORIGINAL PCS. HEURISTICALLY, *W MEANS - ANY COMPOUND STRUCTURE OF OBJECTS WITH A PCS W . THIS STATUS IS LEFT UNCHANGED WHEN WE FURTHER INCORPORATE THIS VARIABLE INTO MORE COMPOUND OBJECTS, BUT WHEN AN EXTRACTION IS MADE OUT OF THIS VARIABLE, WE TAKE INTO ACCOUNT ALL RELEVANT INTERPRETATIONS OF THE ≠*≠. FOR EXAMPLE (ASSUMING M = 2).

A) GCS(X) = ≤ ≠*ST≠ ≥

Y FROM X;

THEN GCS(Y) = ≤ ≠*ST≠,≠ST≠ ≥

B) GCS(X) = ≤ ≠*TS≠ ≥

P = X(K);

THEN GCS(P) = ≤ ≠*TS≠,≠TS≠,≠S≠ ≥

THE THIRD PCS APPEARS BECAUSE X(K) MAY BE A MAP RETRIEVAL IN WHICH CASE ≠*≠ MIGHT REPRESENT ≠S≠, OR IN OTHER WORDS X MIGHT BE A SET OF TUPLES (OF A MAP) OF SETS OF PROCEDURE VARIABLES, AND THUS P MAY BE A SET OF PROCEDURE VARIABLES.

2) WHEN A CONTAMINATED VARIABLE IS ASSIGNED AS A PARAMETER TO SOME PROCEDURE, THEN
   A) IF IT IS A KNOWN INTERNAL PROCEDURE THEN WE HAVE TO UPDATE THE CORRESPONDING FORMAL PARAMETER OF THE PROCEDURE.
   B) OTHERWISE, WE HAVE TO UPDATE THE FORMAL PARAMETERS OF ALL THE INTERNAL PROCEDURES WITH THE SAME NUMBER OF PARAMETERS. (SEE 4B BELOW FOR ADDITIONAL DETAILS CONCERNING THIS CASE).

   A SIMILAR TECHNIQUE APPLIES IF A WRITE FORMAL PARAMETER OF A PROCEDURE BECOMES CONTAMINATED.

3) AN INTERNAL FUNCTION CONSTANT WILL BECOME CONTAMINATED IF ITS RETURN VALUE BECOMES CONTAMINATED. IN THIS CASE WE HAVE TO CONTAMINATE THE OVARIABLES OF ALL THE CALLS TO THIS FUNCTION, AND ALSO THE OVARIABLES OF ALL CALLS TO FUNCTION VARIABLES WITH THE SAME NUMBER OF PARAMETERS. HOWEVER, WE DO NOT WANT TO UPDATE THE OVARIABLES TO WHICH THIS FUNCTION NAME IS ASSIGNED AS A CONSTANT. SEE ALSO PART B) OF THE NEXT REMARK.

4) WHEN WE ENCOUNTER A VARIABLE PROCEDURE CALL (I.E. FIND A DIRECTLY CONTAMINATED VARIABLE WHICH APPEARS AS A PROCEDURE NAME IN WHAT MAY BE A PROCEDURE CALL) THEN

   A) IN THE PRESENCE OF CONTAMINATED EXTERNALS, WE MUST GIVE TO ALL GLOBAL PUBLIC VARIABLES, ALL LOCAL STATIC VARIABLES WITHIN PUBLIC PROCEDURES OR PROCEDURES WHICH MIGHT BE CALLED BY SUCH PROCEDURES, ALL WRITE PARAMETERS OF THE PROCEDURE CALL INCLUDING THE OVARIABLE IF A FUNCTION CALL, THE UNIVERSAL SET OF PCS≠S AS A GCS. FOR THE FIRST TWO CLASSES OF VARIABLES, THIS COULD BE DONE ONLY ONCE, AFTER THE FIRST VARIABLE PROCEDURE CALL IS ENCOUNTERED. A GLOBAL FLAG SHOULD BE MAINTAINED TO INDICATE WHETHER SUCH A CALL HAS ALREADY BEEN ENCOUNTERED. WE MIGHT EVEN GO FURTHER IN THE ATTEMPT TO IDENTIFY THE CASES IN WHICH WE HAVE POTENTIAL PROCEDURE VARIABLES, BUT NONE OF THEM IS CONTAMINATED VIA AN EXTERNAL NAME, NOR ARE THERE ANY EXTERNAL CALLS. IN SOME CASES THIS WOULD ALLOW

   US TO AVOID THE OVERALL CONTAMINATION OF THE GLOBAL PUBLIC VARIABLES AND THE LOCAL STATIC ONES AS MENTIONED ABOVE, AND THIS WOULD IN SOME CASES GIVE US A SUBSTANTIALLY MORE ACCURATE PICTURE OF THE CONTAMINATION STATUS.

   TO DO THIS, WE CAN DISTINGUISH BETWEEN TWO CASES OF DIRECT CONTAMINATION - BY AN EXTERNAL VARIABLE OR BY AN INTERNAL PROCEDURE CONSTANT. DENOTE THESE STATES BY THE STRINGS ≠E≠ AND ≠I≠ RESPECTIVELY, AND CARRY THE REST OF THE ANALYSIS THE SAME WAY AS BEFORE, ADJUSTING THE PCS IN AN APPROPRIATE MANNER. THEN, IF NO PROCEDURE CALL HAS A ≠DIRECTLY CONTAMINATED BY AN EXTERNAL≠ NAME, THE OVERALL CONTAMINATION ACTION DESCRIBED JUST ABOVE CAN BE AVOIDED.

   B) IF A PROCEDURE VARIABLE IS NOT CONTAMINATED BY EXTERNALS THEN WE ONLY HAVE TO UPDATE THE WRITE PARAMETERS OF THE PROCEDURE CALL, AND ALSO THE OVARIABLE IN THE CASE OF A FUNCTION CALL. FOR EXAMPLE, TO UPDATE THE OVARIABLE WE ASSIGN TO IT THE UNION OF THE GCS≠S OF THE RETURN VALUES OF ALL INTERNAL COMPATIBLE PROCEDURE NAMES (I.E. ALL COMPATIBLE CONTAMINATED INTERNAL FUNCTIONS. SEE REMARK 3 ABOVE.) AND PROCEED SIMILARLY WITH A WRITE PARAMETER.

5) NOTE THAT IN PRINCIPLE IT IS POSSIBLE TO ANALYZE CONTAMINATION IN MUCH MORE DETAIL, E.G. WE COULD HAVE A DIFFERENT DIRECT CONTAMINATION STATUS INDUCED BY EACH INTERNAL ROUTINE, AND IN THIS WAY WOULD BE APLE TO KEEP TRACK OF THE POINTS AT WHICH SPECIFIC INTERNAL PROCEDURE CONSTANTS MIGHT APPEAR AS VARIABLE VALUES. HOWEVER, SINCE MOST SETL PROGRAMS INVOLVE NO PROCEDURE VARIABLES, SO ELABORATE AN APPROACH DOES NOT SEEM JUSTIFIED.